I. Introduction

In: Štefan Schwabik (author); Milan Tvrdý (author); Otto Vejvoda (author): Differential and Integral Equations. Boundary Value Problems and Adjoints. (English). Praha: Academia, 1979. pp. 9–74.

Persistent URL: http://dml.cz/dmlcz/400396

I. *Introduction*

This chapter provides some auxiliary results and notations needed in the subsequent chapters. As most of them can be easily found in the plentiful literature on linear algebra, real functions, functional analysis etc. we give only the necessary references without including their proofs. More attention is paid only to the Perron-Stieltjes integral in sections 4, 5 and 6.

1. Preliminaries

1.1. Basic notations. By R we denote the set of all real numbers. For $a < b$ we denote by $[a, b]$ and (a, b) respectively the closed and the open interval with the endpoints a, b. Similarly $[a, b)$, $(a, b]$ means the corresponding halfopen intervals.

A matrix with m rows and n columns is called an $m \times n$-matrix, $n \times 1$-matrices are called column n-vectors and $1 \times m$-matrices are called row m-vectors.

Matrices which in general do not reduce to vectors are denoted by capitals while vectors are denoted by lower-case letters. Given an $m \times n$-matrix $\mathbf{A}$, its element in the j-th row and k-th column is usually denoted by $a_{j,k}$ ($\mathbf{A} = (a_{j,k})$, $j = 1, \ldots, m$, $k = 1, \ldots, n$). Furthermore, $\mathbf{A}^*$ denotes the transpose of $\mathbf{A}$ ($\mathbf{A}^* = (a_{k,j})$, $k = 1, \ldots, n$, $j = 1, \ldots, m$),

$$|\mathbf{A}| = \max_{j=1,\ldots,m} \sum_{k=1}^{n} |a_{j,k}|,$$

rank$(\mathbf{A})$ is the rank of $\mathbf{A}$ and det$(\mathbf{A})$ denotes the value of the determinant of $\mathbf{A}$. If $m = n$ and $\det(\mathbf{A}) \neq 0$, then $\mathbf{A}^{-1}$ denotes the inverse of $\mathbf{A}$. $\mathbf{I}_m$ is the identity $m \times m$-matrix and $\mathbf{0}_{m,n}$ is the zero $m \times m$-matrix ($\mathbf{I}_m = (\delta_{j,k})$ $j, k = 1, \ldots, m$, where $\delta_{j,k} = 1$ if $j = k$, $\delta_{j,k} = 0$ if $j \neq k$ and $\mathbf{0}_{m,n} = (n_{j,k})$ $j = 1, \ldots, m$, $k = 1, \ldots, n$, where $n_{j,k} = 0$ for all $j = 1, \ldots, m$ and $k = 1, \ldots, n$). Usually, if no confusion may arise, the indices are omitted. The addition and multiplication on the space of matrices are defined in the obvious way and the usual notation

$$\mathbf{A} + \mathbf{B}\,, \ \mathbf{AB}\,, \ \lambda\mathbf{A} \qquad (\lambda \in R)$$

is used. Let the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be of the types $m \times n$, $m \times p$ and $q \times n$, respectively. Then $\boldsymbol{D} = [\boldsymbol{A}, \boldsymbol{B}]$ is the $m \times (n + p)$-matrix with $d_{j,k} = a_{j,k}$ for $j = 1, \ldots, m$, $k = 1, \ldots, n$ and $d_{j,k} = b_{j,k-n}$ for $j = 1, \ldots, m$, $k = n + 1, n + 2, \ldots, n + p$. Analogously

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}$$

is the $(m + q) \times n$-matrix with $h_{j,k} = a_{j,k}$ if $j \leq m$ and $h_{j,k} = c_{j-m,k}$ if $j > m$.

R_n is the space of all real column n-vectors and R_n^* is the space of all real row n-vectors, $R_1 = R_1^* = R$. For $\boldsymbol{x} \in R_n$, $\boldsymbol{x}^* \in R_n^*$ we write

$$|\boldsymbol{x}| = \max_{j=1,\ldots,n} |x_j|$$

and

$$|\boldsymbol{x}^*| = \sum_{j=1}^{n} |x_j| \, .$$

Given an $m \times n$-matrix $\boldsymbol{A}$, $\boldsymbol{x} \in R_n$ and $\boldsymbol{y} \in R_m$, then $|\boldsymbol{A}\boldsymbol{x}| \leq |\boldsymbol{A}| \, |\boldsymbol{x}|$ and $|\boldsymbol{y}^*\boldsymbol{A}| \leq$ $\leq |\boldsymbol{y}^*| \, |\boldsymbol{A}|$. The Euclidean norm in R_n is denoted by $|\,.\,|_e$

$$\boldsymbol{x} \in R_n \to |\boldsymbol{x}|_e = (\boldsymbol{x}^*\boldsymbol{x})^{1/2} = \left(\sum_{j=1}^{n} x_j^2 \right)^{1/2} .$$

It is easy to see that any $\boldsymbol{x} \in R_n$ satisfies $|\boldsymbol{x}|_e = |\boldsymbol{x}^*|_e$ and $|\boldsymbol{x}| \leq |\boldsymbol{x}|_e \leq |\boldsymbol{x}^*| \leq n|\boldsymbol{x}|$.

The space of all real $m \times n$-matrices is denoted by $L(R_n, R_m)$ $(L(R_n, R_n) = L(R_n))$.

If M, N are sets and f is a mapping defined on M with values in N then we write $f\colon M \to N$ or $x \in M \to f(x) \in N$. For example, if f is a real function defined on an interval $[a, b]$, we write simply $f\colon [a, b] \to R$.

The words "measure", "measurable" without specification stand always for Lebesgue measure in R_n and measurability with respect to Lebesgue measure.

1.2. Linear spaces. A nonempty set X is called a (real) linear space if for every $\boldsymbol{x}, \boldsymbol{y} \in X$ and $\lambda \in R$ the sum $\boldsymbol{x} + \boldsymbol{y} \in X$ and the product $\lambda\boldsymbol{x} \in X$ are defined and the operations satisfy the usual axioms of a linear space. The zero element in X is denoted by $\boldsymbol{0}$.

A subset $L \subset X$ is a linear subspace of X if L is a linear space with respect to the sum and product with a real number given in X.

The elements $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of X are called *linearly independent* if $\alpha_1\boldsymbol{x}_1 + \ldots + \alpha_n\boldsymbol{x}_n = \boldsymbol{0}$, $\alpha_i \in R$, $i = 1, \ldots, n$ implies $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$. Otherwise the elements $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are *linearly dependent*.

If X is a linear space and a norm $\boldsymbol{x} \in X \to \|\boldsymbol{x}\| \in R$ is defined, X is called a *normed linear space.* If X is a normed linear space which is complete with respect to the metric induced by the norm, then X is called a *Banach space.*

A real linear space X is called an *inner product space* (or pre-Hilbert space) if on $X \times X$ a real function $(\boldsymbol{x}_1, \boldsymbol{x}_2)_X$ is defined $((\boldsymbol{x}_1, \boldsymbol{x}_2) \in X \times X \to (\boldsymbol{x}_1, \boldsymbol{x}_2)_X \in R)$ such that for all $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in X$

$$(\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{x}_3)_X = (\boldsymbol{x}_1, \boldsymbol{x}_3)_X + (\boldsymbol{x}_2, \boldsymbol{x}_3)_X,$$
$$(\boldsymbol{x}_1, \boldsymbol{x}_2)_X = (\boldsymbol{x}_2, \boldsymbol{x}_1)_X,$$
$$(\alpha \boldsymbol{x}_1, \boldsymbol{x}_2)_X = \alpha(\boldsymbol{x}_1, \boldsymbol{x}_2)_X,$$
$$(\boldsymbol{x}, \boldsymbol{x})_X \geq 0 \quad \text{and} \quad (\boldsymbol{x}, \boldsymbol{x})_X \neq 0 \qquad \text{for} \quad \boldsymbol{x} \neq \boldsymbol{0}.$$

The real function $(.,\ .)_X$ is called an *inner product* on X.

