Introduction

In: Antonín Slavík (author): Product integration. Its history and applications. (English). Praha: Matfyzpress, 2007. pp. 3–12.

Persistent URL: http://dml.cz/dmlcz/401132

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 $(i,j = 1,\dots,n)$ 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

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

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 $y_1^0, \dots, y_n^0 \in \mathbf{C}$ 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 $p_1, \dots, p_n, q$ 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

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

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

Thus we see that the solutions of Equation (1.1.1), whose coefficients $p_1, \dots, p_n, q$ 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 $p_1, \dots, p_n, q$ 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 $y_1, \dots, y_n$ 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 $\Delta t_i = t_i - t_{i-1}$, $i = 1, \ldots, m$. 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

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

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 $\nu(D) = \max\{\Delta t_i, i = 1, \ldots, m\}$), 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 $F(t) = P(T \leq t)$ 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

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

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

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

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

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

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

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 $s = u_0 < u_1 < \cdots < u_m = t$ 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 $\mathrm{d}u$.