NEČAS CENTER FOR MATHEMATICAL MODELING, Volume 1

HISTORY OF MATHEMATICS, Volume 29

PRODUCT INTEGRATION, ITS HISTORY AND APPLICATIONS

Antonín Slavík

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = I + \int_a^b A(x)\,\mathrm{d}x + \int_a^b \int_a^{x_2} A(x_2)A(x_1)\,\mathrm{d}x_1\,\mathrm{d}x_2 + \cdots$$

MATFYZPRESS

NEČAS CENTER FOR MATHEMATICAL MODELING, Volume 1

HISTORY OF MATHEMATICS, Volume 29

PRODUCT INTEGRATION, ITS HISTORY AND APPLICATIONS

Antonín Slavík

matfyzpress
VYDAVATELSTVÍ
MATEMATICKO-FYZIKÁLNÍ FAKULTY
UNIVERZITY KARLOVY V PRAZE

PRAGUE 2007

The publication was supported by Nečas Center for Mathematical Modeling and by grant MSM 0021620839 of the Czech Ministery of Education.

ISBN 80-7378-006-2

Table of contents

Preface

This publication is devoted to the theory of product integral, its history and applications. The text represents an English translation of my dissertation with numerous corrections and several complements.

The definition of product integral appeared for the first time in the work of Vito Volterra at the end of the 19th century. Although it is a rather elementary concept, it is almost unknown among mathematicians.

Whereas the ordinary integral of a function A provides solution of the equation

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

the product integral helps us to find solutions of the equation

$$y'(x) = A(x)y(x).$$

The function A can be a scalar function, but product integration is most useful when A is a matrix function; in the latter case, y is a vector function and the above equation represents in fact a system of linear differential equations of the first order.

Volterra was trying (on the whole successfully) to create analogy of infinitesimal calculus for the product integral. However, his first papers didn't meet with a great response. Only the development of Lebesgue integral and the birth of functional analysis in the 20th century was followed by the revival of interest in product integration. The attempts to generalize the notion of product integral followed two directions: Product integration of matrix functions whose entries are not Riemann integrable, and integration of more general objects than matrix functions (e.g. operator-valued functions).

In the 1930's, the ideas of Volterra were taken up by Ludwig Schlesinger, who elaborated Volterra's results and introduced the notion of Lebesgue product integral. Approximately at the same time, a Czech mathematician and physicist Bohuslav Hostinský proposed a definition of product integral for functions whose values are integral operators on the space of continuous functions.

New approaches to ordinary integration were often followed by similar theories of product integration; one of the aims of this work is to document this progress. It can be also used as an introductory textbook of product integration. Most of the text should be comprehensible to everyone with a good knowledge of calculus. Parts of Section 1.1 and Section 2.8 require a basic knowledge of analytic functions in complex domain, but both may be skipped. Sections 3.5 to 3.8 assume that the reader is familiar with the basics of Lebesgue integration theory, and Chapters 4 and 5 use some elementary facts from functional analysis.

Almost every text about product integration contains references to the works of V. Volterra, B. Hostinský, L. Schlesinger and P. Masani, who have strongly influenced the present state of product integration theory. The largest part of this publication is devoted to the discussion of their work. There were also other pioneers of product integration such as G. Rasch and G. Birkhoff, whose works [GR] and [GB] didn't have such a great influence and will not be treated here. The readers with a deeper interest in product integration should consult the monograph [DF], which includes an exhausting list of references.

All theorems and proofs that were taken over from another work include a reference to the original source. However, especially the results of V. Volterra were reformulated in the language of modern mathematics. Some of his proofs contained gaps, which I have either filled, or suggested a different proof.

Since the work was originally written in Czech, it includes references to several Czech monographs and articles; I have also provided a reference to an equivalent work written in English if possible.

I am indebted especially to professor Štefan Schwabik, who supervised my dissertation on product integration, suggested valuable advices and devoted a lot of time to me. I also thank to ing. Tomáš Hostinský, who has kindly provided a photograph of his grandfather Bohuslav Hostinský.

Chapter 1

Introduction

This chapter starts with a brief look on the prehistory of product integration; the first section summarizes some results concerning ordinary differential equations that were obtained prior the discovery of product integral. The next part provides a motivation for the definition of product integral, and the last two sections describe simple applications of product integration in physics and in probability theory.

1.1 Ordinary differential equations in the 19th century

The notion of product integral was introduced by Vito Volterra in connection with the differential equation of the n-th order

$$y^{(n)}(x) + p_1(x)y^{(n-1)}(x) + \cdots + p_n(x)y(x) = q(x). \tag{1.1.1}$$

Such an equation can be converted (see Example 2.5.5) into a system of n linear differential equations of the first order

$$y_i'(x) = \sum_{j=1}^{n} a_{ij}(x)y_j(x) + b_i(x), \quad i = 1, \ldots, n,$$

which can be also written in the vector form

$$y'(x) = A(x)y(x) + b(x). \tag{1.1.2}$$

Volterra was initially interested in solving this equation in the real domain: Given the functions $A : [a, b] \to \mathbf{R}^{n\times n}$ (where $\mathbf{R}^{n\times n}$ denotes the set of all real $n \times n$ matrices) and $b : [a, b] \to \mathbf{R}^n$, we have to find all solutions $y : [a, b] \to \mathbf{R}^n$ of the system (1.1.2). Later Volterra considered also the complex case, where $y : G \to \mathbf{C}^n$, $A : G \to \mathbf{C}^{n\times n}$ and $b : G \to \mathbf{C}^n$, where $G \subseteq \mathbf{C}$ and $\mathbf{C}^{n\times n}$ denotes the set of all $n \times n$ matrices with complex entries.

To be able to appreciate Volterra's results, let's have a brief look on the theory of Equations (1.1.1) and (1.1.2) as developed at the end of the 19th century. A more detailed discussion can be found e.g. in the book [Kl] (Chapters 21 and 29).

A large amount of problems in physics and in geometry leads to differential equations; mathematicians were thus forced to solve differential equations already since the invention of infinitesimal calculus. The solutions of many differential equations have been obtained (often in an ingenious way) in a closed form, i.e. expressed as combinations of elementary functions.

Leonhard Euler proposed a method for solving Equation (1.1.1) in case when the p_i are constants. Substituting $y(x) = \exp(\lambda x)$ in the corresponding homogeneous equation yields the characteristic equation

$$\lambda^n + p_1\lambda^{n-1} + \cdots + p_n = 0.$$

If the equation has n distinct real roots, then we have obtained a fundamental system of solutions. Euler knew how to proceed even in the case of multiple or complex roots and was also able to solve inhomogeneous equations. The well-known method of finding a particular solution using the variation of constants (which works even in the case of non-constant coefficients) was introduced by Joseph Louis Lagrange.

More complicated equations of the form (1.1.1) can be often solved using the power series method: Assuming that the solution can be expressed as $y(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$ and substituting to the differential equation we obtain a recurrence relation for the coefficients a_n. Of course, this procedure works only in case when the solution can be indeed expressed as a power series. Consequently, mathematicians began to be interested in the problems of existence of solutions.

The pioneering result was due to Augustin Louis Cauchy, who proved in 1820's the existence of a solution of the equation

$$\begin{aligned} y'(x) &= f(x, y(x)) \\ y(x_0) &= y_0 \end{aligned} \tag{1.1.3}$$

under the assumption that f and $\frac{\partial f}{\partial y}$ are continuous functions. The statement is also true for vector functions y, and thus the linear Equation (1.1.2) is a special case of the Equation (1.1.3). Rudolf Lipschitz later replaced the assumption of continuity of $\frac{\partial f}{\partial y}$ by a weaker condition

$$\|f(x, y_1) - f(x, y_2)\| < K \cdot \|y_1 - y_2\|$$

(now known as the Lipschitz condition).

Today, the existence and uniqueness of solution of Equation (1.1.3) is usually proved using the Banach fixed point theorem: We put

$$\begin{aligned} y_1(x) &= y_0 + \int_{x_0}^{x} f(t, y_0)\,\mathrm{d}t, \\ y_n(x) &= y_0 + \int_{x_0}^{x} f(t, y_{n-1}(t))\,\mathrm{d}t, \quad n \geq 2. \end{aligned}$$

If f is continuous and satisfies the Lipschitz condition, then the successive approximations $\{y_n\}_{n=1}^{\infty}$ converge to a function y which solves Equation (1.1.3). The method of successive approximations was already known to Joseph Liouville and was used by Émile Picard.

Around 1840 Cauchy proved the existence of a solution of Equation (1.1.3) in complex domain using the so-called majorant method (see [VJ, EH]). We are looking for the solution of Equation (1.1.3) in the neighbourhood of a point $x_0 \in \mathbf{C}$; the solution is a holomorphic function and thus can be expressed in the form

$$y(x) = \sum_{n=0}^{\infty} c_n(x - x_0)^n \tag{1.1.4}$$

in a certain neighbourhood of x_0. Suppose that f is holomorphic for $|x - x_0| \le a$ and $|y - y_0| \le b$, i.e. that

$$f(x,y) = \sum_{i,j=0}^{\infty} a_{ij}(x-x_0)^i(y-y_0)^j. \tag{1.1.5}$$

Substituting the power series (1.1.4) and (1.1.5) to Equation (1.1.3) gives an equation for the unknown coefficients c_n; it is however necessary to prove that the function (1.1.4) converges in the neighbourhood of x_0. We put

$$M = \sup\{|f(x,y)|,\ |x-x_0| \le a,\ |y-y_0| \le b\}$$

and define

$$A_{ij} = \frac{M}{a^i b^j},$$

$$F(x,y) = \sum_{i,j=0}^{\infty} A_{ij}(x-x_0)^i(y-y_0)^j = \frac{M}{(1-(x-x_0)/a)(1-(y-y_0)/b)}. \tag{1.1.6}$$

The coefficients a_{ij} can be expressed using the Cauchy's integral formula

$$a_{ij} = \frac{1}{i!j!}\frac{\partial^{i+j} f}{\partial x^i \partial y^j} = \frac{1}{(2\pi i)^2}\int_{\varphi_a}\int_{\varphi_b} \frac{f(x,y)}{(x-x_0)^{i+1}(y-y_0)^{j+1}}\,\mathrm{d}y\,\mathrm{d}x,$$

where φ_a is a circle centered at x_0 with radius $a > 0$ and φ_b is a circle centered at y_0 with radius $b > 0$. The last equation leads to the estimate $|a_{ij}| \le A_{ij}$, i.e. the infinite series (1.1.6) is a majorant to the series (1.1.5). Cauchy proved that there exists a solution of the equation

$$Y'(x) = F(x, Y(x))$$

that can be expressed in the form $Y(x) = \sum_{n=0}^{\infty} C_n(x-x_0)^n$ in a neighbourhood of x_0 and such that $|c_n| \le C_n$. Consequently the series (1.1.4) is also convergent in a neighbourhood of x_0.

In particular, for the system of linear equations (1.1.2) Cauchy arrived at the following result:

Theorem 1.1.1. Consider functions a_{ij}, b_j $(i,j = 1,\ldots,n)$ that are holomorphic in the disk $B(x_0,r) = \{x \in \mathbf{C}; |x-x_0| < r\}$. Then there exists exactly one system of functions

$$y_i(x) = y_i^0 + \sum_{j=1}^{\infty} c_{ij}(x-x_0)^i \qquad (i = 1,\ldots,n)$$

defined in $B(x_0,r)$ that satisfies

$$y_i'(x) = \sum_{j=1}^{n} a_{ij}(x)y_j(x) + b_i(x),$$
$$y_i(x_0) = y_i^0,$$

where $y_1^0, \ldots, y_n^0 \in \mathbf{C}$ are given numbers.

As a consequence we obtain the following theorem concerning linear differential equations of the n-th order:

Theorem 1.1.2. Consider functions $p_1, \ldots, p_n, q$ that are holomorphic in the disk $B(x_0, r) = \{x \in \mathbf{C}; |x - x_0| < r\}$. Then there exists exactly one holomorphic function

$$y(x) = \sum_{k=0}^{\infty} c_k (x - x_0)^k$$

defined in $B(x_0, r)$ that satisfies the differential equation

$$y^{(n)}(x) + p_1(x) y^{(n-1)}(x) + \cdots + p_n(x) y(x) = q(x)$$

and the initial conditions

$$y(x_0) = y_0,\ y'(x_0) = y_0', \ldots, y^{(n-1)}(x_0) = y_0^{(n-1)},$$

where $y_0,\ y_0', \ldots, y_0^{(n-1)} \in \mathbf{C}$ are given complex numbers.

Thus we see that the solutions of Equation (1.1.1), whose coefficients $p_1, \ldots, p_n, q$ are holomorphic functions, can be indeed obtained by the power series method.

However, it is often necessary to solve Equation (1.1.1) in case when the coefficients $p_1, \ldots, p_n, q$ have an isolated singularity. For example, separation of variables in the wave partial differential equation leads to the Bessel equation

$$y''(x) + \frac{1}{x} y'(x) + \left(1 - \frac{n^2}{x^2}\right) y(x) = 0,$$

whose coefficients have a singularity at 0. Similarly, separation of variables in the Laplace equation gives the Legendre differential equation

$$y''(x) - \frac{2x}{1 - x^2} y'(x) + \frac{n(n+1)}{1 - x^2} y(x) = 0$$

with singularities at -1 and 1.

The behaviour of solutions in the neighbourhood of a singularity has been studied by Bernhard Riemann and after 1865 also by Lazarus Fuchs. Consider the homogeneous equation

$$y^{(n)}(x) + p_1(x) y^{(n-1)}(x) + \cdots + p_n(x) y(x) = 0 \tag{1.1.7}$$

in the neighbourhood of an isolated singularity at $x_0 \in \mathbf{C}$; we assume that the functions p_i are holomorphic in the ring $P(x_0, R) = \{x \in \mathbf{C}; 0 < |x - x_0| < r\}$. If we choose an arbitrary $a \in P(x_0, R)$, then the functions p_i are holomorphic in $U(a, r)$ (where $r = |a - x_0|$), and Equation (1.1.7) has n linearly independent holomorphic solutions $y_1, \ldots, y_n$ in $U(a, r)$. We now continue these functions along the circle

$$\varphi(t) = x_0 + (a - x_0) \exp(it),\ t \in [0, 2\pi]$$

centered at x_0 and passing through a. We thus obtain a different system of solutions $Y_1, \dots, Y_n$ in $U(a,r)$. Since both systems are fundamental, we must have $Y_i = \sum_{j=1}^n M_{ij} y_j$, or in the matrix notation $Y = My$. By a clever choice of the system $y_1, \dots, y_n$ it can be achieved that M is a Jordan matrix. Using these facts, Fuchs was able to prove the existence of a fundamental system of solutions of Equation (1.1.7) in $P(x_0, R)$ that consists of analytic functions of the form

$$(x - x_0)^{\lambda_i} \left(\varphi_0^i(x) + \varphi_1^i(x) \log(x - x_0) + \cdots + \varphi_{n_i}^i(x) \log^{n_i}(x - x_0)\right),$$

$i = 1, \dots, n$, where φ_k^j are holomorphic functions in $P(x_0, R)$ and $\lambda_i \in \mathbf{C}$ is such that $\exp(2\pi i \lambda_i)$ is an eigenvalue of M with multiplicity n_i.

Moreover, if p_i has a pole of order at most i at x_0 for $i \in \{1, \dots, n\}$, then the Fuchs theorem guarantees that φ_k^j has a pole (i.e. not an essential singularity) at x_0. This result implies that Equation (1.1.7) has at least one solution in the form $y(x) = (x - x_0)^r \sum_{k=0}^\infty a_k (x - x_0)^k$; the numbers r and a_k can be calculated by substituting the solution to Equation (1.1.7) (this is the Frobenius method).

We have now recapitulated some basic facts from the theory of ordinary differential equations. In later chapters we will see that many of them can be also obtained using the theory of product integration.

1.2 Motivation to the definition of product integral

The theory of product integral is rather unknown among mathematicians. The following text should provide a motivation for the following chapters.

We consider the ordinary differential equation

$$y'(t) = f(t, y(t)) \tag{1.2.1}$$

$$y(a) = y_0 \tag{1.2.2}$$

where $f : [a,b] \times \mathbf{R}^n \to \mathbf{R}^n$ is a given function. Thus, we are seeking a solution $y : [a,b] \to \mathbf{R}^n$ that satisfies (1.2.1) on $[a,b]$ (one-sided derivatives are taken at the endpoints of $[a,b]$) as well as the initial condition (1.2.2).

An approximate solution can be obtained using the Euler method, which is based on the observation that for small Δt,

$$y(t + \Delta t) \doteq y(t) + y'(t)\Delta t = y(t) + f(t, y(t))\Delta t.$$

We choose a partition $D : a = t_0 < t_1 < \cdots < t_m = b$ of interval $[a,b]$ and put

$$\begin{aligned}
y(t_0) &= y_0 \\
y(t_1) &= y(t_0) + f(t_0, y(t_0))\Delta t_1 \\
y(t_2) &= y(t_1) + f(t_1, y(t_1))\Delta t_2 \\
&\cdots \\
y(t_m) &= y(t_{m-1}) + f(t_{m-1}, y(t_{m-1}))\Delta t_m,
\end{aligned}$$

where $\Delta t_i = t_i - t_{i-1}$, $i = 1, \dots, m$. We expect that the finer partition D we choose, the better approximation we get (provided that f is a "well-behaved", e.g. continuous, function).

We now turn to the special case $f(t, y(t)) = A(t)y(t)$, where $A(t) \in \mathbf{R}^{n\times n}$ is a square matrix for every $t \in [a, b]$. The Euler method applied to the linear equation

$$\begin{aligned} y'(t) &= A(t)y(t) \\ y(a) &= y_0 \end{aligned} \tag{1.2.3}$$

yields

$$\begin{aligned} y(t_0) &= y_0, \\ y(t_1) &= (I + A(t_0)\Delta t_1)y(t_0) = (I + A(t_0)\Delta t_1)y_0, \\ y(t_2) &= (I + A(t_1)\Delta t_2)y(t_1) = (I + A(t_1)\Delta t_2)(I + A(t_0)\Delta t_1)y_0, \\ &\dots \\ y(t_m) &= (I + A(t_{m-1})\Delta t_m) \cdots (I + A(t_1)\Delta t_2)(I + A(t_0)\Delta t_1)y_0, \end{aligned}$$

where I denotes the identity matrix. Put

$$P(A, D) = (I + A(t_{m-1})\Delta t_k) \cdots (I + A(t_1)\Delta t_2)(I + A(t_0)\Delta t_1).$$

Provided the entires of A are continuous functions, it is possible to prove (as will be done in the following chapters) that, if $\nu(D) \to 0$ (where $\nu(D) = \max\{\Delta t_i, i = 1, \dots, m\}$), then $P(A, D)$ converges to a certain matrix; this matrix will be denoted by the symbol

$$\prod_a^b (I + A(x)\,\mathrm{d}x)$$

and will be called the left product integral of the matrix function A over the interval $[a, b]$. Moreover, the function

$$Y(t) = \prod_a^t (I + A(x)\,\mathrm{d}x)$$

satisfies

$$\begin{aligned} Y'(t) &= A(t)Y(t) \\ Y(a) &= I \end{aligned}$$

Consequently, the vector function

$$y(t) = \prod_a^t (I + A(x)\,\mathrm{d}x)\, y_0$$

is the solution of Equation (1.2.3).

1.3 Product integration in physics

The following example shows that product integration also finds applications outside mathematical analysis, particularly in fluid mechanics (a more general treatment is given in [DF]).

Consider a fluid whose motion is described by a function $S : [t_0, t_1] \times \mathbf{R}^3 \to \mathbf{R}^3$; the value $S(t, x)$ corresponds to the position (at the moment t) of the particle that was at position x at the moment t_0. Thus, for every $t \in [t_0, t_1]$, S can be viewed as an operator on $\mathbf{R}^3$; we emphasize this fact by writing $S(t)(x)$ instead of $S(t, x)$.

If x is a position of a certain particle at the moment t, it will move to $S(t + \Delta t) \cdot S(t)^{-1}(x)$ (where $\cdot$ denotes the composition of two operators) during the interval $[t, t + \Delta t]$. Consequently, its instantaneous velocity at the moment t is given by

$$V(t)(x) = \lim_{\Delta t \to 0} \frac{S(t + \Delta t) \cdot S(t)^{-1}(x) - x}{\Delta t} = \left(\lim_{\Delta t \to 0} \frac{S(t + \Delta t) \cdot S(t)^{-1} - I}{\Delta t} \right)(x),$$

where I denotes the identity operator. The velocity $V(t)$ is an operator on $\mathbf{R}^3$ for every $t \in [t_0, t_1]$; in the following chapters it will be called the left derivative of the operator S.

Given the velocity operator V, how to reconstruct the position operator S? For small Δt we have

$$S(t + \Delta t)(x) \doteq (I + V(t)\Delta t) \cdot S(t)(x).$$

If we choose a sufficiently fine partition $D : t_0 = u_0 < u_1 < \cdots < u_m = t$ of interval $[t_0, t]$, we obtain

$$S(t)(x) \doteq (I + V(u_{m-1})\Delta u_m) \cdots (I + V(u_0)\Delta u_1)(x),$$

where $\Delta u_i = u_i - u_{i-1}$, $i = 1, \ldots, m$. The above product (or composition) resembles the product encountered in the previous section. Indeed, passing to the limit $\nu(D) \to 0$, we see that S is the left product integral of operator V, i.e.

$$S(t) = \prod_{t_0}^{t} (I + V(u)\,\mathrm{d}u), \quad t \in [t_0, t_1].$$

In a certain sense, the left derivative and the left product integral are inverse operations.

1.4 Product integration in probability theory

Some results of probability theory can be elegantly expressed in the language of product integration. We present two examples concerning survival analysis and Markov processes; both are inspired by [Gil].

Example 1.4.1. Let T be a non-negative continuous random variable with distribution function $F(t) = P(T \le t)$ and probability density function $f(t) = F'(t)$.

For example, T can be interpreted as the service life of a certain component (or the length of life of a person etc.). The probability of failure in the interval $[t, t + \Delta t]$ is

$$P(t \le T \le t + \Delta t) = F(t + \Delta t) - F(t).$$

We remind that the survival function is defined as

$$S(t) = 1 - F(t) = P(T > t)$$

and the failure rate (or the hazard rate) is

$$a(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -\frac{S'(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t). \tag{1.4.1}$$

The name "failure rate" stems from the fact that

$$\lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t \,|\, T > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t)}{P(T > t)\Delta t} =$$

$$= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{S(t)\Delta t} = \frac{f(t)}{S(t)} = a(t),$$

i.e. for small Δt, the conditional probability of failure during the interval $[t, t + \Delta t]$ is approximately $a(t)\Delta t$.
Given the function a, Equation (1.4.1) tells us how to calculate S:

$$S(t) = \exp\left(-\int_0^t a(u)\,\mathrm{d}u\right). \tag{1.4.2}$$

We can also proceed in a different way: If we choose an arbitrary partition

$$D : 0 = t_0 < t_1 < \cdots < t_m = t,$$

then

$$S(t) = P(T > t) = P(T > t_0)P(T > t_1 \,|\, T > t_0) \cdots P(T > t_m \,|\, T > t_{m-1}) =$$

$$= \prod_{i=1}^{m} P(T > t_i \,|\, T > t_{i-1}) = \prod_{i=1}^{m} (1 - P(T \le t_i \,|\, T > t_{i-1})).$$

In case the partition is sufficiently fine, the last product is approximately equal to

$$\prod_{i=1}^{m} (1 - a(t_i)\Delta t_i).$$

This product is similar to the one used in the definition of left product integral, but the factors are reversed. Its limit for $\nu(D) \to 0$ is called the right product integral of the function $-a$ on interval $[0, t]$ and will be denoted by the symbol

$$S(t) = (1 - a(u)\,\mathrm{d}u) \prod_0^t. \tag{1.4.3}$$

Comparing Equations (1.4.2) and (1.4.3) we obtain the result

$$(1 - a(u)\,\mathrm{d}u)\prod_{0}^{t} = \exp\left(-\int_0^t a(u)\,\mathrm{d}u\right),$$

which will be proved in Chapter 2 (see Example 2.5.6). The product integral representation of S has the advantage that it can be intuitively viewed as the product of probabilities $1 - a(u)\,\mathrm{d}u$ that correspond to infinitesimal intervals of length $\mathrm{d}u$.

The last example corresponds in fact to a simple Markov process with two states s_1 ("the component is operating") and s_2 ("the component is broken"). The process starts in the state s_1 and goes over to the state s_2 at time T. We now generalize our calculation to Markov processes with more than two states; before that we recall the definition of a Markov process.
A stochastic process X on interval $[0,\infty)$ is a random function $t \mapsto X(t)$, where $X(t)$ is a random variable for every $t \in [0,\infty)$. We say that the process is in the state $X(t)$ at time t. A Markov process is a stochastic process such that the range of X is either finite or countably infinite and such that for every choice of numbers $n \in \mathbf{N}$, $n > 1$, $0 \le t_1 < t_2 < \cdots < t_n$, we have

$$P(X(t_n){=}x_n|X(t_{n-1}){=}x_{n-1},\ldots,X(t_1){=}x_1) = P(X(t_n){=}x_n|X(t_{n-1}){=}x_{n-1}),$$

where $x_1,\ldots,x_n$ are arbitrary states (i.e. values from the range of X). The above condition means that the conditional probability distribution of the process at time t_n depends only on the last observation at t_{n-1} and not on the whole history.

Example 1.4.2. Let $\{X(t); t \ge 0\}$ be a Markov process with a finite number of states $S = \{s_1,\ldots,s_n\}$. For example, we can imagine that $X(t)$ determines the number of patients in physician's waiting room (whose capacity is of course finite). Suppose that the limit

$$a_{ij}(t) = \lim_{\Delta t \to 0+} \frac{P(X(t+\Delta t) = s_j \mid X(t) = s_i)}{\Delta t}$$

exists for every $i,j = 1,\ldots,n$, $i \ne j$ and for every $t \in [0,\infty)$. The number $a_{ij}(t)$ is called the transition rate from state i to state j at time t. For sufficiently small Δt we have

$$P(X(t+\Delta t) = s_j \mid X(t) = s_i) \doteq a_{ij}(t)\Delta t, \quad i \ne j, \tag{1.4.4}$$

$$P(X(t+\Delta t) = s_i \mid X(t) = s_i) \doteq 1 - \sum_{j \ne i} a_{ij}(t)\Delta t. \tag{1.4.5}$$

We also define

$$a_{ii}(t) = -\sum_{j \ne i} a_{ij}(t), \quad i = 1,\ldots,n$$

and denote $A(t) = \{a_{ij}(t)\}_{i,j=1}^n$. Given the matrix A, we are interested in calculating the probabilities

$$p_i(t) = P(X(t) = s_i), \quad t \in [0,\infty),\ i = 1,\ldots,n,$$

and

$$p_{ij}(s,t) = P(X(t) = s_j \mid X(s) = s_i), \quad 0 \le s < t,\ i,j = 1,\dots,n.$$

The total probability theorem gives

$$p_j(t) = \sum_{i=1}^{n} p_i(0) p_{ij}(0,t).$$

The probabilities $p_i(0)$, $i = 1,\dots,n$ are usually given and it is thus sufficient to calculate the probabilities $p_{ij}(0,t)$, or generally $p_{ij}(s,t)$. Putting $P(s,t) = \{p_{ij}(s,t)\}_{i,j=1}^{n}$ we can rewrite Equations (1.4.4) and (1.4.5) to the matrix form

$$P(t, t+\Delta t) \doteq I + A(t)\Delta t \tag{1.4.6}$$

for sufficiently small Δt.

Using the total probability theorem once more we obtain

$$p_{ij}(s,u) = \sum_{k=1}^{n} p_{ik}(s,t) p_{kj}(t,u), \tag{1.4.7}$$

for $0 \le s < t < u$, $i,j = 1,\dots,n$. This is equivalent to the matrix equation

$$P(s,u) = P(s,t)P(t,u). \tag{1.4.8}$$

If we choose a sufficiently fine partition $s = u_0 < u_1 < \dots < u_m = t$ of interval $[s,t]$, then Equations (1.4.6) and (1.4.8) imply

$$P(s,t) = \prod_{i=1}^{m} P(u_{i-1}, u_i) \doteq \prod_{i=1}^{m} (I + A(u_i)\Delta u_i).$$

Passing to the limit for $\nu(D) \to 0$ we obtain a matrix which is called the right product integral of the function A over interval $[s,t]$:

$$P(s,t) = (I + A(u)\,\mathrm{d}u) \prod_{s}^{t}.$$

The last result can be again intuitively interpreted as the product of matrices $I + A(u)\,\mathrm{d}u$ which correspond to transition probabilities in the infinitesimal time intervals of length $\mathrm{d}u$.

Chapter 2

The origins of product integration

The notion of product integral has been introduced by Vito Volterra at the end of the 19th century. We start with a short biography of this eminent Italian mathematician and then proceed to discuss his work on product integration.

Vito Volterra was born at Ancona on 3rd May 1860. His father died two years later; Vito moved in with his mother Angelica to Alfonso, Angelica's brother, who supported them and was like a boy's father. Because of their financial situation, Angelica and Alfonso didn't want Vito to study his favourite subject, mathematics, at university, but eventually Edoardo Almagià, Angelica's cousin and a railroad engineer, helped to persuade them. An important role was also played by Volterra's teacher Ròiti, who secured him a place of assistant in a physical laboratory.

Vito Volterra[1]

In 1878 Volterra entered the University of Pisa; among his professors was the famous Ulisse Dini. In 1879 he passed the examination to Scuola Normale Superiore of Pisa. Under the influence of Enrico Betti, his interest shifted towards mathematical physics. In 1882 he offered a thesis on hydrodynamics, graduated doctor of physics and became Betti's assistent. Shortly after, in 1883, the young Volterra won the competition for the vacant chair of rational mechanics and was promoted

[1] Photo from [McT]

to professor of the University of Pisa. After Betti's death he took over his course in mathematical physics. In 1893 he moved to the University of Turin, but eventually settled in Rome in 1900. The same year he married Virginia Almagià (the daughter of Edoardo Almagià).

During the first quarter of the 20th century Volterra not only represented the leading figure of Italian mathematics, but also became involved in politics and was nominated a Senator of the Kingdom in 1905.

When Italy entered the world war in 1915, Volterra volunteered the Army Corps of Engineers and engaged himself in perfecting of airships and firing from them; he also promoted the collaboration with French and English scientists. After the end of the war he returned to scientific work and teaching at the university.

Volterra strongly opposed the Mussolini regime which came to power in 1922. As one of the few professors who refused to take an oath of loyalty imposed by the fascists in 1931, he was forced to leave the University of Rome and other scientific institutions. After then he spent a lot of time abroad (giving lectures e.g. in France, Spain, Czechoslovakia or Romania) and also at his country house in Ariccia. Volterra, who was of Jewish descent, was also affected by the antisemitic racial laws of 1938. Although he began to suffer from phlebitis, he still devoted himself actively to mathematics. He died in isolation on 11th October 1940 without a greater interest of Italian scientific community.

Despite the fact that Volterra is best known as a mathematician, he was a man of universal interests and devoted himself also to physics, biology and economy. His mathematical research often had origins in physical problems. Volterra was also an enthusiastic bibliophile and his collection, which reached nearly seven thousand volumes and is now deposited in the United States, included rare copies of scientific papers e.g. by Galileo, Brahe, Tartaglia, Fermat etc. The monograph [JG] contains a wealth of information about Volterra's life and times.

Volterra's name is closely connected with integral equations. He contributed the method of successive approximations for solving integral equations of the second kind, and also noted that an integral equation might be considered as a limiting case of a system of algebraic linear equations; this observation was later utilized by Ivar Fredholm (see also the introduction to Chapter 4).

His investigations in calculus of variations led him to the study of functionals (he called them "functions of lines"); in fact he built a complete calculus including the definitions of continuity, derivative and integral of a functional. Volterra's pioneering work on integral equations and functionals is often regarded as the dawn of functional analysis. An overview of his achievements in this field can be obtained from the book [VV5].

Volterra was also one of the founders of mathematical biology. The motivation came from his son-in-law Umberto D'Ancona, who was studying the statistics of Adriatic fishery. He posed to Volterra the problem of explaining the relative increase of predator fishes, as compared with their prey, during the period of First World War (see e.g. [MB]). Volterra interpreted this phenomenon with the help of mathematical models of struggle between two species; from mathematical point of view, the models were combinations of differential and integral equations. Volterra's correspondence concerning mathematical biology was published in the book [IG].

A more detailed description of Volterra's activites (his work on partial differential equations, theory of elasticity) can be found in the biographies [All, JG] and also in the books [IG, VV5]. An interesting account of Italian mathematics and its intertwining with politics in the first half of the 20th century is given in [GN].

2.1 Product integration in the work of Vito Volterra

Volterra's first work devoted to product integration [VV1] was published in 1887 and was written in Italian. It introduces the two basic concepts of the multiplicative calculus, namely the derivative of a matrix function and the product integral. The topics discussed in [VV1] are essentially the same as in Sections 2.3 to 2.6 of the present chapter. The publication [VV1] was followed by a second part [VV2] printed in 1902, which is concerned mainly with matrix functions of a complex variable. It includes results which are recapitulated in Sections 2.7 and 2.8, and also a treatment of product integration on Riemann surfaces. Volterra also published two short Italian notes, [VV3] from 1887 and [VV4] from 1888, which summarize the results of [VV1, VV2] but don't include proofs.

Volterra's final treatment of product integration is represented by the book *Opérations infinitésimales linéaires* [VH] written together with a Czech mathematician Bohuslav Hostinský. The publication appeared in the series *Collection de monographies sur la théorie des fonctions* directed by Émile Borel in 1938.

More than two hundred pages of [VH] are divided into eighteen chapters. The first fifteen chapters represent a French translation of [VV1, VV2] with only small changes and complements. The remaining three chapters, whose author is Bohuslav Hostinský, will be discussed in Chapter 4.

As Volterra notes in the book's preface, the publication of [VH] was motivated by the results obtained by Bohuslav Hostinský, as well as by an increased interest in matrix theory among mathematicians and physicists. As the bibliography of [VH] suggests, Volterra was already aware of the papers [LS1, LS2] by Ludwig Schlesinger, who linked up to Volterra's first works (see Chapter 3).

The book [VH] is rather difficult to read for contemporary mathematicians. One of the reasons is a somewhat cumbersome notation. For example, Volterra uses the same symbol to denote additive as well as multiplicative integration: The sign $\int$ applied to a matrix function denotes the product integral, while the same sign applied to a scalar function stands for the ordinary (additive) integral. Calculations with matrices are usually written out for individual entries, whereas using the matrix notation would have greatly simplified the proofs. Moreover, Volterra didn't hesitate to calculate with infinitesimal quantities, he interchanges the order of summation and integration or the order of partial derivatives without any justification etc. The conditions under which individual theorems hold (e.g. continuity or differentiability of the given functions) are often omitted and must be deduced from the proof. This is certainly surprising, since the rigorous foundations of mathematical analysis were already laid out at the end of the 19th century, and even Volterra contributed to them during his studies by providing an example of a function which is not Riemann integrable but has an antiderivative.

The following sections summarize Volterra's achievements in the field of product integration. Our discussion is based on the text from [VH], but the results are stated in the language of contemporary mathematics (with occasional comments on Volterra's notation). Proofs of most theorems are also included; they are generally based on Volterra's original proofs except for a few cases where his calculations with infinitesimal quantities were replaced by a different, rigorous argument.

2.2 Basic results of matrix theory

The first four chapters of the book [VH] recapitulate some basic results of matrix theory. Most of them are now taught in linear algebra courses and we repeat only some of them for reader's convenience, as we will refer to them in subsequent chapters.

Volterra refers to matrices as to substitutions, because they can be used to represent a linear change of variables. A composition of two substitutions then corresponds to multiplication of matrices: If

$$x_i' = \sum_{j=1}^n a_{ij} x_j \quad \text{and} \quad x_i'' = \sum_{j=1}^n b_{ij} x_j',$$

then

$$x_i'' = \sum_{j=1}^n c_{ij} x_j,$$

where

$$c_{ij} = \sum_{k=1}^n b_{ik} a_{kj}, \qquad (2.2.1).$$

We will use the symbol $\mathbf{R}^{n\times n}$ to denote the set of all square matrices with n rows and n columns. If

$$A = \{a_{ij}\}_{i,j=1}^n, \quad B = \{b_{ij}\}_{i,j=1}^n, \quad C = \{c_{ij}\}_{i,j=1}^n,$$

we can write Equation (2.2.1) in the form $C = B \cdot A$.

A matrix $A = \{a_{ij}\}_{i,j=1}^n$ is called regular if it has a nonzero determinant. If $i \in \{1, \ldots, n\}$, the theorem on expansion by minors gives

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \sum_{k=1}^n a_{ik} A_{ik}, \qquad (2.2.2)$$

where

$$A_{ik} = (-1)^{i+k} M_{ik}$$

is the so-called cofactor corresponding to minor M_{ik}; the minor is defined as the determinant of a matrix obtained from A by deleting the i-th row and k-th column. Since the determinant of a matrix with two or more identical rows is zero, it follows that

$$\sum_{k=1}^{n} a_{jk} A_{ik} = \delta_{ij} \det A. \tag{2.2.3}$$

for each pair of numbers $i, j \in \{1, \ldots, n\}$; recall that δ_{ij} is the Kronecker symbol

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

If we thus define the matrix

$$A^{-1} = \left\{ \frac{A_{ji}}{\det A} \right\}_{i,j=1}^{n},$$

then Equation (2.2.3) yields

$$AA^{-1} = I = A^{-1}A,$$

where

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is the identity matrix. The matrix A^{-1} is called the inverse of A.

The definition of matrix multiplication gives the following rule for multiplication of block matrices:

Theorem 2.2.1.[1] Consider a matrix that is partitioned into m^2 square blocks A_{ij} and a matrix partitioned into m^2 square blocks B_{ij} such that A_{ij} has the same dimensions as B_{ij} for every $i, j \in \{1, \ldots, m\}$. Then

$$\begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mm} \end{pmatrix} \begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{pmatrix} = \begin{pmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mm} \end{pmatrix},$$

where $C_{ik} = \sum_j A_{ij} B_{jk}$.

We will be often dealing with block diagonal matrices, i.e. with matrices of the form

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix}$$

[1] [VH], p. 27

composed of smaller square matrices $A_1, A_2, \ldots, A_m$; Volterra denotes such a matrix by the symbols

$$\{\underline{\overline{A_1 \cdot A_2 \cdots A_m}}\} \quad \text{or} \quad \left\{\underline{\overline{\prod_{i=1}^{m} A_i}}\right\},$$

but we don't follow his notation.

The following theorem expresses the fact that every square matrix can be transformed to a certain canonical form called the Jordan normal form. Volterra proves the theorem by induction on the dimension of the matrix; we refer the reader to any good linear algebra textbook.

Theorem 2.2.2.[1] To every matrix $A \in \mathbf{R}^{n\times n}$ there exist matrices $C, J \in \mathbf{R}^{n\times n}$ such that

$$A = C^{-1}JC$$

and J has the form

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_m \end{pmatrix}, \quad \text{where } J_i = \begin{pmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_i \end{pmatrix}$$

and $\{\lambda_1, \ldots, \lambda_m\}$ are (not necessarily distinct) eigenvalues of A.

Recall that if $A = C^{-1}BC$ for some regular matrix C, then the matrices A, B are called similar. Thus the previous theorem says that every square matrix is similar to a certain Jordan matrix.

The next two theorems concern the properties of block diagonal matrices and are simple consequences of Theorem 2.2.1.

Theorem 2.2.3. If A_i is a square matrix of the same dimensions as B_i for every $i \in \{1, \ldots, m\}$, then

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix} \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_m \end{pmatrix} = \begin{pmatrix} A_1B_1 & 0 & \cdots & 0 \\ 0 & A_2B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_mB_m \end{pmatrix}.$$

Theorem 2.2.4. The inverse of a block diagonal matrix composed of invertible matrices $A_1, \ldots, A_m$ is equal to

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix}^{-1} = \begin{pmatrix} A_1^{-1} & 0 & \cdots & 0 \\ 0 & A_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m^{-1} \end{pmatrix}.$$

[1] [VH], p. 20–24

2.3 Derivative of a matrix function

In this section we focus on first of the basic operations of Volterra's matrix calculus, which is the derivative of a matrix function.

A matrix function $A : [a, b] \to \mathbf{R}^{n\times n}$ will be called differentiable at a point $x \in (a, b)$ if all the entries a_{ij}, $i, j \in \{1, \dots, n\}$ of A are differentiable at x; in this case we denote

$$A'(x) = \{a'_{ij}(x)\}_{i,j=1}^{n}.$$

We also define $A'(x)$ for the endpoints $x = a$ and $x = b$ as the matrix of the corresponding one-sided derivatives (provided they exist).

Definition 2.3.1. Let $A : [a, b] \to \mathbf{R}^{n\times n}$ be a matrix function that is differentiable and regular at a point $x \in [a, b]$. We define the left derivative of A at x as

$$\frac{\mathrm{d}}{\mathrm{d}x}A(x) = A'(x)A^{-1}(x) = \lim_{\Delta x\to 0}\frac{A(x+\Delta x)A^{-1}(x) - I}{\Delta x}$$

and the right derivative of A at x as

$$A(x)\frac{\mathrm{d}}{\mathrm{d}x} = A^{-1}(x)A'(x) = \lim_{\Delta x\to 0}\frac{A^{-1}(x)A(x+\Delta x) - I}{\Delta x}.$$

Volterra doesn't use the matrix notation and instead writes out the individual entries:

$$\frac{\mathrm{d}}{\mathrm{d}x}\{a_{ij}\} = \left\{\lim_{\Delta x\to 0}\sum_{k=1}^{n}\frac{a_{ik}(x+\Delta x) - a_{ik}(x)}{\Delta x}A_{jk}(x)\right\},$$

$$\{a_{ik}\}\frac{\mathrm{d}}{\mathrm{d}x} = \left\{\lim_{\Delta x\to 0}\sum_{k=1}^{n}A_{ki}(x)\frac{a_{kj}(x+\Delta x) - a_{kj}(x)}{\Delta x}\right\},$$

where $\{A_{ji}\}_{i,j=1}^{n}$ denote the entries of A^{-1}. He also defines the left differential as

$$\mathrm{d}\{a_{ij}\} = A(x+\mathrm{d}x)A(x)^{-1} = I + A'(x)A(x)^{-1}\,\mathrm{d}x = \left\{\delta_{ij} + \sum_{k=1}^{n}a'_{ik}(x)A_{jk}(x)\,\mathrm{d}x\right\}$$

and the right differential as

$$\{a_{ij}\}\mathrm{d} = A(x)^{-1}A(x+\mathrm{d}x) = I + A(x)^{-1}A'(x)\,\mathrm{d}x = \left\{\delta_{ij} + \sum_{k=1}^{n}A_{ki}(x)a'_{kj}(x)\,\mathrm{d}x\right\},$$

where $\mathrm{d}x$ is an infinitesimal quantity. Both differentials are considered as matrices that differ infinitesimally from the identity matrix.

Volterra uses infinitesimal quantities without any scruples, which sometimes leads to very unreadable proofs. This is also the case of the following theorem; Volterra's justification has been replaced by a rigorous proof.