If X is an inner product space then the relation

$$(*) \qquad \boldsymbol{x} \in X \to \|\boldsymbol{x}\|_X = (\boldsymbol{x}, \boldsymbol{x})_X^{1/2} \in R$$

defines a norm on X.

A real inner product space X which is complete with respect to the norm defined by $(*)$ is called a real Hilbert space. Consequently a Hilbert space is a Banach space whose norm is induced by an inner product on X.

1.3. Function spaces. We shall deal with some usual spaces of real functions on an interval $[a, b]$, $-\infty < a < b < +\infty$. The sum of two functions and the product of a scalar and a function is defined in the usual way. For more detailed information see e.g. Dunford, Schwartz [1].

(i) We denote by $C_n[a, b]$ the space of all continuous column n-vector functions $\boldsymbol{f}$: $[a, b] \to R_n$ and define

$$\boldsymbol{f} \in C_n[a, b] \to \|\boldsymbol{f}\|_{C_n[a,b]} = \sup_{t \in [a,b]} |\boldsymbol{f}(t)|.$$

$\|\cdot\|_{C_n[a,b]}$ is a norm on $C_n[a, b]$; $C_n[a, b]$ with respect to this norm forms a Banach space. The zero element in $C_n[a, b]$ is the function vanishing identically on $[a, b]$.

(ii) If $1 \leq p < \infty$ we denote by $L_n^p[a, b]$ the space of all measurable functions $\boldsymbol{f}$: $[a, b] \to R_n$ such that

$$\int_a^b |\boldsymbol{f}(t)|^p \, dt < \infty.$$

We set

$$\boldsymbol{f} \in L_n^p[a, b] \to \|\boldsymbol{f}\|_{L_n^p[a,b]} = \left(\int_a^b |\boldsymbol{f}(t)|^p \, dt \right)^{1/p}.$$

The elements of $L_n^p[a, b]$ are classes of functions which are equal to one another almost everywhere (a. e.)*) on $[a, b]$. For the purposes of this text it is not restrictive

*) If a statement is true except possibly on a set of measure zero then we say that the statement is true almost everywhere (a.e.).

if we consider functions instead of classes of functions which are equal a.e. on $[a, b]$.

$L_n^p[a, b]$ with respect to the norm $\|\cdot\|_{L_n^p[a,b]}$ is a Banach space. By $L_n^\infty[a, b]$ we denote the space of all measurable essentially bounded functions $\boldsymbol{f}: [a, b] \to R_n$ with the norm defined by

$$\boldsymbol{f} \in L_n^\infty[a, b] \to \|\boldsymbol{f}\|_{L_n^\infty[a,b]} = \sup_{t \in [a,b]} \operatorname{ess} |\boldsymbol{f}(t)| .$$

$L_n^\infty[a, b]$ is a Banach space with respect to the norm $\|\cdot\|_{L_n^\infty[a,b]}$. The zero element in $L_n^p[a, b]$ $(1 \leq p \leq \infty)$ is the class of functions which vanish a.e. on $[a, b]$.

(iii) We denote by $BV_n[a, b]$ the space of all functions $\boldsymbol{f}: [a, b] \to R_n$ of bounded variation $\operatorname{var}_a^b \boldsymbol{f} < \infty$ where

$$\operatorname{var}_a^b \boldsymbol{f} = \sup \sum_{i=1}^{k} |\boldsymbol{f}(t_i) - \boldsymbol{f}(t_{i-1})|$$

and the supremum is taken over all finite subdivisions of $[a, b]$ of the form $a = t_0 < t_1 < \dots < t_k = b$. Let $c \in [a, b]$ then

$$\operatorname{var}_a^b \boldsymbol{f} = \operatorname{var}_a^c \boldsymbol{f} + \operatorname{var}_c^b \boldsymbol{f} .$$

If we define

$$\boldsymbol{f} \in BV_n[a, b] \to \|\boldsymbol{f}\|_{BV_n[a,b]} = |\boldsymbol{f}(a)| + \operatorname{var}_a^b \boldsymbol{f}$$

then $\|\cdot\|_{BV_n[a,b]}$ is a norm on $BV_n[a, b]$ and $BV_n[a, b]$ is a Banach space with respect to this norm.

By $NBV_n[a, b]$ the subspace of $BV_n[a, b]$ is denoted such that $\boldsymbol{f} \in NBV_n[a, b]$ if $\boldsymbol{f}$ is continuous from the right at every point of (a, b) and $\boldsymbol{f}(a) = \boldsymbol{0}$. The norm in $NBV_n[a, b]$ is defined by

$$\boldsymbol{f} \in NBV_n[a, b] \to \|\boldsymbol{f}\|_{NBV_n[a,b]} = \operatorname{var}_a^b \boldsymbol{f} .$$

A function $\boldsymbol{f}: [a, b] \to R_n$ is called absolutely continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sum_{i=1}^{k} |\boldsymbol{f}(b_i) - \boldsymbol{f}(a_i)| < \varepsilon$$

where (a_i, b_i), $i = 1, \dots, k$ are arbitrary pairwise disjoint subintervals in $[a, b]$ such that $\sum_{i=1}^{k} |b_i - a_i| < \delta$.

Let $AC_n[a, b]$ be the space of all absolutely continuous functions $\boldsymbol{f}: [a, b] \to R_n$. It is $AC_n[a, b] \subset BV_n[a, b]$ and $AC_n[a, b]$ is a Banach space with respect to the norm of $BV_n[a, b]$, i.e.

$$\boldsymbol{f} \in AC_n[a, b] \to \|\boldsymbol{f}\|_{AC_n[a,b]} = |\boldsymbol{f}(a)| + \operatorname{var}_a^b \boldsymbol{f} .$$

The zero element in $AC_n[a, b]$ and $BV_n[a, b]$ is the function vanishing identically on $[a, b]$.

Given an interval $[a, b]$, we write simply C_n, L_n^p, L_n^∞, BV_n, NBV_n, AC_n instead of $C_n[a, b]$, $L_n^p[a, b]$, $L_n^\infty[a, b]$, $BV_n[a, b]$, $NBV_n[a, b]$, $AC_n[a, b]$ if no misunderstanding may arise. If $n = 1$ then the index n is omitted, e.g. $C_1[a, b] = C[a, b]$, $L_1^p[a, b] = L^p[a, b]$ etc. The index n is also sometimes omitted in symbols for the norms, i.e. instead of $\|\cdot\|_{C_n}$, $\|\cdot\|_{BV_n}$, $\|\cdot\|_{L_n^p}$ we write $\|\cdot\|_C$, $\|\cdot\|_{BV}$, $\|\cdot\|_{L^p}$, respectively.

A matrix valued function $\mathbf{F}$: $[a, b] \to L(R_n, R_m)$ is said to be measurable or continuous or of bounded variation or absolutely continuous or essentially bounded on $[a, b]$ if any of the functions

$$t \in [a, b] \to f_{i,j}(t) \in R \qquad (i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n)$$

is measurable or continuous or of bounded variation or absolutely continuous or essentially bounded on $[a, b]$, respectively.

Let us mention that

$$\operatorname{var}_a^b \mathbf{F} = \sup \sum_{i=1}^{k} |\mathbf{F}(t_i) - \mathbf{F}(t_{i-1})|$$

where the supremum is taken over all finite subdivisions of $[a, b]$ of the form

$$a = t_0 < t_1 < \dots < t_k = b$$

and

$$\max_{\substack{j=1,2,\dots,m \\ l=1,2,\dots,n}} (\operatorname{var}_a^b f_{j,l}) \le \operatorname{var}_a^b \mathbf{F} \le \sum_{j=1}^{m} \sum_{l=1}^{n} \operatorname{var}_a^b f_{j,l}\,.$$

We denote $\|\mathbf{F}\|_{L^\infty} = \sup\operatorname{ess}_{t \in [a,b]} |\mathbf{F}(t)|$ and $\|\mathbf{F}\|_{L^p} = (\int_a^b |\mathbf{F}(t)|^p \,\mathrm{d}t)^{1/p}$ for $1 \le p < \infty$. If $\mathbf{F}$: $[a, b] \to L(R_n, R_m)$ is measurable and $\|\mathbf{F}\|_{L^p} < \infty$ $(1 \le p \le \infty)$, then the matrix valued function $\mathbf{F}$: $[a, b] \to L(R_n, R_m)$ is said to be L^p-integrable on $[a, b]$. (Instead of L^1-integrable we write simply L-integrable.)

1.4. Properties of functions of bounded variation. If $f \in BV[a, b]$ then the limits $\lim_{t \to t_0+} f(t) = f(t_0+)$, $t_0 \in [a, b)$, $\lim_{t \to t_0-} f(t) = f(t_0-)$, $t_0 \in (a, b]$ exist and the set of discontinuity points of f in $[a, b]$ is at most countable.

If $f \in BV[a, b]$ then $f(t) = p(t) - n(t)$, $t \in [a, b]$ where p, n: $[a, b] \to R$ are nondecreasing functions on $[a, b]$. Let a sequence $t_1, t_2, \dots$ of points in $[a, b]$, $t_i \neq t_j$, $i \neq j$ and two sequences of real numbers $c_1, c_2, \dots, d_1, d_2, \dots$ be given such that $t_n = a$ implies $c_n = 0$ and $t_n = b$ implies $d_n = 0$. Assume that the series $\sum_n c_n$, $\sum_n d_n$ converge absolutely. Define on $[a, b]$ a function s: $[a, b] \to R$ by the relation

$$s(t) = \sum_{t_n \le t} c_n + \sum_{t_n < t} d_n\,.$$

Every function of this type is called a *break function* on $[a, b]$. Clearly $s(t_n+) - s(t_n) = d_n$ and $s(t_n) - s(t_n-) = c_n$, $n = 1, 2, \dots$ and $s(t+) = s(t) = s(t-)$ if $t \in [a, b]$, $t \neq t_n$, $n = 1, 2, \dots$. Further $s \in BV[a, b]$ and $\operatorname{var}_a^b s = \sum_n (|c_n| + |d_n|)$.

If $f \in BV[a, b]$ then there exist uniquely determined functions $f_c \in BV[a, b]$, $f_b \in BV[a, b]$ such that f_c is a continuous function on $[a, b]$, f_b is a break function on $[a, b]$ and $f = f_c + f_b$ (the Jordan decomposition of $f \in BV[a, b]$).

If $f \in BV[a, b]$ then the derivative f' of f exists a.e. on $[a, b]$.

If $f \in BV[a, b]$ then f is expressible in the form

$$f = f_{ac} + f_s + f_b$$

where $f_{ac} \in AC[a, b]$, f_b is a break function on $[a, b]$ and f_s; $[a, b] \to R$ is continuous on $[a, b]$ with the derivative $f_s' = 0$ a.e. on $[a, b]$ (the Lebesgue decomposition of $f \in BV[a, b]$).

If $f \in AC[a, b]$ then the derivative f' exists a.e. on $[a, b]$ and $f' \in L^1[a, b]$, i.e. $\int_a^b |f'(t)| \, dt < \infty$ and $\operatorname{var}_a^b f = \int_a^b |f'(t)| \, dt$.

The following statement is important:

Helly's Choice Theorem. *Let an infinite family F of real functions on $[a, b]$ be given. If there is $K \geq 0$ such that*

$$|f(t)| \leq K \quad \text{for} \quad t \in [a, b] \quad \text{and} \quad \operatorname{var}_a^b f \leq K \quad \text{for every} \quad f \in F$$

then the family F contains a sequence $\{f_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} f_n(t) = \varphi(t)$ for every $t \in [a, b]$ and $\varphi \in BV[a, b]$, i.e. the sequence $f_n(t)$ converges pointwise to a function φ: $[a, b] \to R$ which is also of bounded variation.

On functions of bounded variation see e.g. Natanson [1], Aumann [1].

2. Linear algebraical equations and generalized inverse matrices

Let us consider linear algebraical equations for $\boldsymbol{x} \in R_n$ and $\boldsymbol{y}^* \in R_m^*$

(2,1) $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

(2,2) $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

and

(2,3) $$\boldsymbol{y}^*\boldsymbol{A} = \boldsymbol{0},$$

where $\boldsymbol{A}$ is an $m \times n$-matrix $(\boldsymbol{A} \in L(R_n, R_m))$ and $\boldsymbol{b} \in R_m$.

By $N(\boldsymbol{A})$ we denote the set of all solutions to (2,2). Obviously, $N(\boldsymbol{A})$ is a linear subspace in R_n, i.e. if $\boldsymbol{x}_1, \boldsymbol{x}_2 \in N(\boldsymbol{A})$ and $\alpha_1, \alpha_2 \in R$, then $\boldsymbol{x}_1\alpha_1 + \boldsymbol{x}_2\alpha_2 \in N(\boldsymbol{A})$. It is well-known that

(2,4) $$\dim N(\boldsymbol{A}) = n - \operatorname{rank}(\boldsymbol{A}),$$

i.e. either (2,2) possesses only the trivial solution $\boldsymbol{x} = \boldsymbol{0}$ (if $\operatorname{rank}(\boldsymbol{A}) = n$) or $N(\boldsymbol{A})$ contains a subset of $k = n - \operatorname{rank}(\boldsymbol{A})$ elements $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ which are linearly independent, while any subset of $k + 1$ its elements is linearly dependent. (We say also that the homogeneous equation (2,2) has exactly $k = n - \operatorname{rank}(\boldsymbol{A})$ linearly independent solutions.) The set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ forms a basis of $N(\boldsymbol{A})$ and any $\boldsymbol{x} \in N(\boldsymbol{A})$ can be expressed as their linear combination

$$\boldsymbol{x} = \sum_{j=1}^{k} \boldsymbol{x}_j \alpha_j, \qquad \text{where} \quad \alpha_j \in R \quad (j = 1, 2, \ldots, k).$$

As (2,3) is equivalent to $\boldsymbol{A}^*\boldsymbol{y} = \boldsymbol{0}$, $N(\boldsymbol{A}^*)$ denotes the linear subspace in R_m^* of all solutions to (2,3) and

$$\text{(2,5)} \qquad \dim N(\boldsymbol{A}^*) = m - \operatorname{rank}(\boldsymbol{A}^*) = m - \operatorname{rank}(\boldsymbol{A}).$$

Furthermore, the equation (2,1) possesses a solution if and only if (2,3) implies $\boldsymbol{y}^*\boldsymbol{b} = 0$. In particular, (2,1) possesses a solution for any $\boldsymbol{b} \in R_m$ if and only if (2,3) implies $\boldsymbol{y}^* = \boldsymbol{0}$ ($\dim N(\boldsymbol{A}^*) = 0$).