Theorem 2.3.2.[1] If $A, B : [a, b] \to \mathbf{R}^{n\times n}$ are differentiable and regular matrix functions at $x \in [a, b]$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}(AB) = \frac{\mathrm{d}}{\mathrm{d}x}A + A\left(\frac{\mathrm{d}}{\mathrm{d}x}B\right)A^{-1} = A\left(A\frac{\mathrm{d}}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}B\right)A^{-1},$$

$$(AB)\frac{\mathrm{d}}{\mathrm{d}x} = B\frac{\mathrm{d}}{\mathrm{d}x} + B^{-1}\left(A\frac{\mathrm{d}}{\mathrm{d}x}\right)B = B^{-1}\left(A\frac{\mathrm{d}}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}B\right)B,$$

where all derivatives are taken at the given point x.

Proof. The definition of the left derivative gives

$$\frac{\mathrm{d}}{\mathrm{d}x}(AB) = (AB)'(AB)^{-1} = (A'B + AB')B^{-1}A^{-1} = A'A^{-1} + AB'B^{-1}A^{-1},$$

where the expression on the right hand side is equal to

$$\frac{\mathrm{d}}{\mathrm{d}x}A + A\left(\frac{\mathrm{d}}{\mathrm{d}x}B\right)A^{-1},$$

but can be also transformed to the form

$$AA^{-1}A'A^{-1} + AB'B^{-1}A^{-1} = A\left(A\frac{\mathrm{d}}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}B\right)A^{-1}.$$

The second part is proved in a similar way. □

Corollary 2.3.3.[2] Consider a matrix function $A : [a, b] \to \mathbf{R}^{n\times n}$ that is differentiable and regular on $[a, b]$. Then for an arbitrary regular matrix $C \in \mathbf{R}^{n\times n}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}x}(AC) = \frac{\mathrm{d}}{\mathrm{d}x}A.$$

The corollary can be expressed like this: *The left derivative of a matrix function doesn't change, if the function is multiplied by a constant matrix from right.* It is also easy to prove a dual statement: *The right derivative of a matrix function doesn't change, if the function is multiplied by a constant matrix from left.* Symbolically written,

$$(CA)\frac{\mathrm{d}}{\mathrm{d}x} = A\frac{\mathrm{d}}{\mathrm{d}x}.$$

As Volterra notes, this is a general principle: Each statement concerning matrix functions remains true, if we replace all occurences of the word "left" by the word

[1] [VH], p. 43
[2] [VH], p. 39

"right" and vice versa. A precise formulation and justification of this duality principle is due to P. R. Masani and will be given in Chapter 5, Remark 5.2.2.

Theorem 2.3.4.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is differentiable and regular at $x \in [a,b]$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}(A^{-1}) = -A\frac{\mathrm{d}}{\mathrm{d}x}, \quad (A^{-1})\frac{\mathrm{d}}{\mathrm{d}x} = -\frac{\mathrm{d}}{\mathrm{d}x}A,$$

where all derivatives are taken at the given point x.

Proof. Differentiating the equation $AA^{-1} = I$ yields $A'A^{-1} + A(A^{-1})' = 0$, and consequently $(A^{-1})' = -A^{-1}A'A^{-1}$. The statement follows easily. □

Corollary 2.3.5.[2] If $A, B : [a,b] \to \mathbf{R}^{n\times n}$ are differentiable and regular matrix functions at $x \in [a,b]$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}(A^{-1}B) = A^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}B - \frac{\mathrm{d}}{\mathrm{d}x}A\right)A,$$

$$(AB^{-1})\frac{\mathrm{d}}{\mathrm{d}x} = B\left(A\frac{\mathrm{d}}{\mathrm{d}x} - B\frac{\mathrm{d}}{\mathrm{d}x}\right)B^{-1},$$

where all derivatives are taken at the given point x.

Proof. A simple consequence of Theorems 2.3.2 and 2.3.4. □

Theorem 2.3.6.[3] Consider functions $A, B : [a,b] \to \mathbf{R}^{n\times n}$ that are differentiable and regular on $[a,b]$. If

$$\frac{\mathrm{d}}{\mathrm{d}x}A = \frac{\mathrm{d}}{\mathrm{d}x}B$$

on $[a,b]$, then there exists a matrix $C \in \mathbf{R}^{n\times n}$ such that $B(x) = A(x)C$ for every $x \in [a,b]$.

Proof. Define $C(x) = A^{-1}(x)B(x)$ for $x \in [a,b]$. Corollary 2.3.5 gives

$$\frac{\mathrm{d}}{\mathrm{d}x}C = A^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}B - \frac{\mathrm{d}}{\mathrm{d}x}A\right)A = 0,$$

which implies that $0 = C'(x)$ for every $x \in [a,b]$. This means that C is a constant function. □

A combination of Theorem 2.3.6 and Corollary 2.3.3 leads to the following statement: *Two matrix functions have the same left derivative on a given interval, if and only if one of the functions is obtained by multiplying the other by a constant matrix from right.* This is the fundamental theorem of Volterra's differential calculus; a dual statement is again obtained by interchanging the words "left" and

[1] [VH], p. 41
[2] [VH], p. 44
[3] [VH], p. 46

"right". Both statements represent an analogy of the well-known theorem: *Two functions have the same derivative if and only if they differ by a constant.*

Theorem 2.3.7.[1] Consider a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$ that is differentiable and regular on $[a,b]$. Then for an arbitrary regular matrix $C \in \mathbf{R}^{n\times n}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}x}(CA) = C\left(\frac{\mathrm{d}}{\mathrm{d}x}A\right)C^{-1}.$$

Proof. A simple consequence of Theorem 2.3.2. □

Theorem 2.3.8. Let $A : [a,b] \to \mathbf{R}^{n\times n}$ be a matrix function of the form

$$A(x) = \begin{pmatrix} A_1(x) & 0 & \cdots & 0 \\ 0 & A_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k(x) \end{pmatrix},$$

where $A_1, \ldots, A_k$ are square matrix functions. If

$$\frac{\mathrm{d}}{\mathrm{d}x}A_i(x) = B_i(x), \qquad i = 1, \ldots, k,$$

then

$$\frac{\mathrm{d}}{\mathrm{d}x}A(x) = \begin{pmatrix} B_1(x) & 0 & \cdots & 0 \\ 0 & B_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k(x) \end{pmatrix}.$$

Proof. The statement follows from the definition of left derivative and from Theorems 2.2.3 and 2.2.4. □

2.4 Product integral of a matrix function

Consider a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$ with entries $\{a_{ij}\}_{i,j=1}^{n}$. For every tagged partition

$$D:\ a = t_0 \le \xi_1 \le t_1 \le \xi_2 \le \cdots \le t_{m-1} \le \xi_m \le t_m = b$$

of interval $[a,b]$ with division points t_i and tags ξ_i we denote

$$\Delta t_i = t_i - t_{i-1},\ \ i = 1, \ldots, m,$$

$$\nu(D) = \max_{1\le i\le m} \Delta t_i.$$

[1] [VH], p. 41

We also put

$$P(A,D)=\prod_{i=m}^{1}(I+A(\xi_i)\Delta t_i)=(I+A(\xi_m)\Delta t_m)\cdots(I+A(\xi_1)\Delta t_1),$$

$$P^*(A,D)=\prod_{i=1}^{m}(I+A(\xi_i)\Delta t_i)=(I+A(\xi_1)\Delta t_1)\cdots(I+A(\xi_m)\Delta t_m).$$

Volterra now defines the left integral of A as the matrix

$$\int_a^b \{a_{ij}\}=\lim_{\nu(D)\to 0}P(A,D)$$

(in case the limit exists) and the right integral as

$$\{a_{ij}\}\int_a^b=\lim_{\nu(D)\to 0}P^*(A,D)$$

(again if the limit exists). Volterra isn't very precise about the meaning of the limit taken with respect to partitions; we make the following agreement:

If $M(D)$ is a matrix which is dependent on the choice of a tagged partition D of interval $[a,b]$, then the equality

$$\lim_{\nu(D)\to 0}M(D)=M$$

means that for every $\varepsilon>0$ there exists $\delta>0$ such that $|M(D)_{ij}-M_{ij}|<\varepsilon$ for every tagged partition D of interval $[a,b]$ satisfying $\nu(D)<\delta$ and for $i,j=1,\dots,n$.

The following definition also introduces a different notation to better distinguish between ordinary and product integrals.

Definition 2.4.1. Consider function $A:[a,b]\to\mathbf{R}^{n\times n}$. If the limit

$$\lim_{\nu(D)\to 0}P(A,D)\quad\text{or}\quad\lim_{\nu(D)\to 0}P^*(A,D)$$

exists, it is called the left (or right) product integral of A over the interval $[a,b]$. We use the notation

$$\prod_a^b(I+A(t)\,\mathrm{d}t)=\lim_{\nu(D)\to 0}P(A,D)$$

for the left product integral and

$$(I+A(t)\,\mathrm{d}t)\prod_a^b=\lim_{\nu(D)\to 0}P^*(A,D)$$

for the right product integral.

We note that in the case when the upper limit of integration coincides with the lower limit, then

$$\prod_a^a (I + A(t)\,\mathrm{d}t) = (I + A(t)\,\mathrm{d}t)\prod_a^a = I.$$

In the subsequent text we use the following convention: A function $A : [a,b] \to \mathbf{R}^{n\times n}$ is called Riemann integrable, if its entries a_{ij} are Riemann integrable functions on $[a,b]$. In this case we put

$$\int_a^b A(t)\,\mathrm{d}t = \left\{\int_a^b a_{ij}(t)\,\mathrm{d}t\right\}_{i,j=1}^n.$$

We will often encounter integrals of the type

$$\int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

These integrals should be interpreted as iterated integrals, where $x_k \in [a,b]$ and $x_i \in [a, x_{i+1}]$ for $i \in \{1,\dots,k-1\}$.

Lemma 2.4.2. Let $A : [a,b] \to \mathbf{R}^{n\times n}$ be a Riemann integrable function such that $A(x)A(y) = A(y)A(x)$ for every $x, y \in [a,b]$. Then

$$\int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k =$$

$$= \frac{1}{k!}\int_a^b \int_a^b \cdots \int_a^b A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

for every $k \in \mathbf{N}$.

Proof. If $P(k)$ denotes all permutations of the set $\{1,\dots,k\}$ and

$$M_\pi = \{(x_1,\dots,x_k) \in [a,b]^k;\ x_{\pi(1)} \le x_{\pi(2)} \le \cdots \le x_{\pi(k)}\}$$

for every $\pi \in P(k)$, then

$$\int_a^b \int_a^b \cdots \int_a^b A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k =$$

$$= \sum_{\pi\in P(k)} \int\!\!\int \cdots \int_{M_\pi} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

The assumption of commutativity implies

$$\int\!\!\int \cdots \int_{M_\pi} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k = \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

for every permutation $\pi \in P(k)$, which completes the proof. $\square$

Theorem 2.4.3.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then both product integrals exist and

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{k=1}^{\infty} \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k,$$

$$(I + A(x)\,\mathrm{d}x) \prod_a^b = I + \sum_{k=1}^{\infty} \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_1)\cdots A(x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

Proof. Volterra's proof goes as follows: Expanding the product $P(A,D)$ gives

$$\prod_{i=m}^{1} (I + A(\xi_i)\Delta t_i) = I + \sum_{k=1}^{m} \left(\sum_{1\le i_1 < \cdots < i_k \le m} A(\xi_{i_k})\cdots A(\xi_{i_1})\Delta t_{i_1}\cdots \Delta t_{i_k} \right).$$

Volterra now argues that for $\nu(D) \to 0$ we obtain

$$\sum_{1\le i_1 \le m} A(\xi_{i_1})\Delta t_{i_1} \to \int_a^b A(x_1)\,\mathrm{d}x_1,$$

$$\sum_{1\le i_1 < i_2 \le m} A(\xi_{i_2})A(\xi_{i_1})\Delta t_{i_1}\Delta t_{i_2} \to \int_a^b \int_a^{x_2} A(x_2)A(x_1)\,\mathrm{d}x_1\,\mathrm{d}x_2,$$

and generally

$$\sum_{1\le i_1 < \cdots < i_k \le m} A(\xi_{i_k})\cdots A(\xi_{i_1})\Delta t_{i_1}\cdots \Delta t_{i_k} \to \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

for every $k \in \{1,\ldots,m\}$. Using the fact that $m \to \infty$ for $\nu(D) \to 0$, Volterra arrived at the result

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{k=1}^{\infty} \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

The proof for the right product integral is carried out in a similar way. $\square$

The infinite series expressing the value of the product integrals are often referred to as the Peano series. They were discussed by Giuseppe Peano in his paper [GP] from 1888 dealing with systems of linear differential equations.

Remark 2.4.4. The proof of Theorem 2.4.3 given by Volterra is somewhat unsatisfactory. First, he didn't justify that

$$\sum_{1\le i_1 < \cdots < i_k \le m} A(\xi_{i_k})\cdots A(\xi_{i_1})\Delta t_{i_1}\cdots \Delta t_{i_k} \to \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

[1] [VH], p. 49–52

for $\nu(D) \to 0$. This can be done as follows: Define

$$X^k = \{(x_1, \dots, x_k) \in \mathbf{R}^k; a \le x_1 < x_2 < \dots < x_k \le b\}$$

and let χ^k be the characteristic function of the set X^k. Then

$$\lim_{\nu(D)\to 0} \sum_{1\le i_1<\dots<i_k\le m} A(\xi_{i_k}) \cdots A(\xi_{i_1}) \Delta t_{i_1} \cdots \Delta t_{i_k} =$$

$$= \lim_{\nu(D)\to 0} \sum_{i_1,\dots,i_k=1}^{m} A(\xi_{i_k}) \cdots A(\xi_{i_1}) \chi(\xi_{i_1}, \dots, \xi_{i_k}) \Delta t_{i_1} \cdots \Delta t_{i_k} =$$

$$= \int_a^b \int_a^b \cdots \int_a^b A(x_k) \cdots A(x_1) \chi(x_1, \dots, x_k) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k =$$

$$= \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k) \cdots A(x_1) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

The second problem is that Volterra didn't explain the equality

$$\lim_{\nu(D)\to 0} \left(I + \sum_{k=1}^{m} \sum_{1\le i_1<\dots<i_k\le m} A(\xi_{i_k}) \cdots A(\xi_{i_1}) \Delta t_{i_1} \cdots \Delta t_{i_k} \right) =$$

$$I + \sum_{k=1}^{\infty} \lim_{\nu(D)\to 0} \sum_{1\le i_1<\dots<i_k\le m} A(\xi_{i_k}) \cdots A(\xi_{i_1}) \Delta t_{i_1} \cdots \Delta t_{i_k}.$$

We postpone its justification to Chapter 5, Lemma 5.5.9.

Theorem 2.4.5.[1] If $A : [a, b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then the infinite series

$$\prod_a^x (I + A(t)\, \mathrm{d}t) = I + \sum_{k=1}^{\infty} \int_a^x \int_a^{x_k} \cdots \int_a^{x_2} A(x_k) \cdots A(x_1) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k,$$

$$(I + A(t)\, \mathrm{d}t) \prod_a^x = I + \sum_{k=1}^{\infty} \int_a^x \int_a^{x_k} \cdots \int_a^{x_2} A(x_1) \cdots A(x_k) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k$$

converge absolutely and uniformly for $x \in [a, b]$.

Proof. We give only the proof for the first series: Its sum is a matrix whose (i, j)-th entry is the number

$$\sum_{k=1}^{\infty} \left(\sum_{l_1,\dots,l_{k-1}=1}^{n} \int_a^x \int_a^{x_k} \cdots \int_a^{x_2} a_{i,l_1}(x_k) \cdots a_{l_{k-1},j}(x_1) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k \right). \qquad (2.4.1)$$

[1] [VH], p. 51–52

The functions a_{ij} are Riemann integrable, therefore bounded: There exists a positive number $M \in \mathbf{R}$ such that $|a_{ij}(t)| \le M$ for $i, j \in \{1, \ldots, n\}$ and $t \in [a, b]$. Using Lemma 2.4.2 we obtain the estimate

$$\left| \sum_{l_1,\ldots,l_{k-1}=1}^{n} \int_a^x \int_a^{x_k} \cdots \int_a^{x_2} a_{i,l_1}(x_k) \cdots a_{l_{k-1},j}(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k \right| \le$$

$$\le n^{k-1} M^k \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} \mathrm{d}x_1 \cdots \mathrm{d}x_k = \frac{1}{n} \frac{(nM(b-a))^k}{k!}$$

for every $x \in [a, b]$. Since

$$\sum_{k=1}^{\infty} \frac{1}{n} \frac{(nM(b-a))^k}{k!} = \frac{1}{n} e^{nM(b-a)},$$

we see that (according to the Weierstrass M-test) the infinite series (2.4.1) converges uniformly and absolutely on $[a, b]$. □

Theorem 2.4.6.[1] If $A : [a, c] \to \mathbf{R}^{n \times n}$ is a Riemann integrable function, then

$$\prod_a^c (I + A(x)\,\mathrm{d}x) = \prod_b^c (I + A(x)\,\mathrm{d}x) \cdot \prod_a^b (I + A(x)\,\mathrm{d}x)$$

and

$$(I + A(x)\,\mathrm{d}x) \prod_a^c = (I + A(x)\,\mathrm{d}x) \prod_a^b \cdot (I + A(x)\,\mathrm{d}x) \prod_b^c$$

for every $c \in [a, b]$.

Proof. Take two sequences of tagged partitions $\{D_k^1\}_{k=1}^{\infty}$ of interval $[a, b]$ and $\{D_k^2\}_{k=1}^{\infty}$ of interval $[b, c]$ such that

$$\lim_{k\to\infty} \nu(D_k^1) = \lim_{k\to\infty} \nu(D_k^2) = 0.$$

If we put $D_k = D_k^1 \cup D_k^2$, we obtain a sequence of tagged partitions $\{D_k\}_{k=1}^{\infty}$ of interval $[a, c]$ such that $\lim_{k\to\infty} \nu(D_k) = 0$. Consequently

$$\prod_a^c (I + A(x)\,\mathrm{d}x) = \lim_{k\to\infty} P(A, D_k) = \lim_{k\to\infty} P(A, D_k^2) \cdot \lim_{k\to\infty} P(A, D_k^1) =$$

$$= \prod_b^c (I + A(x)\,\mathrm{d}x) \cdot \prod_a^b (I + A(x)\,\mathrm{d}x).$$

The second part is proved in the same way. □

[1] [VH], p. 54–56

Remark 2.4.7. Volterra also offers a different proof of Theorem 2.4.6, which goes as follows[1]: Denote

$$D(i) = \{x \in \mathbf{R}^i;\ a \le x_1 \le \cdots \le x_i \le c\},$$

$$D(j,i) = \{x \in \mathbf{R}^i;\ a \le x_1 \le \cdots \le x_j \le b \le x_{j+1} \le \cdots \le x_i \le c\}$$

for each pair of numbers $i \in \mathbf{N}$ and $j \in \{0, \ldots, i\}$. Clearly

$$D(i) = D(0,i) \cup \left(\bigcup_{j=1}^{i-1} D(i-j, i-j) \times D(0,j) \right) \cup D(i,i) \tag{2.4.2}$$

for every $i \in \mathbf{N}$. We have

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} \int_{D(i,i)} A(x_i) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_i,$$

$$\prod_b^c (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} \int_{D(0,i)} A(x_i) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_i.$$

Since both infinite series converge absolutely, their product is equal to the Cauchy product:

$$\prod_b^c (I + A(x)\,\mathrm{d}x) \prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} \int_{D(i,i)} A(x_i) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_i +$$

$$+ \sum_{i=1}^{\infty} \int_{D(0,i)} A(x_i) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_i +$$

$$+ \sum_{i=1}^{\infty} \left(\sum_{j=1}^{i-1} \left(\int_{D(0,j)} A(x_j) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_j \right) \cdot \right.$$

$$\left. \cdot \left(\int_{D(i-j,i-j)} A(x_{i-j}) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_{i-j} \right) \right) =$$

$$= I + \sum_{i=1}^{\infty} \int_{D(i)} A(x_i) \cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_i = \prod_a^c (I + A(x)\,\mathrm{d}x)$$

(we have used Equation (2.4.2)).

If a and b are two real numbers such that $a < b$, we usually define

$$\int_b^a f = -\int_a^b f.$$

[1] [VH], p. 54–56

The following definition assigns a meaning to product integral whose lower limit is greater than its upper limit.

Definition 2.4.8. For any function $A:[a,b]\to\mathbf{R}^{n\times n}$ we define

$$\prod_b^a (I+A(t)\,\mathrm{d}t)=\lim_{\nu(D)\to 0}\prod_{i=1}^m (I-A(\xi_i)\Delta t_i)=(I-A(t)\,\mathrm{d}t)\prod_a^b$$

and

$$(I+A(t)\,\mathrm{d}t)\prod_b^a=\lim_{\nu(D)\to 0}\prod_{i=m}^1 (I-A(\xi_i)\Delta t_i)=\prod_a^b (I-A(t)\,\mathrm{d}t),$$

provided that the integrals on the right hand sides exist.

Corollary 2.4.9. If $A:[a,b]\to\mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\prod_b^a (I+A(t)\,\mathrm{d}t)=I+\sum_{k=1}^{\infty}(-1)^k\int_a^b\int_a^{x_k}\cdots\int_a^{x_2}A(x_1)\cdots A(x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k,$$

$$(I+A(t)\,\mathrm{d}t)\prod_b^a=I+\sum_{k=1}^{\infty}(-1)^k\int_a^b\int_a^{x_k}\cdots\int_a^{x_2}A(x_k)\cdots A(x_1)\,\mathrm{d}x_1\cdots\mathrm{d}x_k.$$

The following statement represents a generalized version of Theorem 2.4.6.

Theorem 2.4.10.[1] If $A:[p,q]\to\mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\prod_a^c (I+A(x)\,\mathrm{d}x)=\prod_b^c (I+A(x)\,\mathrm{d}x)\cdot\prod_a^b (I+A(x)\,\mathrm{d}x),$$

$$(I+A(x)\,\mathrm{d}x)\prod_a^c=(I+A(x)\,\mathrm{d}x)\prod_a^b\cdot(I+A(x)\,\mathrm{d}x)\prod_b^c$$

for every $a,b,c\in[p,q]$.

Proof. If $a\le b\le c$, then the statement reduces to Theorem 2.4.6. Let's have a look at the case $b<a=c$: Denote

$$E(j,i)=\{x\in\mathbf{R}^i;\ b\le x_1\le\cdots\le x_j\le a\text{ and }a\ge x_{j+1}\ge\cdots\ge x_i\ge b\}$$

for each pair of numbers $i\in\mathbf{N}$ and $j\in\{0,\ldots,i\}$. A simple observation reveals that

$$E(j,i)=E(j,j)\times E(0,i-j)\tag{2.4.3}$$

for every $i\in\mathbf{N}$ and $j\in\{1,\ldots,i-1\}$. We also assert that

$$E(0,i)\cup E(2,i)\cup\cdots=E(1,i)\cup E(3,i)\cup\cdots\tag{2.4.4}$$

[1] [VH], p. 56–58

for every $i \in \mathbf{N}$. Indeed, if $x \in \mathbf{R}^i$ is a member of the union on the left side, then $x \in E(2k, i)$ for some k. If $2k < i$ and $x_{2k} \le x_{2k+1}$, then $x \in E(2k+1, i)$. If $2k = i$, or $2k < i$ and $x_{2k+1} < x_{2k}$, then $x \in E(2k-1, i)$. In any case, x is a member of the union on the right side; the reverse inclusion is proved similarly.

Now, the Peano series expansions might be written as

$$\prod_b^a (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} \int_{E(i,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i,$$

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} (-1)^i \int_{E(0,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i$$

(we have used Corollary 2.4.9). Since both infinite series converge absolutely, their product is equal to the Cauchy product:

$$\prod_b^a (I + A(x)\,\mathrm{d}x) \cdot \prod_a^b (I + A(x)\,\mathrm{d}x) = I + \sum_{i=1}^{\infty} \int_{E(i,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i +$$

$$+ \sum_{i=1}^{\infty} (-1)^i \int_{E(0,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i +$$

$$+ \sum_{i=1}^{\infty} \left(\sum_{j=1}^{i-1} \left(\int_{E(j,j)} A(x_1)\cdots A(x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k \right) \cdot \right.$$

$$\left. \cdot \left((-1)^{i-j} \int_{E(0,i-j)} A(x_1)\cdots A(x_{i-j})\,\mathrm{d}x_1\cdots\mathrm{d}x_{i-j} \right) \right) =$$

$$= I + \sum_{i=1}^{\infty} \sum_{j=0}^{i} (-1)^{i-j} \int_{E(j,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i,$$

where the last equality is a consequence of Equation (2.4.3). Equation (2.4.4) implies

$$\sum_{j=0}^{i} (-1)^{i-j} \int_{E(j,i)} A(x_1)\cdots A(x_i)\,\mathrm{d}x_1\cdots\mathrm{d}x_i = 0$$

for every positive number i, which proves that

$$\prod_b^a (I + A(x)\,\mathrm{d}x) \cdot \prod_a^b (I + A(x)\,\mathrm{d}x) = I.$$

We see that our statement is true is even in the case $b > a = c$.

The remaining cases are now simple to check: For example, if $a < c < b$, then

$$\prod_a^c (I + A(x)\,\mathrm{d}x) = \prod_b^c (I + A(x)\,\mathrm{d}x) \cdot \prod_c^b (I + A(x)\,\mathrm{d}x) \cdot \prod_a^c (I + A(x)\,\mathrm{d}x) =$$

$$= (I + A(x)\,\mathrm{d}x)\prod_a^b \cdot (I + A(x)\,\mathrm{d}x)\prod_b^c .$$

The prove the second part, we calculate

$$(I + A(x)\,\mathrm{d}x)\prod_a^c = \prod_c^a (I - A(x)\,\mathrm{d}x) = \prod_b^a (I - A(x)\,\mathrm{d}x) \cdot \prod_c^b (I - A(x)\,\mathrm{d}x) =$$

$$= (I + A(x)\,\mathrm{d}x)\prod_a^b \cdot (I + A(x)\,\mathrm{d}x)\prod_b^c .$$

$\square$

Corollary 2.4.11. If $A : [a, b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\prod_b^a (I + A(x)\,\mathrm{d}x) = \left(\prod_a^b (I + A(x)\,\mathrm{d}x)\right)^{-1},$$

$$(I + A(x)\,\mathrm{d}x)\prod_b^a = \left((I + A(x)\,\mathrm{d}x)\prod_a^b\right)^{-1}.$$

Theorem 2.4.12.[1] If $A : [a, b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then the functions

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t),$$

$$Z(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x$$

satisfy the integral equations

$$Y(x) = I + \int_a^x A(t)Y(t)\,\mathrm{d}t,$$

$$Z(x) = I + \int_a^x Z(t)A(t)\,\mathrm{d}t$$

for every $x \in [a, b]$.

Proof. Theorem 2.4.3 implies

$$A(t)Y(t) = A(t) + \sum_{k=1}^{\infty} \int_a^t \int_a^{x_k} \cdots \int_a^{x_2} A(t)A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k. \qquad (2.4.5)$$

1 [VH], p. 52–53

The Peano series converges uniformly and the entries of A are bounded, therefore the series (2.4.5) also converges uniformly for $t \in [a,b]$ and might be integrated term by term to obtain

$$\int_a^x A(t)Y(t)\,\mathrm{d}t = \int_a^x A(t)\,\mathrm{d}t +$$

$$+\sum_{k=1}^{\infty} \int_a^x \int_a^t \int_a^{x_k} \cdots \int_a^{x_2} A(t)A(x_k)\cdots A(x_1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k\,\mathrm{d}t = Y(x) - I.$$

The other integral equation is deduced similarly. □

2.5 Continuous matrix functions

Volterra is now ready to state and prove the fundamental theorem of calculus for product integral. Recall that the ordinary fundamental theorem has two parts:
1) If f is a continuous function on $[a,b]$, then the function $F(x) = \int_a^x f(t)\,\mathrm{d}t$ satisfies $F'(x) = f(x)$ for every $x \in [a,b]$.
2) If f is a continuous function on $[a,b]$ and F its antiderivative, then

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a).$$

The function $\int_a^x f(t)\,\mathrm{d}t$ is usually referred to as the indefinite integral of f; similarly, the functions $\prod_a^x (I + A(t)\,\mathrm{d}t)$ and $(I + A(t)\,\mathrm{d}t)\prod_a^x$ are called the indefinite product integrals of A.

Before proceeding to the fundamental theorem we make the following agreement: A matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$ is called continuous, if the entries a_{ij} of A are continuous functions on $[a,b]$.

Theorem 2.5.1.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuous matrix function, then the indefinite product integrals

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t) \quad \text{and} \quad Z(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x$$

satisfy the equations

$$Y'(x) = A(x)Y(x),$$
$$Z'(x) = Z(x)A(x)$$

for every $x \in [a,b]$.

Proof. The required statement is easily deduced by differentiating the integral equations obtained in Theorem 2.4.12. □

[1] [VH], p. 60–61

The differential equations from the previous theorem can be rewritten in the form

$$\frac{\mathrm{d}}{\mathrm{d}x}\prod_a^x (I + A(t)\,\mathrm{d}t) = A(x), \quad (I + A(t)\,\mathrm{d}t)\prod_a^x \frac{\mathrm{d}}{\mathrm{d}x} = A(x),$$

which closely resembles the first part of the ordinary fundamental theorem. We see that the left (or right) derivative is in a certain sense inverse operation to the left (or right) product integral.

Remark 2.5.2. A function $Y : [a,b] \to \mathbf{R}^{n\times n}$ is a solution of the equation

$$Y'(x) = A(x)Y(x), \quad x \in [a,b]$$

and satisfies $Y(a) = I$ if and only if Y solves the integral equation

$$Y(x) = I + \int_a^x A(t)Y(t)\,\mathrm{d}t, \quad x \in [a,b]. \tag{2.5.1}$$

This is a special type of equation of the form

$$y(x) = f(x) + \int_a^x K(x,t)y(t)\,\mathrm{d}t,$$

which is today called the Volterra's integral equation of the second kind. Volterra proved (see e.g. [Kl, VV4]) that such equations may be solved by the method of successive approximations; in case of Equation (2.5.1) we obtain the solution

$$Y(x) = I + \sum_{k=1}^{\infty} \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_k)\cdots A(x_1)\,\mathrm{d}x_1\cdots\mathrm{d}x_k,$$

which is exactly the Peano series.

Theorem 2.5.3.[1] Consider a continuous matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$. If there exists a function $Y : [a,b] \to \mathbf{R}^{n\times n}$ such that

$$\frac{\mathrm{d}}{\mathrm{d}x} Y(x) = A(x)$$

for every $x \in [a,b]$, then

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = Y(b)Y(a)^{-1}.$$

Similarly, if there exists a function $Z : [a,b] \to \mathbf{R}^{n\times n}$ such that

$$Z(x)\frac{\mathrm{d}}{\mathrm{d}x} = A(x)$$

[1] [VH], p. 62–63

for every $x \in [a,b]$, then

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = Z(a)^{-1}Z(b).$$

Proof. We prove the first part: The functions $\prod_a^x (I + A(t)\,\mathrm{d}t)$ and $Y(x)$ have the same left derivative for every $x \in [a,b]$. Theorem 2.3.6 implies the existence of a matrix C such that

$$\prod_a^x (I + A(t)\,\mathrm{d}t) = Y(x)C$$

for every $x \in [a,b]$. Substituting $x = a$ yields $C = Y(a)^{-1}$. □

Theorem 2.5.4.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuous function, then the function

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t)$$

is the fundamental matrix of the system of differential equations

$$y_i'(x) = \sum_{j=1}^n a_{ij}(x)y_j(x), \quad i = 1,\dots,n. \tag{2.5.2}$$

Proof. Let y^k denote the k-th column of Y, i.e.

$$y^k(x) = \prod_a^x (I + A(t)\,\mathrm{d}t)\cdot e_k,$$

where e_k is the k-th vector from the canonical basis of $\mathbf{R}^n$. Theorem 2.5.1 implies that each of the vector functions y^k, $k = 1,\dots,n$ yields a solution of the system (2.5.2). Since $y^k(a) = e_k$, the system of functions $\{y^k\}_{k=1}^n$ is linearly independent and represents a fundamental set of solutions of the system (2.5.2). □

Example 2.5.5.[2] Volterra now shows the familiar method of converting a linear differential equation of the n-th order

$$y^{(n)}(x) = p_1(x)y^{n-1}(x) + p_2(x)y^{n-2}(x) + \cdots + p_n(x)y(x)$$

to a system of equations of the first order. If we introduce the functions $z_0 = y$, $z_1 = z_0'$, $z_2 = z_1', \dots, z_{n-1} = z_{n-2}'$, then the above given n-th order equation is equivalent to the system of equations written in matrix form as

$$\begin{pmatrix} z_0' \\ z_1' \\ \vdots \\ z_{n-2}' \\ z_{n-1}' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ p_n & p_{n-1} & p_{n-2} & \cdots & p_1 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{pmatrix}.$$

[1] [VH], p. 69
[2] [VH], p. 70

The fundamental matrix of this system can be calculated using the product integral and the solution of the original equation (which corresponds to the function z_0) is represented by the first column of the matrix (we obtain a set of n linearly independent solutions).

Example 2.5.6. If $n = 1$, then the function $A : [a, b] \to \mathbf{R}^{n\times n}$ is in fact a scalar function, and we usually write $\prod_a^b(1 + A(t)\,\mathrm{d}t)$ instead of $\prod_a^b(I + A(t)\,\mathrm{d}t)$. Using Theorem 2.4.3 and Lemma 2.4.2 we obtain

$$y(x) = \prod_a^x(1 + A(t)\,\mathrm{d}t) = 1 + \sum_{k=1}^{\infty}\frac{1}{k!}\left(\int_a^x A(t)\,\mathrm{d}t\right)^k = \exp\left(\int_a^x A(t)\,\mathrm{d}t\right),$$

which is indeed a solution of the differential equation $y'(x) = A(x)y(x)$ and satisfies $y(a) = 1$.

Example 2.5.7.[1] Recall that if $A \in \mathbf{R}^{n\times n}$, then the exponential of A is defined as

$$\exp A = \sum_{k=0}^{\infty}\frac{A^k}{k!} \qquad (2.5.3).$$

The fundamental matrix of the system of equations

$$y_i'(x) = \sum_{j=1}^{n} a_{ij}y_j(x), \quad i = 1, \dots, n$$

is given by

$$Y(x) = \prod_a^x(I + A\,\mathrm{d}t) = I + \sum_{k=1}^{\infty}\int_a^b\int_a^{x_k}\cdots\int_a^{x_2} A^k\,\mathrm{d}x_1\cdots\mathrm{d}x_k =$$

$$I + \sum_{k=1}^{\infty}\frac{(x-a)^kA^k}{k!} = e^{(x-a)A}$$

(we have used Theorem 2.4.3 and Lemma 2.4.2), which is a well-known result from the theory of differential equations. We also remark that a similar calculation leads to the relation

$$(I + A\,\mathrm{d}t)\prod_a^x = e^{(x-a)A}, \quad x \in [a, b].$$

Example 2.5.8.[2] Volterra is also interested in actually calculating the matrix $e^{A(x-a)}$. Convert A to the Jordan normal form

$$A = C^{-1}\begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix} C, \quad \text{where } J_i = \begin{pmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_i \end{pmatrix}$$

[1] [VH], p. 70–71
[2] [VH], p. 66–68

for $i \in \{1, \ldots, k\}$. If

$$S_i(x) = \begin{pmatrix} e^{\lambda_i x} & 0 & 0 & \cdots & 0 & 0 \\ \frac{x^1}{1!} e^{\lambda_i x} & e^{\lambda_i x} & 0 & \cdots & 0 & 0 \\ \frac{x^2}{2!} e^{\lambda_i x} & \frac{x^1}{1!} e^{\lambda_i x} & e^{\lambda_i x} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}$$

is a square matrix which has the same dimensions as J_i, it is easily verified that

$$S_i(x)^{-1} = S_i(-x),$$

$$\frac{\mathrm{d}}{\mathrm{d}x} S_i(x) = S_i(x)' S_i(x)^{-1} = J_i.$$

Applying Theorem 2.3.8 to matrix

$$S(x) = \begin{pmatrix} S_1(x) & 0 & \cdots & 0 \\ 0 & S_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(x) \end{pmatrix}$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x} S(x) = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}$$

and Theorem 2.3.7 gives

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(C^{-1} S(x) \right) = C^{-1} \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix} C = A.$$

Theorem 2.3.6 implies the existence of a matrix $D \in \mathbf{R}^{n \times n}$ such that $e^{(x-a)A} = C^{-1} S(x) D$ for every $x \in [a, b]$; substituting $x = a$ yields $D = S(a)^{-1} C$.

Remark 2.5.9. Volterra gives no indication how to "guess" the calculation in the previous example. We may proceed as follows: Let again $A = C^{-1} J C$, where

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}, \quad J_i = \begin{pmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_i \end{pmatrix}.$$

The definition of matrix exponential implies

$$\exp(Ax) = \exp(C^{-1} J x C) = C^{-1} \exp(Jx) C$$

for every $x \in \mathbf{R}$ and it suffices to calculate $\exp(Jx)$. We see that

$$J_i x = x \begin{pmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix} + x \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

It is easy to calculate an arbitrary power of the matrices on the right hand side (the second one is a nilpotent matrix); the definition of matrix exponential then gives

$$\exp(J_i x) = \begin{pmatrix} e^{\lambda_i x} & 0 & 0 & \cdots & 0 & 0 \\ \frac{x^1}{1!} e^{\lambda_i x} & e^{\lambda_i x} & 0 & \cdots & 0 & 0 \\ \frac{x^2}{2!} e^{\lambda_i x} & \frac{x^1}{1!} e^{\lambda_i x} & e^{\lambda_i x} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \end{pmatrix},$$

$$\exp(Jx) = \begin{pmatrix} \exp(J_1 x) & 0 & \cdots & 0 \\ 0 & \exp(J_2 x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp(J_k x) \end{pmatrix}.$$

Theorem 2.5.10.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuous function and $\varphi : [c,d] \to [a,b]$ a continuously differentiable function such that $\varphi(c) = a$ and $\varphi(d) = b$, then

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = \prod_c^d (I + A(\varphi(t))\varphi'(t)\,\mathrm{d}t).$$

Proof. Define

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t)$$

for every $x \in [a,b]$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}(Y \circ \varphi) = Y'(\varphi(t))\varphi'(t)Y(\varphi(t))^{-1} = A(\varphi(t))\varphi'(t)$$

for every $t \in [c,d]$. The fundamental theorem for product integral gives

$$\prod_a^b (I + A(x)\,\mathrm{d}x) = Y(b)Y(a)^{-1} = Y(\varphi(d))Y(\varphi(c))^{-1} = \prod_c^d (I + A(\varphi(t))\varphi'(t)\,\mathrm{d}t).$$

$\square$

[1] [VH], p. 65

Theorem 2.5.11.[1] If $A : [a, b] \to \mathbf{R}^{n\times n}$ is a continuous function, then

$$\det\left(\prod_a^b (I + A(x)\,\mathrm{d}x)\right) = \exp\left(\int_a^b \sum_{i=1}^n a_{ii}(x)\,\mathrm{d}x\right).$$

Proof. Denote

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t), \quad x \in [a, b].$$

The determinant of $Y(x)$ might be interpreted as a function of the n^2 entries $y_{ij}(x)$, $i, j \in \{1, \ldots, n\}$. The chain rule therefore gives

$$(\det Y)'(x) = \sum_{i,j=1}^n \frac{\partial(\det Y)}{\partial y_{ij}} y'_{ij}(x).$$

Formula (2.2.2) for the expansion of determinant by minors implies

$$\frac{\partial(\det Y)}{\partial y_{ij}} = Y_{ij},$$

and consequently

$$(\det Y)'(x) = \sum_{i,j=1}^n y'_{ij}(x)Y_{ij}(x) = \sum_{i,j,k=1}^n a_{ik}(x)y_{kj}(x)Y_{ij}(x) =$$

$$= \sum_{i,j,k=1}^n a_{ik}(x)\delta_{ik}(x)\det Y(x) = \left(\sum_{i=1}^n a_{ii}(x)\right)\det Y(x)$$

(we have used Theorem 2.5.1 and Equation (2.2.3)). However, the differential equation

$$(\det Y)'(x) = \left(\sum_{i=1}^n a_{ii}(x)\right)\det Y(x)$$

has a unique solution that satisfies

$$\det Y(a) = \det\, I = 1.$$

It is given by

$$\det Y(x) = \exp\left(\int_a^x \sum_{i=1}^n a_{ii}(t)\,\mathrm{d}t\right), \quad x \in [a, b].$$

□

[1] [VH], p. 61–62

Theorem 2.5.12.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuous function and $C \in \mathbf{R}^{n\times n}$ a regular matrix, then

$$\prod_a^b (I + C^{-1}A(x)C\,\mathrm{d}x) = C^{-1}\prod_a^b (I + A(x)\,\mathrm{d}x)C.$$

Proof. Define

$$Y(x) = \prod_a^x (I + A(t)\,\mathrm{d}t)$$

for every $x \in [a,b]$. Theorem 2.3.7 gives

$$\frac{\mathrm{d}}{\mathrm{d}x}(C^{-1}Y) = C^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)C = C^{-1}AC,$$

and therefore

$$\prod_a^b (I + C^{-1}A(x)C\,\mathrm{d}x) = C^{-1}Y(b)(C^{-1}Y(a))^{-1} = C^{-1}\prod_a^b (I + A(x)\,\mathrm{d}x)C.$$

□

2.6 Multivariable calculus

In this section we turn our attention to matrix functions of several variables, i.e. to functions $A : \mathbf{R}^m \to \mathbf{R}^{n\times n}$, where $m, n \in \mathbf{N}$. We introduce the notation

$$\frac{\partial A}{\partial x_k}(x) = \left\{\frac{\partial a_{ij}}{\partial x_k}(x)\right\}_{i,j=1}^{n},$$

provided the necessary partial derivatives exist.

Definition 2.6.1. Let G be a domain in $\mathbf{R}^m$ and $x \in G$. Consider a function $A : G \to \mathbf{R}^{n\times n}$ that is regular at x and such that $\frac{\partial A}{\partial x_k}(x)$ exists. We define the left partial derivative of A at x with respect to the k-th variable as

$$\frac{\mathrm{d}}{\mathrm{d}x_k}A(x) = \frac{\partial A}{\partial x_k}(x)A^{-1}(x).$$

Remark 2.6.2. Volterra also introduces the left differential of A as the matrix

$$\mathrm{d}A = A(x_1 + \mathrm{d}x_1, \ldots, x_m + \mathrm{d}x_m)A^{-1}(x_1, \ldots, x_m) = I + \sum_{k=1}^{m}\left(\frac{\mathrm{d}}{\mathrm{d}x_k}A(x)\right)\mathrm{d}x_k, \tag{2.6.1}$$

[1] [VH], p. 63

which differs infinitesimally from the identity matrix. He also claims that

$$\mathrm{d}A = \prod_{k=1}^{m} \left(I + \left(\frac{\mathrm{d}}{\mathrm{d}x_k} A(x) \right) \mathrm{d}x_k \right),$$

since the product of infinitesimal quantities can be neglected.