The equation (2,4) is said to be an adjoint equation to (2,1).

The concept of a generalized inverse matrix introduced by R. Penrose (Penrose [1] and [2]) enables us to express the solutions to (2,1) if they exist.

The following assertion is helpful.

2.1. Lemma. $\boldsymbol{BAA}^* = \boldsymbol{CAA}^*$ *implies* $\boldsymbol{BA} = \boldsymbol{CA}$ *and* $\boldsymbol{BA}^*\boldsymbol{A} = \boldsymbol{CA}^*\boldsymbol{A}$ *implies* $\boldsymbol{BA}^* = \boldsymbol{CA}^*$.

Proof. If $\boldsymbol{BAA}^* = \boldsymbol{CAA}^*$, then $\boldsymbol{0} = (\boldsymbol{BAA}^* - \boldsymbol{CAA}^*)(\boldsymbol{B} - \boldsymbol{C})^* = (\boldsymbol{BA} - \boldsymbol{CA})(\boldsymbol{A}^*\boldsymbol{B}^* - \boldsymbol{A}^*\boldsymbol{C}^*)$, whence $\boldsymbol{BA} = \boldsymbol{CA}$ immediately follows. (Given a matrix $\boldsymbol{D}$, $\boldsymbol{DD}^* = \boldsymbol{0}$ if and only if $\boldsymbol{D} = \boldsymbol{0}$.) As $(\boldsymbol{A}^*)^* = \boldsymbol{A}$, the latter implication is a consequence of the former one.

2.2. Theorem. *Given* $\boldsymbol{A} \in L(R_n, R_m)$, *there exists a unique matrix* $\boldsymbol{X} \in L(R_m, R_n)$ *such that*

$$\text{(2,6)} \qquad \boldsymbol{AXA} = \boldsymbol{A},$$

$$\text{(2,7)} \qquad \boldsymbol{XAX} = \boldsymbol{X},$$

$$\text{(2,8)} \qquad \boldsymbol{X}^*\boldsymbol{A}^* = \boldsymbol{AX},$$

$$\text{(2,9)} \qquad \boldsymbol{A}^*\boldsymbol{X}^* = \boldsymbol{XA}.$$

Proof. (a) Putting (2,8) into (2,7) we obtain

$$\text{(2,10)} \qquad \boldsymbol{XX}^*\boldsymbol{A}^* = \boldsymbol{X}.$$

On the other hand, if (2,10) holds, then $\boldsymbol{AX} = \boldsymbol{AXX}^*\boldsymbol{A}^*$. Since $(\boldsymbol{AXX}^*\boldsymbol{A}^*)^* = \boldsymbol{AXX}^*\boldsymbol{A}^*$, this means that $(\boldsymbol{AX})^* = \boldsymbol{AX}$ and (2,8) holds. Moreover, (2,8) and (2,10) yields $\boldsymbol{X} = \boldsymbol{XX}^*\boldsymbol{A}^* = \boldsymbol{XAX}$, i.e. the couple of equations (2,7), (2,8) is equivalent to (2,10).

(b) Analogously, the system (2,6), (2,9) is equivalent to

$$\text{(2,11)} \qquad \boldsymbol{XAA}^* = \boldsymbol{A}^*.$$

(c) Furthermore, to find a solution $\mathbf{X}$ to the system (2,10), (2,11) it is sufficient to find a solution $\mathbf{B}$ to the equation

$$\mathbf{B}\mathbf{A}^*\mathbf{A}\mathbf{A}^* = \mathbf{A}^* . \tag{2,12}$$

In fact, (2,12) implies immediately that $\mathbf{X} = \mathbf{B}\mathbf{A}^*$ satisfies (2,11) and consequently also (2,9). Hence

$$\mathbf{A}^*\mathbf{X}^*\mathbf{A}^* = \mathbf{X}\mathbf{A}\mathbf{A}^* = \mathbf{A}^* \quad \text{and} \quad \mathbf{X}\mathbf{X}^*\mathbf{A}^* = \mathbf{B}\mathbf{A}^*\mathbf{X}^*\mathbf{A}^* = \mathbf{B}\mathbf{A}^* = \mathbf{X} .$$

(d) Now, let us consider the set of $n \times n$-matrices $(\mathbf{A}^*\mathbf{A})^j$ ($j = 1, 2, \ldots$). Since the dimension of the space of all real $n \times n$-matrices is finite (n^2), there exist a natural number k and real numbers $\lambda_1, \lambda_2, \ldots, \lambda_k$ such that $|\lambda_1| + |\lambda_2| + \ldots + |\lambda_k| > 0$ and

$$\lambda_1\mathbf{A}^*\mathbf{A} + \lambda_2(\mathbf{A}^*\mathbf{A})^2 + \ldots + \lambda_k(\mathbf{A}^*\mathbf{A})^k = \mathbf{0} . \tag{2,13}$$

Let r be the smallest natural number such that $\lambda_r \neq 0$. If we put

$$\mathbf{B} = -\lambda_r^{-1}\{\lambda_{r+1}\mathbf{I} + \lambda_{r+2}\mathbf{A}^*\mathbf{A} + \ldots + \lambda_k(\mathbf{A}^*\mathbf{A})^{k-r-1}\} , \tag{2,14}$$

then according to (2,13)

$$\mathbf{B}(\mathbf{A}^*\mathbf{A})^{r+1} = (\mathbf{A}^*\mathbf{A})^r .$$

Hence if $r \geq 2$, $\mathbf{B}(\mathbf{A}^*\mathbf{A})^r\,\mathbf{A}^*\mathbf{A} = (\mathbf{A}^*\mathbf{A})^{r-1}\,\mathbf{A}^*\mathbf{A}$ and according to 2.1

$$\mathbf{B}(\mathbf{A}^*\mathbf{A})^r = (\mathbf{A}^*\mathbf{A})^{r-1} .$$

In this way we can successively obtain

$$\mathbf{B}(\mathbf{A}^*\mathbf{A})^j = (\mathbf{A}^*\mathbf{A})^{j-1} \qquad \text{for} \quad j = 2, 3, \ldots, r .$$

In particular, $\mathbf{B}(\mathbf{A}^*\mathbf{A})^2 = \mathbf{A}^*\mathbf{A}$ and by 2.1 $\mathbf{B}\mathbf{A}^*\mathbf{A}\mathbf{A}^* = \mathbf{A}^*$. The matrix $\mathbf{B}$ defined in (2,14) satisfies (2,12) and hence $\mathbf{X} = \mathbf{B}\mathbf{A}^*$ verifies the system (2,6)–(2,9).

(e) It remains to show that this $\mathbf{X}$ is unique. Let us notice that by (2,9) and (2,7)

$$\mathbf{A}^*\mathbf{X}^*\mathbf{X} = \mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{X}$$

and by (2,8) and (2,6)

$$\mathbf{A}^*\mathbf{A}\mathbf{X} = \mathbf{A}^*\mathbf{X}^*\mathbf{A}^* = (\mathbf{A}\mathbf{X}\mathbf{A})^* = \mathbf{A}^*$$

Now, let us assume that $\mathbf{Y} \in L(R_m, R_n)$ is such that

$$\mathbf{A}^*\mathbf{Y}^*\mathbf{Y} = \mathbf{Y}, \qquad \mathbf{A}^*\mathbf{A}\,\mathbf{Y} = \mathbf{A}^* . \tag{2,15}$$

Then, according to (2,10) and (2,11)

$$\mathbf{X} = \mathbf{X}\mathbf{X}^*\mathbf{A}^* = \mathbf{X}\mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathbf{Y} = \mathbf{X}\mathbf{A}\mathbf{Y} = \mathbf{X}\mathbf{A}\mathbf{A}^*\mathbf{Y}^*\mathbf{Y} = \mathbf{A}^*\mathbf{Y}^*\mathbf{Y} = \mathbf{Y} .$$

2.3. Definition. The unique solution $\mathbf{X}$ of the system (2,6)–(2,9) will be called the *generalized inverse matrix* to $\mathbf{A}$ and written $\mathbf{X} = \mathbf{A}^{\#}$.

2.4. Remark. By the definition and by the proof of 2.2 $\boldsymbol{A}^{\#}$ fulfils the relations

(2,16) $\boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{A}$, $\quad \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{A}^{\#}$, $\quad (\boldsymbol{A}^{\#})^{*}\boldsymbol{A}^{*} = \boldsymbol{A}\boldsymbol{A}^{\#}$, $\quad \boldsymbol{A}^{*}(\boldsymbol{A}^{\#})^{*} = \boldsymbol{A}^{\#}\boldsymbol{A}$

and

(2,17) $\boldsymbol{A}^{\#}(\boldsymbol{A}^{\#})^{*}\boldsymbol{A}^{*} = \boldsymbol{A}^{\#}$, $\quad \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{A}^{*} = \boldsymbol{A}^{*}$, $\quad \boldsymbol{A}^{*}(\boldsymbol{A}^{\#})^{*}\boldsymbol{A}^{\#} = \boldsymbol{A}^{\#}$, $\quad \boldsymbol{A}^{*}\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{A}^{*}$

(cf. (2,6)–(2,11) and (2,15)).

2.5. Remark. If $m = n$ and $\boldsymbol{A}$ possesses an inverse matrix $\boldsymbol{A}^{-1}$, then evidently $\boldsymbol{A}^{-1}$ is a generalized inverse matrix to $\boldsymbol{A}$.

2.6. Proposition. *Let $\boldsymbol{A} \in L(R_n, R_m)$, $\boldsymbol{B} \in L(R_p, R_m)$. Then the equation for $\boldsymbol{X} \in L(R_p, R_n)$*

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B} \tag{2,18}$$

possesses a solution if and only if

$$(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{\#})\,\boldsymbol{B} = \boldsymbol{0}\,. \tag{2,19}$$

If this is true, any solution $\boldsymbol{X}$ of (2,18) is of the form

$$\boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{A}^{\#}\boldsymbol{B}\,, \tag{2,20}$$

where $\boldsymbol{X}_0$ is an arbitrary solution of the matrix equation

$$\boldsymbol{A}\boldsymbol{X}_0 = \boldsymbol{0}_{m,p}\,.$$

Proof. Let $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}$, then by (2,6) $(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^{\#})\,\boldsymbol{B} = (\boldsymbol{A} - \boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{X} = \boldsymbol{0}$. If (2,19) holds, then $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{B}$ and (2,18) is equivalent to $\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{A}^{\#}\boldsymbol{B}) = \boldsymbol{0}$, i.e. to $\boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{A}^{\#}\boldsymbol{B}$, where $\boldsymbol{A}\boldsymbol{X}_0 = \boldsymbol{0}$.

2.7. Proposition. *Let $\boldsymbol{A} \in L(R_n, R_m)$. Then $\boldsymbol{A}\boldsymbol{X}_0 = \boldsymbol{0}_{m,p}$ if and only if there exists $\boldsymbol{C} \in L(R_p, R_n)$ such that $\boldsymbol{X}_0 = (\boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{C}$.*

Proof. $\boldsymbol{A}(\boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{C} = (\boldsymbol{A} - \boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{C} = \boldsymbol{0}$ for any $\boldsymbol{C} \in L(R_p, R_n)$. If $\boldsymbol{A}\boldsymbol{X}_0 = \boldsymbol{0}$, then $\boldsymbol{X}_0 = \boldsymbol{X}_0 - \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{X}_0 = (\boldsymbol{I} - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{X}_0$.

Some further properties of generalized inverse matrices are listed in the following lemma.

2.8. Lemma. *Given $\boldsymbol{A} \in L(R_n, R_m)$,*

$$\boldsymbol{A}^{\#\#} = (\boldsymbol{A}^{\#})^{\#} = \boldsymbol{A}\,, \tag{2,21}$$

$$(\boldsymbol{A}^{*})^{\#} = (\boldsymbol{A}^{\#})^{*}\,, \tag{2,22}$$

$$(\lambda\boldsymbol{A})^{\#} = \lambda^{-1}\boldsymbol{A}^{\#} \quad \textit{for any } \lambda \in R\,,\ \lambda \neq 0 \textit{ and } \boldsymbol{0}_{m,n}^{\#} = \boldsymbol{0}_{n,m}\,, \tag{2,23}$$

$$(\boldsymbol{A}^{*}\boldsymbol{A})^{\#} = \boldsymbol{A}^{\#}(\boldsymbol{A}^{\#})^{*}\,, \qquad (\boldsymbol{A}\boldsymbol{A}^{*})^{\#} = (\boldsymbol{A}^{\#})^{*}\boldsymbol{A}^{\#}\,. \tag{2,24}$$

(The relations (2,21)–(2,24) may be easily verified by substituting their right-hand sides in the defining relations for the required generalized inverse.)