Recall the following well-known theorem of multivariable calculus: If $f_1, \ldots, f_m : \mathbf{R}^m \to \mathbf{R}$ are functions that have continuous partial derivatives with respect to all variables, then the following statements are equivalent:

(1) There exists a function $F : \mathbf{R}^m \to \mathbf{R}$ such that $\frac{\partial F}{\partial x_i} = f_i$ for $i = 1, \ldots, m$.

(2) $\dfrac{\partial f_i}{\partial x_j} - \dfrac{\partial f_j}{\partial x_i} = 0$ for $i, j = 1, \ldots, m$, $i \neq j$.

Volterra proceeds to the formulation of a similar theorem concerning left derivatives.

Definition 2.6.3. Let $A, B : \mathbf{R}^m \to \mathbf{R}^{n \times n}$ be matrix functions that possess partial derivatives with respect to the i-th and j-th variable. We define

$$\Delta(A, B)_{x_i, x_j} = \frac{\partial B}{\partial x_i} - \frac{\partial A}{\partial x_j} + BA - AB.$$

Volterra's proof of the following lemma has been slightly modified to make it more readable. We also require the equality of mixed partial derivatives, whereas Volterra supposes that the mixed derivatives can be interchanged without any comment.

Lemma 2.6.4.[1] Let $m \in \mathbf{N}$, $i, j \in \{1, \ldots, m\}$, $i \neq j$. Let G be an open set in $\mathbf{R}^m$ and $x \in G$. Consider a pair of matrix functions $X, Y : G \to \mathbf{R}^{n \times n}$ that possess partial derivatives with respect to x_i and x_j at x, and a function $S : G \to \mathbf{R}^{n \times n}$ that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}x_i} S(x) = X(x), \tag{2.6.2}$$

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

Then the equality

$$\frac{\partial}{\partial x_i} \left(S^{-1} \left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j} S \right) S \right) = S^{-1} \Delta(X, Y)_{x_i, x_j} S$$

holds at the point x.

Proof. Using the formula for the derivative of an inverse matrix and the assumption (2.6.2) we calculate

$$\frac{\partial S^{-1}}{\partial x_i} \left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j} S \right) S = -S^{-1} \frac{\partial S}{\partial x_i} S^{-1} \left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j} S \right) S =$$

[1] [VH], p. 81

$$= -S^{-1}X\left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j}S\right)S = S^{-1}\left(-XY + X\left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)\right)S,$$

further

$$S^{-1}\frac{\partial}{\partial x_i}\left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j}S\right)S = S^{-1}\left(\frac{\partial Y}{\partial x_i} - \frac{\partial}{\partial x_i}\left(\frac{\partial S}{\partial x_j}S^{-1}\right)\right)S =$$

$$= S^{-1}\left(\frac{\partial Y}{\partial x_i} - \frac{\partial^2 S}{\partial x_i \partial x_j}S^{-1} - \frac{\partial S}{\partial x_j}\frac{\partial S^{-1}}{\partial x_i}\right)S =$$

$$= S^{-1}\left(\frac{\partial Y}{\partial x_i} - \frac{\partial (XS)}{\partial x_j}S^{-1} - \frac{\partial S}{\partial x_j}S^{-1}\frac{\partial S}{\partial x_i}S^{-1}\right)S =$$

$$= S^{-1}\left(\frac{\partial Y}{\partial x_i} - \frac{\partial X}{\partial x_j} - X\left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right) - \left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)X\right)S =$$

$$= S^{-1}\left(\frac{\partial Y}{\partial x_i} - \frac{\partial X}{\partial x_j} - X\left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right) - \left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)X\right)S,$$

and finally

$$S^{-1}\left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j}S\right)\frac{\partial S}{\partial x_i} = S^{-1}\left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j}S\right)XS = S^{-1}\left(YX - \left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)X\right)S.$$

The product rule for differentiation gives (using the previous three equations)

$$\frac{\partial}{\partial x_i}\left(S^{-1}\left(Y - \frac{\mathrm{d}}{\mathrm{d}x_j}S\right)S\right) = S^{-1}\Bigg(-XY + X\left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right) + \frac{\partial Y}{\partial x_i} - \frac{\partial X}{\partial x_j} -$$

$$-X\left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right) - \left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)X + YX - \left(\frac{\mathrm{d}}{\mathrm{d}x_j}S\right)X\Bigg)S = S^{-1}\Delta(X,Y)_{x_i,x_j}S.$$

□

Theorem 2.6.5.[1] If $B_1, \ldots, B_m : \mathbf{R}^m \to \mathbf{R}^{n\times n}$ are continuously differentiable with respect to all variables, then the following statements are equivalent:
(1) There exists a function $A : \mathbf{R}^m \to \mathbf{R}^{n\times n}$ such that $B_k = \frac{\mathrm{d}}{\mathrm{d}x_k}A$ for $k = 1, \ldots, m$.
(2) $\Delta(B_i, B_j)_{x_i,x_j} = 0$ for $i, j = 1, \ldots, m$, $i \neq j$.
Proof. We start with the implication (1) ⇒ (2):

$$\frac{\partial B_i}{\partial x_j} - \frac{\partial B_j}{\partial x_i} = \frac{\partial}{\partial x_j}\left(\frac{\partial A}{\partial x_i}A^{-1}\right) - \frac{\partial}{\partial x_i}\left(\frac{\partial A}{\partial x_j}A^{-1}\right) =$$

$$= \frac{\partial}{\partial x_j}\left(\frac{\partial A}{\partial x_i}\right)A^{-1} + \frac{\partial A}{\partial x_i}\frac{\partial A^{-1}}{\partial x_j} - \frac{\partial}{\partial x_i}\left(\frac{\partial A}{\partial x_j}\right)A^{-1} - \frac{\partial A}{\partial x_j}\frac{\partial A^{-1}}{\partial x_i} =$$

[1] [VH], p. 78–85

$$= \frac{\partial A}{\partial x_i}\frac{\partial A^{-1}}{\partial x_j} - \frac{\partial A}{\partial x_j}\frac{\partial A^{-1}}{\partial x_i} = -\frac{\partial A}{\partial x_i}A^{-1}\frac{\partial A}{\partial x_j}A^{-1} + \frac{\partial A}{\partial x_j}A^{-1}\frac{\partial A}{\partial x_i}A^{-1} = B_jB_i - B_iB_j$$

(statement (1) implies that the mixed partial derivatives of A are continuous, and therefore interchangeable).

The reverse implication (2) $\Rightarrow$ (1) is first proved for $m = 2$: Suppose that the function $A : \mathbf{R}^2 \to \mathbf{R}^{n\times n}$ from (2) exists. Choose $x_0 \in \mathbf{R}$ and define

$$S(x, y) = \prod_{x_0}^{x}(I + B_1(t, y)\,\mathrm{d}t).$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}x}S = B_1 = \frac{\mathrm{d}}{\mathrm{d}x}A,$$

which implies the existence of a matrix function $T : \mathbf{R} \to \mathbf{R}^{n\times n}$ such that $A(x, y) = S(x, y)T(y)$ (T is independent on x). We calculate

$$\frac{\mathrm{d}}{\mathrm{d}y}T = \frac{\mathrm{d}}{\mathrm{d}y}(S^{-1}A) = S^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}y}A - \frac{\mathrm{d}}{\mathrm{d}y}S\right)S = S^{-1}\left(B_2 - \frac{\mathrm{d}}{\mathrm{d}y}S\right)S.$$

We now relax the assumption that the function A exists; the function on the right hand side of the last equation is nevertheless independent on x, because Lemma 2.6.4 gives

$$\frac{\partial}{\partial x}\left(S^{-1}\left(B_2 - \frac{\mathrm{d}}{\mathrm{d}y}S\right)S\right) = S^{-1}\Delta(B_1, B_2)_{x,y}S = 0.$$

Thus we define

$$T(y) = \prod_{y_0}^{y}\left(I + S^{-1}(x,t)\left(B_2(x,t) - \frac{\mathrm{d}}{\mathrm{d}y}S(x,t)\right)S(x,t)\,\mathrm{d}t\right)$$

(where x is arbitrary) and $A = ST$. Since

$$\frac{\mathrm{d}}{\mathrm{d}x}A = \frac{\mathrm{d}}{\mathrm{d}x}(ST) = \frac{\mathrm{d}}{\mathrm{d}x}S = B_1$$

and

$$\frac{\mathrm{d}}{\mathrm{d}y}A = \frac{\mathrm{d}}{\mathrm{d}y}(ST) = \frac{\mathrm{d}}{\mathrm{d}y}S + S\left(\frac{\mathrm{d}}{\mathrm{d}y}T\right)S^{-1} = B_2,$$

the proof is finished; we now proceed to the case $m > 2$ by induction: Choose $x_0 \in \mathbf{R}$ and define

$$S(x_1, \ldots, x_m) = \prod_{x_0}^{x_1}(I + B_1(t, x_2, \ldots, x_m)\,\mathrm{d}t).$$

If the function $A : \mathbf{R}^m \to \mathbf{R}^{n\times n}$ exists, we must have

$$\frac{\mathrm{d}}{\mathrm{d}x_1} S = B_1 = \frac{\mathrm{d}}{\mathrm{d}x_1} A$$

and consequently

$$A(x_1,\ldots,x_m) = S(x_1,\ldots,x_m)T(x_2,\ldots,x_m)$$

for some matrix function $T : \mathbf{R}^{m-1} \to \mathbf{R}^{n\times n}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}x_k} T = \frac{\mathrm{d}}{\mathrm{d}x_k}(S^{-1}A) = S^{-1}\left(B_k - \frac{\mathrm{d}}{\mathrm{d}x_k} S\right) S, \quad k = 2,\ldots,m.$$

We now relax the assumption that the function A exists and define

$$U_k = S^{-1}\left(B_k - \frac{\mathrm{d}}{\mathrm{d}x_k} S\right) S, \quad k = 2,\ldots,m.$$

Each of these functions U_k is indeed independent on x_1, because Lemma 2.6.4 gives

$$\frac{\partial U_k}{\partial x_1} = S^{-1}\Delta(B_1,B_k)_{x_1,x_k} S = 0.$$

Since

$$\Delta(U_i,U_j)_{x_i,x_j} = S^{-1}\Delta(B_i,B_j)_{x_i,x_j} S = 0, \quad i,j = 2,\ldots,m,$$

the induction hypothesis implies the existence of a function T of $m-1$ variables $x_2,\ldots,x_m$ such that

$$\frac{\mathrm{d}}{\mathrm{d}x_k} T = U_k, \quad k = 2,\ldots,m.$$

We now let $A = ST$ and obtain

$$\frac{\mathrm{d}}{\mathrm{d}x_1} A = \frac{\mathrm{d}}{\mathrm{d}x_1}(ST) = \frac{\mathrm{d}}{\mathrm{d}x_1} S = B_1$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x_k} A = \frac{\mathrm{d}}{\mathrm{d}x_k}(ST) = \frac{\mathrm{d}}{\mathrm{d}x_k} S + S\left(\frac{\mathrm{d}}{\mathrm{d}x_k} T\right) S^{-1} = B_k$$

for $k = 2,\ldots,m$, which completes the proof. □

Remark 2.6.6. Volterra's proof of Theorem 2.6.5 contains a deficiency: We have applied Lemma 2.6.4 to the function

$$S(x_1,\ldots,x_m) = \prod_{x_0}^{x_1}(I + B_1(t,x_2,\ldots,x_m)\,\mathrm{d}t)$$

without verifying that

$$\frac{\partial^2 S}{\partial x_i \partial x_1}(x) = \frac{\partial^2 S}{\partial x_1 \partial x_i}(x), \quad i \in \{2,\ldots,m\}.$$

This equality follows from the well-known theorem of multivariable calculus provided the derivatives $\frac{\partial S}{\partial x_i}$ exist in some neighbourhood of x for every $i \in \{1, \dots, m\}$, and the derivatives

$$\frac{\partial^2 S}{\partial x_1 \partial x_i}(x)$$

are continuous at x for every $i \in \{2, \dots, m\}$. We have

$$\frac{\partial S}{\partial x_1} = B_1$$

and consequently

$$\frac{\partial^2 S}{\partial x_1 \partial x_i} = \frac{\partial B_1}{\partial x_i},$$

which is a continuous function for every $i \in \{2, \dots, m\}$. The existence of the derivatives $\frac{\partial S}{\partial x_i}$ for $i \in \{2, \dots, m\}$ is certainly not obvious but follows from Theorem 3.6.14 on differentiating the product integral with respect to a parameter, which will be proved in Chapter 3.

Remark 2.6.7. An analogy of Theorem 2.6.5 holds also for right derivatives; the condition $\Delta(B_i, B_j)_{x_i,x_j} = 0$ must be replaced by $\Delta^*(B_i, B_j)_{x_i,x_j} = 0$, where

$$\Delta^*(A, B)_{x_i,x_j} = \frac{\partial B}{\partial x_i} - \frac{\partial A}{\partial x_j} + AB - BA.$$

The second fundamental notion of multivariable calculus is the contour integral. While Volterra introduces only product integrals along a contour φ in $\mathbf{R}^2$, which he denotes by

$$\int_\varphi X\,\mathrm{d}x \cdot Y\,\mathrm{d}y,$$

we give a general definition for curves in $\mathbf{R}^m$; we also use a different notation.

We will always consider curves that are given using a parametrization $\varphi : [a, b] \to \mathbf{R}^m$ that is piecewise continuously differentiable, which means that $\varphi'_-(x)$ exists for $x \in (a, b]$, $\varphi'_+(x)$ exists for $x \in [a, b)$, and $\varphi'_-(x) = \varphi'_+(x)$ except a finite number of points in (a, b).

The image of the curve is then defined as

$$\langle\varphi\rangle = \varphi([a, b]) = \{\varphi(t);\ t \in [a, b]\}.$$

Definition 2.6.8. Consider a piecewise continuously differentiable function $\varphi : [a, b] \to \mathbf{R}^m$ and a system of m matrix functions $B_1, \dots, B_m : \langle\varphi\rangle \to \mathbf{R}^{n\times n}$. The contour product integral of these functions along φ is defined as

$$\prod_\varphi (I+B_1\,\mathrm{d}x_1+\cdots+B_m\,\mathrm{d}x_m) = \prod_a^b (I+(B_1(\varphi(t))\varphi_1'(t) + \cdots + B_m(\varphi(t))\varphi_m'(t))\,\mathrm{d}t).$$

Given an arbitrary curve $\varphi : [a, b] \to \mathbf{R}^m$, we define the curve $-\varphi$ as

$$(-\varphi)(t) = \varphi(-t), t \in [-b, -a].$$

This curve has the same image as the original curve, but is traversed in the opposite direction.
For any pair of curves $\varphi_1 : [a_1, b_1] \to \mathbf{R}^m$, $\varphi_2 : [a_2, b_2] \to \mathbf{R}^m$ such that $\varphi_1(b_1) = \varphi_2(a_2)$ we define the composite curve $\varphi_1 + \varphi_2$ by

$$(\varphi_1 + \varphi_2)(t) = \begin{cases} \varphi_1(t), & t \in [a_1, b_1], \\ \varphi_2(t - b_1 + a_2), & t \in [b_1, b_1 + b_2 - a_2]. \end{cases}$$

Theorem 2.6.9.[1] Contour product integral has the following properties:
(1) If $\varphi_1 + \varphi_2$ is a curve obtained by joining two curves φ_1 and φ_2, then

$$\prod_{\varphi_1+\varphi_2} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m) =$$

$$= \prod_{\varphi_2} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m) \cdot \prod_{\varphi_1} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m).$$

(2) If $-\varphi$ is a curve obtained by reversing the orientation of φ, then

$$\prod_{-\varphi} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m) = \left(\prod_{\varphi} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m) \right)^{-1}.$$

Proof. Let $\varphi_1 : [a_1, b_1] \to \mathbf{R}^m$, $\varphi_2 : [a_2, b_2] \to \mathbf{R}^m$. Then

$$\prod_{-\varphi} (I + B_1 \,\mathrm{d}x_1 + \cdots + B_m \,\mathrm{d}x_m) =$$

$$\prod_{b_1}^{b_1+b_2-a_2} (I + (B_1(\varphi(t - b_1 + a_2))\varphi_1'(t) + \cdots + B_m(\varphi(t - b_1 + a_2))\varphi_m'(t)) \,\mathrm{d}t) \cdot$$

$$\cdot \prod_{a_1}^{b_1} (I + (B_1(\varphi(t))\varphi_1'(t) + \cdots + B_m(\varphi(t))\varphi_m'(t)) \,\mathrm{d}t).$$

The change of variables Theorem 2.5.10 gives

$$\prod_{b_1}^{b_1+b_2-a_2} (I + (B_1(\varphi(t - b_1 + a_2))\varphi_1'(t) + \cdots + B_m(\varphi(t - b_1 + a_2))\varphi_m'(t)) \,\mathrm{d}t) =$$

[1] [VH], p. 91

$$= \prod_{a_2}^{b_2} (I + (B_1(\varphi(t))\varphi_1'(t) + \cdots + B_m(\varphi(t))\varphi_m'(t)) \, \mathrm{d}t),$$

which proves the first statement. The second one is also a direct consequence of Theorem 2.5.10. Note that the change of variables theorem was proved only for continuously differentiable functions, while our contours are piecewise continuously differentiable. It is however always possible to partition the integration interval in such a way that the integrated functions are continuously differentiable on every subinterval. □

Definition 2.6.10. Let G be a subset of $\mathbf{R}^m$ and $B_1, \dots, B_m : G \to \mathbf{R}^{n\times n}$. The contour product integral $\prod(I + B_1 \, \mathrm{d}x_1 + \cdots + B_m \, \mathrm{d}x_m)$ is called path-independent in G if

$$\prod_{\varphi} (I + B_1 \, \mathrm{d}x_1 + \cdots + B_m \, \mathrm{d}x_m) = \prod_{\psi} (I + B_1 \, \mathrm{d}x_1 + \cdots + B_m \, \mathrm{d}x_m)$$

for each pair of curves $\varphi, \psi : [a, b] \to G$ such that $\varphi(a) = \psi(a)$ and $\varphi(b) = \psi(b)$.

Using Theorem 2.6.9 it is easy to see that the contour product integral is path-independent in G if and only if

$$\prod_{\varphi} (I + B_1 \, \mathrm{d}x_1 + \cdots + B_m \, \mathrm{d}x_m) = I$$

for every closed curve φ in G.

As already mentioned, Volterra is concerned especially with curves in $\mathbf{R}^2$. His effort is directed towards proving the following theorem:

Theorem 2.6.11.[1] Let G be a simply connected domain in $\mathbf{R}^2$. Consider a pair of functions $A, B : G \to \mathbf{R}^{n\times n}$ such that $\Delta(A, B)_{x,y} = 0$ at every point of G. Then

$$\prod_{\varphi} (I + A \, \mathrm{d}x + B \, \mathrm{d}y) = I$$

for every piecewise continuously differentiable closed curve φ in G.

Although Volterra's proof is somewhat incomplete, we try to indicate its main steps in the rest of the section. Theorem 2.6.11 is of great importance for Volterra as he uses it to prove an analogue of Cauchy theorem for product integral in complex domain; this topic will be discussed in the next section.

Definition 2.6.12. A set S in $\mathbf{R}^2$ is called simple in the x-direction, if the set $S \cap \{(x, y_0);\, x \in \mathbf{R}\}$ is connected for every $y_0 \in \mathbf{R}$. Similarly, S is simple in the y-direction, if the set $S \cap \{(x_0, y);\, y \in \mathbf{R}\}$ is connected for every $x_0 \in \mathbf{R}$.

---

[1] [VH], p. 95

Equivalently said, the intersection of S and a line parallel to the x-axis (or the y-axis) is either an interval (possibly degenerated to a single point), or an empty set.

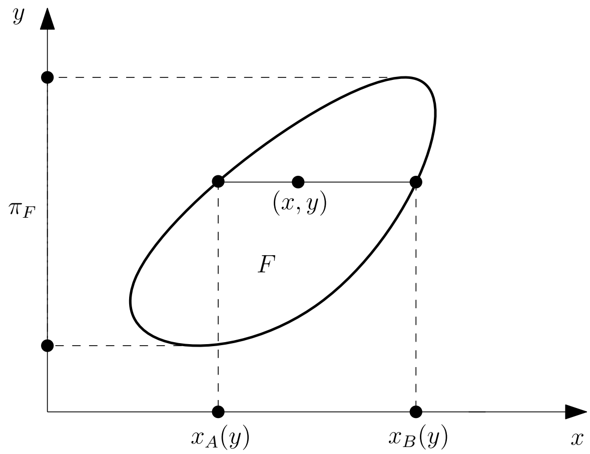
Definition 2.6.13. Let F be a closed bounded subset of $\mathbf{R}^2$ that is simple in the x-direction. For every $y \in \mathbf{R}$ denote

$$\pi_F = \{y \in \mathbf{R};\ \text{there exists } x \in \mathbf{R} \text{ such that } (x,y) \in G\}. \tag{2.6.3}$$

Further, for every $y \in \pi_F$ let

$$x_A(y) = \inf\{x;\ (x,y) \in F\}, \quad x_B(y) = \sup\{x;\ (x,y) \in F\}.$$

The meaning of these symbols is illustrated by the following figure. Note that the segment $[x_A(y), x_B(y)] \times \{y\}$ is contained in F for every $y \in \pi_F$, i.e. the set F is enclosed between the graphs of the functions $y \mapsto x_A(y)$ and $y \mapsto x_B(y)$, $y \in \pi_F$.



Definition 2.6.14. Let F be a closed bounded subset of $\mathbf{R}^2$ that is simple in the x-direction. The double product integral of a continuous function $A : F \to \mathbf{R}^{n\times n}$ is defined as

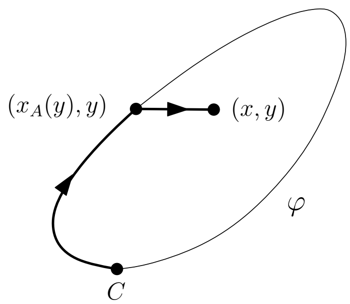
$$\prod_F (I + A(x,y)\,\mathrm{d}x\,\mathrm{d}y) = \prod_{\inf \pi_F}^{\sup \pi_F} \left(I + \left(\int_{x_A(y)}^{x_B(y)} A(x,y)\,\mathrm{d}x \right) \mathrm{d}y \right).$$

Before proceeding to the next theorem we recall that a Jordan curve is a closed curve with no self-intersections. Formally written, it is a curve with parametrization $\varphi : [a,b] \to \mathbf{R}^2$ that is injective on $[a,b)$ and $\varphi(a) = \varphi(b)$. It is known that a Jordan curve divides the plane in two components – the interior and the exterior of φ. In the following text we denote the interior of φ by $\operatorname{Int}\varphi$.

Theorem 2.6.15.[1] Consider a piecewise continuously differentiable Jordan curve $\varphi : [a,b] \to \mathbf{R}^2$ such that the set $F = \langle\varphi\rangle \cup \operatorname{Int}\varphi$ is simple in the x-direction. Assume that φ starts at a point $C = (c_x, c_y)$ such that $c_y = \inf \pi_F$ and $c_x = x_A(c_y)$. Denote

$$S(x,y) = \prod_{x_A(y)}^{x} (I + X(t,y)\,\mathrm{d}t) \prod_{C}^{(x_A(y),y)} (I + X\,\mathrm{d}x + Y\,\mathrm{d}y),$$

where the second integral is taken over that part of φ that joins the points C and $(x_A(y), y)$ (see the figure below).



Let G be an open neighbourhood of the set F. Then the equation

$$\prod_{\varphi}(I + X\,\mathrm{d}x + Y\,\mathrm{d}y) = \prod_{F}(I + S^{-1}\Delta(X,Y)_{x,y} S\,\mathrm{d}x\,\mathrm{d}y)$$

holds for each pair of continuously differentiable functions $X, Y : G \to \mathbf{R}^{n\times n}$.

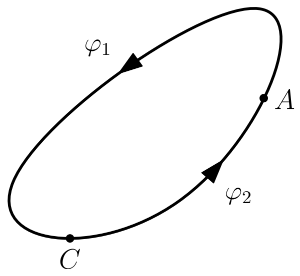
Theorem 2.6.15 might be regarded as an analogy of Green's theorem, since it provides a relationship between the double product integral over F and the contour product integral over the boundary of F. The proof in [VH] is somewhat obscure (mainly because of Volterra's calculations with infinitesimal quantities) and will not be reproduced here. A statement similar to Theorem 2.6.15 will be proved in Chapter 3, Theorem 3.7.4.

Theorem 2.6.16.[2] Consider a piecewise continuously differentiable Jordan curve $\varphi : [a,b] \to \mathbf{R}^2$ such that the set $F = \langle\varphi\rangle \cup \operatorname{Int}\varphi$ is simple in the x-direction. Let G be an open neighbourhood of the set F. If $A, B : G \to \mathbf{R}^{n\times n}$ is a pair of continuously differentiable functions such that $\Delta(A,B)_{x,y} = 0$ at every point of G, then

$$\prod_{\varphi}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I.$$

[1] [VH], p. 92–94
[2] [VH], p. 95

Proof. Let $C = (c_x, c_y)$ be the point such that $c_y = \inf \pi_F$ and $c_x = x_A(c_y)$. If A is the starting point of φ, we may write $\varphi = \varphi_1 + \varphi_2$, where φ_1 is the part of the curve between the points A, C and φ_2 is the part between C, A provided we travel along φ in direction of its orientation.



Theorem 2.6.15 gives

$$\prod_{\varphi_2+\varphi_1} (I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I,$$

and consequently

$$\prod_{\varphi}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = \prod_{\varphi_1+\varphi_2} (I + A\,\mathrm{d}x + B\,\mathrm{d}y) =$$

$$= \prod_{\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) \prod_{\varphi_2+\varphi_1} (I + A\,\mathrm{d}x + B\,\mathrm{d}y) \left(\prod_{\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y)\right)^{-1} = I.$$

□

Remark 2.6.17. In case the set G in statement of the last theorem is simple both in the x direction and in the y direction, there is a simpler alternative proof of Theorem 2.6.16. It is based on Theorem 2.6.5, which we proved for $G = \mathbf{R}^2$, but the proof is exactly the same also for sets G which are simple in x as well as in y direction. Consequently, the assumption $\Delta(A,B)_{x,y} = 0$ and Theorem 2.6.5 imply the existence of a function $T : G \to \mathbf{R}^2$ such that

$$A(x,y) = \frac{\mathrm{d}}{\mathrm{d}x} T(x,y), \quad B(x,y) = \frac{\mathrm{d}}{\mathrm{d}y} T(x,y)$$

for every $(x,y) \in G$. Thus for arbitrary closed curve $\varphi : [a,b] \to G$ we have

$$\prod_{\varphi}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = \prod_{\varphi}\left(I + \frac{\mathrm{d}}{\mathrm{d}x} T\,\mathrm{d}x + \frac{\mathrm{d}}{\mathrm{d}y} T\,\mathrm{d}y\right) =$$

$$= \prod_a^b \left(I + \left(\frac{\partial T}{\partial x}(\varphi(t))\, T(\varphi(t))^{-1}\varphi_1'(t) + \frac{\partial T}{\partial y}(\varphi(t))\, T(\varphi(t))^{-1}\varphi_2'(t)\right)\mathrm{d}t\right) =$$

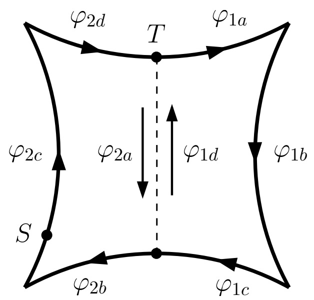
$$= \prod_a^b \left(I + \frac{\mathrm{d}}{\mathrm{d}t}(T \circ \varphi)(t)\,\mathrm{d}t \right) = T(\varphi(b))T(\varphi(a))^{-1} = I.$$

Both the statement and its proof is easily generalized to the case of curves in $\mathbf{R}^m$, $m > 2$.

Volterra now comes to the justification of Theorem 2.6.12: Let G be a simply connected domain in $\mathbf{R}^2$ and $A, B : G \to \mathbf{R}^{n\times n}$ such that $\Delta(A,B)_{x,y} = 0$ at every point of G. We have to verify that

$$\prod_{\varphi}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I \tag{2.6.4}$$

for every piecewise continuously differentiable closed curve φ in G. Theorem 2.6.16 ensures that the statement is true, if the set $F = \langle\varphi\rangle \cup \operatorname{Int}\varphi$ is simple in the x-direction. Volterra first notes that it remains true, if F can be split by a curve in two parts each of which is simple in the x-direction.



Indeed, using the notation from the above figure, if $\varphi_1 = \varphi_{1a} + \varphi_{1b} + \varphi_{1c} + \varphi_{1d}$ and $\varphi_2 = \varphi_{2a} + \varphi_{2b} + \varphi_{2c} + \varphi_{2d}$, then

$$\prod_{\varphi_1}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I, \quad \prod_{\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I,$$

and thus

$$\prod_{\varphi_1+\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = \prod_{\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) \cdot \prod_{\varphi_1}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I.$$

Now if S denotes the initial point of φ, then

$$\prod_{\varphi}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) =$$

$$= \prod_S^T(I + A\,\mathrm{d}x + B\,\mathrm{d}y) \cdot \prod_{\varphi_1+\varphi_2}(I + A\,\mathrm{d}x + B\,\mathrm{d}y) \cdot \prod_T^S(I + A\,\mathrm{d}x + B\,\mathrm{d}y) = I,$$

where $\prod_S^T$ denotes the contour product integral taken along the part of φ that connects the points S, T, and $\prod_T^S$ is taken along the same curve with reversed orientation (it is thus the inverse matrix of $\prod_S^T$).
By induction it follows that (2.6.4) holds if the set $F = \langle\varphi\rangle \cup \operatorname{Int}\varphi$ can be decomposed into a finite number of subsets which are simple in the x-direction. Volterra now states that this is possible for every curve φ in consideration, and so Theorem 2.6.12 is proved. He however gave no justification of the last statement, so his proof remains incomplete.

2.7 Product integration in complex domain

So far we have been concerned with real matrix functions defined on a real interval, i.e. with functions $A : [a,b] \to \mathbf{R}^{n\times n}$. Most of our results can be, without greater effort, generalized to complex-valued matrix functions, i.e. to functions $A : [a,b] \to \mathbf{C}^{n\times n}$. However, in the following two sections, we will focus our interest to matrix functions of a complex variable, i.e. $A : G \to \mathbf{C}^{n\times n}$, where G is a subset of the complex plane.
A matrix function $A = \{a_{jk}\}_{j,k=1}^n$ will be called differentiable at a point $z \in \mathbf{C}$, if its entries a_{jk} are differentiable at that point. We will use the notation

$$A'(z) = \{a'_{jk}(z)\}_{j,k=1}^n.$$

The function A is called holomorphic in an open domain $G \subseteq \mathbf{C}$, if it is differentiable everywhere in G.

Definition 2.7.1. The left derivative of a complex matrix function A at a point $z \in \mathbf{C}$ is defined as

$$\frac{\mathrm{d}}{\mathrm{d}z}A = A'(z)A^{-1}(z),$$

provided that A is differentiable and regular at the point z.

Each matrix function A of a complex variable z might be interpreted as a function of two real variables x, y, where $z = x + iy$. The Cauchy-Riemann equation states that

$$A'(z) = \frac{\partial A}{\partial x}(x+iy) = \frac{1}{i}\frac{\partial A}{\partial y}(x+iy),$$

thus the left derivative satisfies

$$\frac{\mathrm{d}}{\mathrm{d}z}A = \frac{\mathrm{d}}{\mathrm{d}x}A = \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}y}A,$$

provided all the derivatives exist.
We now proceed to the definition of product integral along a contour in the complex domain. We again restrict ourselves to contours with a piecewise continuously differentiable parametrization $\varphi : [a,b] \to \mathbf{C}$, i.e. $\varphi'_-(x)$ exists for all $x \in (a,b]$,

$\varphi'_+(x)$ exists for all $x \in [a,b)$, and $\varphi'_-(x) = \varphi'_+(x)$ except a finite number of points in (a,b).

Definition 2.7.2. Let $\varphi : [a,b] \to \mathbf{C}$ be a piecewise continuously differentiable contour in the complex plane and A a matrix function which is defined and continuous on $\langle\varphi\rangle$. The left product integral along φ is defined as

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = \prod_a^b (I + A(\varphi(t))\varphi'(t)\,\mathrm{d}t). \tag{2.7.1}$$

As Volterra remarks, the left contour product integral is equal to the limit

$$\lim_{\nu(D)\to 0} P(A,D),$$

where D is a tagged partition of $[a,b]$ with division points t_i, tags $\xi_i \in [t_{i-1}, t_i]$ and

$$P(A,D) = \prod_{i=m}^{1} \left(I + A(\varphi(\xi_i))(\varphi(t_i) - \varphi(t_{i-1}))\right). \tag{2.7.2}$$

Instead of our $\prod_{\varphi}(I + A(z)\,\mathrm{d}z)$ he uses the notation $\int_{\varphi} A(z)\,\mathrm{d}z$.
The product integral $\prod_{\varphi}(I + A(z)\,\mathrm{d}z)$ can be converted to a contour product integral taken along a contour $\tilde{\varphi}$ in $\mathbf{R}^2$ with the parametrization

$$\tilde{\varphi}(t) = (\mathrm{Re}\ \varphi(t), \mathrm{Im}\ \varphi(t)), \quad t \in [a,b].$$

Indeed, define $A_1(x,y) = A(x+iy)$ and $A_2(x,y) = iA(x+iy)$. Then

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = \prod_a^b (I + A(\varphi(t))\varphi'(t)\,\mathrm{d}t) =$$

$$\prod_a^b (I + (A_1(\varphi(t))\mathrm{Re}\ \varphi'(t) + A_2(\varphi(t))\mathrm{Im}\ \varphi'(t))\,\mathrm{d}t),$$

thus

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = \prod_{\tilde{\varphi}}(I + A(x+iy)\,\mathrm{d}x + iA(x+iy)\,\mathrm{d}y). \tag{2.7.3}$$

The following theorem is an analogy of Theorem 2.6.9. It can be proved directly in the same way as Theorem 2.6.9, or alternatively by using the relation (2.7.3) and applying Theorem 2.6.9.

Theorem 2.7.3.[1] The left contour product integral has the following properties:

[1] [VH], p. 107

(1) If $\varphi_1 + \varphi_2$ is a curve obtained by joining two curves φ_1 and φ_2, then

$$\prod_{\varphi_1+\varphi_2} (I + A(z)\,\mathrm{d}z) = \prod_{\varphi_2}(I + A(z)\,\mathrm{d}z) \cdot \prod_{\varphi_1}(I + A(z)\,\mathrm{d}z).$$

(2) If $-\varphi$ is a curve obtained by reversing the orientation of φ, then

$$\prod_{-\varphi}(I + A(z)\,\mathrm{d}z) = \left(\prod_{\varphi}(I + A(z)\,\mathrm{d}z)\right)^{-1}.$$

Our interest in product integral of a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$ stems from the fact that it provides a solution of the differential equation (or a system of equations)

$$y'(x) = A(x)y(x),$$

where $y : [a,b] \to \mathbf{R}^n$. The situation is similar in the complex domain: Since the contour product integral is a limit of the products (2.7.2), we expect that the solution of the differential equation

$$y'(z) = A(z)y(z)$$

that satisfies $y(z_0) = y_0$ will be given by

$$y(z) = \left(\prod_{\varphi}(I + A(w)\,\mathrm{d}w)\right) y_0,$$

where $\varphi : [a,b] \to \mathbf{C}$ is a contour connecting the points z_0 and z. However, this definition of y is correct only if the product integral is independent on the choice of a particular contour, i.e. if

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = \prod_{\psi}(I + A(z)\,\mathrm{d}z),$$

whenever φ and ψ are two curves with the same initial points and the same endpoints. From Theorem 2.7.3 we see that $\prod_{\varphi+(-\psi)}(I + A(z)\,\mathrm{d}z)$ should be the identity matrix. Equivalently said,

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = I$$

should hold for every closed contour φ.
Volterra proves that the last condition is satisfied in every simply connected domain G provided that the function A is holomorphic in G. He first uses the formula (2.7.3) to convert the integral in complex domain to an integral in $\mathbf{R}^2$. Then, since

$$\Delta(A, iA)_{x,y} = \frac{\partial iA}{\partial x} - \frac{\partial A}{\partial y} + iAA - AiA = 0,$$

Theorem 2.6.11 implies that the contour product integral along any closed curve in G is equal to the identity matrix. Because we didn't prove Theorem 2.6.11, we offer a different justification taken over from [DF].

Theorem 2.7.4.[1] If $G \subseteq \mathbf{C}$ is a simply connected domain and $A : G \to \mathbf{C}^{n\times n}$ a holomorphic function in G, then the contour product integral of A is path-independent in G.

Proof. Let $\varphi : [a,b] \to G$ be a curve in G. We expand the product integral of A along φ to the Peano series

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = \prod_{a}^{b}(I + A(\varphi(t))\varphi'(t)\,\mathrm{d}t) =$$

$$= I + \int_a^b A(\varphi(t))\varphi'(t)\,\mathrm{d}t + \int_a^b \int_a^{t_2} A(t_2)A(t_1)\varphi'(t_2)\varphi'(t_1)\,\mathrm{d}t_1\,\mathrm{d}t_2 + \cdots$$

This infinite series might be written as

$$\prod_{\varphi}(I + A(z)\,\mathrm{d}z) = I + \int_{\varphi(a)}^{\varphi(b)} A(z)\,\mathrm{d}z + \int_{\varphi(a)}^{\varphi(b)} \int_{\varphi(a)}^{z_2} A(z_2)A(z_1)\,\mathrm{d}z_1\,\mathrm{d}z_2 + \cdots$$

where the contour integrals are all taken along φ (or its initial segment). However, since ordinary contour integrals of holomorphic functions are path-independent in G, the sum of the infinite series depends only on the endpoints of φ. $\square$

In case the product integral is path-independent in a given domain G, we will occasionally use the symbol

$$\prod_{z_1}^{z_2}(I + A(z)\,\mathrm{d}z)$$

to denote product integral taken along an arbitrary curve in G with initial point z_1 and endpoint z_2.

Volterra now claims that if G is a simply connected domain and A is a holomorphic matrix function in G, then the function

$$Y(z) = \prod_{z_0}^{z}(I + A(w)\,\mathrm{d}w)$$

provides a solution of the differential equation $Y'(z) = A(z)Y(z)$ in G.

Theorem 2.7.5. If $G \subseteq \mathbf{C}$ is a simply connected domain and $A : G \to \mathbf{C}^{n\times n}$ a holomorphic function in G, then the function

$$Y(z) = \left(\prod_{z_0}^{z}(I + A(w)\,\mathrm{d}w)\right)$$

[1] [DF], p. 62–63

satisfies $Y'(z) = A(z)Y(z)$ in G.

Proof. The statement is obtained by differentiating the series

$$\prod_{z_0}^{z}(I + A(w)\,\mathrm{d}w) = I + \int_{z_0}^{z} A(w)\,\mathrm{d}w + \int_{z_0}^{z}\int_{z_0}^{w_2} A(w_2)A(w_1)\,\mathrm{d}w_1\,\mathrm{d}w_2 + \cdots$$

with respect to z. $\square$

Corollary 2.7.6. Let $G \subseteq \mathbf{C}$ be a simply connected domain and $A : G \to \mathbf{C}^{n\times n}$ a holomorphic function. If $z_0 \in G$ and $y_0 \in \mathbf{C}^n$, then the function $y : G \to \mathbf{C}^n$ defined by

$$y(z) = \left(\prod_{z_0}^{z}(I + A(w)\,\mathrm{d}w)\right) y_0$$

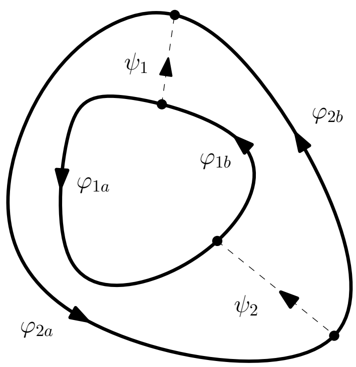
satisfies $y'(z) = A(z)y(z)$ in G and $y(z_0) = y_0$.

Theorem 2.7.7.[1] Let $G \subseteq \mathbf{C}$ be a domain and $A : G \to \mathbf{C}^{n\times n}$ a holomorphic matrix function in G. If $\varphi_1, \varphi_2 : [a,b] \to G$ are two positively oriented Jordan curves such that $\varphi_1 \subset \operatorname{Int}\varphi_2$ and $\operatorname{Int}\varphi_2 \setminus \operatorname{Int}\varphi_1 \subset G$, then

$$\prod_{\varphi_1}(I + A(z)\,\mathrm{d}z) \quad\text{and}\quad \prod_{\varphi_2}(I + A(z)\,\mathrm{d}z)$$

are similar matrices.

Proof. We introduce two disjoint auxiliary segments ψ_1, ψ_2 that connect the curves φ_1, φ_2 (see the figure).



$$\varphi_1 = \varphi_{1a} + \varphi_{1b},$$
$$\varphi_2 = \varphi_{2a} + \varphi_{2b}$$

Theorem 2.7.4 gives

$$\prod_{\varphi_{2a}}(I + A(z)\,\mathrm{d}z)\cdot\prod_{\psi_1}(I + A(z)\,\mathrm{d}z)\cdot\left(\prod_{\varphi_{1a}}(I + A(z)\,\mathrm{d}z)\right)^{-1} = \left(\prod_{\psi_2}(I + A(z)\,\mathrm{d}z)\right)^{-1}$$

[1] [VH], p. 114–116

and

$$\left(\prod_{\varphi_{1b}}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\left(\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\prod_{\varphi_{2b}}(I+A(z)\,\mathrm{d}z)=\prod_{\psi_2}(I+A(z)\,\mathrm{d}z).$$

Multiplying the first equality by the second from left yields

$$\left(\prod_{\varphi_{1b}}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\left(\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\prod_{\varphi_{2b}}(I+A(z)\,\mathrm{d}z)\cdot\prod_{\varphi_{2a}}(I+A(z)\,\mathrm{d}z)\cdot$$

$$\cdot\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)\cdot\left(\prod_{\varphi_{1a}}(I+A(z)\,\mathrm{d}z)\right)^{-1}=I,$$

which can be simplified to

$$\left(\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\prod_{\varphi_2}(I+A(z)\,\mathrm{d}z)\cdot\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)=\prod_{\varphi_1}(I+A(z)\,\mathrm{d}z). \quad (2.7.4)$$

□

Remark 2.7.8. Volterra offers a slightly different proof of the previous theorem: From Theorem 2.7.4 he deduces that

$$\left(\prod_{\varphi_2}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\left(\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)\right)^{-1}\cdot\prod_{\varphi_2}(I+A(z)\,\mathrm{d}z)\prod_{\psi_1}(I+A(z)\,\mathrm{d}z)=I,$$

which implies (2.7.4). This argument is however incorrect, because the domain bounded by $\psi_1+\varphi_2-\psi_1-\varphi_2$ need not be simply connected.