2.9. Lemma. *Let* $\mathbf{A} \in L(R_n, R_m)$ *and let* $\mathbf{U} \in L(R_m, R_n)$ *and* $\mathbf{V} \in L(R_n, R_m)$ *be such that*

$$\mathbf{A}^*\mathbf{A}\mathbf{U} = \mathbf{A}^* \quad \text{and} \quad \mathbf{A}\mathbf{A}^*\mathbf{V} = \mathbf{A}.$$

Then

$$\mathbf{V}^*\mathbf{A}\mathbf{U} = \mathbf{A}^{\#}.$$

Proof. Let $\mathbf{A}^*\mathbf{A}\mathbf{U} = \mathbf{A}^*$ and $\mathbf{A}\mathbf{A}^*\mathbf{V} = \mathbf{A}$. Then by 2.6

$$\mathbf{U} = \mathbf{U}_0 + (\mathbf{A}^*\mathbf{A})^{\#}\,\mathbf{A}^* \quad \text{and} \quad \mathbf{V} = \mathbf{V}_0 + (\mathbf{A}\mathbf{A}^*)^{\#}\,\mathbf{A},$$

where $\mathbf{A}^*\mathbf{A}\mathbf{U}_0 = \mathbf{0}$ and $\mathbf{A}\mathbf{A}^*\mathbf{V}_0 = \mathbf{0}$. It follows from 2.1 that $\mathbf{A}^*\mathbf{A}\mathbf{U}_0 = \mathbf{0}$ (i.e. $\mathbf{U}_0^*\mathbf{A}^*\mathbf{A} = \mathbf{0}\mathbf{A}^*\mathbf{A}$) and $\mathbf{A}\mathbf{A}^*\mathbf{V}_0 = \mathbf{0}$ (i.e. $\mathbf{V}_0^*\mathbf{A}\mathbf{A}^* = \mathbf{0}\mathbf{A}\mathbf{A}^*$) implies $\mathbf{A}\mathbf{U}_0 = \mathbf{0}$ and $\mathbf{V}_0^*\mathbf{A} = \mathbf{0}$, respectively. Furthermore, by (2,22) and (2,24)

$$((\mathbf{A}\mathbf{A}^*)^{\#})^* = (\mathbf{A}\mathbf{A}^*)^{\#} = (\mathbf{A}^{\#})^*\,\mathbf{A}^{\#} \quad \text{and} \quad (\mathbf{A}^*\mathbf{A})^{\#} = \mathbf{A}^{\#}(\mathbf{A}^{\#})^*.$$

Hence by the definition of $\mathbf{A}^{\#}$ (cf. 2.4)

$$\mathbf{V}^*\mathbf{A}\mathbf{U} = [\mathbf{A}^*(\mathbf{A}^{\#})^*]\,[\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#}]\,[(\mathbf{A}^{\#})^*\,\mathbf{A}^*] = \mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}^{\#}.$$

2.10. Lemma. *Given* $\mathbf{A} \in L(R_n, R_m)$, *there exist* $\mathbf{U} \in L(R_m, R_n)$ *and* $\mathbf{V} \in L(R_n, R_m)$ *such that*

$$\mathbf{A}^*\mathbf{A}\mathbf{U} = \mathbf{A}^*, \qquad \mathbf{A}\mathbf{A}^*\mathbf{V} = \mathbf{A}. \tag{2,25}$$

Proof. By (2,24) and (2,17)

$$(\mathbf{A}^*\mathbf{A})^{\#}\,\mathbf{A}^* = \mathbf{A}^{\#}(\mathbf{A}^{\#})^*\,\mathbf{A}^* = \mathbf{A}^{\#}$$

and by (2,16) and (2,22) $\mathbf{A}\mathbf{A}^{\#} = (\mathbf{A}^{\#})^*\,\mathbf{A}^* = (\mathbf{A}^*)^{\#}\,\mathbf{A}^*$. Thus

$$[\mathbf{I} - (\mathbf{A}^*\mathbf{A})(\mathbf{A}^*\mathbf{A})^{\#}]\,\mathbf{A}^* = \mathbf{A}^* - \mathbf{A}^*\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}^* - \mathbf{A}^*(\mathbf{A}^*)^{\#}\,\mathbf{A}^* = \mathbf{0}.$$

Since $(\mathbf{A}^*)^* = \mathbf{A}$, this implies also

$$[\mathbf{I} - (\mathbf{A}\mathbf{A}^*)(\mathbf{A}\mathbf{A}^*)^{\#}]\,\mathbf{A} = \mathbf{0}.$$

The proof follows now from 2.6.

2.11. Remark. Let us notice that from the relations (2,16) defining the generalized inverse of $\mathbf{A}$, only $\mathbf{A}\mathbf{A}^{\#}\mathbf{A} = \mathbf{A}$ was utilized in the proofs of 2.6 and 2.7. Some authors (see e.g. Reid [1]) define any matrix $\mathbf{X}$ fulfilling $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$ to be a generalized inverse of $\mathbf{A}$.

Let $\mathbf{A} \in L(R_n, R_m)$ and $h = \operatorname{rank}(\mathbf{A})$. If $h = n$, then $\mathbf{A}\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$. Let us assume $h < n$. By (2,4) there exist an $n \times (n-h)$-matrix $\mathbf{X}_0$ such that its columns form a basis in $N(\mathbf{A})$, i.e. $\mathbf{A}\mathbf{x} = \mathbf{0}$ if and only if there exists $\mathbf{c} \in R_{n-h}$ such that $\boldsymbol{x} = \boldsymbol{X}_0\boldsymbol{c}$. Consequently $\boldsymbol{X} \in L(R_p, R_n)$ fulfils $\boldsymbol{AX} = \boldsymbol{0}_{m,p}$ if and only if there exists $\boldsymbol{C} \in L(R_p, R_{n-h})$ such that $\boldsymbol{X} = \boldsymbol{X}_0\boldsymbol{C}$. In particular, there exists $\boldsymbol{C}_0 \in L(R_n, R_{n-h})$ such that

$$(2,26) \qquad \boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{X}_0\boldsymbol{C}_0 .$$

Furthermore, let $h = \operatorname{rank}(\boldsymbol{A}) < m$. Then by (2,5) there exists $\boldsymbol{Y}_0 \in L(R_m, R_{m-h})$ such that its rows form a basis in $N(\boldsymbol{A}^*)$. Consequently $\boldsymbol{Y} \in L(R_m, R_p)$ fulfils $\boldsymbol{YA} = \boldsymbol{0}_{p.n}$ if and only if there exists $\boldsymbol{D} \in L(R_{m-h}, R_p)$ such that $\boldsymbol{Y} = \boldsymbol{DY}_0$. In particular, there exists $\boldsymbol{D}_0 \in L(R_{m-h}, R_m)$ such that

$$(2,27) \qquad \boldsymbol{I}_m - \boldsymbol{AA}^{\#} = \boldsymbol{D}_0\boldsymbol{Y}_0 .$$

(If $h = m$, then $\boldsymbol{y}^*\boldsymbol{A} = \boldsymbol{0}$ if and only if $\boldsymbol{y}^* = \boldsymbol{0}$.)

2.12. Proposition. *Let* $\boldsymbol{A} \in L(R_n, R_m)$ *and* $\boldsymbol{X} = L(R_m, R_n)$. *Then* $\boldsymbol{AXA} = \boldsymbol{A}$ *if and only if there exist* $\boldsymbol{H}$ *and* $\boldsymbol{D} \in L(R_m, R_n)$ *such that*

$$\boldsymbol{X} = \boldsymbol{A}^{\#} + (\boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{H} + \boldsymbol{D}(\boldsymbol{I}_m - \boldsymbol{AA}^{\#})$$

or equivalently if and only if

$$\boldsymbol{X} = \boldsymbol{A}^{\#} + \boldsymbol{X}_0\boldsymbol{K} + \boldsymbol{LY}_0 ,$$

where $\boldsymbol{X}_0 \in L(R_{n-h}, R_n)$ *and* $\boldsymbol{Y}_0 \in L(R_m, R_{m-h})$ $(h = \operatorname{rank}(\boldsymbol{A}))$ *were defined above,* $\boldsymbol{K} \in L(R_m, R_{n-h})$ *and* $\boldsymbol{L} \in L(R_{m-h}, R_n)$ *are arbitrary, the term* $\boldsymbol{X}_0\boldsymbol{K}$ *vanishes if* $h = n$ *and the term* $\boldsymbol{LY}_0$ *vanishes if* $h = m$.