Definition 2.7.9. Let $R>0$, $G=\{z\in\mathbf{C};0<|z-z_0|<R\}$. Suppose $A:G\to\mathbf{C}^{n\times n}$ is holomorphic in G. Let $\varphi:[a,b]\to G$ be a positively oriented Jordan curve, $z_0\in\operatorname{Int}\varphi$. Then

$$\prod_{\varphi}(I+A(z)\,\mathrm{d}z)=CJC^{-1},$$

where J is certain Jordan matrix, which is, according to Theorem 2.7.7, independent on the choice of φ. This Jordan matrix is called the residue of A at the point z_0.

Example 2.7.10.[1] We calculate the residue of a matrix function

$$T(z)=\frac{A}{z-z_0}+B(z)$$

[1] [VH], p. 117–120

at the point z_0, where $A \in \mathbf{C}^{n\times n}$ and B is a matrix function holomorphic in the neighbourhood $U(z_0, R)$ of point z_0. We take the product integral along the circle $\varphi_r(t) = z_0 + re^{it}$, $t \in [0, 2\pi]$, $r < R$ and obtain

$$\prod_{\varphi_r}(I + T(z)\,\mathrm{d}z) = \prod_0^{2\pi}(I + iA + ire^{it}B(z_0 + re^{it})\,\mathrm{d}t).$$

Volterra suggests the following procedure (which is however not fully correct, see Remark 2.7.11): Since $iA + ire^{it}B(z_0 + re^{it}) \to iA$ for $r \to 0$, we have

$$\prod_{\varphi_r}(I + T(z)\,\mathrm{d}z) \to \prod_0^{2\pi}(I + iA\,\mathrm{d}t).$$

The integrals $\prod_{\varphi_r}(I+T(z)\,\mathrm{d}z)$, $r \in (0, R)$, are all similar to a single Jordan matrix. Their limit $\prod_0^{2\pi}(I + iA\,\mathrm{d}t)$ is thus similar to the same matrix and it is sufficient to find its Jordan normal form. By the way, this integral is equal to $e^{2\pi iA}$, giving an analogy of the residue theorem:

$$\text{The matrix } \prod_{\varphi_r}(I + T(z)\,\mathrm{d}z) \text{ is similar to } e^{2\pi iA}. \tag{2.7.5}$$

Consider the Jordan normal form of A:

$$A = C^{-1}\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix} C, \quad \text{where } A_j = \begin{pmatrix} \lambda_j & 0 & \cdots & 0 & 0 \\ 1 & \lambda_j & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_j \end{pmatrix}.$$

Using the result of Example 2.5.8 we obtain

$$\prod_0^{2\pi}(I + iA\,\mathrm{d}t) = C^{-1}\begin{pmatrix} S_1(2\pi) & 0 & \cdots & 0 \\ 0 & S_2(2\pi) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(2\pi) \end{pmatrix}\cdot$$

$$\cdot\left(C^{-1}\begin{pmatrix} S_1(0) & 0 & \cdots & 0 \\ 0 & S_2(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(0) \end{pmatrix}\right)^{-1} = C^{-1}\begin{pmatrix} S_1(2\pi) & 0 & \cdots & 0 \\ 0 & S_2(2\pi) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(2\pi) \end{pmatrix} C,$$

where

$$S_j(x) = \begin{pmatrix} e^{i\lambda_j x} & 0 & 0 & \cdots & 0 & 0 \\ \frac{x^1}{1!}e^{i\lambda_j x} & e^{i\lambda_j x} & 0 & \cdots & 0 & 0 \\ \frac{x^2}{2!}e^{i\lambda_j x} & \frac{x^1}{1!}e^{i\lambda_j x} & e^{i\lambda_j x} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}$$

is a square matrix of the same dimensions as A_j. The Jordan normal form of the matrix

$$\begin{pmatrix} S_1(2\pi) & 0 & \cdots & 0 \\ 0 & S_2(2\pi) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(2\pi) \end{pmatrix},$$

and therefore also the demanded residue, is

$$\begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_k \end{pmatrix}, \text{ where } V_j = \begin{pmatrix} e^{2\pi i\lambda_j} & 0 & \cdots & 0 & 0 \\ 1 & e^{2\pi i\lambda_j} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & e^{2\pi i\lambda_j} \end{pmatrix}$$

is a square matrix of the same dimensions as S_j.

Remark 2.7.11. The calculation from the previous example contains two deficiencies: First, Volterra interchanges the order of limit and product integral to obtain

$$\lim_{r\to 0} \prod_0^{2\pi} (I + iA + ire^{it}B(z_0 + re^{it})\,\mathrm{d}t) = \prod_0^{2\pi} (I + iA\,\mathrm{d}t)$$

without any further comment. However, the convergence $iA+ire^{it}B(z_0+re^{it}) \to iA$ for $r \to 0$ is uniform and in this case, as we will prove in Chapter 5, Theorem 5.6.4, the statement is in fact true.

The second deficiency is more serious. Volterra seems to have assumed that if some matrices $S(r)$, $r \in (0, R)$ (in our case $S(r)$ is the product integral taken along φ_r), are all similar to a single Jordan matrix J, then $\lim_{r\to 0} S(r)$ is also similar to J. This statement is incorrect, as demonstrated by the example

$$S(r) = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix},$$

where $S(r)$ is similar to

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

for $r > 0$, but

$$\lim_{r\to 0} S(r) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

isn't. The mentioned statement can be proved only under additional assumptions on $S(r)$. For example, if the matrices $S(r)$, $r > 0$, have n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then the limit matrix $\lim_{r\to 0} S(r)$ has the same eigenvalues, because

$$\det(S(0) - \lambda I) = \lim_{r\to 0} \det(S(r) - \lambda I) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

This means that all the matrices $S(r)$, $r \geq 0$, are similar to a single Jordan matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

A more detailed discussion of the residue theorem for product integral can be found in the book [DF]; for example, if the set

$$\sigma(A) - \sigma(A) = \{\lambda_1 - \lambda_2;\ \lambda_1 \text{ and } \lambda_2 \text{ are eigenvalues of matrix } A\}$$

doesn't contain any positive integers, then the statement (2.7.5) is shown to be true.

2.8 Linear differential equations at a singular point

In this section we assume that the reader is familiar with the basics of the theory of analytic functions (see e.g. [EH] or [VJ]). We are interested in studying the differential equation

$$Y'(z) = A(z)Y(z), \tag{2.8.1}$$

where the function A is holomorphic in the ring $P(z_0, R) = \{z \in \mathbf{C}; 0 < |z - z_0| < R\}$ and $R > 0$. If we choose $z_1 \in P(z_0, R)$ and denote $r = \min(|z_1 - z_0|, R - |z_1 - z_0|)$, then the function

$$Y_1(z) = \prod_{z_1}^{z} (I + A(w)\,\mathrm{d}w)$$

provides a solution of (2.8.1) in $B(z_1, r) = \{z \in \mathbf{C}; |z - z_1| < r\}$; the product integral is path-independent, because A is holomorphic in $U(z_1, r)$. The holomorphic function Y_1 can be continued along an arbitrary curve φ v $P(z_0, R)$; this procedure leads to a (multiple-valued) analytic function $\mathcal{Y}$, which will be denoted by

$$\mathcal{Y}(z) = \prod_{z_1}^{z} (I + A(w)\,\mathrm{d}w), \quad z \in P(z_0, R).$$

If the element $(z_1, Y_1) \in \mathcal{Y}$ is continued along a curve φ in $P(z_0, R)$ to an element $(z_2, Y_2) \in \mathcal{Y}$ (we write this as $(z_1, Y_1) \xrightarrow{\varphi} (z_2, Y_2)$), then (using $Y_1(z_1) = I$)

$$Y_2(z_2) = \prod_{\varphi} (I + A(w)\,\mathrm{d}w) \cdot Y_1(z_1) = \prod_{\varphi} (I + A(w)\,\mathrm{d}w).$$

Let φ be the circle with center z_0 which passes through the point z_1, i.e.

$$\varphi(t) = z_0 + (z_1 - z_0)\exp(it), \qquad t \in [0, 2\pi].$$

If $(z_1, Y_2) \in \mathcal{Y}$ is the element such that $(z_1, Y_1) \xrightarrow{\varphi} (z_1, Y_2)$, then

$$\frac{\mathrm{d}}{\mathrm{d}z} Y_1 = \frac{\mathrm{d}}{\mathrm{d}z} Y_2 = A(z)$$

for $z \in B(z_1, r)$. Consequently, there is a matrix $C \in \mathbf{C}^{n \times n}$ such that $Y_2(z) = Y_1(z) \cdot C$. Substituting $z = z_1$ gives

$$C = \prod_{\varphi} (I + A(w)\,\mathrm{d}w).$$

Volterra refers[1] to the point z_0 as *point de ramification abélien* of the analytic function $\mathcal{Y}$; this means that it is the branch point of $\mathcal{Y}$, but not of its derivative A, which is a single-valued function. Volterra proceeds to prove that $\mathcal{Y}$ can be written in the form

$$\mathcal{Y} = \mathcal{S}_1 \cdot \mathcal{S}_2,$$

where $\mathcal{S}_1$ is single-valued in $P(z_0, R)$ and $\mathcal{S}_2$ is an analytic function that is uniquely determined by the matrix $C = \prod_{\varphi} (I + A(w)\,\mathrm{d}w)$.
Here is the proof[2]: We write $C = M^{-1} T M$, where T is a Jordan matrix. Then

$$T = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_k \end{pmatrix}, \quad \text{kde } T_j = \begin{pmatrix} e^{2\pi i \lambda_j} & 0 & \cdots & 0 & 0 \\ 1 & e^{2\pi i \lambda_j} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & e^{2\pi i \lambda_j} \end{pmatrix},$$

where we have expressed the eigenvalues of T in the form $\exp(2\pi i \lambda_j)$; this is certainly possible as the matrices C and consequently also T are regular, and thus have nonzero eigenvalues. We now define the analytic function

$$\mathcal{V}(z) = \prod_{z_1}^{z} \left(I + \frac{U}{w - z_0}\,\mathrm{d}w \right), \quad z \in P(z_0, R),$$

where

$$U = \begin{pmatrix} U_1 & 0 & \cdots & 0 \\ 0 & U_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U_k \end{pmatrix} \quad \text{and} \quad U_j = \begin{pmatrix} \lambda_j & 0 & \cdots & 0 & 0 \\ 1 & \lambda_j & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_j \end{pmatrix}$$

and U_j has the same dimensions as T_j for $j \in \{1, \ldots, k\}$. Consider a function element (z_1, V_1) of $\mathcal{V}$; what happens if we continue it along the circle φ? As in the case of function $\mathcal{Y}$ we obtain the result

$$(z_1, V_1) \xrightarrow{\varphi} (z_1, V_1 \cdot D),$$

[1] [VH], p. 121
[2] [VH], p. 122–124

where

$$D = \prod_{\varphi} \left(I + \frac{U}{w - z_0} \, \mathrm{d}w \right) = \prod_{0}^{2\pi} (I + iU \, \mathrm{d}w).$$

In Example 2.7.10 we have calculated the result

$$D = S(2\pi) \cdot S(0)^{-1} = S(2\pi) = \begin{pmatrix} S_1(2\pi) & 0 & \cdots & 0 \\ 0 & S_2(2\pi) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(2\pi) \end{pmatrix},$$

where

$$S_j(x) = \begin{pmatrix} e^{i\lambda_j x} & 0 & 0 & \cdots & 0 & 0 \\ \frac{x^1}{1!} e^{i\lambda_j x} & e^{i\lambda_j x} & 0 & \cdots & 0 & 0 \\ \frac{x^2}{2!} e^{i\lambda_j x} & \frac{x^1}{1!} e^{i\lambda_j x} & e^{i\lambda_j x} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \end{pmatrix}.$$

The matrix D is similar to the Jordan matrix T; thus

$$D = N^{-1} T N,$$

and consequently

$$(z_1, V_1) \xrightarrow{\varphi} (z_1, V_1 N^{-1} T N),$$

$$(z_1, Y_1) \xrightarrow{\varphi} (z_1, Y_1 M^{-1} T M).$$

We now consider the analytic function

$$\mathcal{S}_1(z) = \mathcal{Y}(z) M^{-1} N \mathcal{V}(z)^{-1}$$

and continue its element along φ:

$$(z_1, Y_1 M^{-1} N V_1^{-1}) \xrightarrow{\varphi} (z_1, Y_1 M^{-1} T M M^{-1} N (V_1 N^{-1} T N)^{-1}) = (z_1, Y_1 M^{-1} N V_1^{-1}).$$

Thus the analytic function $\mathcal{S}_1$ is in fact single-valued in $P(z_0, R)$. The proof is finished by putting

$$\mathcal{S}_2(z) = \mathcal{V}(z) N^{-1} M.$$

Consequently

$$\mathcal{Y} = \mathcal{S}_1 \cdot \mathcal{S}_2$$

and $\mathcal{S}_2$ is uniquely determined by the matrix C.

We now briefly turn our attention to the analytic function $\mathcal{V}$. Assume that

$$(z_1, V_1) \xrightarrow{\psi} (z, V_2),$$

where $\psi : [a,b] \to P(z_0, R)$, $\psi(a) = z_1$, $\psi(b) = z$. It is known from complex analysis that given the curve φ, we can find a function a function $g : [a,b] \to \mathbf{C}$ such that

$$\exp(g(t)) = \psi(t) - z_0,$$

for every $t \in [a,b]$; g is a continuous branch of logarithm of the function $\psi - z_0$. For convenience we will use the notation

$$g(t) = \log(\psi(t) - z_0)$$

with the understanding that g is defined as above. We also have

$$g'(t) = \frac{\psi'(t)}{\psi(t) - z_0}$$

for every $t \in [a,b]$. We now calculate

$$V_2(z) = \prod_{\psi} \left(I + \frac{U}{z - z_0}\,\mathrm{d}z \right) = \prod_{a}^{b} \left(I + \frac{U}{\psi(t) - z_0}\,\psi'(t)\,\mathrm{d}t \right).$$

Substituting $v = g(t)$ gives

$$V_2(z) = \prod_{g(a)}^{g(b)} (I + U\,\mathrm{d}v) = S(g(b))S(g(a))^{-1},$$

where

$$S(z) = \begin{pmatrix} S_1(z) & 0 & \cdots & 0 \\ 0 & S_2(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k(z) \end{pmatrix}$$

is a block diagonal matrix composed of the matrices

$$S_j(z) = \begin{pmatrix} e^{\lambda_j z} & 0 & 0 & \cdots & 0 & 0 \\ \frac{z^1}{1!}e^{\lambda_j z} & e^{\lambda_j z} & 0 & \cdots & 0 & 0 \\ \frac{z^2}{2!}e^{\lambda_j z} & \frac{z^1}{1!}e^{\lambda_j z} & e^{\lambda_j z} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}.$$

Consequently, the solution of Equation (2.8.1), i.e. the analytic function $\mathcal{Y}$, can be expressed as

$$\mathcal{Y}(z) = \mathcal{S}_1(z)\mathcal{S}_2(z) = \mathcal{S}_1(z)S(g(b))S(g(a))^{-1}N^{-1}M, \tag{2.8.2}$$

where

$$S(g(b)) = S(\log(z - z_0)),$$

$$S_j(\log(z-z_0)) =$$

$$= \begin{pmatrix} (z-z_0)^{\lambda_j} & 0 & 0 & \cdots & 0 & 0 \\ (z-z_0)^{\lambda_j}\log(z-z_0) & (z-z_0)^{\lambda_j} & 0 & \cdots & 0 & 0 \\ (z-z_0)^{\lambda_j}\frac{\log^2(z-z_0)}{2!} & (z-z_0)^{\lambda_j}\log(z-z_0) & (z-z_0)^{\lambda_j} & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \end{pmatrix}.$$

The above result can be applied to obtain the general form of solution of the differential equation

$$y^{(n)}(z) + p_1(z)y^{(n-1)}(z) + \cdots + p_n(z)y(z) = 0, \tag{2.8.3}$$

where the functions p_i are holomorphic in $P(z_0, R)$. We have seen that this equation of the n-th order is easily converted to a system of linear differential equations of first order, which can be written in the vector form as

$$y'(z) = A(z)y(z),$$

where A is a holomorphic matrix function in $P(z_0, R)$. The fundamental matrix of this system is given by (2.8.2); the first row of this matrix then yields the fundamental system of solutions (composed of n analytic functions) of Equation (2.8.3). From the form of Equation (2.8.2) we infer that every solution of Equation (2.8.3) can be expressed as a linear combination of analytic functions of the form

$$(z-z_0)^{\lambda_j}\left(\varphi_0^j(z) + \varphi_1^j(z)\log(z-z_0) + \cdots + \varphi_{n_j}^j(z)\log^{n_j}(z-z_0)\right),$$

where φ_k^j are holomorphic functions in $P(z_0, R)$.

Thus we see that Volterra was able to obtain the result of Lazarus Fuchs (see Chapter 1) using the theory of product integration. The next chapters of Volterra's book [VH] are concerned with the study of analytic functions on Riemann surfaces; the topic is rather special and we don't follow Volterra's treatment here.

Chapter 3

Lebesgue product integration

While it is sufficient to use the Riemann integral in applications, it is rather unsatisfactory from the viewpoint of theoretical mathematics. The generalization of Riemann integral due to Henri Lebesgue is based on the notion of measure. The problem of extending Volterra's definition of product integral in a similar way has been solved by Ludwig Schlesinger.

Volterra's and Schlesinger's works differ in yet another way: Volterra did not worry about using infinitesimal quantities, and it is not always easy to translate his ideas into the language of modern mathematics. Schlesinger's proofs are rather precise and can be read without greater effort except for occasionally strange notation. The foundations of mathematical analysis in 1930's were firmer than in 1887; moreover, Schlesinger inclined towards theoretical mathematics, as opposed to Volterra, who always kept applications in mind.

Ludwig Schlesinger[1]

Schlesinger's biographies can be found in [Lex, McT]: Ludwig (Lajos in Hungarian) Schlesinger was born on the 1st November 1864 in a Hungarian town Trnava (Nagyszombat), which now belongs to Slovakia. He studied mathematics and physics at the universities of Heidelberg and Berlin, where he received a doctorate in 1887.

[1] Photo from [McT]

The advisors of his thesis (which was concerned with homogeneous linear differential equations of the fourth order) were Lazarus Fuchs (who later became his father-in-law) and Leopold Kronecker. Two years later Schlesinger became an associate professor in Berlin and in 1897 an invited professor at the University of Bonn. During the years 1897 to 1911 he served as an ordinary professor and also as the head of the department of higher mathematics at the University of Kolozsvár (now Cluj in Romania). In 1911 he moved to Giessen in Germany where he continued to teach until his retirement in 1930. Ludwig Schlesinger died on the 16th December 1933.

Schlesinger devoted himself especially to complex function theory and linear differential equations; he also made valuable contributions to the history of mathematics. He translated Descartes' *Geometrie* into German, and was one of the organizers of the centenary festivities dedicated to the hundredth anniversary of János Bolyai, one of the pioneers of non-Euclidean geometry. The most important works of Schlesinger include *Handbuch der Theorie der linearen Differentialgleichungen* (1895–98), *J. Bolyai in Memoriam* (1902), *Vorlesungen über lineare Differentialgleichungen* (1908) and *Raum, Zeit und Relativitätstheorie* (1920).

Schlesinger's paper on product integration called *Neue Grundlagen für einen Infinitesimalkalkul der Matrizen* [LS1] was published in 1931. The author links up to Volterra's theory of product integral. He starts with the Riemann-type definition and establishes the basic properties of the product integral. His proofs are nevertheless original – while Volterra proved most of his statements using the Peano series expansion, Schlesinger prefers the "$\varepsilon - \delta$" proofs. He then proceeds to define the Lebesgue product integral (as a limit of product integrals of step functions) and explores its properties.

A continuation of this paper appeared in 1932 under the title *Weitere Beiträge zum Infinitesimalkalkul der Matrizen* [LS2]. Schlesinger again studies the properties of Lebesgue product integral and is also concerned with contour product integration in $\mathbf{R}^2$ and in $\mathbf{C}$.

This chapter summarizes the most important results from both Schlesinger's papers; the final section then presents a generalization of Schlesinger's definition of the Lebesgue product integral.

3.1 Riemann integrable matrix functions

When dealing with product integral we need to work with sequences of matrices and their limits. Volterra was mainly working with the individual entries of the matrices and convergence of a sequence of matrices was for him equivalent to convergence of all entries.

Schlesinger chooses a different approach: He defines the norm of a matrix $A = \{a_{ij}\}_{i,j=1}^n$ by

$$[A] = n \cdot \max_{1 \le i,j \le n} |a_{ij}|.$$

He also mentions another norm

$$\Omega_A = \max\{|\lambda|;\ \lambda \text{ is an eigenvalue of } A\}$$

and states that

$$\Omega_A \leq \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} \leq [A].$$

The second inequality is obvious, the first is proved in [LS1][1].

Schlesinger's norm $[A]$ has the nice property that $[A \cdot B] \leq [A] \cdot [B]$ for every $A, B \in \mathbf{R}^{n\times n}$, but its disadvantage is that $[I] = n$. In the following text we will use the operator norm

$$\|A\| = \sup\{\|Ax\|;\ \|x\| \leq 1\},$$

where $\|Ax\|$ and $\|x\|$ denote the Euclidean norms of vectors Ax, $x \in \mathbf{R}^n$. This simplifies Schlesinger's proofs slightly, because $\|I\| = 1$ and

$$\|A \cdot B\| \leq \|A\| \cdot \|B\|$$

still holds for every $A, B \in \mathbf{R}^{n\times n}$. It should be noted that the space $\mathbf{R}^{n\times n}$ is finite-dimensional, therefore it doesn't matter which norm we choose since they are all equivalent.

The convergence of a sequence of matrices and the limit of a matrix function is now defined in a standard way using the norm introduced above.

For an arbitrary matrix function $A : [a, b] \to \mathbf{R}^{n\times n}$ and a tagged partition

$$D:\ a = t_0 \leq \xi_1 \leq t_1 \leq \xi_2 \leq \cdots \leq t_{m-1} \leq \xi_m \leq t_m = b$$

of interval $[a, b]$ with division points t_i and tags ξ_i we denote

$$P(A, D) = \prod_{k=1}^{m} (I + A(\xi_k)\Delta t_k),$$

where $\Delta t_k = t_k - t_{k-1}$.

Schlesinger is now interested in the limit value of $P(A, D)$ as the lengths of the intervals $[t_{k-1}, t_k]$ approach zero (if the limit exists independently on the choice of $\xi_k \in [t_{k-1}, t_k]$). Clearly, the limit is nothing else than Volterra's right product integral.

Definition 3.1.1. Consider function $A : [a, b] \to \mathbf{R}^{n\times n}$. In case the limit

$$\lim_{\nu(D)\to 0} P(A, D)$$

exists, it is called the product integral of function A on interval $[a, b]$ and denoted by the symbol

$$(I + A(t)\,\mathrm{d}t) \prod_{a}^{b}.$$

[1] [LS1], p. 34–35

Remark 3.1.2. Schlesinger in fact defines the product integral as the limit of the products

$$P(A,D) = Y^0 \prod_{k=1}^{m} (I + A(\xi_k)\Delta t_k),$$

where Y^0 is an arbitrary regular matrix (which plays the role of an "integration constant"). In the following text we assume for simplicity that $Y^0 = I$. Also, instead of Schlesinger's notation

$$\overset{b}{\underset{a}{\widehat{\int}}} (I + A(x)\,\mathrm{d}x)$$

we use the symbol $(I + A(x)\,\mathrm{d}x)\prod_a^b$ to denote the product integral.

Lemma 3.1.3.[1] Let $A_1, A_2, \ldots, A_m \in \mathbf{R}^{n\times n}$ be arbitrary matrices. Then

$$\|(I + A_1)(I + A_2)\cdots(I + A_m)\| \le \exp\left(\sum_{k=1}^{m} \|A_k\|\right).$$

Proof. A simple consequence of the inequalities

$$\|I + A_k\| \le 1 + \|A_k\| \le \exp\|A_k\|.$$

□

Corollary 3.1.4.[2] If $\|A(x)\| \le M$ for every $x \in [a,b]$, then $\|P(A,D)\| \le e^{M(b-a)}$ for every tagged partition D of interval $[a,b]$.

Corollary 3.1.5. If the function $A : [a,b] \to \mathbf{R}^{n\times n}$ is product integrable and $\|A(x)\| \le M$ for every $x \in [a,b]$, then

$$\left\| (I + A(x)\,\mathrm{d}x)\prod_a^b \right\| \le e^{M(b-a)}.$$

Schlesinger's first task is to prove the existence of product integral for Riemann integrable matrix functions, i.e. functions $A : [a,b] \to \mathbf{R}^{n\times n}$ whose entries a_{ij} are Riemann integrable on $[a,b]$. The proof is substantially different from the proof given by Volterra; the technique is similar to Cauchy's proof of the existence of $\int_a^b f$ for a continuous function f (see [CE, SŠ]).

Definition 3.1.6. Consider function $A : [a,b] \to \mathbf{R}^{n\times n}$ and let $[c,d] \subseteq [a,b]$. The oscillation of A on interval $[c,d]$ is the number

$$\operatorname{osc}(A,[c,d]) = \sup\{\|A(\xi_1) - A(\xi_2)\|;\ \ \xi_1, \xi_2 \in [c,d]\}.$$

[1] [LS1], p. 37
[2] [LS1], p. 38

The following characterization of Riemann integrable function will be needed in subsequent proofs:

Lemma 3.1.7. If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\lim_{\nu(D)\to 0} \sum_{k=1}^{m} \operatorname{osc}(A, [t_{k-1}, t_k])\Delta t_k = 0.$$

Proof. The statement follows easily from Darboux's definition of the Riemann integral which is based on upper and lower sums; it is in fact equivalent to Riemann integrability of the given function (see e.g. [Sch2]). $\square$

Definition 3.1.8. We say that a tagged partition D' is a refinement of a tagged partition D (we write $D' \prec D$), if every division point of D is also a division point of D' (no condition being imposed on the tags).

Lemma 3.1.9.[1] Let the function $A : [a,b] \to \mathbf{R}^{n\times n}$ be such that $\|A(x)\| \le M$ for every $x \in [a,b]$. Then for every pair of tagged partitions D, D' of interval $[a,b]$ such that $D' \prec D$ we have

$$\|P(A,D) - P(A,D')\| \le e^{M(b-a)} \sum_{k=1}^{m} (\operatorname{osc}(A, [t_{k-1}, t_k])\Delta t_k + (M\Delta t_k)^2 e^{M\Delta t_k}),$$

where t_i, $i = 0, \ldots, m$ are division points of the partition D.

Proof. Let the partition D consist of division points and tags

$$D: \; a = t_0 \le \xi_1 \le t_1 \le \xi_2 \le \cdots \le t_{m-1} \le \xi_m \le t_m = b.$$

First, we refine it only on the subinterval $[t_{k-1}, t_k]$, i.e. we consider a partition D^* which contains division points and tags

$$t_{k-1} = u_0 \le \eta_1 \le u_1 \cdots \le u_{l-1} \le \eta_l \le u_l = t_k$$

and coincides with the partition D on the rest of interval $[a,b]$. Then

$$\|P(A,D^*) - P(A,D)\| \le \left\| \prod_{i=1}^{k-1} (I + A(\xi_i)\Delta t_i) \right\| \cdot$$

$$\cdot \left\| \prod_{j=1}^{l} (I + A(\eta_j)\Delta u_j) - I - A(\xi_k)\Delta t_k \right\| \cdot \left\| \prod_{i=k+1}^{m} (I + A(\xi_i)\Delta t_i) \right\|.$$

We estimate

$$\left\| \prod_{i=1}^{k-1} (I + A(\xi_i)\Delta t_i) \right\| \cdot \left\| \prod_{i=k+1}^{m} (I + A(\xi_i)\Delta t_i) \right\| \le e^{M(b-a)}$$

---

[1] [LS1], p. 39–41

and

$$\left\|\prod_{j=1}^{l}(I+A(\eta_j)\Delta u_j)-I-A(\xi_k)\Delta t_k\right\| \le \left\|\sum_{j=1}^{l}(A(\eta_j)-A(\xi_k))\Delta u_j\right\| +$$

$$+\left\|\sum_{p=2}^{l}\sum_{1\le r_1<\cdots<r_p\le l} A(\eta_{r_1})\cdots A(\eta_{r_p})\Delta u_{r_1}\cdots\Delta u_{r_p}\right\| \le$$

$$\le \operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k + \sum_{p=2}^{l}\sum_{1\le r_1<\cdots<r_p\le l} M^p\,\Delta u_{r_1}\cdots\Delta u_{r_p} =$$

$$= \operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k + \prod_{j=1}^{l}(1+M\Delta u_j)-1-\sum_{j=1}^{l}M\Delta u_j \le$$

$$\le \operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k + e^{M\Delta t_k}-1-M\Delta t_k \le \operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k+(M\Delta t_k)^2 e^{M\Delta t_k}.$$

Therefore we conclude that

$$\|P(A,D)-P(A,D^*)\| \le e^{M(b-a)}(\operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k + (M\Delta t_k)^2 e^{M\Delta t_k}).$$

Now, since the given partition D' can be obtained from D by successively refining the subintervals $[t_0,t_1],\dots,[t_{m-1},t_m]$, we obtain

$$\|P(A,D)-P(A,D')\| \le e^{M(b-a)}\sum_{k=1}^{m}(\operatorname{osc}(A,[t_{k-1},t_k])\Delta t_k + (M\Delta t_k)^2 e^{M\Delta t_k}).$$

□

Corollary 3.1.10.[1] Consider a Riemann integrable function $A:[a,b]\to\mathbf{R}^{n\times n}$. Then for every $\varepsilon>0$ there exists $\delta>0$ such that

$$\|P(A,D)-P(A,D')\| < \varepsilon$$

whenever $\nu(D)<\delta$ and $D'\prec D$.

Proof. The statement follows from the previous lemma, Lemma 3.1.7 and the estimate

$$\sum_{k=1}^{m}(M\Delta t_k)^2 e^{M\Delta t_k} \le \nu(D)M^2 e^{M\nu(D)}\sum_{k=1}^{m}\Delta t_k = (b-a)\nu(D)M^2 e^{M\nu(D)}.$$

□

[1] [LS1], p. 39–41

Theorem 3.1.11.[1] The product integral $(I+A(x)\,\mathrm{d}x)\prod_a^b$ exists for every Riemann integrable function $A:[a,b]\to\mathbf{R}^{n\times n}$.

Proof. Take $\varepsilon>0$. Corollary 3.1.10 guarantees the existence of a $\delta>0$ such that

$$\|P(A,D)-P(A,D')\|<\varepsilon/2$$

whenever $\nu(D)<\delta$ and $D'\prec D$. Consider a pair of tagged partitions D_1, D_2 of interval $[a,b]$ satisfying $\nu(D_1)<\delta$ and $\nu(D_2)<\delta$. These partitions have a common refinement, i.e. a partition D such that $D\prec D_1$, $D\prec D_2$ (the tags in D can be chosen arbitrarily). Then

$$\|P(A,D_1)-P(A,D_2)\|\le\|P(A,D_1)-P(A,D)\|+\|P(A,D)-P(A,D_2)\|<\varepsilon.$$

We have proved that every Riemann integrable function $A:[a,b]\to\mathbf{R}^{n\times n}$ satisfies a certain Cauchy condition and this is also the end of Schlesinger's proof; the existence of product integral follows from the Cauchy condition in the same way as in the analoguous theorem for the ordinary Riemann integral (see e.g. [Sch2]). $\square$

Theorem 3.1.12.[2] Consider a Riemann integrable function $A:[a,b]\to\mathbf{R}^{n\times n}$. If $c\in[a,b]$, then

$$(I+A(x)\,\mathrm{d}x)\prod_a^b=(I+A(x)\,\mathrm{d}x)\prod_a^c\cdot(I+A(x)\,\mathrm{d}x)\prod_c^b.$$

Proof. As Schlesinger remarks, the proof follows directly from the definition of product integral (see the proof in Chapter 2). $\square$

3.2 Matrix exponential function

Let $A:[a,b]\to\mathbf{R}^{n\times n}$ be a constant function. If D_m is a partition of $[a,b]$ to m subintervals of length $(b-a)/m$, then

$$P(A,D_m)=\left(I+\frac{b-a}{m}A\right)^m.$$

Since $\nu(D_m)\to 0$ as $m\to\infty$, we have

$$(I+A(x)\,\mathrm{d}x)\prod_a^b=\lim_{m\to\infty}\left(I+\frac{b-a}{m}A\right)^m=e^{(b-a)A}.$$

The last equality follows from the fact that $e^A=\lim_{m\to\infty}(I+A/m)^m$ for every $A\in\mathbf{R}^{n\times n}$; recall that the matrix exponential was defined in Chapter 2 using the series

$$e^A=\sum_{m=0}^{\infty}\frac{A^m}{m!}. \tag{3.2.1}$$

[1] [LS1], p. 41
[2] [LS1], p. 41

Lemma 3.2.1. If $A_1, \dots, A_m \in \mathbf{R}^{n\times n}$ and $B_1, \dots, B_m \in \mathbf{R}^{n\times n}$, then

$$\prod_{i=1}^m A_i - \prod_{i=1}^m B_i = \sum_{i=1}^m \left(\prod_{j=1}^{i-1} B_j \cdot (A_i - B_i) \cdot \prod_{j=i+1}^m A_j \right).$$

Proof.

$$\prod_{i=1}^m A_i - \prod_{i=1}^m B_i = \sum_{i=1}^m (B_1 \cdots B_{i-1} A_i \cdots A_m - B_1 \cdots B_i A_{i+1} \cdots A_m) =$$

$$= \sum_{i=1}^m \left(\prod_{j=1}^{i-1} B_j \cdot (A_i - B_i) \cdot \prod_{j=i+1}^m A_j \right).$$

□

Theorem 3.2.2.[1] Consider a Riemann integrable function $A : [a,b] \to \mathbf{R}^{n\times n}$. Then

$$\lim_{\nu(D)\to 0} \prod_{k=1}^m e^{A(\xi_k)\Delta t_k} = \lim_{\nu(D)\to 0} \prod_{k=1}^m (I + A(\xi_k)\Delta t_k) = (I + A(t)\,\mathrm{d}t) \prod_a^b .$$

Proof. Since every Riemann integrable function is bounded, we have $\|A(x)\| \le M$ for some $M \in \mathbf{R}$ and for every $x \in [a,b]$. The definition of matrix exponential (3.2.1) implies

$$\left\| e^{A(\xi_k)\Delta t_k} - (I + A(\xi_k)\Delta t_k) \right\| \le (\|A(\xi_k)\|\Delta t_k)^2 e^{\|A(\xi_k)\|\Delta t_k} \le (M\Delta t_k)^2 e^{M\Delta t_k}$$

for $k = 1, \dots, m$. According to Lemma 3.2.1,

$$\left\| \prod_{k=1}^m e^{A(\xi_k)\Delta t_k} - \prod_{k=1}^m (I + A(\xi_k)\Delta t_k) \right\| =$$

$$= \left\| \sum_{j=1}^m \left(\prod_{k=1}^{j-1} (I + A(\xi_k)\Delta t_k) \cdot (e^{A(\xi_j)\Delta t_j} - I - A(\xi_j)\Delta t_j) \cdot \prod_{k=j+1}^m e^{A(\xi_k)\Delta t_k} \right) \right\| \le$$

$$\le e^{M(b-a)} \sum_{j=1}^m \left\| e^{A(\xi_j)\Delta t_j} - I - A(\xi_j)\Delta t_j \right\| \le e^{M(b-a)} M^2 \sum_{j=1}^m (\Delta t_j)^2 e^{M\Delta t_j} \le$$

$$\le e^{M(b-a)} M^2 \nu(D) e^{M\nu(D)} \sum_{j=1}^m \Delta t_j = (b-a) e^{M(b-a)} M^2 \nu(D) e^{M\nu(D)}.$$

[1] [LS1], p. 42

By choosing a sufficiently fine partition D of $[a,b]$, the last expression can be made arbitrarily small. □

Definition 3.2.3. The trace of a matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ is the number

$$\operatorname{Tr} A = \sum_{i=1}^{n} a_{ii}.$$

Theorem 3.2.4.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\det\left((I + A(x)\,\mathrm{d}x)\prod_a^b\right) = \exp\left(\int_a^b \operatorname{Tr} A(x)\,\mathrm{d}x\right).$$

Proof.

$$\det\left((I + A(x)\,\mathrm{d}x)\prod_a^b\right) = \det\left(\lim_{\nu(D)\to 0}\prod_{k=1}^{m} e^{A(\xi_k)\Delta t_k}\right) = \lim_{\nu(D)\to 0}\prod_{k=1}^{m}\det\, e^{A(\xi_k)\Delta t_k} =$$

$$= \lim_{\nu(D)\to 0}\prod_{k=1}^{m} e^{\operatorname{Tr} A(\xi_k)\Delta t_k} = \lim_{\nu(D)\to 0}\exp\left(\sum_{k=1}^{m}\operatorname{Tr} A(\xi_k)\Delta t_k\right) = \exp\left(\int_a^b \operatorname{Tr} A(x)\,\mathrm{d}x\right)$$

(we have used a theorem from linear algebra: $\det \exp A = \exp \operatorname{Tr} A$). □

Remark 3.2.5. This formula (sometimes called the Jacobi formula) appeared already in Volterra's work. Schlesinger employs a different proof and his statement is also more general – it requires only the Riemann integrability of A, in contrast to Volterra's assumption that A is continuous.

Corollary 3.2.6. If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then the product integral $(I + A(x)\,\mathrm{d}x)\prod_a^b$ is a regular matrix.

Recall that Volterra has also assigned meaning to product integrals whose lower limit is greater than the upper limit; his definition for the right integral was

$$(I + A(t)\,\mathrm{d}t)\prod_b^a = \lim_{\nu(D)\to 0}\prod_{k=m}^{1}(I - A(\xi_k)\Delta t_k).$$

If A is Riemann integrable, we know that this is equivalent to

$$(I + A(t)\,\mathrm{d}t)\prod_b^a = \lim_{\nu(D)\to 0}\prod_{k=m}^{1} e^{-A(\xi_k)\Delta t_k}.$$

[1] [LS1], p. 43–44

Thus

$$I = \lim_{\nu(D)\to 0}\left(\prod_{k=1}^{m} e^{A(\xi_k)\Delta t_k}\cdot\prod_{k=m}^{1} e^{-A(\xi_k)\Delta t_k}\right) =$$

$$= \lim_{\nu(D)\to 0}\prod_{k=1}^{m} e^{A(\xi_k)\Delta t_k}\cdot\lim_{\nu(D)\to 0}\prod_{k=m}^{1} e^{-A(\xi_k)\Delta t_k} =$$

$$(I + A(t)\,\mathrm{d}t)\prod_{a}^{b}\cdot(I + A(t)\,\mathrm{d}t)\prod_{b}^{a},$$

which proves that $(I+A(t)\,\mathrm{d}t)\prod_b^a$ is the inverse matrix of $(I+A(t)\,\mathrm{d}t)\prod_a^b$; compare with Volterra's proof of Theorem 2.4.10.

3.3 The indefinite product integral

Schlesinger now proceeds to study the properties of the indefinite product integral, i.e. of the function $Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x$.

Theorem 3.3.1.[1] If $A : [a, b] \to \mathbf{R}^{n\times n}$ is Riemann integrable, then the function $Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x$ is continuous on $[a, b]$.

Proof. We prove the right-continuity of Y at $x_0 \in [a, b)$; continuity from left is proved similarly. Let $x_0 \le x_0 + h \le b$. The function A is bounded: $\|A(x)\| \le M$ for some $M \in \mathbf{R}$. We now employ the inequality from Lemma 3.1.9. Let D' be a partition of interval $[x_0, x_0 + h]$. Then

$$\|I + A(x_0)h - P(A, D')\| \le e^{Mh}(\mathrm{osc}(A, [x_0, x_0 + h])h + (Mh)^2 e^{Mh}).$$

Passing to the limit $\nu(D') \to 0$ we obtain

$$\left\| I + A(x_0)h - (I + A(t)\,\mathrm{d}t)\prod_{x_0}^{x_0+h}\right\| \le e^{Mh}(\mathrm{osc}(A, [x_0, x_0 + h])h + (Mh)^2 e^{Mh}),$$

which implies

$$\lim_{h\to 0+}(I + A(t)\,\mathrm{d}t)\prod_{x_0}^{x_0+h} = I.$$

Therefore

$$\lim_{h\to 0+}(Y(x_0 + h) - Y(x_0)) = Y(x_0)\left((I + A(t)\,\mathrm{d}t)\prod_{x_0}^{x_0+h} - I\right) = 0.$$

□

[1] [LS1], p. 44–46

Theorem 3.3.2.[1] If $A : [a,b] \to \mathbf{R}^{n\times n}$ is Riemann integrable, then the function

$$Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x,$$

satisfies the integral equation

$$Y(x) = I + \int_a^x Y(t)A(t)\,\mathrm{d}t, \quad x \in [a,b].$$

Proof. It is sufficient to prove the statement for $x = b$. Let

$$D:\ a = t_0 \le \xi_1 \le t_1 \le \xi_2 \le \cdots \le t_{m-1} \le \xi_m \le t_m = b$$

be a tagged partition of interval $[a,b]$. We define

$$Y^k = \prod_{i=1}^{k}(I + A(\xi_i)\Delta t_i), \quad k = 0,\ldots,m.$$

Then

$$Y^k - Y^{k-1} = Y^{k-1}A(\xi_k)\Delta t_k, \quad k = 1,\ldots,m. \tag{3.3.1}$$

Since $Y^0 = I$ and $Y^m = P(A,D)$, adding the equalities (3.3.1) for $k = 1,\ldots,m$ yields

$$P(A,D) - I = \sum_{k=1}^{m} Y^{k-1}A(\xi_k)\Delta t_k.$$

The function A is bounded: $\|A(x)\| \le M$ for some $M \in \mathbf{R}$. We estimate

$$\left\|Y(b) - I - \int_a^b Y(t)A(t)\,\mathrm{d}t\right\| \le \|Y(b) - P(A,D)\| +$$

$$+\left\|P(A,D) - I - \int_a^b Y(t)A(t)\,\mathrm{d}t\right\| \le \|Y(b) - P(A,D)\| +$$

$$+\left\|\sum_{k=1}^{m}(Y^{k-1} - Y(t_{k-1}))A(\xi_k)\Delta t_k\right\| + \left\|\sum_{k=1}^{m}(Y(t_{k-1}) - Y(\xi_k))A(\xi_k)\Delta t_k\right\| +$$

$$+\left\|\sum_{k=1}^{m} Y(\xi_k)A(\xi_k)\Delta t_k - \int_a^b Y(t)A(t)\,\mathrm{d}t\right\|. \tag{3.3.2}$$

[1] [LS1], p. 46–47

Using the inequalities

$$\left\|\sum_{k=1}^{m}(Y^{k-1}-Y(t_{k-1}))A(\xi_k)\Delta t_k\right\| \le M\sum_{k=1}^{m}\|Y^{k-1}-Y(t_{k-1})\|\Delta t_k \le$$

$$\le M\sum_{k=1}^{m}e^{M(b-a)}\Delta t_k\left(\sum_{j=1}^{m}(\operatorname{osc}(A,[t_{j-1},t_j])\Delta t_j+(M\Delta t_j)^2e^{M\Delta t_j})\right)\le$$

$$\le Me^{M(b-a)}(b-a)\left(\sum_{j=1}^{m}\operatorname{osc}(A,[t_{j-1},t_j])\Delta t_j+M^2\nu(D)e^{M\nu(D)}\right)$$

(we have used Lemma 3.1.9) and

$$\left\|\sum_{k=1}^{m}(Y(t_{k-1})-Y(\xi_k))A(\xi_k)\Delta t_k\right\| \le M\sum_{k=1}^{m}\operatorname{osc}(Y,[t_{k-1},t_k])\Delta t_k,$$

we see that all terms on the right-hand side of (3.3.2) can be made arbitrarily small if the partition D is sufficiently fine. □

Corollary 3.3.3.[1] If $A:[a,b]\to\mathbf{R}^{n\times n}$ is continuous, then the function

$$Y(x)=(I+A(t)\,\mathrm{d}t)\prod_{a}^{x}$$

provides a solution of the differential equation

$$Y'(x)=Y(x)A(x),\quad x\in[a,b]$$

and satisfies the initial condition $Y(a)=I$.