Proof. Let us assume $h < m$ and $h < n$. Let both $\boldsymbol{AX}_1\boldsymbol{A} = \boldsymbol{A}$ and $\boldsymbol{AX}_2\boldsymbol{A} = \boldsymbol{A}$. Then $\boldsymbol{A}(\boldsymbol{X}_1 - \boldsymbol{X}_2)\,\boldsymbol{A} = \boldsymbol{0}_{m,n}$ and hence $(\boldsymbol{X}_1 - \boldsymbol{X}_2)\,\boldsymbol{A} = (\boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{C}$ with some $\boldsymbol{C} \in L(R_n)$. By 2.6 and 2.7 this is possible if and only if

$$\boldsymbol{X}_1 - \boldsymbol{X}_2 = (\boldsymbol{I}_n - \boldsymbol{A}^{\#}\boldsymbol{A})\,\boldsymbol{CA}^{\#} + \boldsymbol{D}(\boldsymbol{I}_m - \boldsymbol{AA}^{\#})$$

or by (2,26) and (2,27) if and only if

$$\boldsymbol{X}_1 - \boldsymbol{X}_2 = \boldsymbol{X}_0[\boldsymbol{C}_0\boldsymbol{CA}^{\#}] + [\boldsymbol{DD}_0]\,\boldsymbol{Y}_0 .$$

Putting $\boldsymbol{CA}^{\#} = \boldsymbol{H}$, $\boldsymbol{C}_0\boldsymbol{CA}^{\#} = \boldsymbol{K}$ and $\boldsymbol{DD}_0 = \boldsymbol{L}$ we obtain the desired relations. The modification of the proof in the case that $h = m$ and/or $h = n$ is obvious.

2.13. Lemma. *Let* $\boldsymbol{A} \in L(R_n, R_m)$. *If* $\operatorname{rank}(\boldsymbol{A}) = m$, *then* $\det(\boldsymbol{AA}^*) \neq 0$. *If* $\operatorname{rank}(\boldsymbol{A}) = n$, *then* $\det(\boldsymbol{A}^*\boldsymbol{A}) \neq 0$.

Proof. Let $\operatorname{rank}(\boldsymbol{A}) = m$. Then by (2,5) $\boldsymbol{A}^*\boldsymbol{y} = \boldsymbol{0}$ if and only if $\boldsymbol{y} = \boldsymbol{0}$. Now, since $\boldsymbol{A}^* = \boldsymbol{A}^{\#}\boldsymbol{AA}^*$ (cf. (2,17)), $\boldsymbol{AA}^*\boldsymbol{y} = \boldsymbol{0}$ implies $\boldsymbol{A}^*\boldsymbol{y} = \boldsymbol{A}^{\#}\boldsymbol{AA}^*\boldsymbol{y} = \boldsymbol{0}$ and hence $\boldsymbol{y} = \boldsymbol{0}$. This implies that $\operatorname{rank}(\boldsymbol{AA}^*) = m$ (cf. (2,4)).

If $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^*) = n$, then by the first assertion of the lemma $\operatorname{rank}(\boldsymbol{A}^*\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^*(\boldsymbol{A}^*)^*) = n$.

2.14. Remark. It is well known that $\operatorname{rank}(\boldsymbol{AX}) = \min(\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{X}))$ whenever the product $\boldsymbol{AX}$ of the matrices $\boldsymbol{A}, \boldsymbol{X}$ is defined. Hence for a given $\boldsymbol{A} \in L(R_n, R_m)$ there exists $\boldsymbol{X} \in L(R_m, R_n)$ such that $\boldsymbol{AX} = \boldsymbol{I}_m$ only if $\operatorname{rank}(\boldsymbol{A}) = m$. Analogously, there exists $\boldsymbol{X} \in L(R_m, R_n)$ such that $\boldsymbol{XA} = \boldsymbol{I}_n$ only if $\operatorname{rank}(\boldsymbol{A}) = n$.

2.15. Lemma. *Let* $\boldsymbol{A} \in L(R_n, R_m)$. *If* $\operatorname{rank}(\boldsymbol{A}) = m$, *then* $\boldsymbol{AA}^{\#} = \boldsymbol{I}_m$. *If* $\operatorname{rank}(\boldsymbol{A}) = n$, *then* $\boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{I}_n$.

Proof. (a) Let $\operatorname{rank}(\boldsymbol{A}) = m$. Then by 2.13 $(\boldsymbol{AA}^*)$ possesses an inverse $(\boldsymbol{AA}^*)^{-1}$ and according to the relation $\boldsymbol{A}^{\#}\boldsymbol{AA}^* = \boldsymbol{A}^*$ (cf. (2,17))

$$\boldsymbol{A}^{\#} = \boldsymbol{A}^*(\boldsymbol{AA}^*)^{-1} \tag{2,28}$$

and hence $\boldsymbol{AA}^{\#} = \boldsymbol{I}_m$.

(b) If $\operatorname{rank}(\boldsymbol{A}) = n$, then the relation $\boldsymbol{A}^*\boldsymbol{AA}^{\#} = \boldsymbol{A}^*$ from (2,17) and 2.13 imply

$$\boldsymbol{A}^{\#} = (\boldsymbol{A}^*\boldsymbol{A})^{-1}\,\boldsymbol{A} \tag{2,29}$$

and hence $\boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{I}_n$.

2.16. Lemma. *Let* $\boldsymbol{A} \in L(R_m)$, $\boldsymbol{B} \in L(R_n, R_m)$ *and* $\boldsymbol{C} \in L(R_n)$. *If* $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{B}) = m$, *then* $(\boldsymbol{AB})^{\#} = \boldsymbol{B}^{\#}\boldsymbol{A}^{-1}$. *If* $\operatorname{rank}(\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{C}) = n$, *then* $(\boldsymbol{BC})^{\#} = \boldsymbol{C}^{-1}\boldsymbol{B}^{\#}$.

Proof. Let $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{B}) = m$. Then by 2.15 $\boldsymbol{BB}^{\#} = \boldsymbol{I}$. Consequently $\boldsymbol{ABB}^{\#}\boldsymbol{A}^{-1} = \boldsymbol{I}$. Furthermore, $(\boldsymbol{B}^{\#}\boldsymbol{A}^{-1})(\boldsymbol{AB}) = \boldsymbol{B}^{\#}\boldsymbol{B} = \boldsymbol{B}^*(\boldsymbol{B}^{\#})^* = \boldsymbol{B}^*\boldsymbol{A}^*(\boldsymbol{A}^{-1})^*(\boldsymbol{B}^{\#})^* = (\boldsymbol{AB})^*(\boldsymbol{B}^{\#}\boldsymbol{A}^{-1})^*$. This completes the proof of the former assertion. The latter one could be proved analogously.

For some more details about generalized inverse matrices see e.g. Reid [1] (Appendix B), Moore [1], Nashed [1] and "*Proceedings of Symposium on the Theory and Applications of Generalized Inverses of Matrices*" held at the Texas Technological College, Lubbock, Texas, March 1968, Texas Technological College Math. Series, No. 4.