Remark 3.3.4. The function Y is therefore the fundamental matrix of the system

$$y_i'(x)=\sum_{j=1}^{n}a_{ji}(x)y_j(x),\quad i=1,\dots,n.$$

Schlesinger uses the notation $D_xY(x)=A(x)$, where

$$D_xY=Y^{-1}Y',$$

i.e. D_x is exactly Volterra's right derivative of a matrix function.

3.4 Product integral inequalities

In this section we summarize various inequalities that will be useful later.

[1] [LS1], p. 47–48

Lemma 3.4.1.[1] If $A : [a, b] \to \mathbf{R}^{n\times n}$ is a Riemann integrable function, then

$$\left\| (I + A(x)\,\mathrm{d}x) \prod_a^b \right\| \le \exp\left(\int_a^b \|A(x)\|\,\mathrm{d}x \right).$$

Proof. Lemma 3.1.3 implies that

$$\left\| \prod_{i=1}^m (I + A(\xi_i)\Delta t_i) \right\| \le \exp\left(\sum_{i=1}^m \|A(\xi_i)\|\Delta t_i \right)$$

for every tagged partition D of interval $[a, b]$; the proof is completed by passing to the limit $\nu(D) \to 0$. □

Lemma 3.4.2.[2] Let $m \in \mathbf{N}$, $A_k, B_k \in \mathbf{R}^{n\times n}$ for every $k = 1, \ldots, m$. Then

$$\left\| \prod_{k=1}^m (I + B_k) - \prod_{k=1}^m (I + A_k) \right\| \le \exp\left(\sum_{k=1}^m \|A_k\| \right) \left(\exp\left(\sum_{k=1}^m \|B_k - A_k\| \right) - 1 \right).$$

Proof. Define

$$Y^k = \prod_{i=1}^k (I + A_i), \quad Z^k = \prod_{i=1}^k (I + B_i), \quad k = 0, \ldots, m$$

(where the empty product for $k = 0$ equals the identity matrix). Then

$$Y^k - Y^{k-1} = Y^{k-1}A_k,$$

$$Z^k - Z^{k-1} = Z^{k-1}B_k,$$

for $k = 1, \ldots, m$. This implies

$$Z^k - Y^k = (Z^{k-1} - Y^{k-1})(I + B_k) + E_k, \tag{3.4.1}$$

where

$$E_k = Y^{k-1}(B_k - A_k).$$

Applying the equality (3.4.1) m times on the difference $Z^m - Y^m$ we obtain

$$Z^m - Y^m = \sum_{k=1}^{m-1} E_k (I + B_{k+1}) \cdots (I + B_m) + E_m.$$

[1] [LS1], p. 51
[2] [LS1], p. 52–53

We also estimate

$$\|E_k\| \le \exp\left(\sum_{i=1}^{k-1} \|A_i\|\right) \|B_k - A_k\|$$

(the empty sum for $k=0$ equals zero),

$$\|Z^m - Y^m\| \le \sum_{k=1}^{m-1} \exp\left(\sum_{i=1}^{k-1} \|A_i\|\right) \|B_k - A_k\| \cdot \exp\left(\sum_{i=k+1}^{m} (\|B_i - A_i\| + \|A_i\|)\right) +$$

$$+ \exp\left(\sum_{i=1}^{m-1} \|A_i\|\right) \|B_m - A_m\| =$$

$$= \sum_{k=1}^{m-1} \exp\left(\sum_{i\neq k} \|A_i\|\right) \|B_k - A_k\| \cdot \exp\left(\sum_{i=k+1}^{m} \|B_i - A_i\|\right) +$$

$$+ \exp\left(\sum_{i=1}^{m-1} \|A_i\|\right) \|B_m - A_m\| \le$$

$$\le \sum_{k=1}^{m} \exp\left(\sum_{i=1}^{m} \|A_i\|\right) \|B_k - A_k\| \cdot \exp\left(\sum_{i=k+1}^{m} \|B_i - A_i\|\right).$$

Since

$$\|B_k - A_k\| \le \exp\left(\|B_k - A_k\|\right) - 1,$$

we conclude that

$$\left\|\prod_{k=1}^{m} (I + B_k) - \prod_{k=1}^{m} (I + A_k)\right\| = \|Z^m - Y^m\| \le$$

$$\le \exp\left(\sum_{i=1}^{m} \|A_i\|\right) \sum_{k=1}^{m} \left(\left(\exp\left(\|B_k - A_k\|\right) - 1\right) \exp\left(\sum_{i=k+1}^{m} \|B_i - A_i\|\right)\right) =$$

$$= \exp\left(\sum_{i=1}^{m} \|A_i\|\right) \sum_{k=1}^{m} \left(\exp\left(\sum_{i=k}^{m} \|B_i - A_i\|\right) - \exp\left(\sum_{i=k+1}^{m} \|B_i - A_i\|\right)\right) =$$

$$= \exp\left(\sum_{i=1}^{m} \|A_i\|\right) \left(\exp\left(\sum_{i=1}^{m} \|B_i - A_i\|\right) - 1\right).$$

□

Corollary 3.4.3.[1] If A, $B : [a,b] \to \mathbf{R}^{n\times n}$ are Riemann integrable functions, then

$$\left\|(I + B(x)\,\mathrm{d}x) \prod_a^b - (I + A(x)\,\mathrm{d}x) \prod_a^b\right\| \le$$

[1] [LS1], p. 53

$$\leq \exp\left(\int_a^b \|A(x)\|\,\mathrm{d}x\right)\left(\exp\left(\int_a^b \|B(x)-A(x)\|\,\mathrm{d}x\right)-1\right).$$

Proof. The previous lemma ensures that for every tagged partition D of interval $[a,b]$ we have

$$\|P(B,D)-P(A,D)\| = \left\|\prod_{k=1}^{m}(I+B(\xi_k)\Delta t_k) - \prod_{k=1}^{m}(I+A(\xi_k)\Delta t_k)\right\| \leq$$

$$\leq \exp\left(\sum_{k=1}^{m}\|A(\xi_k)\|\Delta t_k\right)\left(\exp\left(\sum_{k=1}^{m}\|B(\xi_k)-A(\xi_k)\|\Delta t_k\right)-1\right).$$

The proof is completed by passing to the limit $\nu(D)\to 0$. □

Remark 3.4.4. Lemma 3.4.2 is not present in Schlesinger's work, he proves directly the Corollary 3.4.3; our presentation is perhaps more readable.

3.5 Lebesgue product integral

The most valuable contribution of Schlesinger's paper is his generalized definition of product integral which is applicable to all matrix functions with bounded and measurable (i.e. bounded Lebesgue integrable) entries.

From a historical point of view, such a generalization certainly wasn't a straightforward one. Recall the original Lebesgue's definition: To compute the integral $\int_a^b f$ of a bounded measurable function $f:[a,b]\to[m,M]$, we choose a partition

$$D:\ m=m_0<m_1<\cdots<m_p=M,$$

then form the sets

$$E_0=\{x\in[a,b];\, f(x)=m\},$$
$$E_j=\{x\in[a,b];\, m_{j-1}<f(x)\leq m_j\},\quad j=1,\ldots,p,$$

and compute the lower and upper sums

$$s(f,D)=m_0\mu_0+\sum_{j=1}^{p}m_{j-1}\mu_j,\quad S(f,D)=m_0\mu_0+\sum_{j=1}^{p}m_j\mu_j, \tag{3.5.1}$$

where $\mu_j=\mu(E_j)$ is the Lebesgue measure of the set E_j. Since

$$S(f,D)-s(f,D)=\sum_{j=1}^{p}(m_j-m_{j-1})\mu_j\leq\nu(D)(b-a),$$

the sums in (3.5.1) approach a common limit as $\nu(D) \to 0$ and we define

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\nu(D)\to 0} s(f,D) = \lim_{\nu(D)\to 0} S(f,D).$$

Similar procedure cannot be used to define product integral of a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$, because $\mathbf{R}^{n\times n}$ is not an ordered set. Schlesinger was instead inspired by an equivalent definition of Lebesgue integral which is due to Friedrich Riesz (see [FR, KZ]): A bounded function $f : [a,b] \to \mathbf{R}$ is integrable, if and only if there exists a uniformly bounded sequence of step (i.e. piecewise-constant) functions $\{f_n\}_{n=1}^{\infty}$ such that $f_n \to f$ almost everywhere on $[a,b]$; in this case,

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x.$$

To proceed to the definition of product integral we first recall that (see Theorem 3.2.2)

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{\nu(D)\to 0} \prod_{k=1}^{m} e^{A(\xi_k)\Delta t_k}$$

for every Riemann integrable function $A : [a,b] \to \mathbf{R}^{n\times n}$. The product on the right side might be interpreted as

$$\prod_{k=1}^{m} e^{A(\xi_k)\Delta t_k} = (I + A_D(t)\,\mathrm{d}t)\prod_a^b,$$

where A_D is a step function defined by

$$A_D(t) = A(\xi_k), \quad t \in (t_{k-1}, t_k)$$

(the values $A(t_k)$, $k = 0,\ldots,m$, might be chosen arbitrarily). If $\{D_k\}_{k=1}^{\infty}$ is a sequence of tagged partitions of $[a,b]$ such that $\lim_{k\to\infty} \nu(D_k) = 0$, it is easily proved that

$$\lim_{k\to\infty} A_{D_k}(t) = A(t) \tag{3.5.2}$$

at every point $t \in [a,b]$ at which A is continuous. Since Riemann integrable functions are continuous almost everywhere, the Equation (3.5.2) holds a.e. on $[a,b]$. We are therefore led to the following generalized definition of product integral:

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty} (I + A_k(x)\,\mathrm{d}x)\prod_a^b,$$

where $\{A_k\}_{k=1}^{\infty}$ is a suitably chosen sequence of matrix step functions that converge to A almost everywhere.

Definition 3.5.1. A function $A : [a,b] \to \mathbf{R}^{n\times n}$ is called a step function if there exist numbers

$$a = t_0 < t_1 < \cdots < t_m = b$$

such that A is a constant function on every interval (t_{k-1}, t_k), $k = 1, \dots, m$.

Clearly, a matrix function $A = \{a_{ij}\}_{i,j=1}^n$ is a step function if and only if all the entries a_{ij} are step functions.

Definition 3.5.2. A sequence of functions $A_k : [a,b] \to \mathbf{R}^{n\times n}$, $k \in \mathbf{N}$, is called uniformly bounded if there exists a number $M \in \mathbf{R}$ such that $\|A_k(x)\| \le M$ for every $k \in \mathbf{N}$ and every $x \in [a,b]$.

Definition 3.5.3. A function $A : [a,b] \to \mathbf{R}^{n\times n}$ is called measurable if all the entries a_{ij} are measurable functions.

Lemma 3.5.4. Let $A_k : [a,b] \to \mathbf{R}^{n\times n}$, $k \in \mathbf{N}$, be a uniformly bounded sequence of measurable functions such that

$$\lim_{k\to\infty} A_k(x) = A(x)$$

a.e. on $[a,b]$. Then $A_k \to A$ in the norm of the space L^1, i.e.

$$\lim_{k\to\infty} \int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x = 0.$$

Proof. Choose $\varepsilon > 0$. As $\|A_k(x)\| \le M$ for every $k \in \mathbf{N}$ and every $x \in [a,b]$, we can estimate

$$\int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x \le \varepsilon(b-a) + 2M\mu(\{x; \|A_k(x) - A(x)\| \ge \varepsilon\}).$$

The convergence $A_k \to A$ a.e. implies convergence in measure[1], i.e. for every $\varepsilon > 0$ we have

$$\lim_{k\to\infty} \mu(\{x; \|A(x) - A_k(x)\| \ge \varepsilon\}) = 0.$$

Therefore

$$\lim_{k\to\infty} \int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x \le \varepsilon(b-a)$$

for every $\varepsilon > 0$. □

Theorem 3.5.5.[2] Let $A_k : [a,b] \to \mathbf{R}^{n\times n}$, $k \in \mathbf{N}$, be a sequence of step functions such that

$$\lim_{k\to\infty} \int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x = 0.$$

Then the limit

$$\lim_{k\to\infty} (I + A_k(x)\,\mathrm{d}x) \prod_a^b$$

[1] [IR], Proposition 8.3.3, p. 256

[2] [LS1], p. 55–56

exists and is independent on the choice of the sequence $\{A_k\}_{k=1}^{\infty}$.

Proof. We verify that $(I + A_k(x)\,\mathrm{d}x)\prod_a^b$ is a Cauchy sequence. According to Corollary 3.4.3 we have

$$\left\|(I + A_l(x)\,\mathrm{d}x)\prod_a^b - (I + A_m(x)\,\mathrm{d}x)\prod_a^b\right\| \le$$

$$\le \exp\left(\int_a^b \|A_m(x)\|\,\mathrm{d}x\right)\left(\exp\left(\int_a^b \|A_l(x) - A_m(x)\|\,\mathrm{d}x\right) - 1\right).$$

The assumption of our theorem implies that the sequence of numbers $\int_a^b \|A_m(x)\|\,\mathrm{d}x$ is bounded and that

$$\lim_{l,m\to\infty} \int_a^b \|A_l(x) - A_m(x)\|\,\mathrm{d}x = 0,$$

which proves the existence of the limit. To verify the uniqueness consider two sequences of step functions $\{A_k\}$, $\{B_k\}$ that satisfy the assumption of the theorem. We construct a sequence $\{C_k\}$, where $C_{2k-1} = A_k$ and $C_{2k} = B_k$. Then $C_k \to A$ a.e. and

$$\lim_{k\to\infty} \int_a^b \|C_k(x) - A(x)\|\,\mathrm{d}x = 0,$$

which means that $\lim_{k\to\infty}(I + C_k(x)\,\mathrm{d}x)\prod_a^b$ exists. Every subsequence of $\{C_k\}$ must have the same limit, therefore

$$\lim_{k\to\infty}(I + A_k(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}(I + B_k(x)\,\mathrm{d}x)\prod_a^b.$$

□

Definition 3.5.6. Consider function $A : [a,b] \to \mathbf{R}^{n\times n}$. Assume there exists a uniformly bounded sequence of step functions $A_k : [a,b] \to \mathbf{R}^{n\times n}$ such that

$$\lim_{k\to\infty} A_k(x) = A(x)$$

a.e. on $[a,b]$. Then the function A is called product integrable and we define

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}(I + A_k(x)\,\mathrm{d}x)\prod_a^b.$$

We use the symbol $L^*([a,b],\mathbf{R}^{n\times n})$ to denote the set of all product integrable functions.

Remark 3.5.7. The correctness of the previous definition is guaranteed by Lemma 3.5.4 and Theorem 3.5.5. Every function $A \in L^*([a,b],\mathbf{R}^{n\times n})$ is clearly bounded and measurable (step functions are measurable and the limit of measurable functions is again measurable). Assume on the contrary that $A : [a, b] \to \mathbf{R}^{n\times n}$ is a measurable function on $[a, b]$ such that

$$\|a_{ij}(x)\| \le M, \quad x \in [a, b], \;\; i, j = 1, \ldots, n.$$

There exists[1] a sequence of step functions $\{A_k\}_{k=1}^{\infty}$ which converge to A in the L^1 norm. This sequence contains[2] a subsequence $\{B_k\}_{k=1}^{\infty}$ of matrix functions $B_k = \{b_{ij}^k\}_{i,j=1}^n$ such that $B_k \to A$ a.e. on $[a, b]$. Without loss of generality we can assume that the sequence $\{B_k\}_{k=1}^{\infty}$ is uniformly bounded (otherwise consider the functions $\min(\max(-M, b_{ij}^k), M)$). We have thus found a uniformly bounded sequence of step functions which converge to A a.e. on $[a, b]$. This means that

$$L^*([a, b], \mathbf{R}^{n\times n}) = \left\{A : [a, b] \to \mathbf{R}^{n\times n};\; A \text{ is measurable and bounded}\right\}.$$

Schlesinger remarks that it is possible to further extend the definition of product integral to encompass all matrix functions with Lebesgue integrable (not necessarily bounded) entries, but he doesn't give any details. We return to this question at the end of the chapter.

3.6 Properties of Lebesgue product integral

After having defined the Lebesgue product integral in [LS1], Schlesinger carefully studies its properties. Interesting results may be found also in [LS2].

Lemma 3.6.1. Assume that $\{A_k\}_{k=1}^{\infty}$ is a uniformly bounded sequence of functions from $L^*([a, b], \mathbf{R}^{n\times n})$, and that $A_k \to A$ a. e. on $[a, b]$. Then

$$\int_a^b \|A(x)\|\,\mathrm{d}x = \lim_{k\to\infty} \int_a^b \|A_k(x)\|\,\mathrm{d}x.$$

Proof. According to the Lebesgue's dominated convergence theorem,

$$\lim_{k\to\infty} \int_a^b \|A_k(x)\|\,\mathrm{d}x = \int_a^b \lim_{k\to\infty} \|A_k(x)\|\,\mathrm{d}x = \int_a^b \|A(x)\|\,\mathrm{d}x$$

(we have used continuity of the norm). □

Corollary 3.6.2. Inequalities 3.4.1 and 3.4.3 are satisfied for all step functions. As a consequence of the previous lemma we see they are valid even for functions from $L^*([a, b], \mathbf{R}^{n\times n})$.

The next statement represents a dominated convergence theorem for the Lebesgue product integral.

---

[1] [RG], Corollary 3.29, p. 47

[2] [IR], Theorem 8.4.14, p. 267, and Theorem 8.3.6, p. 257

Theorem 3.6.3.[1] Assume that $\{A_k\}_{k=1}^{\infty}$ is a uniformly bounded sequence of functions from $L^*([a,b],\mathbf{R}^{n\times n})$ such that $A_k \to A$ a. e. on $[a,b]$. Then

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}(I + A_k(x)\,\mathrm{d}x)\prod_a^b.$$

Proof. The function A is measurable and bounded, therefore $A \in L^*([a,b],\mathbf{R}^{n\times n})$. To complete the proof we use Corollary 3.4.3 in the form

$$\left\|(I + A(x)\,\mathrm{d}x)\prod_a^b - (I + A_k(x)\,\mathrm{d}x)\prod_a^b\right\| \le$$
$$\exp\left(\int_a^b \|A(x)\|\,\mathrm{d}x\right)\left(\exp\left(\int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x\right) - 1\right)$$

and Lemma 3.5.4. □

Remark 3.6.4. The previous theorem holds also for Riemann product integral in case we add an extra assumption that the limit function A is Riemann product integrable.

Definition 3.6.5. If M is a measurable subset of $[a,b]$ and $A \in L^*([a,b],\mathbf{R}^{n\times n})$, we define

$$(I + A(x)\,\mathrm{d}x)\prod_M = (I + \chi_M(x)A(x)\,\mathrm{d}x)\prod_a^b$$

(where χ_M is the characteristic function of the set M).

The previous definition is correct, because the product $\chi_M A$ is obviously a measurable bounded function.

Remark 3.6.6. The following theorem is proved in the theory of Lebesgue integral[2]: For every $f \in L^1([a,b])$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left\|\int_M f(x)\,\mathrm{d}x\right\| < \varepsilon$$

whenever M is a measurable subset of $[a,b]$ and $\mu(M) < \delta$. Schlesinger proceeds to prove an analoguous theorem for the product integral (he speaks about "total continuity").

Theorem 3.6.7.[3] For every $A \in L^*([a,b],\mathbf{R}^{n\times n})$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left\|(I + A(x)\,\mathrm{d}x)\prod_M - I\right\| < \varepsilon$$

[1] [LS1], p. 57–58
[2] [RG], theorem 3.26, p. 46
[3] [LS1], p. 59

whenever M is a measurable subset of $[a,b]$ and $\mu(M) < \delta$.

Proof. Substituting $B = 0$ to Corollary 3.4.3 we obtain

$$\left\| (I + A(x)\,\mathrm{d}x) \prod_{M} - I \right\| \le \exp\left(\int_M \|A(x)\|\,\mathrm{d}x \right) \left(\exp\left(\int_M \|A(x)\|\,\mathrm{d}x \right) - 1 \right),$$

which completes the proof (see Remark 3.6.6). □

Schlesinger now turns his attention to the indefinite product integral. Recall that if $f \in L^1([a,b])$, then the indefinite integral

$$F(x) = \int_a^x f(t)\,\mathrm{d}t, \quad x \in [a,b]$$

is an absolutely continuous function and $F'(x) = f(x)$ a. e. on $[a,b]$. Before looking at a product analogy of this theorem we state the following lemma.

Lemma 3.6.8. If $A \in L^*([a,b], \mathbf{R}^{n\times n})$, then

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} \|A(t) - A(x)\|\,\mathrm{d}t = 0$$

for almost all $x \in (a,b)$.

Proof. If $f \in L^1([a,b])$, then[1]

$$\lim_{h\to 0} \frac{1}{h} \int_0^{h} |f(x+t) - f(x)|\,\mathrm{d}t = 0$$

for almost all $x \in (a,b)$ (every such x is called the Lebesgue point of f). Applying this equality to the entries of A we obtain

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} \|A(t) - A(x)\|\,\mathrm{d}t = \lim_{h\to 0} \frac{1}{h} \int_0^{h} \|A(x+t) - A(x)\|\,\mathrm{d}t = 0$$

for almost all $x \in (a,b)$. □

Theorem 3.6.9.[2] If $A \in L^*([a,b], \mathbf{R}^{n\times n})$, then the indefinite integral

$$Y(x) = (I + A(t)\,\mathrm{d}t) \prod_a^x .$$

satisfies $Y^{-1}(x)Y'(x) = A(x)$ for almost all $x \in [a,b]$.

Proof. According to the definition of derivative,

$$Y^{-1}(x)Y'(x) = \lim_{h\to 0} \frac{Y^{-1}(x)Y(x+h) - I}{h}.$$

[1] [IR], Theorem 6.3.2, p. 194

[2] [LS1], p. 60–61

We now prove that

$$\lim_{h\to 0+}\frac{Y^{-1}(x)Y(x+h)-I}{h}=\lim_{h\to 0+}\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-I\right)=A(x) \qquad (3.6.1)$$

for almost all $x\in[a,b]$; the procedure is similar for the limit from left. We estimate

$$\left\|\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-I\right)-A(x)\right\|\le\left\|\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-e^{A(x)h}\right)\right\|+$$

$$+\left\|\frac{1}{h}\sum_{k=2}^{\infty}\frac{A^k(x)h^k}{k!}\right\|\le\left\|\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-e^{A(x)h}\right)\right\|+\|A(x)\|^2|h|e^{\|A(x)\|h} \qquad (3.6.2)$$

Since $\|A(x)\|\le M$ for some $M\in\mathbf{R}$, the Corollary 3.4.3 yields

$$\left\|\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-e^{A(x)h}\right)\right\|=\left\|\frac{1}{h}\left((I+A(t)\,\mathrm{d}t)\prod_{x}^{x+h}-(I+A(x)\,\mathrm{d}t)\prod_{x}^{x+h}\right)\right\|\le$$

$$\le\frac{1}{|h|}\exp\left(\int_x^{x+h}\|A(x)\|\,\mathrm{d}t\right)\left(\exp\left(\int_x^{x+h}\|A(t)-A(x)\|\,\mathrm{d}t\right)-1\right)=$$

$$=\exp(\|A(x)\|h)\,\frac{1}{|h|}\sum_{k=1}^{\infty}\frac{1}{k!}\left(\int_x^{x+h}\|A(t)-A(x)\|\,\mathrm{d}t\right)^k\le$$

$$\le\exp(Mh)\left(\frac{1}{|h|}\int_x^{x+h}\|A(t)-A(x)\|\,\mathrm{d}t+(2M)^2h\exp(2Mh)\right). \qquad (3.6.3)$$

Equations (3.6.2), (3.6.3), and Lemma 3.6.8 imply Equation (3.6.1). □

Remark 3.6.10. In the previous theorem we have tacitly assumed that the matrix

$$Y(x)=(I+A(t)\,\mathrm{d}t)\prod_{a}^{x}$$

is regular for every $x\in[a,b]$. Schlesinger proved it only for $A\in R([a,b],\mathbf{R}^{n\times n})$ (see Corollary 3.2.6), but the proof is easily adjusted to $A\in L^*([a,b],\mathbf{R}^{n\times n})$: If $\{A_k\}_{k=1}^{\infty}$ is a uniformly bounded sequence of step functions such that $A_k\to A$ a. e. on $[a,b]$, then (using 3.2.4 and Lebesgue's dominated convergence theorem)

$$\det(I+A(t)\,\mathrm{d}t)\prod_{a}^{b}=\det\lim_{k\to\infty}(I+A_k(t)\,\mathrm{d}t)\prod_{a}^{b}=\lim_{k\to\infty}\det(I+A_k(t)\,\mathrm{d}t)\prod_{a}^{b}=$$

$$=\lim_{k\to\infty}\exp\left(\int_a^b\operatorname{Tr}A_k(t)\,\mathrm{d}t\right)=\exp\left(\int_a^b\operatorname{Tr}A(t)\,\mathrm{d}t\right)>0.$$

Theorem 3.6.11.[1] If $A \in L^*([a,b], \mathbf{R}^{n\times n})$, then

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = I + \sum_{k=1}^{\infty} \int_a^b \int_a^{x_k} \cdots \int_a^{x_2} A(x_1)\cdots A(x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

(where the integrals on the right side are taken in the sense of Lebesgue).

Proof. Let $\{A_k\}_{k=1}^{\infty}$ be a uniformly bounded sequence of step functions such that $A_k \to A$ a. e. on $[a,b]$. Every function A_k is associated with a partition

$$D_k : a = t_0^k < t_1^k < \cdots < t_{m(k)}^k = b$$

such that

$$A_k(x) = A_j^k, \quad x \in (t_{j-1}^k, t_j^k).$$

According to the definition of Lebesgue product integral,

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}(I + A_k(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}\prod_{j=1}^{m(k)} \exp(A_j^k \Delta t_j^k).$$

Schlesinger proves[2] first that the product integral might be also calculated as

$$(I + A(x)\,\mathrm{d}x)\prod_a^b = \lim_{k\to\infty}\prod_{i=1}^{m(k)}(I + A_i^k \Delta t_i^k), \tag{3.6.4}$$

provided that

$$\lim_{k\to\infty} \nu(D_k) = 0 \tag{3.6.5}$$

(which can be assumed without loss of generality); note that if (3.6.5) is not satisfied, (3.6.4) need not hold (consider $A = A_k = I$ and the partitions $a = t_0^k < t_1^k = b$ for every $k \in \mathbf{N}$). Schlesinger's proof of (3.6.4) seems too complicated and even faulty; we instead argue similarly as in the proof of Theorem 3.2.2: Take a positive number M such $\|A_k(x)\| \le M$ for every $k \in \mathbf{N}$ and $x \in [a,b]$. Then

$$\left\|\exp(A_j^k \Delta t_j^k) - I - A_j^k \Delta t_j^k\right\| \le (M\Delta t_j^k)^2 e^{M\Delta t_j^k}$$

for every $k \in \mathbf{N}$ and $j = 1, \ldots, m(k)$. According to Lemma 3.2.1,

$$\left\|\prod_{j=1}^{m(k)} \exp(A_j^k \Delta t_j^k) - \prod_{j=1}^{m(k)}(I + A_j^k \Delta t_j^k)\right\| =$$

$$= \left\|\sum_{j=1}^{m(k)}\left(\prod_{l=1}^{j-1}(I + A_l^k \Delta t_l^k)\cdot(\exp(A_j^k \Delta t_j^k) - I - A_j^k \Delta t_j^k)\cdot\prod_{l=j+1}^{m(k)} \exp(A_l^k \Delta t_l^k)\right)\right\| \le$$

[1] [LS2], p. 487

[2] [LS2], p. 485–486

$$\le e^{M(b-a)} \sum_{j=1}^{m(k)} \left\| \exp(A_j^k \Delta t_j^k) - I - A_j^k \Delta t_j^k \right\| \le e^{M(b-a)} M^2 \sum_{j=1}^{m(k)} (\Delta t_j^k)^2 e^{M \Delta t_j^k} \le$$

$$\le e^{M(b-a)} M^2 \nu(D_k) e^{M\nu(D_k)} \sum_{j=1}^{m(k)} \Delta t_j^k = e^{M(b-a)} M^2 \nu(D_k) e^{M\nu(D_k)} (b-a).$$

This completes the proof of (3.6.4). Schlesinger now states that

$$\prod_{i=1}^{m(k)} (I + A_i^k \Delta t_i^k) = I + \sum_{s=1}^{m(k)} \sum_{1 \le i_1 < \cdots < i_s \le m(k)} A_{i_1}^k \cdots A_{i_s}^k \Delta t_{i_1}^k \cdots \Delta t_{i_s}^k$$

and concludes the proof saying that

$$\lim_{k\to\infty} \sum_{1 \le i_1 < \cdots < i_s \le m(k)} A_{i_1}^k \cdots A_{i_s}^k \Delta t_{i_1}^k \cdots \Delta t_{i_s}^k =$$

$$= \int_a^b \int_a^{x_s} \cdots \int_a^{x_2} A(x_1) \cdots A(x_s) \, \mathrm{d}x_1 \cdots \mathrm{d}x_s$$

The last step perhaps deserves a better explanation: Denote

$$X^s = \{(x_1, \ldots, x_s) \in \mathbf{R}^s; a \le x_1 < x_2 < \cdots < x_s \le b\},$$

and

$$X_k^s = \bigcup_{1 \le i_1 < \cdots < i_s \le m(k)} [t_{i_1-1}, t_{i_1}] \times [t_{i_2-1}, t_{i_2}] \times \cdots \times [t_{i_s-1}, t_{i_s}],$$

where s and k are arbitrary positive integers. If χ^s and χ_k^s denote the characteristic functions of X^s and X_k^s, then $\chi_k^s \to \chi^s$ for $k \to \infty$. Consequently

$$\lim_{k\to\infty} \sum_{1 \le i_1 < \cdots < i_s \le m(k)} A_{i_1}^k \cdots A_{i_s}^k \Delta t_{i_1}^k \cdots \Delta t_{i_s}^k =$$

$$= \lim_{k\to\infty} \int_a^b \int_a^b \cdots \int_a^b A_k(x_1) \cdots A_k(x_s) \chi_k^s(x_1, \ldots, x_s) \, \mathrm{d}x_1 \cdots \mathrm{d}x_s =$$

$$= \int_a^b \int_a^b \cdots \int_a^b A(x_1) \cdots A(x_s) \chi^s(x_1, \ldots, x_s) \, \mathrm{d}x_1 \cdots \mathrm{d}x_s =$$

$$= \int_a^b \int_a^{x_s} \cdots \int_a^{x_2} A(x_1) \cdots A(x_s) \, \mathrm{d}x_1 \cdots \mathrm{d}x_s$$

(we have used the dominated convergence theorem). □

Remark 3.6.12. The deficiency in the previous proof is that Schlesinger didn't justify the equality

$$\lim_{k\to\infty} \left(I + \sum_{s=1}^{m(k)} \sum_{1 \le i_1 < \cdots < i_s \le m(k)} A_{i_1}^k \cdots A_{i_s}^k \Delta t_{i_1}^k \cdots \Delta t_{i_s}^k \right) =$$

$$I+\sum_{s=1}^{\infty}\lim_{k\to\infty}\left(\sum_{1\le i_1<\cdots<i_s\le m(k)} A_{i_1}^k\cdots A_{i_s}^k\Delta t_{i_1}^k\cdots\Delta t_{i_s}^k\right).$$

We have already encountered a similar inaccuracy when discussing Volterra's proof of the Peano series expansion theorem for product integral; see also Masani's proof of Theorem 5.5.10.

Remark 3.6.13. Recall that, according to Theorem 2.3.5, the right derivative of a matrix function satisfies

$$(CD^{-1})\frac{\mathrm{d}}{\mathrm{d}x}=D\left(C\frac{\mathrm{d}}{\mathrm{d}x}-D\frac{\mathrm{d}}{\mathrm{d}x}\right)D^{-1}.$$

Consider two continuous matrix functions A,B defined on $[a,b]$. Using the previous formula and also the convention that

$$(I+A(t)\,\mathrm{d}t)\prod_y^x=\left((I+A(t)\,\mathrm{d}t)\prod_x^y\right)^{-1}$$

for $y>x$, we infer the equality

$$\left((I+B(t)\,\mathrm{d}t)\prod_b^x(I+A(t)\,\mathrm{d}t)\prod_x^b\right)\frac{\mathrm{d}}{\mathrm{d}x}=$$

$$=(I+A(t)\,\mathrm{d}t)\prod_b^x(B(x)-A(x))(I+A(t)\,\mathrm{d}t)\prod_x^b$$

for every $x\in[a,b]$. Denoting $S(x)=(I+A(t)\,\mathrm{d}t)\prod_b^x$ we obtain

$$\left((I+B(t)\,\mathrm{d}t)\prod_b^x(I+A(t)\,\mathrm{d}t)\prod_x^b\right)\frac{\mathrm{d}}{\mathrm{d}x}=S(x)(B(x)-A(x))S^{-1}(x),$$

and consequently (since the left hand side is equal to I for $x=b$)

$$(I+B(t)\,\mathrm{d}t)\prod_b^x(I+A(t)\,\mathrm{d}t)\prod_x^b=(I+S(t)(B(t)-A(t))S^{-1}(t)\,\mathrm{d}t)\prod_b^x.$$

Substituting $x=a$ and inverting both sides of the equation yields

$$(I+A(t)\,\mathrm{d}t)\prod_b^a(I+B(t)\,\mathrm{d}t)\prod_a^b=(I+S(t)(B(t)-A(t))S^{-1}(t)\,\mathrm{d}t)\prod_a^b. \quad (3.6.6)$$

A similar theorem (concerning the left product integral) was already present in Volterra's work[1]. Schlesinger proves[2] that the statement remains true even if A, $B\in L^*([a,b],\mathbf{R}^{n\times n})$. The proof is rather technical and we don't reproduce it here.

[1] [VH], p. 85–86

[2] [LS2], p. 488–489

Theorem 3.6.14.[1] Let $A : [a,b] \times [c,d] \to \mathbf{R}^{n\times n}$ be such that the integral

$$P(t) = (I + A(x,t)\,\mathrm{d}x)\prod_a^b$$

exists for every $t \in [c,d]$ and that

$$\left\| \frac{\partial A}{\partial t}(x,t) \right\| \le M, \quad x \in [a,b],\ t \in [c,d],$$

for some $M \in \mathbf{R}$. Then

$$P\frac{\mathrm{d}}{\mathrm{d}t} = P^{-1}(t)P'(t) = \int_a^b S(x,t)\frac{\partial A}{\partial t}(x,t)S^{-1}(x,t)\,\mathrm{d}x,$$

where $S(x,t) = (I + A(u,t)\,\mathrm{d}u)\prod_b^x$.

Proof. The definition of derivative gives

$$P^{-1}(t)P'(t) = \lim_{h\to 0}\frac{1}{h}\left((I + A(x,t)\,\mathrm{d}x)\prod_b^a (I + A(x,t+h)\,\mathrm{d}x)\prod_a^b - I \right).$$

Using Equation (3.6.6) we convert the above limit to

$$\lim_{h\to 0}\frac{1}{h}\left((I + S(x,t)(A(x,t+h) - A(x,t))S^{-1}(x,t)\,\mathrm{d}x)\prod_a^b - I \right).$$

Expanding the product integral to Peano series (see Theorem 3.6.11) we obtain

$$\lim_{h\to 0}\frac{1}{h}\sum_{k=1}^{\infty}\int_a^b\int_a^{x_k}\cdots\int_a^{x_2}\Delta(x_1,t,h)\cdots\Delta(x_k,t,h)\,\mathrm{d}x_1\cdots dx_k, \tag{3.6.7}$$

where

$$\Delta(x,t,h) = S(x,t)(A(x,t+h) - A(x,t))S^{-1}(x,t).$$

As the Peano series converges uniformly (the Weierstrass M-test, see Theorem 2.4.5), we can interchange the order of limit and summation. According to the mean value theorem there is a $\xi(h) \in [t, t+h]$ such that

$$\left\| \frac{A(x,t+h) - A(x,t)}{h} \right\| = \left\| \frac{\partial A}{\partial t}(x,\xi(h)) \right\| \le M.$$

The dominated convergence theorem therefore implies

$$\lim_{h\to 0}\frac{1}{h}\int_a^b \Delta(x_1,t,h)\,\mathrm{d}x_1 = \int_a^b \lim_{h\to 0}\frac{\Delta(x_1,t,h)}{h}\,\mathrm{d}x_1 =$$

[1] [LS2], p. 490–491

$$= \int_a^b S(x_1,t)\frac{\partial A}{\partial t}(x_1,t)S^{-1}(x_1,t)\,\mathrm{d}x_1,$$

and for $k \geq 2$

$$\lim_{h\to 0}\frac{1}{h}\int_a^b\int_a^{x_k}\cdots\int_a^{x_2}\Delta(x_1,t,h)\cdots\Delta(x_k,t,h)\,\mathrm{d}x_1\cdots dx_k =$$

$$= \int_a^b\int_a^{x_k}\cdots\int_a^{x_2}\lim_{h\to 0}\left(h^{k-1}\frac{\Delta(x_1,t,h)}{h}\cdots\frac{\Delta(x_k,t,h)}{h}\right)\mathrm{d}x_1\cdots dx_k = 0,$$

which completes the proof. □

The following statement generalizes Theorem 2.5.12; Schlesinger replaces Volterra's assumption $A \in \mathcal{C}([a,b],\mathbf{R}^{n\times n})$ by a weaker condition $A \in L^*([a,b],\mathbf{R}^{n\times n})$.

Theorem 3.6.15.[1] If $A \in L^*([a,b],\mathbf{R}^{n\times n})$ and $C \in \mathbf{R}^{n\times n}$ is a regular matrix, then

$$(I + C^{-1}A(x)C\,\mathrm{d}x)\prod_a^b = C^{-1}(I + A(x)\,\mathrm{d}x)\prod_a^b C.$$

Proof. Since $(C^{-1}AC)^k = C^{-1}A^kC$ for every $k \in \mathbf{N}$, we have

$$\exp(C^{-1}AC) = C^{-1}\exp(A)C.$$

If A is a step function, then

$$(I + C^{-1}A(x)C\,\mathrm{d}x)\prod_a^b = \prod_{i=1}^m e^{C^{-1}A(\xi_i)C\Delta t_i} =$$

$$= C^{-1}\prod_{i=1}^m e^{A(\xi_i)\Delta t_i}C = C^{-1}(I + A(x)\,\mathrm{d}x)\prod_a^b C.$$

In the general case when $A \in L^*([a,b],\mathbf{R}^{n\times n})$, we rewrite the above equation with simple functions A_k in place of A, and then pass to the limit $k \to \infty$. □

3.7 Double and contour product integrals

A considerable part of the paper [LS2] is devoted to double and contour product integrals (in $\mathbf{R}^2$ as well as in $\mathbf{C}$). Probably the most remarkable achievement is Schlesinger's proof of the "Green's theorem" for product integral, which is reproduced in the following text.

Definition 3.7.1. Let G be the rectangle $[a,b]\times[c,d]$ in $\mathbf{R}^2$ and $A : G \to \mathbf{R}^{n\times n}$ a matrix function on G. The double product integral of A over G is defined as

$$(I + A(x,y)\,\mathrm{d}x\,\mathrm{d}y)\prod_G = \left(I + \left(\int_a^b A(x,y)\,\mathrm{d}x\right)\mathrm{d}y\right)\prod_c^d,$$

[1] [LS2], p. 489

provided both integrals on the right hand side exist (in the sense of Lebesgue).

Definition 3.7.2. Let G be the rectangle $[a,b]\times[c,d]$ in $\mathbf{R}^2$ and $P,Q:G\to\mathbf{R}^{n\times n}$ continuous functions on G. We denote

$$U(x,y)=(I+P(t,c)\,\mathrm{d}t)\prod_a^x (I+Q(x,t)\,\mathrm{d}t)\prod_c^y,$$

$$T(x,y)=(I+Q(a,t)\,\mathrm{d}t)\prod_c^y (I+P(t,y)\,\mathrm{d}t)\prod_a^x$$

for every $x\in[a,b]$, $y\in[c,d]$. The contour product integral over the boundary of rectangle G is defined as the matrix

$$(I+P(x,y)\,\mathrm{d}x+Q(x,y)\,\mathrm{d}y)\prod_{\partial G}=U(b,d)T(b,d)^{-1}. \tag{3.7.1}$$

Remark 3.7.3. Schlesinger refers to the matrices $U(b,d)$ and $T(b,d)$ as to the "integral over the lower step" and "integral over the upper step" of the rectangle G. They are clearly a special case of the contour product integral as defined by Volterra (see definition 2.6.8); the matrix (3.7.1) corresponds to the value of contour product integral along the (anticlockwise oriented) boundary of G.

Theorem 3.7.4.[1] Let G be the rectangle $[a,b]\times[c,d]$ in $\mathbf{R}^2$ and $P,Q:G\to\mathbf{R}^{n\times n}$ continuous matrix functions on G. Assume that the derivatives

$$\frac{\partial P}{\partial y},\quad \frac{\partial Q}{\partial x}$$

exist and are continuous on G. Then

$$(I+P(x,y)\,\mathrm{d}x+Q(x,y)\,\mathrm{d}y)\prod_{\partial G}=(I+T\cdot\Delta^*(P,Q)\cdot T^{-1}\,\mathrm{d}x\,\mathrm{d}y)\prod_G,$$

where

$$\Delta^*(P,Q)=\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}+PQ-QP,$$

$$T(x,y)=(I+Q(a,t)\,\mathrm{d}t)\prod_c^y (I+P(t,y)\,\mathrm{d}t)\prod_a^x.$$

Proof. A simple calculation reveals that (compare to Lemma 2.6.4)

$$T\cdot\Delta^*(P,Q)\cdot T^{-1}=\frac{\partial}{\partial x}\left(T\left(Q-T\frac{\mathrm{d}}{\mathrm{d}y}\right)T^{-1}\right).$$

[1] [LS2], p. 496–497

Taking the product integral over G we obtain

$$(I + T \cdot \Delta^*(P,Q) \cdot T^{-1}\,\mathrm{d}x\,\mathrm{d}y) \prod_G =$$

$$= \left(I + \left[T(x,y)\left(Q(x,y) - T(x,y)\frac{\mathrm{d}}{\mathrm{d}y}\right)T(x,y)^{-1}\right]_a^b \mathrm{d}y\right)\prod_c^d. \tag{3.7.2}$$

According to the rules for differentiating a product of functions (see Theorem 2.3.2),

$$T\frac{\mathrm{d}}{\mathrm{d}y} = (I + P(t,y)\,\mathrm{d}t)\prod_x^a Q(a,y)(I + P(t,y)\,\mathrm{d}t)\prod_a^x + \left((I + P(t,y)\,\mathrm{d}t)\prod_a^x\right)\frac{\mathrm{d}}{\mathrm{d}y}.$$

Theorem 3.6.14 on differentiating the product integral with respect to a parameter yields

$$\lim_{x\to a}\left((I + P(t,y)\,\mathrm{d}t)\prod_a^x\right)\frac{\mathrm{d}}{\mathrm{d}y} = 0,$$

and consequently

$$\lim_{x\to a} T\frac{\mathrm{d}}{\mathrm{d}y} = Q(a,y). \tag{3.7.3}$$

The equalities (3.7.2) and (3.7.3) imply

$$(I + T \cdot \Delta^*(P,Q) \cdot T^{-1}\,\mathrm{d}x\,\mathrm{d}y) \prod_G =$$

$$= \left(I + \lim_{x\to b}\left(T(x,y)\left(Q(x,y) - T(x,y)\frac{\mathrm{d}}{\mathrm{d}y}\right)T(x,y)^{-1}\right)\mathrm{d}y\right)\prod_c^d =$$

$$= \lim_{x\to b}\left(I + T(x,y)\left(Q(x,y) - T(x,y)\frac{\mathrm{d}}{\mathrm{d}y}\right)T(x,y)^{-1}\mathrm{d}y\right)\prod_c^d \tag{3.7.4}$$

(we have used Theorem 3.6.3 on interchanging the order of limit and integral). For every $x \in [a,b]$ we have

$$T(x,y)\frac{\mathrm{d}}{\mathrm{d}y} = \left(T(x,d)^{-1}T(x,y)\right)\frac{\mathrm{d}}{\mathrm{d}y}$$

and also

$$T(x,d)^{-1}T(x,y) = \left(I + T(x,u)\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{d}u\right)\prod_d^y = \left(I + \left(T(x,d)^{-1}T(x,u)\right)\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{d}u\right)\prod_d^y.$$

Using Theorem 3.6.15 and Equation (3.6.6) we arrive at

$$\left(I + T(x,y)\left(Q(x,y) - T(x,y)\frac{\mathrm{d}}{\mathrm{d}y}\right)T(x,y)^{-1}\mathrm{d}y\right)\prod_c^d = T(x,d)\cdot$$

$$\cdot\left(I+T(x,d)^{-1}T(x,y)\left(Q(x,y)-\left(T(x,d)^{-1}T(x,y)\right)\frac{\mathrm{d}}{\mathrm{d}y}\right)T(x,y)^{-1}T(x,d)\,\mathrm{d}y\right)\prod_{c}^{d}\cdot$$

$$\cdot\,T(x,d)^{-1}=T(x,d)\left(I+\left(T(x,d)^{-1}T(x,y)\right)\frac{\mathrm{d}}{\mathrm{d}y}\,\mathrm{d}y\right)\prod_{d}^{c}\cdot$$

$$\cdot(I+Q(x,y)\,\mathrm{d}y)\prod_{c}^{d}T(x,d)^{-1}=T(x,d)\,T(x,d)^{-1}T(x,c)(I+Q(x,y)\,\mathrm{d}y)\prod_{c}^{d}\cdot$$

$$\cdot\,T(x,d)^{-1}=T(x,c)(I+Q(x,y)\,\mathrm{d}y)\prod_{c}^{d}T(x,d)^{-1}.$$

Finally, Equation (3.7.4) gives

$$(I+T\cdot\Delta^*(P,Q)\cdot T^{-1}\,\mathrm{d}x\,\mathrm{d}y)\prod_{G}=\lim_{x\to b}\left(T(x,c)(I+Q(x,y)\,\mathrm{d}y)\prod_{c}^{d}T(x,d)^{-1}\right)=$$

$$=T(b,c)(I+Q(b,y)\,\mathrm{d}y)\prod_{c}^{d}T(b,d)^{-1}=(I+P(x,y)\,\mathrm{d}x+Q(x,y)\,\mathrm{d}y)\prod_{\partial G}.$$

□

Remark 3.7.5. The previous theorem represents an analogy of Green's theorem for the product integral; we have already encountered a similar statement when discussing Volterra's work. Volterra's analogy of the curl operator was

$$\Delta(P,Q)=\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}+QP-PQ,$$

while Schlesinger's curl has the form

$$\Delta^*(P,Q)=\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}+PQ-QP.$$

The reason is that Volterra stated his theorem for the left product integral, while Schlesinger was concerned with the right product integral (see Theorem 2.6.15 and Remark 2.6.7). Whereas Volterra worked with a simply connected domain G (see definition 2.6.12), Schlesinger considers only rectangles.

Consider functions P, Q that satisfy assumptions of Theorem 3.7.4 and such that

$$\Delta^*(P,Q)=0 \tag{3.7.5}$$

everywhere in G. Then

$$(I+P(x,y)\,\mathrm{d}x+Q(x,y)\,\mathrm{d}y)\prod_{\partial G}=I,$$

which in consequence means that the values of contour product integral over the lower step and over the upper step are the same. Schlesinger then denotes the common value of the matrices $U(x,y)$ and $T(x,y)$ (see definition 3.7.2) by the symbol

$$(I + P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y)\prod_{(a,c)}^{(b,d)}.$$

Clearly

$$\left((I + P(u,v)\,\mathrm{d}u + Q(u,v)\,\mathrm{d}v)\prod_{(a,c)}^{(x,y)}\right)\frac{\mathrm{d}}{\mathrm{d}x} = T(x,y)\frac{\mathrm{d}}{\mathrm{d}x} = P(x,y),$$

$$\left((I + P(u,v)\,\mathrm{d}u + Q(u,v)\,\mathrm{d}v)\prod_{(a,c)}^{(x,y)}\right)\frac{\mathrm{d}}{\mathrm{d}y} = U(x,y)\frac{\mathrm{d}}{\mathrm{d}y} = Q(x,y).$$

Schlesinger now proceeds to define product integral along a contour and shows that (in a simply connected domain) the condition (3.7.5) implies that the value of product integral depends only on the endpoints of the contour. His method is almost the same as Volterra's and we don't repeat it here.

At the end of paper [LS2] Schlesinger treats matrix functions of a complex variable. He defines the contour product integral in complex domain and recapitulates the results proved earlier by Volterra (theorems 2.7.4, 2.7.7, and 2.7.6).

3.8 Generalization of Schlesinger's definition

Thanks to the definition proposed by Ludwig Schlesinger it is possible to extend the class of product integrable functions and to work with bounded measurable functions instead of Riemann integrable functions. At this place we remind the notation

$$L^*([a,b],\mathbf{R}^{n\times n}) = \left\{A : [a,b] \to \mathbf{R}^{n\times n};\ A \text{ is measurable and bounded}\right\}.$$

Schlesinger was aware that his definition might be extended to all matrix functions with Lebesgue integrable (not necessarily bounded) entries, i. e. to the class

$$L([a,b],\mathbf{R}^{n\times n}) = \left\{A : [a,b] \to \mathbf{R}^{n\times n};\ (L)\int_a^b \|A(t)\|\,\mathrm{d}t < \infty\right\},$$

where the symbol (L) emphasizes that we are dealing with the Lebesgue integral. Clearly $L^* \subset L$. If $\{A_k\}_{k=1}^{\infty}$ is a uniformly bounded sequence of functions which converge to A almost everywhere, then according to lemma 3.5.4

$$\lim_{k\to\infty}\|A_k - A\|_1 = \lim_{k\to\infty}\int_a^b \|A_k(x) - A(x)\|\,\mathrm{d}x = 0,$$

i.e. A_k converge to A also in the norm of space $L([a,b],\mathbf{R}^{n\times n})$. Taking account of Theorem 3.5.5 it is natural to state the following definition.

Definition 3.8.1. A function $A:[a,b]\to\mathbf{R}^{n\times n}$ is called product integrable if there exists a sequence of step functions $\{A_k\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty}\|A_k-A\|_1=0.$$

We define

$$(I+A(t)\,\mathrm{d}t)\prod_a^b=\lim_{k\to\infty}(I+A_k(t)\,\mathrm{d}t)\prod_a^b.$$

Remark 3.8.2. The correctness of the previous definition is ensured by theorem 3.5.5. Since step functions belong to the space $L([a,b],\mathbf{R}^{n\times n})$, which is complete, every product integrable function also belongs to this space. Moreover, step functions form a dense subset in this space[1], and therefore $(I+A(t)\,\mathrm{d}t)\prod_a^b$ exists iff $A\in L([a,b],\mathbf{R}^{n\times n})$, i. e. iff the integral $(L)\int_a^b A(t)\,\mathrm{d}t$ exists.

Interested readers are referred to the book [DF] for more details about the theory of product integral based on definition 3.8.1. As an interesting example we present the proof of theorem on differentiating the product integral with respect to the upper bound of integration. We start with a preliminary lemma (which follows also from Theorem 3.3.2, but we don't want to use it as we are seeking another way to prove it).

Lemma 3.8.3. If $A:[a,b]\to\mathbf{R}^{n\times n}$ is a step function, then

$$Y(x)=I+\int_a^x Y(t)A(t)\,\mathrm{d}t,\quad x\in[a,b].$$

Proof. There exist a partition $a=t_0<t_1<\cdots<t_m=b$ and matrices $A_1,\ldots,A_m\in\mathbf{R}^{n\times n}$ such that

$$A(x)=A_k,\quad x\in(t_{k-1},t_k).$$

Then

$$Y(x)=(I+A(t)\,\mathrm{d}t)\prod_a^x=e^{A_1(t_2-t_1)}\cdots e^{A_{k-1}(t_{k-1}-t_{k-2})}e^{A_k(x-t_{k-1})}$$

for every $x\in[t_{k-1},t_k]$. The function Y is continuous on $[a,b]$ and differentiable except a finite number of points; we have

$$Y'(x)=Y(x)A(x) \tag{3.8.1}$$

[1] [RG], Corollary 3.29, p. 47

for $x \in [a,b]\backslash\{t_0, t_1, \dots, t_m\}$. This implies

$$Y(x) = I + \int_a^x Y(t)A(t)\,\mathrm{d}t, \quad x \in [a,b].$$

□

Theorem 3.8.4.[1] Consider function $A \in L([a,b], \mathbf{R}^{n\times n})$. For every $x \in [a,b]$ the integral

$$Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x \tag{3.8.2}$$

exists and the function Y satisfies the equation

$$Y(x) = I + \int_a^x Y(t)A(t)\,\mathrm{d}t, \quad x \in [a,b]. \tag{3.8.3}$$

Proof. Let $A_k : [a,b] \to \mathbf{R}^{n\times n}$, $k \in \mathbf{N}$ be a sequence of step functions such that

$$\lim_{k\to\infty} \|A_k - A\|_1 = \lim_{k\to\infty} \int_a^b \|A_k(t) - A(t)\|\,\mathrm{d}t = 0. \tag{3.8.4}$$

Then clearly

$$\lim_{k\to\infty} \int_a^x \|A_k(t) - A(t)\|\,\mathrm{d}t = 0, \quad x \in [a,b],$$

i. e. the definition (3.8.2) is correct. Denote

$$Y_k(x) = (I + A_k(t)\,\mathrm{d}t)\prod_a^x .$$

Because A_k are step functions, Lemma 3.8.3 implies

$$Y_k(x) = I + \int_a^x Y_k(t)A_k(t)\,\mathrm{d}t, \quad x \in [a,b]. \tag{3.8.5}$$

According to Corollary 3.4.3,

$$\|Y_l(x) - Y_m(x)\| \le \exp\left(\int_a^b \|A_l(t)\|\,\mathrm{d}t\right)\left(\exp\left(\int_a^b \|A_l(t) - A_m(t)\|\,\mathrm{d}t\right) - 1\right) =$$

$$= \exp\|A_l\|_1 \left(\exp\|A_l(t) - A_m(t)\|_1 - 1\right).$$

From Equation (3.8.4) we see that $\{A_k\}_{k=1}^{\infty}$ is a bounded and Cauchy sequence with respect to the norm $\|\cdot\|_1$. The previous inequality therefore implies that Y_k converge uniformly to Y, i. e.

$$\|Y_k - Y\|_\infty = \sup_{x\in[a,b]} \|Y_k(x) - Y(x)\| \to 0 \quad \text{pro } k \to \infty.$$

---

[1] [DF], p. 54–55

We now estimate

$$\left\|\int_a^x Y_k(t)A_k(t)\,\mathrm{d}t - \int_a^x Y(t)A(t)\,\mathrm{d}t\right\| \leq \|Y_kA_k - YA\|_1 \leq$$

$$\leq \|(Y_k - Y)A_k\|_1 + \|Y(A_k - A)\|_1 \leq \|A_k\|_1\|Y_k - Y\|_\infty + \|(A_k - A)\|_1\|Y\|_\infty,$$

and consequently

$$\lim_{k\to\infty}\int_a^x Y_k(t)A_k(t)\,\mathrm{d}t = \int_a^x Y(t)A(t)\,\mathrm{d}t.$$

The equality (3.8.3) is obtained by passing to the limit in equation (3.8.5). □

Corollary 3.8.5. If $A \in L([a,b],\mathbf{R}^{n\times n})$, then the function

$$Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x, \quad x \in [a,b],$$

is absolutely continuous on $[a,b]$ and

$$Y(x)^{-1}\cdot Y'(x) = A(x)$$

almost everywhere on $[a,b]$.

Remark 3.8.6. In our proof of the previous theorem we have employed Schlesinger's estimate from Corollary 3.4.3, whose proof is somewhat laborious. The authors of [DF] instead make use of a different inequality, which is easier to demonstrate. Let A, $B : [a,b] \to \mathbf{R}^{n\times n}$ be two step functions. Denoting

$$Y(x) = (I + A(t)\,\mathrm{d}t)\prod_a^x, \quad Z(x) = (I + B(t)\,\mathrm{d}t)\prod_a^x,$$

we see that the function YZ^{-1} is continuous on $[a,b]$ and differentiable except a finite number of points. Using Equation (3.8.1) we calculate

$$(YZ^{-1})' = Y'Z^{-1} + Y(Z^{-1})' = YY^{-1}Y'Z^{-1} - YZ^{-1}Z'Z^{-1} = Y(A-B)Z^{-1},$$

and consequently

$$Y(x)Z^{-1}(x) = I + \int_a^x (YZ^{-1})'(t)\,\mathrm{d}t = I + \int_a^x Y(t)(A(t) - B(t))Z^{-1}(t)\,\mathrm{d}t.$$

Multiplying this equation by Z from right and substituting $x = b$ we obtain

$$\left\|(I + A(t)\,\mathrm{d}t)\prod_a^b - (I + B(t)\,\mathrm{d}t)\prod_a^b\right\| = \|Y(b) - Z(b)\| \leq$$

$$\leq \int_a^b \|Y(t)\|\cdot\|A(t) - B(t)\|\cdot\|Z^{-1}(t)\|\,\mathrm{d}t\cdot\|Z(b)\| \leq e^{2\|B\|_1}e^{\|A\|_1}\|A - B\|_1$$

(we have used Lemma 3.4.1 to estimate $\|Y(t)\|$, $\|Z^{-1}(t)\|$ and $\|Z(b)\|$). The meaning of the last inequality is similar to the meaning of inequality from Corollary 3.4.3: "If two step functions A, B are close with respect to the norm $\|\cdot\|_1$, then the values of their product integrals are also close to each other."

Chapter 4

Operator-valued functions

In the previous chapters we have encountered various definitions of product integral of a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$. It is known that every $n \times n$ matrix represents a linear transformation on the space $\mathbf{R}^n$, and that the composition of two linear transformations corresponds to multiplication of their matrices.

It is thus natural to ask whether it is possible to define the product integral of a function A defined on $[a,b]$ whose values are operators on a certain (possibly infinite-dimensional) vector space X. This pioneering idea (although in a less general scope) can be already found in the second part of the book [VH] written by Bohuslav Hostinský. He studies certain special linear operators on the space of continuous functions (he calls them "linear functional transformations") and calculates their product integrals as well as their left and right derivatives. Hostinský imagines a function as a vector with infinitely many coordinates; the linear operator on $\mathbf{R}^n$ given by

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

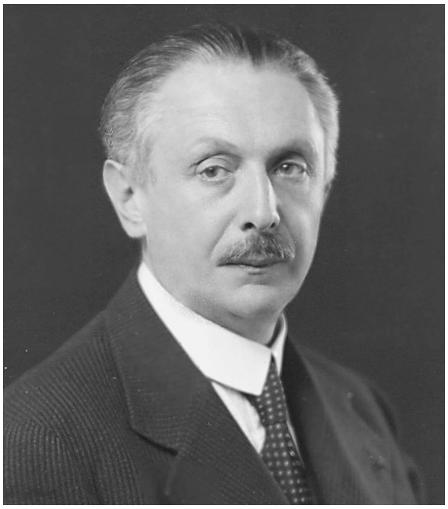
then goes over to the operator on continuous functions given by

$$y(t) = \int_p^q A(t,u)x(u)\,\mathrm{d}u.$$

This idea was not a new one – already Volterra noted[1] that integral equations can be treated as limiting cases of systems of linear algebraic equations. His observation was also later used by Ivar Fredholm, who obtained the solution of an integral equation as a limit of the solutions of linear algebraic equations.

Bohuslav Hostinský was born on the 5th December 1884 in Prague. He studied mathematics and physics at the philosophical faculty of Charles University in Prague and obtained his doctoral degree in 1907 (his dissertation thesis was devoted to geometry). He spent a short time as a high school teacher and visited Paris in 1908–09 (he cooperated especially with Gaston Darboux). Since 1912 he gave lectures as privatdozent at the philosophical faculty in Prague and was promoted to professor of theoretical physics at the faculty of natural sciences in Brno in 1920; he held this position until his death on the 12th April 1951.

[1] [Kl], Chapter 45



Bohuslav Hostinský[1]

Hostinský initially devoted himself to pure mathematics, especially to differential geometry. However, his interest gradually moved to theoretical physics, in particular to the kinetic theory of gases (being influenced mainly by Borel, Ehrenfest and Poincaré). Of course, this discipline requires a good knowledge of probability theory; Hostinský is considered one of the pioneers of the theory of Markov processes and their application in physics. Apart from that he also worked on differential and integral equations and is known for his critical attitude towards the theory of relativity. A good overview of the life and work of Bohuslav Hostinský is given in [Ber].

4.1 Integral operators

In the following discussion we focus our attention to the space $\mathcal{C}([p,q])$ of continuous functions defined on $[p,q]$. Every mapping $T : \mathcal{C}([p,q]) \to \mathcal{C}([p,q])$ is called an operator on $\mathcal{C}([p,q])$; we will often write Tf instead of $T(f)$. The operator is called linear if

$$T(af_1 + bf_2) = aT(f_1) + bT(f_2)$$

for each pair of functions $f_1, f_2 \in \mathcal{C}([p,q])$ and each pair of numbers $a, b \in \mathbf{R}$. The inverse operator of T, which is denoted by T^{-1}, satisfies

$$T^{-1}(T(f)) = T(T^{-1}(f)) = f \tag{4.1.1}$$

for every function $f \in \mathcal{C}([p,q])$; note that the inverse operator need not always exist. The last equation can be shortened to

$$T^{-1} \cdot T = T \cdot T^{-1} = I,$$

[1] Photo provided by Tomáš Hostinský

where $\cdot$ denotes the composition of operators and I is the identity operator, which satisfies $If = f$ for every function f.

Bohuslav Hostinský was concerned especially with integral operators of the first kind

$$Tf(x) = \int_p^q K(x,y)f(y)\,\mathrm{d}y$$

and with the operators of the second kind

$$Tf(x) = f(x) + \int_p^q K(x,y)f(y)\,\mathrm{d}y,$$

where the function K (the so-called kernel) is continuous on $[p,q] \times [p,q]$. Thus if $f \in \mathcal{C}([p,q])$, then also $Tf \in \mathcal{C}([p,q])$. These operators play an important role in the theory of integral equations and have been studied by Vito Volterra, Ivar Fredholm, David Hilbert and others since the end of the 19th century (see [Kl], Chapter 45).

Hostinský starts[1] with a recapitulation of the basic properties of the integral operators. If $K(x,y) = 0$ for $y > x$, we obtain either the Volterra operator of the first kind

$$Tf(x) = \int_p^x K(x,y)f(y)\,\mathrm{d}y,$$

or the Volterra operator of the second kind

$$Tf(x) = f(x) + \int_p^x K(x,y)f(y)\,\mathrm{d}y.$$

Composition of two integral operators of the second kind

$$T_1 f(x) = f(x) + \int_p^q K_1(x,y)f(y)\,\mathrm{d}y,$$

$$T_2 f(x) = f(x) + \int_p^q K_2(x,y)f(y)\,\mathrm{d}y,$$

produces another operator of the second kind

$$(T_2 \cdot T_1)f(x) = f(x) + \int_p^q J(x,y)f(y)\,\mathrm{d}y,$$

whose kernel is

$$J(x,y) = K_1(x,y) + K_2(x,y) + \int_p^q K_2(x,z)K_1(z,y)\,\mathrm{d}z. \tag{4.1.2}$$

[1] [VH], p. 182–186

An important question is the existence of inverse operators (which corresponds to solvability of the corresponding integral equations). The Volterra operator of the second kind always has an inverse operator

$$T^{-1}f(x) = f(x) + \int_p^x N(x,y)f(y)\,\mathrm{d}y,$$

which is again a Volterra operator of the second kind whose kernel $N(x,y)$ (usually called the resolvent kernel) can be found using the method of successive approximations. Several sufficient conditions are known for the existence of the inverse operator to the Fredholm operator of the second kind; we do not dwell into the details and only mention that the inverse operator is again a Fredholm operator of the second kind

$$T^{-1}f(x) = f(x) + \int_p^q N(x,y)f(y)\,\mathrm{d}y,$$

with a kernel $N(x,y)$. Equations (4.1.1) and (4.1.2) imply that the kernel satisfies

$$K(x,y) + N(x,y) = -\int_p^q K(x,t)N(t,y)\,\mathrm{d}t = -\int_p^q N(x,t)K(t,y)\,\mathrm{d}t. \tag{4.1.3}$$

Generally, the operators of the first kind need not have an inverse operator.

4.2 Product integral of an operator-valued function

We now assume that the integral operator kernel depends on a parameter $u \in [a,b]$:

$$T(u)f(x) = f(x) + \int_p^q K(x,y,u)f(y)\,\mathrm{d}y$$

Thus T is a function defined on $[a,b]$ whose values are operators on $(\mathcal{C}([p,q]))$. Hostinský now proceeds[1] to calculate its left derivative. He doesn't state any definition, but his calculation follows Volterra's definition of the left derivative of a matrix function. Assume that K is continuous on $[p,q]\times[p,q]\times[a,b]$ and that the derivative

$$\frac{\partial K}{\partial u}(x,y,u)$$

exists and is continuous for every $u \in [a,b]$ and $x,y \in [p,q]$. We choose a particular $u \in [a,b]$ and assume that the inverse operator $T^{-1}(u)$ exists. According to Equation (4.1.2), the kernel of the operator $T(u+\Delta u)\cdot T^{-1}(u)$ is

$$J(x,y,u,\Delta u) = K(x,y,u+\Delta u) + N(x,y,u) + \int_p^q K(x,t,u+\Delta u)N(t,y,u)\,\mathrm{d}t =$$

$$= K(x,y,u+\Delta u) - K(x,y,u) + K(x,y,u) + N(x,y,u) +$$

[1] [VH], p. 186–188

$$+\int_p^q K(x,t,u+\Delta u)N(t,y,u)\,\mathrm{d}t =$$

$$= K(x,y,u+\Delta u) - K(x,y,u) + \int_p^q \left(K(x,t,u+\Delta u) - K(x,t,u)\right) N(t,y,u)\,\mathrm{d}t$$

(we have used Equation (4.1.3)). Denote

$$M(x,y,u) = \lim_{\Delta u \to 0} \frac{J(x,y,u,\Delta u)}{\Delta u} =$$

$$= \frac{\partial K}{\partial u}(x,y,u) + \int_p^q \frac{\partial K}{\partial u}(x,t,u)N(t,y,u)\,\mathrm{d}t. \tag{4.2.1}$$

According to Bohuslav Hostinský, the left derivative of T at u is an operator of the second kind whose kernel is $M(x,y,u)$. By a similar method he deduces that the right derivative of T at u is an operator of the second kind with the kernel

$$M^*(x,y,u) = \frac{\partial K}{\partial u}(x,y,u) + \int_p^q N(x,t,u)\frac{\partial K}{\partial u}(t,y,u)\,\mathrm{d}t.$$

These statements are somewhat confusing, because the left derivative should be the operator

$$\lim_{\Delta u \to 0} \frac{T(u+\Delta u)\cdot T^{-1}(u) - I}{\Delta u},$$

which is an operator of the *first* kind with kernel $M(x,y,u)$. Similarly, the right derivative should be rather

$$\lim_{\Delta u \to 0} \frac{T^{-1}(u)\cdot T(u+\Delta u) - I}{\Delta u},$$

i.e. an operator of the *first* kind with kernel $M^*(x,y,u)$.

The next problem tackled by Hostinský is the calculation of the left integral of the operator-valued function

$$T(u)f(x) = f(x) + \int_p^q K(x,y,u)f(y)\,\mathrm{d}y.$$

We again assume that the function K is continuous on $[p,q]\times[p,q]\times[a,b]$. As in the case of derivatives, Hostinský provides no definition and starts[1] directly with the calculation, which is again somewhat strange: He chooses a tagged partition $D: a = t_0 < t_1 < \cdots < t_m = b$, $\xi_i \in [t_{i-1}, t_i]$, and forms the operator

$$T(\xi_m)\Delta t_m \cdots T(\xi_1)\Delta t_1,$$

[1] [VH], p. 188–191

where $\Delta t_i = t_i - t_{i-1}$. He now claims it is an operator of the second kind, which is not true. The correct procedure that he probably has in mind is to take the operator-valued function

$$S(u)f(x) = \int_p^q K(x,y,u)f(y)\,\mathrm{d}y$$

and to form the operator

$$P(S,D) = (I + S(\xi_m)\Delta t_m)\cdots(I + S(\xi_1)\Delta t_1),$$

which is an operator of the second kind with the kernel

$$\sum_{s=1}^{m} \sum_{1\le i_1<\cdots<i_s\le m} \int_p^q \cdots \int_p^q K(x,z_1,\xi_{i_s})\cdots K(z_{s-1},y,\xi_{i_1})\Delta t_{i_s}\cdots\Delta t_{i_1}\,\mathrm{d}z_1\cdots\mathrm{d}z_{s-1}$$

(we have used Equation (4.1.2)). Passing to the limit for $\nu(D)\to 0$ we calculate that the left product integral of the function S is an operator of the second kind with the kernel

$$L(x,y,a,b) = \tag{4.2.2}$$
$$= \sum_{s=1}^{\infty} \int_p^q \cdots \int_p^q \int_a^b \int_a^{t_s} \cdots \int_a^{t_2} K(x,z_1,t_s)\cdots K(z_{s-1},y,t_1)\,\mathrm{d}t_1\cdots\mathrm{d}t_s\,\mathrm{d}z_1\cdots\mathrm{d}z_{s-1}.$$

In a similar way can calculate the right product integral of S, which is obtained as the limit of operators

$$P^*(S,D) = (I + S(\xi_1)\Delta t_1)\cdots(I + S(\xi_m)\Delta t_m)$$

for $\nu(D)\to 0$; the result is an operator of the second kind with the kernel

$$R(x,y,a,b) = \tag{4.2.3}$$
$$= \sum_{s=1}^{\infty} \int_p^q \cdots \int_p^q \int_a^b \int_a^{t_s} \cdots \int_a^{t_2} K(x,z_1,t_1)\cdots K(z_{s-1},y,t_s)\,\mathrm{d}t_1\cdots\mathrm{d}t_s\,\mathrm{d}z_1\cdots\mathrm{d}z_{s-1}.$$

In analogy with product integrals of matrix functions we use the symbols

$$\prod_a^b (I + S(t)\,\mathrm{d}t) \quad \text{and} \quad (I + S(t)\,\mathrm{d}t)\prod_a^b$$

to denote the left and right product integrals, respectively. Let us briefly summarize their properties:

If the operator-valued function S is defined on $[a,c]$ and $b\in[a,c]$, then[1]

$$\prod_a^c (I + S(t)\,\mathrm{d}t) = \prod_b^c (I + S(t)\,\mathrm{d}t)\cdot\prod_a^b (I + S(t)\,\mathrm{d}t).$$

[1] [VH], p. 193

The left side of the last equality represents an operator of the second kind whose kernel is $L(x,y,a,c)$, while the right side is a composition of two second-kind operators with kernels $L(x,y,a,b)$ and $L(x,y,b,c)$. Thus according to Equation (4.1.2) we have

$$L(x,y,a,c)=L(x,y,a,b)+L(x,y,b,c)+\int_p^q L(x,z,b,c)L(z,y,a,b)\,\mathrm{d}z. \quad (4.2.4)$$

Similarly, the right integral satisfies

$$(I+S(t)\,\mathrm{d}t)\prod_a^c=(I+S(t)\,\mathrm{d}t)\prod_a^b\cdot(I+S(t)\,\mathrm{d}t)\prod_b^c,$$

and we consequently

$$R(x,y,a,c)=R(x,y,a,b)+R(x,y,b,c)+\int_p^q R(x,z,a,b)R(z,y,b,c)\,\mathrm{d}z. \quad (4.2.5)$$

Hostinský next demonstrates[1] that the derivative and the integral are reverse operations; this means that if we consider the left (right) integral as a function of its upper bound, then its left (right) derivative is the original function.
If f is an arbitrary continuous function, then

$$\lim_{c\to b}\int_b^c f=f(b).$$

Using this result and the definition of L we conclude

$$\lim_{c\to b}\frac{L(x,y,b,c)}{c-b}=K(x,y,b)$$

(the series in the definition of L is uniformly convergent and we can interchange the order of limit and summation). Consequently, Equation (4.2.4) gives

$$\frac{\partial L(x,y,a,b)}{\partial b}=\lim_{c\to b}\frac{L(x,y,a,c)-L(x,y,a,b)}{c-b}=$$

$$=\lim_{c\to b}\frac{1}{c-b}\left(L(x,y,b,c)+\int_p^q L(x,z,b,c)L(z,y,a,b)\,\mathrm{d}z\right)=$$

$$=K(x,y,b)+\int_p^q K(x,z,b)L(z,y,a,b)\,\mathrm{d}z.$$

If $N_L(x,y,a,b)$ is the resolvent kernel corresponding to $L(x,y,a,b)$, then, according to Equation (4.2.1), the left derivative of the left integral of function S is an operator of the first kind with kernel

$$\frac{\partial L}{\partial b}(x,y,a,b)+\int_p^q\frac{\partial L}{\partial b}(x,t,a,b)N_L(t,y,a,b)\,\mathrm{d}t=$$

[1] [VH], p. 199–200

$$= K(x,y,b) + \int_p^q K(x,z,b)L(z,y,a,b)\,\mathrm{d}z +$$

$$+ \int_p^q \left(K(x,t,b) + \int_p^q K(x,z,b)L(z,t,a,b)\,\mathrm{d}z \right) N_L(t,y,a,b)\,\mathrm{d}t = K(x,y,b)$$

(we have applied Equation (4.1.3) to functions L and N_L), which completes the proof.

Finally let us note that in case when S is a constant function, i.e. if the kernel K does not depend on u, we obtain

$$L(x,y,a,b) = R(x,y,a,b) = \sum_{s=1}^{\infty} \frac{(b-a)^s}{s!} K^s(x,y), \tag{4.2.6}$$

where

$$K^s(x,y) = \int_p^q \cdots \int_p^q K(x,z_1)K(z_1,z_2)\cdots K(z_{s-1},y)\,\mathrm{d}z_1 \cdots \mathrm{d}z_{s-1}.$$

Remark 4.2.1. There exists a close relationship between the formulas derived by Hostinský and those obtained by Volterra. Recall that if $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuous matrix function, then

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = I + L(a,b),$$

where $L(a,b)$ is a matrix with components

$$L_{ij}(a,b) = \sum_{s=1}^{\infty} \left(\sum_{z_1,\ldots,z_{s-1}=1}^{n} \int_a^b \int_a^{t_s} \cdots \int_a^{t_2} a_{i,z_1}(t_s) \cdots a_{z_{s-1},j}(t_1)\,\mathrm{d}t_1 \cdots \mathrm{d}t_s \right).$$

This resembles Equation (4.2.2) – the only difference is that in (4.2.2) we integrate over interval $[p,q]$ instead of taking a sum over $\{1,\ldots,n\}$. Similarly, the right integral of a matrix function satisfies

$$(I + A(t)\,\mathrm{d}t) \prod_a^b = I + R(a,b),$$

where

$$R_{ij}(a,b) = \sum_{s=1}^{\infty} \left(\sum_{z_1,\ldots,z_{s-1}=1}^{n} \int_a^b \int_a^{t_s} \cdots \int_a^{t_2} a_{i,z_1}(t_1) \cdots a_{z_{s-1},j}(t_s)\,\mathrm{d}t_1 \cdots \mathrm{d}t_s \right),$$

which is an analogy of Equation (4.2.3).

Also the formula

$$\prod_a^c (I + A(t)\,\mathrm{d}t) = \prod_b^c (I + A(t)\,\mathrm{d}t) \prod_a^b (I + A(t)\,\mathrm{d}t)$$

implies

$$I + L(a,c) = (I + L(b,c))(I + L(a,b)),$$

i.e. the components satisfy

$$L_{ij}(a,c) = L_{ij}(a,b) + L_{ij}(b,c) + \sum_{z=1}^{n} L_{i,z}(b,c)L_{z,j}(a,b).$$

In a similar way we obtain that the components of the right integral satisfy

$$R_{ij}(a,c) = R_{ij}(a,b) + R_{ij}(b,c) + \sum_{z=1}^{n} R_{i,z}(a,b)R_{z,j}(b,c).$$

These relations resemble Equations (4.2.4) and (4.2.5).

Remark 4.2.2. Product integration of operator-valued functions can be used to solve certain integro-differential equations. Hostinský considers[1] the equation

$$\frac{\partial f}{\partial t}(x,t) = \int_p^q K(x,y)f(y,t)\,\mathrm{d}t,$$

whose kernel is independent of t; as he remarks, its solution was found by Volterra (without using product integration) in 1914. More generally, we can consider the integro-differential equation

$$\frac{\partial f}{\partial t}(x,t) = \int_p^q K(x,y,t)f(y,t)\,\mathrm{d}t$$

with the initial condition

$$f(x,t_0) = f_0(x), \quad x \in [p,q].$$

This equation can be rewritten using the operator notation to

$$\frac{\partial f}{\partial t}(t) = S(t)f(t),$$

where $S(t)$ is an integral operator of the first kind for every t. Thus, for small Δt we have

$$f(t+\Delta t) = f(t) + \Delta t S(t)f(t) = (I + S(t)\Delta t)f(t).$$

[1] [VH, p. 192]

Arguing similarly as in the case of matrix functions, we conclude that

$$f(t)=\prod_{t_0}^{t}(I+S(u)\,\mathrm{d}u)f_0,$$

i.e.

$$f(x,t)=f_0(x)+\int_p^q L(x,y,t_0,t)f_0(y)\,\mathrm{d}t.$$

However, the effort of Bohuslav Hostinský is not primarily directed towards integro-differential equations. He shows that the right product integral kernel $R(x,y,a,b)$ can be used to obtain the solution of a certain differential equation known from the work of Jacques Hadamard.
The last chapter of [VH] is devoted to integral operators of the first kind. By an appropriate choice of the kernel, Hostinský is able to produce infinitesimal operators of the first kind, i.e. operators that differ infinitesimally from the identity operator (thus the kernel is something like the Dirac δ-function). The composition of such infinitesimal operators leads to an operator whose kernel K provides a solution of the equation

$$K(x,y,u,v)=\int_p^q K(x,z,u,w)K(z,y,w,v)\,\mathrm{d}z.$$

This equation is known as the Chapman or Chapman-Kolmogorov equation and is often encountered in the theory of stochastic processes (see also Section 1.4 and Equation (1.4.7), which represents a discrete version of the Chapman equation). The topic is rather special and we don't discuss it here.

4.3 General definition of product integral

Although Hostinský is interested only in integral operators on the space $\mathcal{C}([p,q])$, it is possible to work more generally with linear operators on an arbitrary Banach space X. Before proceeding to the corresponding definitions we recall that the norm of a linear operator T on the space X is defined as

$$\|T\|=\sup\{\|T(x)\|;\ \|x\|=1\}.$$

Let $\mathcal{L}(X)$ denote the space of all bounded linear operators on X, i.e. operators whose norm is finite; $\mathcal{L}(X)$ is a normed vector space.

Definition 4.3.1. Let X be a Banach space, $A:[a,b]\to\mathcal{L}(X)$, $t\in[a,b]$. Assume that $A(t)^{-1}$ exists. We define the left and right derivatives of A at t as

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t)=\lim_{h\to 0}\frac{A(t+h)A(t)^{-1}-I}{h},$$

$$A(t)\frac{\mathrm{d}}{\mathrm{d}t}=\lim_{h\to 0}\frac{A(t)^{-1}A(t+h)-I}{h},$$

provided the limits exist. Of course, at the endpoints of $[a,b]$ we require only the existence of the corresponding one-sided limits.

To every function $A : [a,b] \to \mathcal{L}(X)$ and every tagged partition

$$D : a = t_0 < t_1 < \cdots < t_m = b,\ \xi_i \in [t_{i-1}, t_i],$$

we assign the operators

$$P(A,D) = \prod_{i=m}^{1} (I + A(\xi_i)\Delta t_i) = (I + A(\xi_m)\Delta t_m) \cdots (I + A(\xi_1)\Delta t_1),$$

$$P^*(A,D) = \prod_{i=1}^{m} (I + A(\xi_i)\Delta t_i) = (I + A(\xi_1)\Delta t_1) \cdots (I + A(\xi_m)\Delta t_m).$$

Definition 4.3.2. Consider function $A : [a,b] \to \mathcal{L}(X)$. The left and right product integrals of A are the operators

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \lim_{\nu(D)\to 0} P(A,D),$$

$$(I + A(t)\,\mathrm{d}t) \prod_a^b = \lim_{\nu(D)\to 0} P^*(A,D),$$

provided the limits exist.

Example 4.3.3. Recall that for every $A \in \mathcal{L}(X)$ we define the exponential $e^A \in \mathcal{L}(X)$ by

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!},$$

where $A^0 = I$ is the identity operator, $A^1 = A$ and $A^{n+1} = A^n \cdot A$ for every $n \in \mathbf{N}$. It can be proved (see Theorem 5.5.11) that

$$\prod_a^b (I + A\,\mathrm{d}t) = (I + A\,\mathrm{d}t) \prod_a^b = e^{A(b-a)}$$

(the special case for integral operators with a constant kernel follows also from Equation (4.2.6)).

We state the next theorem without proof; a more general version will be proved in Chapter 5 (see Theorem 5.6.2).

Theorem 4.3.4. Let $A : [a,b] \to \mathcal{L}(X)$ be a continuous function. Then the functions

$$Y(t) = \prod_a^t (I + A(u)\,\mathrm{d}u),$$

$$Z(t) = (I + A(u)\,\mathrm{d}u)\prod_a^t$$

satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t} Y(t) = A(t), \quad Z(t)\frac{\mathrm{d}}{\mathrm{d}t} = A(t)$$

for every $t \in [a, b]$.

Definition 4.3.2 represents a straightforward generalization of the Riemann product integral as defined by Volterra. In the following chapter we provide an even more general definition applicable to functions $A : [a, b] \to X$, where X is a Banach algebra. It is also possible to proceed in a different way and try to generalize the Lebesgue product integral from Chapter 3; the following paragraphs outline this possibility.

Definition 4.3.5. A function $A : [a, b] \to \mathcal{L}(X)$ is called a step function if there exists a partition $a = t_0 < t_1 < \cdots < t_m = b$ and operators $A_1, \ldots, A_m \in \mathcal{L}(X)$ such that $A(t) = A_i$ for $t \in (t_{i-1}, t_i)$, $i = 1, \ldots, m$. For such a step function we define

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \prod_{i=m}^{1} e^{A_i(t_i - t_{i-1})},$$

$$(I + A(t)\,\mathrm{d}t)\prod_a^b = \prod_{i=1}^{m} e^{A_i(t_i - t_{i-1})}.$$

Definition 4.3.6. A function $A : [a, b] \to \mathcal{L}(X)$ is called product integrable if there exists a sequence of step functions $A_n : [a, b] \to \mathcal{L}(X)$, $n \in \mathbf{N}$ such that

$$\lim_{n\to\infty} (L) \int_a^b \|A_n(t) - A(t)\|\,\mathrm{d}t = 0$$

(where (L) denotes the Lebesgue integral of a real function). We then define

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \lim_{n\to\infty} \prod_a^b (I + A_n(t)\,\mathrm{d}t),$$

$$(I + A(t)\,\mathrm{d}t)\prod_a^b = \lim_{n\to\infty} (I + A_n(t)\,\mathrm{d}t)\prod_a^b.$$

The method from Chapter 3 can be again used to show that the value of product integral does not depend on the choice of a particular sequence of step functions $\{A_n\}_{n=1}^{\infty}$. The above defined integral is called the Lebesgue product integral or the Bochner product integral; the class of integrable functions is larger as compared to Definition 4.3.2. More information about this type of product integral can be found in [DF, Sch1].

Chapter 5

Product integration in Banach algebras

A final treatment of Riemann product integration is given in the article [Mas] by Pesi Rustom Masani; it was published in 1947. Consider a matrix-valued function $f : [a, b] \to \mathbf{R}^{n\times n}$ and recall that Volterra defined the product integral of f as the limit of products

$$P(f, D) = \prod_{k=m}^{1} (I + f(\xi_k)\Delta x_k)$$

corresponding to tagged partitions D of interval $[a, b]$. This definition is also applicable to operator-valued functions $f : [a, b] \to \mathcal{L}(X)$, where $\mathcal{L}(X)$ is the space of all bounded linear operators on a Banach space X. It is just sufficient to replace multiplication by composition of operators in the definition of $P(f, D)$; the role of identity matrix is now played by the identity operator I.

Masani's intent was to define the product integral of a function $f : [a, b] \to X$ for the most general space X possible. Let X be a normed vector space equipped with the operation of multiplication. Assuming there is a vector $1 \in X$ such that $1 \cdot x = x \cdot 1 = x$ for every $x \in X$ and $\|1\| = 1$, we let

$$P(f, D) = \prod_{k=m}^{1} (1 + f(\xi_k)\Delta x_k),$$

where D is an arbitrary tagged partition of $[a, b]$. We would like to define the product integral as the limit

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = \lim_{\nu(D)\to 0} P(f, D).$$

To obtain a reasonable theory it is necessary that the space X is complete, i.e. it is a Banach space.

Before giving an overview of Masani's result let's start with a short biography (see also [PRM, IMS]). Pesi Rustom Masani was born in Bombay, 1919. He obtained his doctoral degree at Harvard in 1946; the thesis concerned product integration in Banach algebras and it was supervised by Garrett Birkhoff [1]. During the years 1948–58 Masani held the chairs of professor of mathematics and science researcher in Bombay and then he returned to the United States. In the 1970's he accepted the position at the University of Pittsburgh. Masani was active in mathematics even after his retirement in 1989. He died in Pittsburgh on the 15th October 1999.

[1] G. Birkhoff also devoted himself to product integration, see [GB].

Pesi R. Masani[1]

Masani contributed to the development of integration theory, functional analysis, theory of probability and mathematical statistics. The appendix in [DF] written by Masani also concerns product integration. He collaborated with Norbert Wiener and edited his collected works after Wiener's death. Masani was also interested in history, philosophy, theology and politics.

5.1 Riemann-Graves integral

We begin with a brief recapitulation of facts concerning integration of vector-valued functions (see also [Mas]). The notion of Graves integral is a direct generalization of Riemann integral and was presented by Lawrence M. Graves in 1927.

Let X be a Banach space, $f : [a,b] \to X$. To every tagged partition $D : a = t_0 < t_1 < \cdots < t_m = b$ of interval $[a,b]$ with tags $\xi_i \in [t_{i-1}, t_i]$, $i = 1, \ldots, m$ we assign the sum

$$S(f,D) = \sum_{i=1}^{m} f(\xi_i)\Delta t_i,$$

where $\Delta t_i = t_i - t_{i-1}$. We recall that if $T(D) \in X$ is a vector dependent on the choice of a tagged partition D, then

$$\lim_{\nu(D)\to 0} T(D) = T$$

means that to every $\varepsilon > 0$ there is $\delta > 0$ such that $\|T(D) - T\| < \varepsilon$ for every partition D of $[a,b]$ such that $\nu(D) < \delta$.

Definition 5.1.1. A function $f : [a,b] \to X$ is called integrable if

$$\lim_{\nu(D)\to 0} S(f,D) = S_f$$

[1] Photo from `http://www.york.ac.uk/depts/maths/histstat/people/`

for some $S_f \in X$. We speak of Riemann-Graves integral of function f on $[a,b]$ and denote $S_f = \int_a^b f(t)\,\mathrm{d}t$.

The following theorem provides additional two equivalent characterizations of integrable functions; recall that the notation $D' \prec D$ means that the partition D' is a refinement of partition D (see Definition 3.1.8).

Theorem 5.1.2. Let $f : [a,b] \to X$. The following statements are equivalent:

1) f is integrable and $\int_a^b f(t)\,\mathrm{d}t = S_f$.
2) Every sequence of partitions $\{D_n\}_{n=1}^{\infty}$ of $[a,b]$ such that $\nu(D_n) \to 0$ satisfies $\lim_{n\to\infty} S(f, D_n) = S_f$.
3) For every $\varepsilon > 0$ there is a partition D_ε of $[a,b]$ such that $\|S(f,D) - S_f\| < \varepsilon$ for every $D \prec D_\varepsilon$.

The proof proceeds in the same way as in the case when f is a real function. The rest of this section summarizes the basic results concerning the Riemann-Graves integral; again, the proofs can be carried out in the classical way.

Theorem 5.1.3. Let $f : [a,b] \to X$. Then the following statements are equivalent:

1) f is integrable.
2) For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|S(f,D_1) - S(f,D_2)\| < \varepsilon$ whenever D_1 and D_2 are tagged partitions of $[a,b]$ satisfying $\nu(D_1) < \delta$, $\nu(D_2) < \delta$.

Theorem 5.1.4. If $f : [a,b] \to X$ is an integrable function, then it is bounded and

$$\left\| \int_a^b f(t)\,\mathrm{d}t \right\| \le (b-a) \sup_{t\in[a,b]} \|f(t)\|.$$

Theorem 5.1.5. Let $f : [a,b] \to X$. If the integral $\int_a^b f(t)\,\mathrm{d}t$ exists and if $[c,d] \subset [a,b]$, then the integral $\int_c^d f(t)\,\mathrm{d}t$ exists as well.

Theorem 5.1.6. Let $f : [a,c] \to X$, $a < b < c$. Suppose that the integrals $\int_a^b f(t)\,\mathrm{d}t$ and $\int_b^c f(t)\,\mathrm{d}t$ exists. Then the integral $\int_a^c f(t)\,\mathrm{d}t$ also exists and

$$\int_a^c f(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t + \int_b^c f(t)\,\mathrm{d}t.$$

The following two statements generalize the fundamental theorem of calculus to the case of vector-valued functions $f : [a,b] \to X$. The derivative of such a function is of course defined as

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

provided the limit exists (in the case when x_0 is one of the boundary points of $[a,b]$ we require only existence of the corresponding one-sided limit). Since X is a normed space, the last equation means that

$$\lim_{x\to x_0} \left\| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right\| = 0.$$

Theorem 5.1.7. Let $f : [a, b] \to X$ be integrable and put $F(x) = \int_a^x f(t)\,\mathrm{d}t$. If f is continuous at $x_0 \in [a, b]$, then $F'(x_0) = f(x_0)$.

Theorem 5.1.8. Let $F : [a, b] \to X$ and $F'(t) = f(t)$ for every $t \in [a, b]$. If f is an integrable function, then

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a).$$

Theorem 5.1.9. Let $f, g : [a, b] \to X$ be integrable functions, $\alpha, \beta \in \mathbf{R}$. Then

$$\int_a^b (\alpha f(t) + \beta g(t))\,\mathrm{d}t = \alpha \int_a^b f(t)\,\mathrm{d}t + \beta \int_a^b g(t)\,\mathrm{d}t.$$

The set of all integrable functions is thus a vector space; it is interesting to note that if the space X is equipped with the operation of multiplication (i.e. it is a Banach algebra, see the next section), then a product of two integrable functions need not be an integrable function. Another surprising fact concerning the Riemann-Graves integral is that every bounded function which is almost everywhere continuous is also integrable, but the converse statement is no longer true (it holds only in finite-dimensional spaces X).

5.2 Definition of product integral

Masani turns his attention to the product analogy of the Riemann-Graves integral. In the sequel we assume that X is a Banach algebra (Masani uses the term normed ring), i.e. that

1) X is a Banach space,
2) X is an associative algebra with a unit vector $1 \in X$, $\|1\| = 1$,
3) $\|x \cdot y\| \leq \|x\|\|y\|$ for every $x, y \in X$.

The second condition means that for every pair $x, y \in X$ the product $x \cdot y \in X$ is defined, that the multiplication is associative and that there exists a vector $1 \in X$ such that $1 \cdot x = x \cdot 1 = x$ for every $x \in X$ and $\|1\| = 1$; we use the same symbol 1 to denote the unit vector of X as well as the number $1 \in \mathbf{R}$; the meaning should be always clear from the context.

Let $f : [a, b] \to X$. To every partition D of $[a, b]$ we assign the product

$$P(f, D) = \prod_{i=m}^{1} (1 + f(\xi_i)\Delta t_i) = (1 + f(\xi_m)\Delta t_m) \cdots (1 + f(\xi_1)\Delta t_1).$$

Definition 5.2.1. A function $f : [a, b] \to X$ is called product integrable if there is a vector $P_f \in X$ such that for every $\varepsilon > 0$ there exists a partition D_ε of $[a, b]$ such that

$$\|P(f, D) - P_f\| < \varepsilon$$

whenever $D \prec D_\varepsilon$. The vector P_f is called the (left) product integral of f and we use the notation $\prod_a^b (1 + f(t)\,\mathrm{d}t) = P_f$.

Remark 5.2.2. Masani also defines the right product integral as the limit of the products

$$P^*(f, D) = \prod_{i=1}^{m} (1 + f(\xi_i)\Delta t_i) = (1 + f(\xi_1)\Delta t_1) \cdots (1 + f(\xi_m)\Delta t_m),$$

which are obtained by reversing the order of factors in $P(f, D)$. Masani uses the symbols

$$\int_a^{b\frown} (1 + f(t)\,\mathrm{d}t), \quad \int_a^{b\smile} (1 + f(t)\,\mathrm{d}t)$$

to denote left and right product integrals. As he remarks, it is sufficient to study either the left integral or the right integral, respectively. This is because the following *principle of duality* holds:

To every Banach algebra X there is a dual algebra X^* which is identical with X except the operation of multiplication: We define the product $x \cdot y$ in X^* as the vector $y \cdot x$, where the last multiplication is carried out in X. Every statement C about Banach algebra X has a corresponding dual statement C^*, which is obtained by reversing the order of all products in C. Consequently, every occurence of the term "left product integral" must be replaced by "right product integral" and vice versa. A dual statement C^* is true in X^* if and only if C is true in X. In case C is true in every Banach algebra, the same can be said of C^*.

Theorem 5.2.3.[1] Let $f : [a, b] \to X$ be a bounded function. The following statements are equivalent:

1) f is product integrable and $\prod_a^b (1 + f(t)\,\mathrm{d}t) = P_f$.
2) Every sequence of partitions $\{D_n\}_{n=1}^{\infty}$ of interval $[a, b]$ such that $\nu(D_n) \to 0$ satisfies $\lim_{n\to\infty} P(f, D_n) = P_f$.
3) $\lim_{\nu(D)\to 0} P(f, D) = P_f$.

Proof. The equivalence of statements 2) and 3) is proved in the same way as in the case of ordinary integral. Assume that 3) holds, i.e. to every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|P(f, D) - P_f\| < \varepsilon$ for every partition D of interval $[a, b]$ which satisfies $\nu(D) < \delta$. Let D_ε be such a partition. Then for every $D \prec D_\varepsilon$ we have $\nu(D) \le \nu(D_\varepsilon) < \delta$, and therefore $\|P(f, D) - P_f\| < \varepsilon$; thus we have proved the implication 3) $\Rightarrow$ 1). Masani gives only a brief indication of the proof of 1) $\Rightarrow$ 3), details are left to the reader; boundedness of f is important here. $\square$

The following theorem represents a "Cauchy condition" for the existence of product integral.

Theorem 5.2.4. Let $f : [a, b] \to X$ be bounded. The following statements are equivalent:

[1] [Mas], p. 157–159

1) f is product integrable.
2) To every $\varepsilon > 0$ there is a partition D_ε such that $\|P(f,D) - P(f,D_\varepsilon)\| < \varepsilon$ whenever $D \prec D_\varepsilon$.
3) To every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|P(f,D_1) - P(f,D_2)\| < \varepsilon$ whenever D_1, D_2 are partitions of $[a,b]$ satisfying $\nu(D_1) < \delta$, $\nu(D_2) < \delta$.

Proof. The equivalence of statements 1) and 2) is proved in the same way as in the case of ordinary integral. The statement 3) is clearly equivalent to the statement 3) of the previous theorem. □

5.3 Useful inequalities

We now present five inequalities which will be useful later. Masani didn't prove the first three; we have however met the first two in Chapter 3 – see the Lemmas 3.1.3 and 3.4.2. Although we have proved them only for matrices, the proofs are valid even for elements of an arbitrary Banach algebra X.

Lemma 5.3.1.[1] Let $x_k \in X$ for $k = 1, \dots, m$. Then

$$\left\| \prod_{k=1}^{m} (1 + x_k) \right\| \leq \exp\left(\sum_{k=1}^{m} \|x_k\| \right).$$

Lemma 5.3.2.[2] Let $x_k, y_k \in X$ for $k = 1, \dots, m$. Then

$$\left\| \prod_{k=1}^{m} (1 + x_k) - \prod_{k=1}^{m} (1 + y_k) \right\| \leq \exp\left(\sum_{k=1}^{m} \|x_k\| \right) \left(\exp \sum_{k=1}^{m} \|x_k - y_k\| - 1 \right).$$

Lemma 5.3.3.[3] Let $x_k \in X$ for $k = 1, \dots, m$. Then

$$\left\| \prod_{k=1}^{m} (1 + x_k) - 1 \right\| \leq \exp\left(\sum_{k=1}^{m} \|x_k\| \right) - 1.$$

Proof. Elementary calculation yields

$$\left\| \prod_{k=1}^{m} (1 + x_k) - 1 \right\| = \left\| \sum_{j=1}^{m} \left(\sum_{1 \leq i_1 < \cdots < i_j \leq m} x_{i_1} \cdots x_{i_j} \right) \right\| \leq$$

$$\leq \sum_{j=1}^{m} \left(\sum_{1 \leq i_1 < \cdots < i_j \leq m} \|x_{i_1}\| \cdots \|x_{i_j}\| \right) \leq \sum_{j=1}^{m} \frac{1}{j!} \left(\sum_{i_1, \dots, i_j = 1}^{m} \|x_{i_1}\| \cdots \|x_{i_j}\| \right) =$$

[1] [Mas], p. 153
[2] [Mas], p. 154
[3] [Mas], p. 153

$$= \sum_{j=1}^{m} \frac{1}{j!}(\|x_1\| + \cdots + \|x_m\|)^j \leq \exp\left(\sum_{k=1}^{m} \|x_k\|\right) - 1.$$

□

Lemma 5.3.4.[1] Let $x_k \in X$ for $k = 1, \dots, m$. Then

$$\left\| \prod_{k=1}^{m} (1 + x_k) - \left(1 + \sum_{k=1}^{m} x_k\right) \right\| \leq \left(\exp \sum_{k=1}^{m} \|x_k\| - 1\right) \sum_{k=1}^{m} \|x_k\|.$$

Proof. The statement is a simple consequence of the inequality

$$\prod_{k=1}^{m} (1 + x_k) - \left(1 + \sum_{k=1}^{m} x_k\right) = \sum_{k=1}^{m} x_k \left(\prod_{j=k+1}^{m} (1 + x_j) - 1\right)$$

and Lemma 5.3.3. □

Lemma 5.3.5.[2] Let $m, n \in \mathbf{N}$, $u, v, x_j, y_k \in X$, $\|x_j\|, \|y_k\| \leq 1/2$ for every $j = 1, \dots, m$ and $k = 1, \dots, n$. Then

$$\left\| \prod_{j=1}^{m} (1 + x_j) \cdot (u - v) \cdot \prod_{k=1}^{n} (1 + y_k) \right\| \geq \exp\left(-2\left(\sum_{j=1}^{m} \|x_j\| + \sum_{k=1}^{n} \|y_k\|\right)\right) \|u - v\|.$$

Proof. Define $f(t) = e^{2t} - te^{2t} - 1$. Then

$$f'(t) = e^{2t}(1 - 2t) \geq 0, \quad t \in [0, 1/2],$$

and therefore

$$e^{2t} - te^{2t} - 1 = f(t) \geq f(0) = 0, \quad t \in [0, 1/2].$$

We get

$$1 - t \geq e^{-2t}, \quad t \in [0, 1/2].$$

Now let $x, w \in X$, $\|x\| \leq 1/2$. Then

$$\|w\| \leq \|w + x \cdot w\| + \|x \cdot w\| \leq \|(1 + x) \cdot w\| + \|x\| \cdot \|w\|,$$

which implies

$$\|(1 + x) \cdot w\| \geq \|w\|(1 - \|x\|) \geq \|w\| \exp(-2\|x\|). \tag{5.3.1}$$

For $y \in X$, $\|y\| \leq 1/2$ we obtain in a similar way

$$\|w\| \leq \|w + w \cdot y\| + \|w \cdot y\| \leq \|w \cdot (1 + y)\| + \|y\| \cdot \|w\|,$$

[1] [Mas], p. 153
[2] [Mas], p. 152–153

$$\|w \cdot (1+y)\| \geq \|w\|(1-\|y\|) \geq \|w\| \exp(-2\|y\|). \tag{5.3.2}$$

To complete the proof it is sufficient to use m times the Inequality (5.3.1) and n times the Inequality (5.3.2). □

5.4 Properties of product integral

This section summarizes the basic properties of product integrable functions. We first prove that every product integrable function is necessarily bounded.

Lemma 5.4.1.[1] To every $\Delta : [a,b] \to (0,\infty)$ there exists a tagged partition $D : a = t_0 < t_1 < \cdots < t_m = b$, $\xi_i \in [t_{i-1}, t_i]$, such that $t_i - t_{i-1} \leq \Delta(\xi_i)$.

Proof. The system of intervals $\{(t - \Delta(t)/2, t + \Delta(t)/2),\ t \in [a,b]\}$ forms an open covering of $[a,b]$ and the result follows from the Heine–Borel theorem. It is also a simple consequence of Cousin's lemma (see [Sch2], p. 55 or [RG], Lemma 9.2). □

Theorem 5.4.2.[2] Every product integrable function f is bounded and

$$\left\| \prod_a^b (1 + f(t)\,\mathrm{d}t) \right\| \leq \exp\left((b-a) \sup_{t\in[a,b]} \|f(t)\| \right).$$

Proof. Assume that f is not bounded. Choose $N \in \mathbf{N}$ and $\delta > 0$. Define

$$\Delta(x) = \begin{cases} \min(\delta, (2\|f(x)\|)^{-1}) & \text{if } \|f(x)\| > 0, \\ \delta & \text{if } f(x) = 0. \end{cases}$$

According to Lemma 5.4.1 there exists a tagged partition $D : a = t_0 < t_1 < \cdots < t_m = b$, $\xi_i \in [t_{i-1}, t_i]$, such that

$$t_i - t_{i-1} \leq \Delta(\xi_i). \tag{5.4.1}$$

Clearly $\nu(D) \leq \delta$. Since f is not bounded, we can find a sequence of points $\{x_n\}_{n=1}^\infty$ from $[a,b]$ such that $x_n \to x \in [a,b]$ and $\|f(x_n)\| \geq n$. There must be a point $y \in \{x_n\}_{n=1}^\infty$, which lies in the same interval $[t_{j-1}, t_j]$ as the point x and such that

$$\|f(y) - f(x)\| \geq \|f(y)\| - \|f(x)\| \geq N \cdot \left(\exp(-m) \cdot \min_{1\leq i\leq m} (t_i - t_{i-1}) \right)^{-1}.$$

Let D_1 and D_2 be tagged partitions that are obtained from D by replacing the tag ξ_j by x and y, respectively. Then, according to Lemma 5.3.5 and Inequality (5.4.1),

$$\|P(f,D_1) - P(f,D_2)\| \geq \exp\left(-2 \sum_{i\neq j} \|f(\xi_i)\|(t_i - t_{i-1}) \right) \|f(x) - f(y)\|(t_j - t_{j-1}) \geq$$

[1] [Mas], p. 162
[2] [Mas], p. 163

$$\geq \exp(-m)\|f(x) - f(y)\|(t_j - t_{j-1}) \geq N.$$

Since $\nu(D_1) = \nu(D_2) = \nu(D) \leq \delta$, the number δ can be arbitrarily small and N arbitrarily large, we arrive at a contradiction with Theorem 5.2.4. The second part of the theorem is easily proved using Lemma 5.3.1, which guarantees that

$$\|P(f,D)\| \leq \exp\left(\sum_{i=1}^{m} \|f(\xi_i)\|(t_i - t_{i-1})\right) \leq \exp\left((b-a) \sup_{t\in[a,b]} \|f(t)\|\right)$$

for every tagged partition D of $[a,b]$. □

Theorem 5.4.3.[1] Assume that $\prod_a^b(1+f(t)\,\mathrm{d}t)$ exists. If $[c,d] \subset [a,b]$, then $\prod_c^d(1+f(t)\,\mathrm{d}t)$ exists as well.

Proof. Denote $M = \sup_{t\in[a,b]}\|f(t)\| < \infty$. Let D_1, D_2 be tagged partitions of $[c,d]$, D_A a tagged partition of $[a,c]$ satisfying $\nu(D_A) < 1/(2M)$ and D_B a tagged partition of $[d,b]$ satisfying $\nu(D_B) < 1/(2M)$. Letting

$$D_1^* = D_A \cup D_1 \cup D_B, \quad D_2^* = D_A \cup D_2 \cup D_B,$$

we obtain (using Lemma 5.3.5)

$$\|P(f,D_1^*) - P(f,D_2^*)\| = \|P(f,D_B)\,(P(f,D_1) - P(f,D_2))\,P(f,D_A)\| \geq$$

$$\geq \exp\left(-2\sum_{D_A\cup D_B} f(\xi_i)(t_i - t_{i-1})\right)\|P(f,D_1) - P(f,D_2)\| \geq$$

$$\geq \exp(-2M(b-a))\|P(f,D_1) - P(f,D_2)\|,$$

therefore

$$\|P(f,D_1) - P(f,D_2)\| \leq \exp(2M(b-a))\|P(f,D_1^*) - P(f,D_2^*)\|.$$

Because f is product integrable, to every $\varepsilon > 0$ there is a tagged partition D_ε^* of interval $[a,b]$ such that

$$\|P(f,D^*) - P(f,D_\varepsilon^*)\| < \frac{\varepsilon}{\exp(2M(b-a))}$$

whenever $D^* \prec D_\varepsilon^*$. Without loss of generality assume that $D_\varepsilon^* = D_A \cup D_\varepsilon \cup D_B$, where D_A is a partition of $[a,c]$ satisfying $\nu(D_A) < 1/(2M)$, D_ε is a partition of $[c,d]$ and D_B is a partition of $[d,b]$ satisfying $\nu(D_B) < 1/(2M)$. If $D \prec D_\varepsilon$, we construct the partition $D^* = D_A \cup D \cup D_B$. Then

$$\|P(f,D) - P(f,D_\varepsilon)\| \leq \exp(2M(b-a))\|P(f,D^*) - P(f,D_\varepsilon^*)\| < \varepsilon.$$

□

[1] [Mas], p. 163–165

Theorem 5.4.4.[1] If $a < b < c$ and the integrals $\prod_a^b (1+f(t)\,\mathrm{d}t)$ and $\prod_b^c (1+f(t)\,\mathrm{d}t)$ exist, then the integral $\prod_a^c (1+f(t)\,\mathrm{d}t)$ exists as well and

$$\prod_a^c (1+f(t)\,\mathrm{d}t) = \prod_b^c (1+f(t)\,\mathrm{d}t) \cdot \prod_a^b (1+f(t)\,\mathrm{d}t).$$

Proof. Masani's proof is somewhat incomplete; we present a modified version. The assumptions imply the existence of a tagged partition D_ε^1 of $[a,b]$ and a tagged partition D_ε^2 of $[b,c]$ such that

$$\left\| P(f,D^1) - \prod_a^b (1+f(t)\,\mathrm{d}t) \right\| < \varepsilon,$$

$$\left\| P(f,D^2) - \prod_b^c (1+f(t)\,\mathrm{d}t) \right\| < \varepsilon$$

whenever $D^1 \prec D_\varepsilon^1$ and $D^2 \prec D_\varepsilon^2$. Let $D_\varepsilon = D_\varepsilon^1 \cup D_\varepsilon^2$. Then every tagged partition $D \prec D_\varepsilon$ can be written as $D = D^1 \cup D^2$, where $D^1 \prec D_\varepsilon^1$ and $D^2 \prec D_\varepsilon^2$. We have $P(f,D) = P(f,D^2) \cdot P(f,D^1)$ and

$$\left\| P(f,D) - \prod_b^c (1+f(t)\,\mathrm{d}t) \cdot \prod_a^b (1+f(t)\,\mathrm{d}t) \right\| \le$$

$$\le \left\| P(f,D^2) \left(P(f,D^1) - \prod_a^b (1+f(t)\,\mathrm{d}t) \right) \right\| +$$

$$+ \left\| \left(P(f,D^2) - \prod_b^c (1+f(t)\,\mathrm{d}t) \right) \prod_a^b (1+f(t)\,\mathrm{d}t) \right\| \le$$

$$\le \left(\left\| \prod_b^c (1+f(t)\,\mathrm{d}t) \right\| + \varepsilon \right) \varepsilon + \varepsilon \left\| \prod_a^b (1+f(t)\,\mathrm{d}t) \right\|,$$

which completes the proof. □

Statements similar to Theorem 5.4.4 have already appeared in the work of Volterra and Schlesinger. Their versions are however less general: They assume that f is Riemann integrable on $[a,c]$, which implies the existence of product integral on $[a,c]$ and the rest of the proof is trivial. Masani on the other hand proves that the existence of product integral on $[a,b]$ and on $[b,c]$ implies the existence of product integral on $[a,c]$. The same remark also applies to Theorem 5.4.3. In the following section we prove that the product integral exists if and only if the function is

[1] [Mas], p. 165

(Riemann-Graves) integrable; the proof of this fact is nevertheless based on the use of Theorem 5.4.3.

Lemma 5.4.5. Every $x \in X$ such that $\|x-1\| < 1$ has an inverse element and

$$x^{-1} = \sum_{n=0}^{\infty} (1-x)^n.$$

Proof. The condition $\|x-1\| < 1$ implies that the infinite series given above is absolutely convergent; let x^{-1} be defined as the sum of that series. If

$$s_k = \sum_{n=0}^{k} (1-x)^n,$$

then $s_{k+1} = 1 + (1-x) \cdot s_k = 1 + s_k \cdot (1-x)$.
Passing to the limit $k \to \infty$ we obtain

$$x^{-1} = 1 + (1-x) \cdot x^{-1} = 1 + x^{-1} \cdot (1-x),$$

i.e. $x^{-1} \cdot x = x \cdot x^{-1} = 1$. □

Theorem 5.4.6.[1] If $f : [a,b] \to X$ is a product integrable function, then $\prod_a^b (1 + f(t)\,\mathrm{d}t)$ is an invertible element of the Banach algebra X.

Proof. Denote $M = \sup_{t \in [a,b]} \|f(t)\| < \infty$. Choose $\delta > 0$ such that $\exp(M\delta) < 2$ and a partition $D : a = t_0 < t_1 < \cdots < t_m = b$ such that $\nu(D) \leq \delta$. Then

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = \prod_{i=m}^{1} \prod_{t_{i-1}}^{t_i} (1 + f(t)\,\mathrm{d}t) \tag{5.4.2}$$

Lemma 5.3.3 implies that for every $i = 1, \ldots, m$

$$\left\| \prod_{t_{i-1}}^{t_i} (1 + f(t)\,\mathrm{d}t) - 1 \right\| \leq \exp(M(t_i - t_{i-1})) - 1 < 1,$$

i.e. $\prod_{t_{i-1}}^{t_i} (1 + f(t)\,\mathrm{d}t)$ is (according to Lemma 5.4.5) an invertible element of the algebra X. As a consequence of (5.4.2) we obtain

$$\left(\prod_a^b (1 + f(t)\,\mathrm{d}t) \right)^{-1} = \prod_{i=1}^{m} \left(\prod_{t_{i-1}}^{t_i} (1 + f(t)\,\mathrm{d}t) \right)^{-1}.$$

□

[1] [Mas], p. 165–166

5.5 Integrable and product integrable functions

Masani now proceeds to prove an important theorem which states that the classes of integrable and product integrable functions coincide. The fact that the existence of Riemann integral implies the existence of product integral was already known to Volterra; the reverse implication appears for the first time in Masani's paper.

Lemma 5.5.1. Let $f : [a, b] \to X$ be a bounded function. For every $\varepsilon > 0$ there exists $\delta > 0$ such that if $[c, d] \subseteq [a, b]$, $d - c < \delta$ and D is a tagged partition of $[c, d]$, then

$$\|P(f, D) - (1 + S(f, D))\| \le \varepsilon(d - c).$$

Proof. Denote $M = \sup_{t\in[a,b]} \|f(t)\|$. Choose $\delta > 0$ such that

$$(\exp(M\delta) - 1) < \varepsilon/M.$$

Then according to Lemma 5.3.4

$$\|P(f, D) - (1 + S(f, D))\| \le (\exp(M(d - c)) - 1)M(d - c) \le \varepsilon(d - c).$$

□

Definition 5.5.2. Let $Y \subseteq X$. The diameter of the set Y is the number

$$\operatorname{diam} Y = \sup\{\|y_1 - y_2\|;\ y_1, y_2 \in Y\}.$$

The convex closure of Y is the set

$$\operatorname{conv} Y = \left\{\sum_{i=1}^{k} \alpha_i y_i;\ \ k \in \mathbf{N}, y_i \in Y, \alpha_i \ge 0, \sum_{i=1}^{k} \alpha_i = 1\right\}.$$

Theorem 5.5.3.[1] If $Y \subseteq X$, then

$$\operatorname{diam} \operatorname{conv} Y = \operatorname{diam} Y.$$

Proof. The proof is not difficult, although it's not included in Masani's paper. Since $Y \subseteq \operatorname{conv} Y$, it is sufficient to prove that

$$\operatorname{diam} \operatorname{conv} Y \le \operatorname{diam} Y.$$

Let y^1, $y^2 \in \operatorname{conv} Y$,

$$y^1 = \sum_{i=1}^{l} \alpha_i y_i^1, \quad y^2 = \sum_{j=1}^{m} \beta_j y_j^2,$$

[1] [Mas], p. 159

where $y_i^1, y_j^2 \in Y$, $i = 1, \dots, l$, $j = 1, \dots, m$,

$$\sum_{i=1}^{l} \alpha_i = \sum_{j=1}^{m} \beta_j = 1.$$

Then

$$\|y^1 - y^2\| = \left\| \sum_{j=1}^{m} \beta_j \left(\sum_{i=1}^{l} \alpha_i y_i^1 \right) - \sum_{i=1}^{l} \alpha_i \left(\sum_{j=1}^{m} \beta_j y_j^2 \right) \right\| = \left\| \sum_{i=1}^{l} \sum_{j=1}^{m} \alpha_i \beta_j (y_i^1 - y_j^2) \right\| \leq$$

$$\leq \sum_{i=1}^{l} \sum_{j=1}^{m} \alpha_i \beta_j \|y_i^1 - y_j^2\| \leq \sum_{i=1}^{l} \sum_{j=1}^{m} \alpha_i \beta_j \operatorname{diam} Y = \operatorname{diam} Y.$$

□

Lemma 5.5.4.[1] Let $f : [a, b] \to X$ be a product integrable function. Then for every $\varepsilon > 0$ there is a partition $D:\ a = t_0 < t_1 < \cdots < t_{m-1} < t_m = b$ such that

$$\left\| \prod_{i=m}^{1} (1 + f(\xi_i) \Delta t_i) - \prod_{k=m}^{1} (1 + f(\eta_i) \Delta t_i) \right\| < \varepsilon$$

for every choice of $\xi_i, \eta_i \in [t_{i-1}, t_i]$, $i = 1, \dots, m$.

Proof. Follows from Theorem 5.2.4. □

Remark 5.5.5. Masani notes that the reverse implication is not valid; his counterexample is

$$f(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1/x & \text{if } x \in (0, 1/2), \\ -2 & \text{if } x \in [1/2, 1]. \end{cases}$$

Taking the partition $t_0 = 0 < t_1 = 1/2 < t_2 = 1$ we obtain

$$\prod_{i=2}^{1} (1 + f(\xi_i) \Delta t_i) = 0$$

for every choice of $\xi_i \in [t_{i-1}, t_i]$, but f is not product integrable (because it is not bounded).

Lemma 5.5.6.[2] Consider function $f : [a, b] \to X$. Assume that for every $\varepsilon > 0$ there is a partition $D_\varepsilon:\ a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ such that

$$\left\| \sum_{i=1}^{n} f(\xi_i) \Delta t_i - \sum_{i=1}^{n} f(\eta_i) \Delta t_i \right\| < \varepsilon$$

[1] [Mas], p. 160–161
[2] [Mas], p. 159–160

for every choice of $\xi_i, \eta_i \in [t_{i-1}, t_i]$. Then f is an integrable function.

Proof. If we introduce the notation

$$\sum_{i=1}^{n} f([t_{i-1}, t_i])\Delta t_i = \left\{ \sum_{i=1}^{n} f(\xi_i)\Delta t_i;\ \xi_i \in [t_{i-1}, t_i] \right\},$$

then the assumption of the lemma might be written as

$$\operatorname{diam}\left(\sum_{i=1}^{n} f([t_{i-1}, t_i])\Delta t_i \right) < \varepsilon.$$

To prove that f is integrable it is sufficient to verify that for every partition $D \prec D_\varepsilon$ which consists of division points

$$t_{i-1} = t_0^i < t_1^i < \cdots < t_{m(i)}^i = t_i, \quad i = 1, \ldots, n$$

and for every choice of $\xi_j^i \in [t_{j-1}^i, t_j^i]$, $\eta_i \in [t_{i-1}, t_i]$ we have

$$\|P(f, D) - P(f, D_\varepsilon)\| = \left\| \sum_{i=1}^{n} \sum_{j=1}^{m(i)} f(\xi_j^i)\Delta t_j^i - \sum_{i=1}^{n} f(\eta_i)\Delta t_i \right\| < \varepsilon.$$

But

$$\sum_{i=1}^{n} \sum_{j=1}^{m(i)} f(\xi_j^i)\Delta t_j^i = \sum_{i=1}^{n} \sum_{j=1}^{m(i)} \frac{\Delta t_j^i}{\Delta t_i} f(\xi_j^i)\Delta t_i \in \operatorname{conv}\left(\sum_{i=1}^{n} f([t_{i-1}, t_i])\Delta t_i \right),$$

and the proof is completed by using Theorem 5.5.3. □

Theorem 5.5.7.[1] Every product integrable function $f : [a, b] \to X$ is integrable.

Proof. We verify that the assumption of Theorem 5.5.6 is fulfilled. According to Lemma 5.5.1 it is possible to choose numbers $a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b$ in such a way that

$$\|P(f, D_k) - (1 + S(f, D_k))\| \leq \frac{\varepsilon}{3} \frac{(s_k - s_{k-1})}{(b - a)}$$

for every tagged partition D_k of interval $[s_{k-1}, s_k]$. Since f is product integrable on $[s_{k-1}, s_k]$, there exists (according to Lemma 5.5.4) a partition

$$s_{k-1} = t_0^k < t_1^k < \cdots < t_{m(k)}^k = s_k$$

such that

$$\left\| \prod_{i=m(k)}^{1} (1 + f(\xi_i^k)\Delta t_i^k) - \prod_{i=m(k)}^{1} (1 + f(\eta_i^k)\Delta t_i^k) \right\| < \frac{\varepsilon}{3} \frac{(s_k - s_{k-1})}{(b - a)}$$

[1] [Mas], p. 167–169

for every choice of $\xi_i^k, \eta_i^k \in [t_{i-1}^k, t_i^k]$. For such ξ_i^k, η_i^k we have

$$\left\|\sum_{i=1}^{m(k)} f(\xi_i^k)\Delta t_i^k - \sum_{k=1}^{m(k)} f(\eta_i^k)\Delta t_i^k\right\| \leq \left\|\left(1+\sum_{i=1}^{m(k)} f(\xi_i^k)\Delta t_i^k\right) - \prod_{i=m(k)}^{1}(1+f(\xi_i^k)\Delta t_i^k)\right\| +$$

$$+\left\|\prod_{i=m(k)}^{1}(1+f(\xi_i^k)\Delta t_i^k) - \prod_{i=m(k)}^{1}(1+f(\eta_i^k)\Delta t_i^k)\right\| +$$

$$+\left\|\prod_{i=m(k)}^{1}(1+f(\eta_i^k)\Delta t_i^k) - \left(1+\sum_{k=1}^{m(k)} f(\eta_i^k)\Delta t_i^k\right)\right\| < \frac{\varepsilon(s_k - s_{k-1})}{(b-a)}.$$

Adding these inequalities for $k = 1, \ldots, n$ and using the triangle inequality leads to

$$\left\|\sum_{k=1}^{n}\sum_{i=1}^{m(k)} f(\xi_i^k)\Delta t_i^k - \sum_{k=1}^{n}\sum_{k=1}^{m(k)} f(\eta_i^k)\Delta t_i^k\right\| < \varepsilon.$$

This means that the partition D can be chosen as

$$a = t_0^1 < t_1^1 < \cdots < t_{m(1)}^1 = t_0^2 < \cdots < t_{m(n-1)}^{n-1} = t_0^n < t_1^n < \cdots < t_{m(n)}^n = b.$$

□

We now follow Masani's proof of the reverse implication which says that every integrable function is also product integrable and that the product integral might be expressed using the Peano series. As we know, the history of the theorem can be traced back to Volterra (in the case $X = \mathbf{R}^{n\times n}$). Masani was probably the first one to give a rigorous proof.

We will be working with tagged partitions $D : a = t_0 < t_1 < \cdots < t_{m(D)} = b$, $\xi_i \in [t_{i-1}, t_i]$. For every $n \leq m(D)$ we define

$$T_n(f, D) = \sum_{m(D)\geq i_1 > i_2 > \cdots > i_n \geq 1} f(\xi_{i_1})\cdots f(\xi_{i_n})\Delta t_1 \cdots \Delta t_n$$

and

$$T(f, D) = T_1(f, D) + \cdots + T_{m(D)}(f, D).$$

We state the following lemma without proof; see Remark 2.4.4 for the proof in the finite-dimensional case (the difficulty in the general case is hidden in the fact that the product of two integrable functions need not be integrable).

Lemma 5.5.8.[1] Let $f : [a, b] \to X$ be an integrable function, $n \in \mathbf{N}$. Then the limit $T_n(f) = \lim_{\nu(D)\to 0} T_n(f, D)$ exists and

$$T_n(f) = \int_a^b \int_a^{t_1} \cdots \int_a^{t_{n-1}} f(t_1)\cdots f(t_n)\,\mathrm{d}t_n \cdots \mathrm{d}t_1.$$

[1] [Mas], p. 174–176

Masani refers to the following lemma as to the extension of Tannery's theorem.

Lemma 5.5.9.[1] Consider function $f : [a, b] \to X$ and assume that the following conditions are satisfied:

1) There exists $T_n(f) = \lim_{\nu(D)\to 0} T_n(f, D)$ for every $n \in \mathbf{N}$.
2) $M_n = \sup_D \|T_n(f, D)\| < \infty$ for every $n \in \mathbf{N}$, where the supremum is taken over all partitions D of interval $[a, b]$ which consist of at least n division points.
3) The series $\sum_{n=1}^{\infty} M_n$ is convergent.

Then

$$T(f) = \lim_{\nu(D)\to 0} T(f, D) = \sum_{n=1}^{\infty} T_n(f).$$

Proof. The series $T(f) = \sum_{n=1}^{\infty} T_n(f)$ is convergent, because $\|T_n(f)\| \le M_n$ for every $n \in \mathbf{N}$. We will prove that $T(f) = \lim_{\nu(D)\to 0} T(f, D)$. Choose $\varepsilon > 0$. There exists a number $n(\varepsilon) \in \mathbf{N}$ such that

$$\sum_{k=n(\varepsilon)+1}^{\infty} M_k < \varepsilon/3.$$

According to the first assumption, there exists a $\delta > 0$ such that

$$\|T_k(f, D) - T_k(f)\| < \frac{\varepsilon}{3n(\varepsilon)}, \quad k = 1, \dots, n(\varepsilon),$$

for every tagged partition D of $[a, b]$ that satisfies $\nu(D) < \delta$. Without loss of generality we assume that δ is so small that D consists of at least $n(\varepsilon)$ division points, i.e. $T_1(f, D), \dots, T_{n(\varepsilon)}(f, D)$ are well-defined. Now for every tagged partition D that satisfies $\nu(D) < \delta$ we estimate

$$\|T(f, D) - T(f)\| = \left\| \sum_{k=1}^{m(D)} T_k(f, D) - \sum_{k=1}^{\infty} T_k(f) \right\| \le \sum_{k=1}^{n(\varepsilon)} \|T_k(f, D) - T_k(f)\| +$$

$$+ \sum_{k=n(\varepsilon)+1}^{m(D)} \|T_k(f, D)\| + \sum_{k=n(\varepsilon)+1}^{\infty} \|T_k(f)\| < n(\varepsilon)\frac{\varepsilon}{3n(\varepsilon)} + 2\sum_{k=n(\varepsilon)+1}^{\infty} M_k < \varepsilon.$$

□

Theorem 5.5.10.[2] Let $f : [a, b] \to X$ be an integrable function. Then f is also product integrable and

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = 1 + \sum_{n=1}^{\infty} T_n(f).$$

[1] [Mas], p. 189–191
[2] [Mas], p. 176–177

Proof. Denote $M = \sup_{t\in[a,b]} \|f(t)\| < \infty$. For every partition D of interval $[a,b]$ we have

$$P(f,D) = 1 + T(f,D),$$

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = 1 + \lim_{\nu(D)\to 0} T(f,D),$$

$$\|T_n(f,D)\| \le \frac{(b-a)^n M^n}{n!},$$

$$\sum_{n=1}^{\infty} \frac{(b-a)^n M^n}{n!} = \exp(M(b-a)) - 1 < \infty.$$

The statement of the theorem is therefore a consequence of the preceding two lemmas. □

We have proved that a function is product integrable if and only if it is integrable. Thus, in the rest of this chapter we use the terms “integrable” and “product integrable” as synonyms.

Theorem 5.5.11.[1] Let $f : [a,b] \to X$ be an integrable function. Suppose that $f(x) \cdot f(y) = f(y) \cdot f(x)$ for each pair $x, y \in X$. Then

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = \exp\left(\int_a^b f(t)\,\mathrm{d}t\right).$$

Proof. A simple consequence of Theorem 5.5.10 and the equality

$$\int_a^b \int_a^{t_1} \cdots \int_a^{t_{n-1}} f(t_1)\cdots f(t_n)\,\mathrm{d}t_n \cdots \mathrm{d}t_1 = \frac{1}{n!}\left(\int_a^b f(t)\,\mathrm{d}t\right)^n$$

(see Lemma 2.4.2). □

5.6 Additional properties of product integral

This section is devoted to Masani’s versions of the fundamental theorem of calculus, the uniform convergence theorem, and the change of variables theorem.

Theorem 5.6.1.[2] Let $f : [a,b] \to X$ be an integrable function. Denote

$$Y(x) = \prod_a^x (1 + f(t)\,\mathrm{d}t), \quad x \in [a,b].$$

[1] [Mas], p. 179

[2] [Mas], p. 178

Then

$$Y(x) = 1 + \int_a^x f(t)Y(t)\,\mathrm{d}t, \quad x \in [a,b]. \tag{5.6.1}$$

Proof. Using Theorem 5.5.10 we obtain

$$Y(t) = 1 + \int_a^t f(t_1)\,\mathrm{d}t_1 + \int_a^t \int_a^{t_1} f(t_1)f(t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1 + \cdots. \tag{5.6.2}$$

Since

$$\left\| \int_a^t \int_a^{t_1} \cdots \int_a^{t_{n-1}} f(t_1)\cdots f(t_n)\,\mathrm{d}t_n \cdots \mathrm{d}t_1 \right\| \le \frac{(b-a)^n M^n}{n!},$$

the series (5.6.2) is uniformly convergent. Because f is bounded, the series

$$f(t)Y(t) = f(t) + \int_a^t f(t)f(t_1)\,\mathrm{d}t_1 + \int_a^x \int_a^{t_1} f(t)f(t_1)f(t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1 + \cdots$$

is also uniformly convergent and might be integrated term by term on $[a,x]$; performing this step leads to Equation (5.6.1). $\square$

Corollary 5.6.2.[1] If $f : [a,b] \to X$ is a continuous function, then

$$Y'(x)Y(x)^{-1} = f(x)$$

for every $x \in [a,b]$.

The previous corollary represents an analogy of the formula

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_a^x f(t)\,\mathrm{d}t = f(x)$$

(see Theorem 5.1.7). Also the Newton-Leibniz formula

$$\int_a^b f'(x)\,\mathrm{d}t = f(b) - f(a)$$

(see Theorem 5.1.8) has the following product analogy (whose proof we omit).

Theorem 5.6.3.[2] Assume that $Z : [a,b] \to X$ satisfies $Z'(x)Z(x)^{-1} = f(x)$ for every $x \in [a,b]$. Then

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = Z(b)Z(a)^{-1}$$

provided the function f is integrable.

[1] [Mas], p. 181
[2] [Mas], p. 182

The next theorem establishes a criterion for interchanging the order of limit and product integral, i.e. for the formula

$$\lim_{n\to\infty} \prod_a^b (1 + f_n(t)\,\mathrm{d}t) = \prod_a^b (1 + \lim_{n\to\infty} f_n(t)\,\mathrm{d}t).$$

We have already encountered such a criterion in Chapter 3 when discussing the Lebesgue product integral; Schlesinger's statement represented in fact a product analogy of the Lebesgue dominated convergence theorem. Masani's theorem concerns the Riemann product integral and requires uniform convergence to perform the interchange of limit and integral.

Theorem 5.6.4.[1] Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of integrable functions which converge uniformly to function f on interval $[a, b]$. Then

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = \lim_{n\to\infty} \prod_a^b (1 + f_n(t)\,\mathrm{d}t).$$

Proof. The existence of $\prod_a^b (1+f(t)\,\mathrm{d}t)$ follows from the fact that the limit of a uniformly convergent sequence of integrable functions is again an integrable function. For an arbitrary tagged partition D we can use Lemma 5.3.2 to estimate

$$\|P(f, D) - P(f_n, D)\| \le \exp(M(b-a)) \cdot \left(\exp\left(\sum_i \|f(\xi_i) - f_n(\xi_i)\| \Delta t_i \right) - 1 \right),$$

where $M = \sup_{t\in[a,b]} \|f(t)\|$. Choose $\varepsilon > 0$ and find a corresponding $\varepsilon_0 > 0$ such that

$$\exp(M(b-a)) \cdot (\exp(\varepsilon_0(b-a)) - 1) < \varepsilon/3.$$

Let $n_0 \in \mathbf{N}$ be such that $\|f(t) - f_n(t)\| < \varepsilon_0$ for every $t \in [a, b]$ and $n \ge n_0$. The partition D can be chosen so that the inequalities

$$\left\| P(f, D) - \prod_a^b (1 + f(t)\,\mathrm{d}t) \right\| < \varepsilon/3, \quad \left\| P(f_n, D) - \prod_a^b (1 + f_n(t)\,\mathrm{d}t) \right\| < \varepsilon/3$$

hold. Then for every $n \ge n_0$ we have

$$\left\| \prod_a^b (1 + f_n(t)\,\mathrm{d}t) - \prod_a^b (1 + f(t)\,\mathrm{d}t) \right\| \le \left\| \prod_a^b (1 + f(t)\,\mathrm{d}t) - P(f, D) \right\| +$$

$$+ \left\| P(f, D) - P(f_n, D) \right\| + \left\| P(f_n, D) - \prod_a^b (1 + f_n(t)\,\mathrm{d}t) \right\| < \varepsilon.$$

□

[1] [Mas], p. 171

Masani also proved a generalized version of the change of variables theorem for the product integral (compare to Theorem 2.5.10); we state it without proof.

Theorem 5.6.5.[1] Let $f : [a,b] \to X$ be an integrable function, $\varphi : [\alpha,\beta] \to [a,b]$ increasing, $\varphi(\alpha) = a$, $\varphi(\beta) = b$. If φ' exists and is integrable on $[a,b]$, then

$$\prod_a^b (1 + f(t)\,\mathrm{d}t) = \prod_\alpha^\beta (1 + f(\varphi(u))\varphi'(u)\,\mathrm{d}u).$$

[1] [Mas], p. 187–188

Chapter 6

Kurzweil and McShane product integrals

The introduction of Lebesgue integration signified a revolution in mathematical analysis: Every Riemann integrable function is also Lebesgue integrable, but the class of functions having Lebesgue integral is considerably larger.

However, there exist functions f which are Newton integrable and

$$(N)\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a),$$

where F is an antiderivative of f, but the Lebesgue integral $(L)\int_a^b f(t)\,\mathrm{d}t$ does not exist. Consider for example the function

$$F(x) = \begin{cases} x^2\sin(1/x^2) & \text{if } x \in (0,1], \\ 0 & \text{if } x = 0. \end{cases}$$

This function has a derivative $F'(x) = f(x)$ for every $x \in [0,1]$ and

$$f(x) = \begin{cases} 2x\sin(1/x^2) - (2/x)\cos(1/x^2) & \text{if } x \in (0,1], \\ 0 & \text{if } x = 0. \end{cases}$$

The function f is therefore Newton integrable and

$$(N)\int_0^1 f(t)\,\mathrm{d}t = F(1) - F(0).$$

If we denote

$$a_k = \frac{1}{\sqrt{(k+1/2)\pi}}, \quad b_k = \frac{1}{\sqrt{k\pi}}$$

for every $k \in \mathbf{N}$, then

$$\int_{a_k}^{b_k} |f(t)|\,\mathrm{d}t \geq \left|\int_{a_k}^{b_k} f(t)\,\mathrm{d}t\right| = |F(b_k) - F(a_k)| = \frac{1}{(k+1/2)\pi},$$

which implies that

$$\int_0^1 |f(t)|\,\mathrm{d}t \geq \sum_{k=1}^{\infty} \frac{1}{(k+1/2)\pi} = \infty.$$

The Lebesgue integral $(L)\int_0^1 f(t)\,\mathrm{d}t$ therefore does not exist.

Jaroslav Kurzweil (and later independently Ralph Henstock) introduced a new definition of integral which avoids the above mentioned drawback of the Lebesgue integral. The Kurzweil integral (also known as the gauge integral or the Henstock-Kurzweil integral) encompasses the Newton and Lebesgue (and consequently also Riemann) integrals. Another benefit is that the definition of Kurzweil integral is obtained by a gentle modification of Riemann's definition and is considerably simpler than Lebesgue's definition.

Jaroslav Kurzweil[1]

Edward J. McShane[2]

The definition of integral due to E. J. McShane is similar to Kurzweil's definition and in fact represents an equivalent definition of Lebesgue integral.

In this chapter we first summarize the definitions of Kurzweil and McShane integrals; in the second part we turn our attention to product analogies of these integrals. The proofs in this chapter are often omitted and may be found in the original papers (the references are included).

6.1 Kurzweil and McShane integrals

A finite collection of point-interval pairs $D = \{([t_{i-1}, t_i], \xi_i)\}_{i=1}^m$ is called an M-partition of interval $[a, b]$ if

$$a = t_0 < t_1 < \cdots < t_m = b,$$

$$\xi_i \in [a, b], \quad i = 1, \ldots, m.$$

A K-partition is a M-partition which moreover satisfies

$$\xi_i \in [t_{i-1}, t_i], \quad i = 1, \ldots, m.$$

Given a function $\Delta : [a, b] \to (0, \infty)$ (the so-called gauge on $[a, b]$), a partition D is called Δ-fine if

$$[t_{i-1}, t_i] \subset (\xi_i - \Delta(\xi_i), \xi_i + \Delta(\xi_i)), \quad i = 1, \ldots, m.$$

[1] Photo taken by Š. Schwabik
[2] Photo from [McT]

For a given function f on $[a,b]$ and a M-partition D of $[a,b]$ denote

$$S(f,D) = \sum_{i=1}^{m} f(\xi_i)(t_i - t_{i-1}).$$

Definition 6.1.1. Let X be a Banach space. A vector $S_f \in X$ is called the Kurzweil (McShane) integral of function $f : [a,b] \to X$ if for every $\varepsilon > 0$ there is a gauge $\Delta : [a,b] \to (0,\infty)$ such that

$$\|S(f,D) - S_f\| < \varepsilon$$

for every Δ-fine K-partition (M-partition) D of interval $[a,b]$. We define

$$(K)\int_a^b f(t)\,\mathrm{d}t = S_f \quad \text{or} \quad (M)\int_a^b f(t)\,\mathrm{d}t = S_f, \quad \text{respectively.}$$

We state the following theorems without proofs; they can be found (together with more information about the Kurzweil, McShane and Bochner integrals) in the book [SY]; other good sources are [Sch2, RG].

Theorem 6.1.2. Let X be a Banach space. Then every McShane integrable function $f : [a,b] \to X$ is also Kurzweil integrable (but not vice versa) and

$$(K)\int_a^b f(t)\,\mathrm{d}t = (M)\int_a^b f(t)\,\mathrm{d}t.$$

Theorem 6.1.3. Let X be a Banach space. Then every Lebesgue (Bochner) integrable function $f : [a,b] \to X$ is also McShane integrable and

$$(M)\int_a^b f(t)\,\mathrm{d}t = (L)\int_a^b f(t)\,\mathrm{d}t.$$

The converse statement holds if and only if X is a finite-dimensional space.

6.2 Product integrals and their properties

We now proceed to the definitions of Kurzweil and McShane product integrals. The definition of Kurzweil product integral appeared for the first time in the paper [JK]; the authors speak about the Perron product integral and use the notation

$$(PP)\int_a^b (I + A(t)\,\mathrm{d}t).$$

The McShane product integral was studied in [Sch1, SS].

For the sake of simplicity we confine our exposition only to matrix functions $A : [a,b] \to \mathbf{R}^{n\times n}$ instead of working with operator-valued functions $A : [a,b] \to \mathcal{L}(X)$ or even with functions $A : [a,b] \to X$ with values in a Banach algebra X. For an arbitrary M-partition D of $[a,b]$ and a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$ denote

$$P(A,D) = \prod_{i=m}^{1} (I + A(\xi_i)(t_i - t_{i-1})).$$

Definition 6.2.1. Consider function $A : [a,b] \to \mathbf{R}^{n\times n}$. A matrix $P_A \in \mathbf{R}^{n\times n}$ is called the Kurzweil (McShane) product integral of A if for every $\varepsilon > 0$ there is a gauge $\Delta : [a,b] \to (0,\infty)$ such that

$$\|P(A,D) - P_A\| < \varepsilon$$

for every Δ-fine K-partition (M-partition) D of interval $[a,b]$. We define

$$(K)\prod_a^b (I + A(t)\,\mathrm{d}t) = P_A, \quad \text{or} \quad (M)\prod_a^b (I + A(t)\,\mathrm{d}t) = P_A, \quad \text{respectively.}$$

We also denote

$$KP([a,b],\mathbf{R}^{n\times n}) = \left\{ A : [a,b] \to \mathbf{R}^{n\times n};\ (K)\prod_a^b (I + A(t)\,\mathrm{d}t) \text{ exists} \right\},$$

$$MP([a,b],\mathbf{R}^{n\times n}) = \left\{ A : [a,b] \to \mathbf{R}^{n\times n};\ (M)\prod_a^b (I + A(t)\,\mathrm{d}t) \text{ exists} \right\}.$$

The right product integrals can be introduced using the products

$$P^*(A,D) = \prod_{i=1}^{m} (I + A(\xi_i)(t_i - t_{i-1})),$$

but we limit our discussion to the left integrals.

Example 6.2.2. Assume that the Riemann product integral $(R)\prod_a^b (I + A(t)\,\mathrm{d}t)$ exists, i.e. for every $\varepsilon > 0$ we can find $\delta > 0$ such that

$$\left\| P(A,D) - \prod_a^b (I + A(t)\,\mathrm{d}t) \right\| < \varepsilon$$

for every partition D of $[a,b]$ which such that $\nu(D) < \delta$. If we put

$$\Delta(x) = \frac{\delta}{2}, \quad x \in [a,b],$$

then every Δ-fine K-partition D of $[a,b]$ satisfies $\nu(D) < \delta$. This means that the Kurzweil product integral of A exists and

$$(K)\prod_a^b (I + A(t)\,\mathrm{d}t) = (R)\prod_a^b (I + A(t)\,\mathrm{d}t).$$

Example 6.2.3. Consider the function

$$f(x) = \begin{cases} -1/x & \text{if } x \in (0,1], \\ 0 & \text{if } x = 0. \end{cases}$$

It can be proved (see [SS]) that

$$(M)\prod_0^1 (1 + f(x)\,\mathrm{d}x) = 0.$$

It is worth noting that neither the Riemann integral $(R)\prod_0^1(1 + f(x)\,\mathrm{d}x)$ nor the Lebesgue integral $(L)\prod_0^1(1 + f(x)\,\mathrm{d}x)$ exist; this follows e.g. from Theorem 6.2.10.

Theorem 6.2.4.[1] Consider function $A : [a,b] \to \mathbf{R}^{n\times n}$. Then the following conditions are equivalent:

1) The integral $(K)\prod_a^b(I + A(t)\,\mathrm{d}t)$ exists and is invertible.
2) There exists an invertible matrix P_A such that for every $\varepsilon > 0$ there is a gauge $\Delta : [a,b] \to (0,\infty)$ such that

$$\left\| \prod_{i=m}^{1} e^{A(\xi_i)(t_i - t_{i-1})} - P_A \right\| < \varepsilon$$

whenever $D = \{([t_{i-1}, t_i], \xi_i)\}_{i=1}^m$ is a Δ-fine K-partition of $[a,b]$.
If one of these conditions is fulfilled, then

$$(K)\prod_a^b (I + A(t)\,\mathrm{d}t) = P_A.$$

A similar statement holds also for McShane product integral.

Theorem 6.2.5. Consider function $A : [a,b] \to \mathbf{R}$. The integral $(K)\int_a^b A(t)\,\mathrm{d}t$ exists if and only if the integral $(K)\prod_a^b(1 + A(t)\,\mathrm{d}t)$ exists and is different from zero. In this case the following equality holds:

$$(K)\prod_a^b (1 + A(t)\,\mathrm{d}t) = \exp\left((K)\int_a^b A(t)\,\mathrm{d}t \right).$$

[1] [JK], p. 651, and [SS]

A similar statement holds also for McShane product integral.

Proof. Assume that $(K)\int_a^b A(t)\,\mathrm{d}t = S_A$ exists and choose $\varepsilon > 0$. Since the exponential function is continuous at the point S_A, there is a $\delta > 0$ such that

$$\left|e^x - e^{S_A}\right| < \varepsilon, \quad x \in (S_A - \delta, S_A + \delta). \tag{6.2.1}$$

Let $\Delta : [a,b] \to (0,\infty)$ be a gauge such that

$$|S(A,D) - S_A| < \delta$$

for every Δ-fine K-partition of interval $[a,b]$. Each of these partitions satisfies

$$\left|\prod_{i=m}^{1} e^{A(\xi_i)(t_i - t_{i-1})} - e^{S_A}\right| = \left|e^{S(A,D)} - e^{S_A}\right| < \varepsilon$$

and using Theorem 6.2.4 we obtain

$$(K)\prod_a^b (1 + A(t)\,\mathrm{d}t) = \exp\left((K)\int_a^b A(t)\,\mathrm{d}t\right).$$

The reverse implication is proved in a similar way using the equality

$$S(A,D) = \log\left(\prod_{i=m}^{1} e^{A(\xi_i)(t_i - t_{i-1})}\right).$$

□

Remark 6.2.6. The previous theorem no longer holds for matrix functions $A : [a,b] \to \mathbf{R}^{n\times n}$. Jaroslav Kurzweil and Jiří Jarník constructed[1] two functions $A, B : [-1,1] \to \mathbf{R}^{2\times 2}$ such that

$$(K)\int_{-1}^{1} A(t)\,\mathrm{d}t \text{ exists}, \quad (K)\prod_{-1}^{1}(I + A(t)\,\mathrm{d}t) \text{ doesn't exist},$$

$$(K)\int_{-1}^{1} B(t)\,\mathrm{d}t \text{ doesn't exist}, \quad (K)\prod_{-1}^{1}(I + B(t)\,\mathrm{d}t) \text{ exists}.$$

Theorem 6.2.7. Every McShane product integrable function is also Kurzweil product integrable (but not vice versa) and

$$(K)\prod_a^b (I + A(t)\,\mathrm{d}t) = (M)\prod_a^b (I + A(t)\,\mathrm{d}t).$$

[1] [JK], p. 658

Proof. The inclusion $MP \subseteq KP$ follows from the fact that every K-partition is also a M-partition; the equality of the two product integrals is then obvious. We only have to prove that $MP \neq KP$. For an arbitrary function $f : [a, b] \to \mathbf{R}$ denote

$$A_f(t) = I \cdot f(t),$$

where I is the identity matrix of order n (A_f is therefore a matrix-valued function on $[a, b]$). Then evidently

$$P(A, D) = I \cdot P(f, D)$$

for every partition D of $[a, b]$ and A_f is product integrable (in the Kurzweil or McShane sense) if and only if f is product integrable. Theorem 6.1.2 guarantees the existence of a function $f : [a, b] \to \mathbf{R}$ that is Kurzweil integrable, but not McShane integrable; then (according to Theorem 6.2.5) the corresponding function $A_f : [a, b] \to \mathbf{R}^{n\times n}$ is Kurzweil product integrable, but not McShane product integrable. □

Theorem 6.2.8.[1] Consider function $A : [a, b] \to \mathbf{R}^{n\times n}$. Suppose that the integral $(M)\prod_a^b(I + A(t)\,\mathrm{d}t)$ exists and is invertible. Then for every $x \in (a, b)$ the integral

$$Y(x) = (M)\prod_a^x(I + A(t)\,\mathrm{d}t)$$

exists as well and the function Y satisfies

$$Y'(x) = A(x)Y(x)$$

almost everywhere on $[a, b]$.

Remark 6.2.9. In Chapter 3 we have defined the Lebesgue (or Bochner) product integral $(L)\prod_a^b(I+A(t)\,\mathrm{d}t)$; the definition was based on the approximation of A by a sequence of step functions which converge to A in the norm of space $L([a, b], \mathbf{R}^{n\times n})$. The following theorem describes the relationship between McShane and Lebesgue product integrals.

Theorem 6.2.10.[2] Consider function $A : [a, b] \to \mathbf{R}^{n\times n}$. The following conditions are equivalent:

1) A is Lebesgue (Bochner) integrable.
2) The McShane product integral $(M)\prod_a^b(I + A(t)\,\mathrm{d}t)$ exists and is invertible.

If one of these conditions is fulfilled, then

$$(M)\prod_a^b(I + A(t)\,\mathrm{d}t) = (L)\prod_a^b(I + A(t)\,\mathrm{d}t).$$

[1] [JK], p. 652–656 and [SS]
[2] [Sch1], p. 329–334 and [SS]

Remark 6.2.11. We conclude this chapter by comparing the classes of functions which are integrable according to different definitions presented in the previous text.

Let R, L, M and K be the classes of all functions $A : [a,b] \to \mathbf{R}^{n\times n}$ which are integrable in the sense of Riemann, Lebesgue, McShane and Kurzweil, respectively. In a similar way let RP and LP denote the classes of Riemann product integrable and Lebesgue product integrable functions. Instead of working with the classes KP and MP it is more convenient to concentrate on the classes

$$KP^* = \left\{ A : [a,b] \to \mathbf{R}^{n\times n};\ (K)\prod_a^b (I + A(t)\,\mathrm{d}t) \text{ exists and is invertible} \right\},$$

$$MP^* = \left\{ A : [a,b] \to \mathbf{R}^{n\times n};\ (M)\prod_a^b (I + A(t)\,\mathrm{d}t) \text{ exists and is invertible} \right\}.$$

The following diagram shows the inclusions between the above mentioned classes.

$$\begin{array}{ccccccc} R & \subset & L & = & M & \subset & K \\ = & & = & & = & & \neq \\ RP & \subset & LP & = & MP^* & \subset & KP^* \end{array}$$

Chapter 7

Complements

This final chapter contains additional remarks on product integration theory. The topics discussed here complement the previous chapters; however, most proofs are omitted and the text is intended only to arouse reader's interest (references to other works are included).

7.1 Variation of constants

Product integral enables us to express solution of the differential equation

$$y'(x) = A(x)y(x), \quad x \in [a,b],$$

where $A : [a,b] \to \mathbf{R}^{n\times n}$, $y : [a,b] \to \mathbf{R}^n$. The fundamental matrix of this system is

$$Z(x) = \prod_a^x (I + A(t)\,\mathrm{d}t) = \begin{pmatrix} z_1^1(x) & \cdots & z_n^1(x) \\ \vdots & \ddots & \vdots \\ z_1^n(x) & \cdots & z_n^n(x) \end{pmatrix}$$

and its columns

$$z_i(x) = \begin{pmatrix} z_i^1(x) \\ \vdots \\ z_i^n(x) \end{pmatrix}, \quad i = 1, \dots, n \tag{7.1.1}$$

thus provide a fundamental system of solutions.

We now focus our attention to the inhomogeneous equation

$$\begin{aligned} y'(x) &= A(x)y(x) + f(x), \quad x \in [a,b], \\ y(a) &= y_0. \end{aligned} \tag{7.1.2}$$

A method for solving this system using product integral (based on the well-known method of variation of constants) was first proposed by G. Rasch in the paper [GR]; it can be also found in the monograph [DF].

We assume that the functions $A : [a,b] \to \mathbf{R}^{n\times n}$ and $f : [a,b] \to \mathbf{R}^n$ are continuous, and we try to find the solution of (7.1.2) in the form

$$y(x) = \sum_{i=1}^n z_i(x)c_i(x), \tag{7.1.3}$$

where $c_i : [a,b] \to \mathbf{R}$, $i = 1, \dots, n$ are certain unknown functions. If we denote

$$c(x) = \begin{pmatrix} c_1(x) \\ \vdots \\ c_n(x) \end{pmatrix},$$

then the equations (7.1.1) and (7.1.3) imply

$$y(x) = Z(x)c(x).$$

We obtain

$$y'(x) = Z'(x)c(x) + Z(x)c'(x) = A(x)Z(x)c(x) + Z(x)c'(x) = A(x)y(x) + Z(x)c'(x),$$

and using Equation (7.1.2)

$$f(x) = Z(x)c'(x).$$

Consequently

$$c'(x) = Z(x)^{-1}f(x),$$
$$c(a) = Z(a)^{-1}y(a) = y_0,$$

which implies

$$c(x) = y_0 + \int_a^x Z(t)^{-1}f(t)\,\mathrm{d}t.$$

The solution of the system (7.1.2) is thus given by the explicit formula

$$y(x) = Z(x)c(x) = Z(x)y_0 + Z(x)\int_a^x Z(t)^{-1}f(t)\,\mathrm{d}t =$$

$$= \prod_a^x (I + A(t)\,\mathrm{d}t) y_0 + \prod_a^x (I + A(t)\,\mathrm{d}t) \int_a^x \left(\prod_t^a (I + A(s)\,\mathrm{d}s) f(t) \right) \mathrm{d}t =$$

$$= \prod_a^x (I + A(t)\,\mathrm{d}t) y_0 + \int_a^x \left(\prod_t^x (I + A(s)\,\mathrm{d}s) f(t) \right) \mathrm{d}t.$$

We summarize the result: The solution of the inhomogeneous system (7.1.2) has the form

$$y(x) = \sum_{i=1}^n z_i(x)c_i(x),$$

where $z_1, \dots, z_n : [a,b] \to \mathbf{R}^n$ is the fundamental system of solutions of the corresponding homogeneous equation, the functions $c_i : [a,b] \to \mathbf{R}$, $i = 1, \dots, n$ are continuously differentiable and satisfy

$$\sum_{i=1}^n c_i'(x)z_i(x) = f(x), \quad x \in [a,b].$$

7.2 Equivalent definitions of product integral

Consider a tagged partition $D : a = t_0 < t_1 < \cdots < t_m = b$, $\xi_i \in [t_{i-1}, t_i]$, $i = 1, \dots, m$. Ludwig Schlesinger proved (see Theorem 3.2.2) that the product integral of a Riemann integrable function $A : [a, b] \to \mathbf{R}^{n\times n}$ can be calculated not only as

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \lim_{\nu(D)\to 0}\left(\prod_{k=m}^{1} (I + A(\xi_k)\Delta t_k)\right),$$

but also as

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \lim_{\nu(D)\to 0}\left(\prod_{k=m}^{1} e^{A(\xi_k)\Delta t_k}\right).$$

The equivalence of these definitions can be intuitively explained using the fact that

$$e^{A(\xi_k)\Delta t_k} = 1 + A(\xi_k)\Delta t_k + O((\Delta t_k)^2),$$

and the terms of order $(\Delta t_k)^2$ and higher do not change the value of the integral. We have also encountered a similar theorem applicable to the Kurzweil and McShane integrals (see Theorem 6.2.4).
We now proceed to a more general theorem concerning equivalent definitions of product integral, which was given in [DF].

Definition 7.2.1. A function

$$f(z) = \sum_{k=0}^{\infty} c_k z^k \tag{7.2.1}$$

is called admissible, if the series (7.2.1) has a positive radius of convergence $r > 0$ and

$$f(0) = c_0 = 1, \quad f'(0) = c_1 = 1.$$

For example, the functions $z \mapsto \exp z$, $z \mapsto 1 + z$ and $z \mapsto (1 - z)^{-1}$ are admissible.
For every matrix $A \in \mathbf{R}^{n\times n}$ such that $\|A\| < r$ we put

$$f(A) = \sum_{k=0}^{\infty} c_k A^k.$$

Theorem 7.2.2.[1] If f is an admissible function and $A : [a, b] \to \mathbf{R}^{n\times n}$ a continuous matrix function, then

$$\prod_a^b (I + A(t)\,\mathrm{d}t) = \lim_{\nu(D)\to 0}\left(\prod_{k=m}^{1} f(A(\xi_k)\Delta t_k)\right).$$

According to the previous theorem, the product integral of a function A can be defined as the limit

$$\lim_{\nu(D)\to 0}\left(\prod_{k=m}^{1} f(A(\xi_k)\Delta t_k)\right),$$

[1] [DF], p. 50–53

where f is an admissible function. Product integral defined in this way is usually denoted by the symbol $\prod_a^b f(A(t)\,\mathrm{d}t)$, e.g.

$$\prod_a^b (I + A(t)\,\mathrm{d}t), \quad \prod_a^b e^{A(t)\,\mathrm{d}t}, \quad \prod_a^b (I - A(t)\,\mathrm{d}t)^{-1}$$

etc. The integral $\prod_a^b e^{A(t)\,\mathrm{d}t}$ is taken as a primary definition in the monograph [DF]. We note that it is possible to prove an analogy of Theorem 7.2.2 even for the Kurzweil and McShane product integrals (see [JK, Sch1]).

7.3 Riemann-Stieltjes product integral

Consider two functions $f, g : [a,b] \to \mathbf{R}$. Then the ordinary Riemann-Stieltjes integral is defined as the limit

$$\int_a^b f(x)\,\mathrm{d}g(x) = \lim_{\nu(D)\to 0} \sum_{i=1}^m f(\xi_i)(g(t_i) - g(t_{i-1})), \tag{7.3.1}$$

where $D : a = t_0 < t_1 < \cdots < t_m = b$ is a tagged partition of $[a,b]$ with tags $\xi_i \in [t_{i-1}, t_i]$, $i = 1, \ldots, m$ (provided the limit exists). This integral was introduced in 1894 by Thomas Jan Stieltjes (see [Kl], Chapters 44 and 47), who was working with continuous functions f and non-decreasing functions g. Later in 1909 Friedrich Riesz discovered that the Stieltjes integral can be used to represent continuous linear functionals on the space $\mathcal{C}([a,b])$. Also, if $g(x) = x$, we obtain the ordinary Riemann integral.

Assume that the function g is of bounded variation, i.e. that

$$\sup \left\{ \sum_{i=1}^m |g(t_i) - g(t_{i-1})| \right\} < \infty,$$

where the supremum is taken over all partitions $a = t_0 < t_1 < \cdots < t_m = b$ of interval $[a,b]$. Then (see e.g. [RG]) the Riemann-Stieltjes integral exists for every continuous function f.

In particular, if f is continuous and g is a step function defined as

$$g = g_1\chi_{[t_0,t_1)} + g_2\chi_{[t_1,t_2)} + \cdots + g_{m-1}\chi_{[t_{m-2},t_{m-1})} + g_m\chi_{[t_{m-1},t_m]},$$

where $a = t_0 < t_1 < \cdots < t_m = b$, $g_1, \ldots, g_m \in \mathbf{R}$ and χ_M denotes the characteristic function of a set M, then

$$\int_a^b f(x)\,\mathrm{d}g(x) = f(t_1)(g_2 - g_1) + \cdots + f(t_{m-1})(g_m - g_{m-1}).$$

Now consider a matrix function $A : [a,b] \to \mathbf{R}^{n\times n}$. The product analogy of Riemann-Stieltjes integral can be defined as

$$\prod_a^b (I + \mathrm{d}A(t)) = \lim_{\nu(D)\to 0} \prod_{i=m}^{1} (I + A(t_i) - A(t_{i-1})) \tag{7.3.2}$$

(see e.g. [Sch3, GJ, Gil, DN]), or even more generally as

$$\prod_a^b (I + f(t)\mathrm{d}A(t)) = \lim_{\nu(D)\to 0} \prod_{i=m}^{1} (I + f(\xi_i)(A(t_i) - A(t_{i-1}))),$$

where $f : [a,b] \to \mathbf{R}$ (see the entry "Product integral" in [EM]). We now present some basic statements concerning the Riemann-Stieltjes product integral (7.3.2).

Product integrals of the type (7.3.2) are encountered in survival analysis (when working with the cumulative hazard $A(t) = \int_0^t a(s)\,\mathrm{d}s$ instead of the hazard rate $a(t)$; see Example 1.4.1) and in the theory of Markov processes (when working with cumulative intensities $A_{ij}(t) = \int_0^t a_{ij}(s)\,\mathrm{d}s$ for $i \neq j$ and $A_{ii}(t) = -\sum_{j\neq i} A_{ij}(t)$ instead of the transition rates $a_{ij}(t)$; see Example 1.4.2).

A sufficient condition for the existence of the limit (7.3.2) is again that the variation of A is finite. A different sufficient condition (see [DN]) says that the product integral exists provided A is continuous and its p-variation is finite for some $p \in (0,2)$, i.e.

$$\sup\left\{\sum_{i=1}^{m} \|A(t_i) - A(t_{i-1})\|^p\right\} < \infty,$$

where the supremum is again taken over all partitions $a = t_0 < t_1 < \cdots < t_m = b$ of interval $[a,b]$.

If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a step function defined as

$$A = A_1\chi_{[t_0,t_1)} + A_2\chi_{[t_1,t_2)} + \cdots + A_{m-1}\chi_{[t_{m-2},t_{m-1})} + A_m\chi_{[t_{m-1},t_m]},$$

where $a = t_0 < t_1 < \cdots < t_m = b$ and $A_1, \ldots, A_m \in \mathbf{R}^{n\times n}$, then

$$\prod_a^b (I + \mathrm{d}A(t)) = (I + A_m - A_{m-1}) \cdots (I + A_2 - A_1). \tag{7.3.3}$$

Thus, if $A_{k-1} - A_k = I$ for some $k = 2, \ldots, m$, then

$$\prod_a^b (I + \mathrm{d}A(t)) = 0,$$

i.e. the product integral need not be a regular matrix. Equation (7.3.3) also suggests that the indefinite product integral

$$Y(x) = \prod_a^x (I + \mathrm{d}A(t)), \quad x \in [a,b],$$

need not be a continuous function.

If $A : [a,b] \to \mathbf{R}^{n\times n}$ is a continuously differentiable function, it can be proved that

$$\prod_a^b (I + \mathrm{d}A(t)) = \prod_a^b (I + A'(t)\,\mathrm{d}t).$$

As we have seen in the previous chapters, the Riemann product integral provides a solution of the differential equation

$$\begin{aligned} y'(x) &= A(x)y(x), \\ y(a) &= y_0, \end{aligned}$$

or the equivalent integral equation

$$y(x) - y_0 = \int_a^x A(t)y(t)\,\mathrm{d}t.$$

Similarly, the Riemann-Stieltjes product integral can be used as a tool for solving the generalized differential equation

$$\begin{aligned} \mathrm{d}y(x) &= \mathrm{d}A(x)y(x), \\ y(a) &= y_0, \end{aligned}$$

or the equivalent integral equation

$$y(x) - y_0 = \int_a^x \mathrm{d}A(t)y(t).$$

The details are given in the paper [Sch3].

There exists a definition of product integral (see [JK, Sch1, Sch3]) that generalizes both the Riemann and Riemann-Stieltjes product integrals: Consider a mapping V that assigns a square matrix of order n to every point-interval pair $(\xi, [x,y])$, where $[x,y] \subseteq [a,b]$ and $\xi \in [x,y]$. We define

$$\prod_a^b V(t, \mathrm{d}t) = \lim_{\nu(D)\to 0} \prod_{i=m}^1 V(\xi_i, [t_{i-1}, t_i]),$$

provided the limit exists. The choice

$$V(\xi, [x,y]) = I + A(\xi)(y - x)$$

leads to the Riemann product integral, whereas

$$V(\xi, [x,y]) = I + A(y) - A(x)$$

gives the Riemann-Stieltjes product integral.

We note that it is also possible to define the Kurzweil-Stieltjes and McShane-Stieltjes product integrals (see [Sch3]), whose definitions are based on the notion of Δ-fine M-partitions and K-partitions (see Chapter 6); these integrals generalize the notion of Riemann-Stieltjes product integral.

Bibliography

[All] Allen E. S.: *The scientific work of Vito Volterra.* American Mathematical Monthly 48, 516–519, 1941.

[Ber] Beránek J.: *Bohuslav Hostinský (1884–1951).* Časopis pro pěstování matematiky 109, 442–448, 1984 (in Czech).

[CE] Edwards, C. H.: *The Historical Development of the Calculus.* Springer-Verlag, 1979.

[DF] Dollard J. D., Friedman C. N.: *Product integration with applications to differential equations.* Addison–Wesley Publishing Company, 1979.

[DN] Dudley R. M., Norvaiša R.: *Differentiability of Six Operators on Nonsmooth Functions and p-Variation.* Springer, 1999.

[EH] Hille E.: *Ordinary Differential Equations in the Complex Domain.* Dover Publications, 1997.

[EM] *Encyclopaedia of Mathematics.* `http://eom.springer.de/`

[FR] Riesz F. R.: *Sur l'intégrale de Lebesgue.* Acta Mathematica 42, 191-205, 1919.

[GB] Birkhoff G.: *On product integration.* Journal of Mathematics and Physics 16, 104–132, 1938.

[Gil] Gill R. D.: *Product Integration*, 2001. `http://www.math.uu.nl/people/gill/Preprints/prod_int_1.pdf`

[GJ] Gill R. D., Johansen S.: *A survey of product integration with a view toward application in survival analysis.* The Annals of Statistics vol. 18, no. 4, 1501–1555, 1990.

[GN] Gueraggio A., Nastasi P.: *Italian Mathematics Between the Two World Wars.* Birkhäuser Verlag, 2005.

[GP] Peano G.: *Intégration par séries des équations différentielles linéaires.* Mathematische Annalen 32, 450–456, 1888.

[GR] Rasch G.: *Zur Theorie und Anwendung des Produktintegrals.* Journal für die Reine und Angewandte Mathematik 191, 65–119, 1934.

[IG] Israel G., Gasca A. M.: *The Biology of Numbers.* Birkhäuser Verlag, 2002.

[IMS] *Obituary: Pesi Rustom Masani.* IMS Bulletin 1/32, 2003. `http://www.imstat.org/bulletin/Bulletin32_1.pdf`

[IR] Rana, Inder K.: *An Introduction to Measure and Integration (2nd edition).* American Mathematical Society, 2002.

[JG] Goodstein, Judith R.: *The Volterra Chronicles: The Life and Times of an Extraordinary Mathematician.* American Mathematical Society, 2007.

[JK] Jarník J., Kurzweil J.: *A general form of the product integral and linear ordinary differential equations.* Czechoslovak Mathematical Journal 37, 642–659, 1987.

[Kl] Kline M.: *Mathematical Thought from Ancient to Modern Times.* Oxford University Press, 1990.

[KZ] Žitný K.: *Přínos F. Riesze k teorii Lebesgueova integrálu funkcí jedné reálné proměnné.* In Člověk – Umění – Matematika, Prometheus, Praha, 1996 (in Czech).

[Lex] *Magyar Életrajzi Lexikon 1000-1990.* `http://www.mek.iif.hu/porta/szint/egyeb/lexikon/`

[LS1] Schlesinger L.: *Neue Grundlagen für einen Infinitesimalkalkul der Matrizen.* Mathematische Zeitschrift 33, 33–61, 1931.

[LS2] Schlesinger L.: *Weitere Beiträge zum Infinitesimalkalkul der Matrizen.* Mathematische Zeitschrift 35, 485–501, 1932.

[Mas] Masani P. R.: *Multiplicative Riemann integration in normed rings.* Transactions of the American Mathematical Society 61, 147–192, 1947.

[MB] Braun M.: *Differential Equations and Their Applications.* Springer Verlag, 1975.

[McT] *MacTutor History of Mathematics.* `http://www-gap.dcs.st-and.ac.uk/~history/`

[PRM] *Pesi Rustom Masani.* Pittsburgh University Times 5/32, 1999. `http://www.pitt.edu/utimes/issues/32/991028/07.htm`

[RG] Gordon R. A.: *The integrals of Lebesgue, Denjoy, Perron and Henstock.* American Mathematical Society, 1994.

[Sch1] Schwabik Š.: *Bochner product integration.* Mathematica Bohemica 119, 305–335, 1994.

[Sch2] Schwabik Š.: *Integrace v* **R** *(Kurzweilova teorie).* Univerzita Karlova, nakladatelství Karolinum, 1999 (in Czech).

[Sch3] Schwabik Š.: *The Perron product integral and generalized linear differential equations.* Časopis pro pěstování matematiky 115, 368–404, 1990.

[SS] Schwabik Š., Slavík A.: *Henstock-Kurzweil and McShane product integration; Descriptive definitions.* To appear in Czechoslovak Mathematical Journal.

[SŠ] Schwabik Š., Šarmanová P.: *Malý průvodce historií integrálu.* Prometheus, 1996 (in Czech).

[SY] Schwabik Š., Ye G.: *Topics in Banach Space Integration.* World Scientific, Singapore, 2005.

[VH] Volterra V., Hostinský B.: *Opérations infinitésimales linéaires.* Gauthier-Villars, Paris, 1938.

[VJ] Jarník V.: *Diferenciální rovnice v komplexním oboru.* Academia, 1975 (in Czech).

[VV1] Volterra V.: *Sui fondamenti della teoria delle equazioni differenziali lineari.* Memorie della Società Italiana della Scienze (3) VI, 1887.

[VV2] Volterra V.: *Sui fondamenti della teoria delle equazioni differenziali lineari.* Memorie della Società Italiana della Scienze (3) XII, 3–68, 1902.

[VV3] Volterra V.: *Sulle equazioni differenziali lineari.*
Rendiconti dell' Academia dei Lincei 3, 393–396, 1887.

[VV4] Volterra V.: *Sulla teoria delle equazioni differenziali lineari.*
Rendiconti del Circolo matematico di Palermo 2, 69–75, 1888.

[VV5] Volterra V.: *Theory of functionals and of integral and integro-differential equations.* Dover Publications, 1959.

Previous Volumes of the Series

HISTORY OF MATHEMATICS

1. J. Bečvář, E. Fuchs (ed.): *History of Mathematics I*, 1994
2. J. Bečvář et al.: *Eduard Weyr (1852–1903)*, 1995
3. J. Bečvář, E. Fuchs (ed.): *Mathematics of the 19th Century*, 1996
4. J. Bečvář, E. Fuchs (ed.): *Man – Art – Mathematics*, 1996
5. J. Bečvář (ed.): *Jan Vilém Pexider (1874–1914)*, 1997
6. P. Šarmanová, Š. Schwabik: *A Brief Guide to the History of the Integral*, 1996
7. J. Bečvář, E. Fuchs (ed.): *History of Mathematics II*, 1997
8. P. Šišma: *Graph Theory 1736–1963*, 1997
9. K. Mačák: *On the Origin of the Probability Calculus*, 1997
10. M. Němcová: *František Josef Studnička (1836–1903)*, 1998
11. J. Bečvář, E. Fuchs (ed.): *Mathematics throughout the Ages I*, 1998
12. J. Bečvář, E. Fuchs (ed.): *Mathematics in the 16th and 17th Century*, 1999
13. M. Bečvářová: *On the History of the Association (1862–1869)*, 1999
14. K. Lepka: *History of Fermat Quotients (Fermat – Lerch)*, 2000
15. K. Mačák: *Three Medieval Collections of Mathematical Problems*, 2001
16. J. Bečvář, E. Fuchs (ed.): *Mathematics throughout the Ages II*, 2001
17. E. Fuchs (ed.): *Mathematics throughout the Ages I*, 2001 (Engl.)
18. K. Mačák, G. Schuppener: *Mathematics at the Jesuit Clementinum in the Years 1600–1740*, 2001
19. J. Bečvář et al.: *Mathematics in Medieval Europe*, 2001
20. M. Bečvářová: *Euclid's Elements, its Editions and Translations*, 2002

21. P. Šišma: *Mathematics at the Brno German Technical University*, 2002

22. M. Hykšová: *Karel Rychlík (1885–1968)*, 2003

23. J. Bečvář, M. Bečvářová, H. Vymazalová: *Mathematics in Ancient Egypt and Mesopotamia*, 2003

24. J. Bečvář, E. Fuchs (ed.): *Mathematics throughout the Ages III*, 2004

25. E. Fuchs (ed.): *Mathematics throughout the Ages II*, 2004 (Engl.)

26. K. Mačák: *The Development of Probability Theory in the Czech Lands until 1938*, 2005

27. Z. Kohoutová, J. Bečvář: *Vladimír Kořínek (1899–1981)*, 2005

28. J. Bečvář, M. Bečvářová, J. Škoda: *Emil Weyr and his Italian Stay in the Year 1870/71*, 2006

29. A. Slavík: *Product Integration, its History and Applications*, 2007 (Engl.)

Antonín Slavík

PRODUCT INTEGRATION, ITS HISTORY AND APPLICATIONS

Nečas Center for Mathematical Modeling, Volume 1

History of Mathematics, Volume 29

Vydal
MATFYZPRESS
vydavatelství
Matematicko-fyzikální fakulty
Univerzity Karlovy v Praze
Sokolovská 83, 185 75 Praha 8
jako svou 200. publikaci

Vydání publikace bylo podpořeno
výzkumným záměrem MŠMT č. MSM 0021620839

Z předloh připravených v systému TEX
vytisklo Reprostředisko UK MFF
Sokolovská 83, 186 75 Praha 8

První vydání

Praha 2007

ISBN 80-7378-006-2