3. Functional analysis

Here we review some concepts and results from linear functional analysis used in the subsequent chapters. For more information we mention e.g. Dunford, Schwartz [1], Heuser [1], Goldberg [1], Schechter [1].

Let X be a linear space over the real scalars R. If F, G are linear subspaces of X, then we set

$$F + G = \{\boldsymbol{z} \in X;\ \boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y},\ \boldsymbol{x} \in F,\ \boldsymbol{y} \in G\}.$$

$F + G$ is evidently a linear subspace of X.

$F + G$ is called the direct sum of two linear subspaces F, G if $F \cap G = \{\mathbf{0}\}$. Let the direct sum of F and G be denoted by $F \oplus G$.

If $F \oplus G = X$ then G is called the complementary subspace to F in X.

It can be shown (see e.g. Heuser [1], II.4) that

(1) for any linear subspace $F \subset X$ there exists at least one complementary subspace $G \subset X$

(2) for any two complementary subspaces G_1, G_2 to a given subspace $F \subset X$ we have $\dim G_1 = \dim G_2$ where by dim the usual linear dimension of a linear set is denoted.

This enables us to define the codimension of a linear subspace $F \subset X$ as follows.

Let $X = F \oplus G$; then we set

$$\operatorname{codim} F = \dim G .$$

(If $\dim G = \infty$ or $X = F$, we put $\operatorname{codim} F = \infty$ or $\operatorname{codim} F = 0$, respectively.)

If $F \subset X$ is a linear subspace, then we set $\mathbf{x} \sim \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in X$ if $\mathbf{x} - \mathbf{y} \in F$. By $\sim$ an equivalence relation on X is given. This equivalence relation decomposes X into disjoint classes of equivalent elements of X. If $\mathbf{x} \in X$ belongs to a given equivalence class with respect to the equivalence relation $\sim$ then all elements of this class belong to the set $\mathbf{x} + F$.

Let us denote by X/F the set of all equivalence classes with respect to the given equivalence relation. Let the equivalence class containing $\mathbf{x} \in X$ be denoted by $[\mathbf{x}]$, i.e.

$$[\mathbf{x}] = \mathbf{x} + F .$$

Then

$$X/F = \{[\mathbf{x}] = \mathbf{x} + F;\ \mathbf{x} \in X\} .$$

If we define $[\mathbf{x}] + [\mathbf{y}] = [\mathbf{x} + \mathbf{y}]$, $\alpha[\mathbf{x}] = [\alpha\mathbf{x}]$ where $\mathbf{x} \in [\mathbf{x}]$, $\mathbf{y} \in [\mathbf{y}]$, $\alpha \in R$ then X/F becomes a linear space over R called the quotient space. It can be shown that if $X = F \oplus G$, then there is a one-to-one correspondence between X/F and G (see e.g. Heuser [1], III.20). Hence

$$\operatorname{codim} F = \dim G = \dim (X/F) .$$

Let X and Y be linear spaces over R. We consider linear operators $\mathbf{A}$ which assign a unique element $\mathbf{A}\mathbf{x} = \mathbf{y} \in Y$ to every element $\mathbf{x} \in D(\mathbf{A}) \subset X$. The set $D(\mathbf{A})$ called the domain of $\mathbf{A}$ forms a linear subspace in X and the linearity relation

$$\mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{z}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{z}$$

holds for all $\mathbf{x}, \mathbf{z} \in X$, $\alpha, \beta \in R$.

The set of all linear operators $\mathbf{A}$ with values in Y such that $D(\mathbf{A}) = X$ will be denoted by $L(X, Y)$. If $X = Y$, then we write simply $L(X)$ instead of $L(X, X)$. The identity operator $\boldsymbol{x} \in X \to \boldsymbol{x} \in X$ on X is usually denoted by $\boldsymbol{I}$. For an operator $\boldsymbol{A} \in L(X, Y)$ we use the following notations:

$$R(\boldsymbol{A}) = \{\boldsymbol{y} \in Y;\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x} \in X\}$$

denotes the *range* of $\boldsymbol{A}$, the linear subspace of values of $\boldsymbol{A} \in L(X, Y)$ in Y.

$$N(\boldsymbol{A}) = \{\boldsymbol{x} \in X;\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \in Y\}$$

denotes the *null-space* of $\boldsymbol{A} \in L(X, Y)$; $N(\boldsymbol{A}) \subset X$ is a linear subspace in X. Further we denote

$$\alpha(\boldsymbol{A}) = \dim N(\boldsymbol{A})$$

and

$$\beta(\boldsymbol{A}) = \operatorname{codim} R(\boldsymbol{A}) = \dim (Y/R(\boldsymbol{A}))\,.$$

If $\alpha(\boldsymbol{A})$, $\beta(\boldsymbol{A})$ are not both infinite, then we define the index $\operatorname{ind} \boldsymbol{A}$ of $\boldsymbol{A} \in L(X, Y)$ by the relation

$$\operatorname{ind} \boldsymbol{A} = \beta(\boldsymbol{A}) - \alpha(\boldsymbol{A})\,.$$

The operator $\boldsymbol{A} \in L(X, Y)$ is called one-to-one if for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ we have $\boldsymbol{A}\boldsymbol{x}_1 \neq \boldsymbol{A}\boldsymbol{x}_2$. Evidently $\boldsymbol{A} \in L(X, Y)$ is one-to-one if and only if $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$ (or equivalently $\alpha(\boldsymbol{A}) = 0$).

The inverse operator $\boldsymbol{A}^{-1}$ for $\boldsymbol{A} \in L(X, Y)$ can be defined only if $\boldsymbol{A}$ is one-to-one. By definition $\boldsymbol{A}^{-1}$ is a linear operator from Y to X mapping $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in Y$ to $\boldsymbol{x} \in X$. We have $D(\boldsymbol{A}^{-1}) = R(\boldsymbol{A})$, $R(\boldsymbol{A}^{-1}) = D(\boldsymbol{A}) = X$, $\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in X$, $\boldsymbol{A}(\boldsymbol{A}^{-1}\boldsymbol{y}) = \boldsymbol{y}$ for $\boldsymbol{y} \in R(\boldsymbol{A})$. If $R(\boldsymbol{A}) = Y$ and $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$ (i.e. $\alpha(\boldsymbol{A}) = \beta(\boldsymbol{A}) = 0$) then we can assign to any $\boldsymbol{y} \in Y$ the element $\boldsymbol{A}^{-1}\boldsymbol{y}$ which is the unique solution of the linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\,. \tag{3,1}$$

In this case we have $\boldsymbol{A}^{-1} \in L(Y, X)$. The linear equation (3,1) can be solved in general only for $\boldsymbol{y} \in R(\boldsymbol{A})$.

The linear equation (3,1) for $\boldsymbol{A} \in L(X, Y)$ is called uniquely solvable on $R(\boldsymbol{A})$ if for any $\boldsymbol{y} \in R(\boldsymbol{A})$ there is only one $\boldsymbol{x} \in X$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. The equation (3,1) is uniquely solvable on $R(\boldsymbol{A})$ if and only if $\boldsymbol{A}$ is one-to-one (i.e. $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$).

Let now X, X^+ be linear spaces. Assume that a bilinear form $\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle : X \times X^+ \to R$ is defined on $X \times X^+$ (i.e. $\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{x}^+ \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle + \beta\langle \boldsymbol{y}, \boldsymbol{x}^+ \rangle$, $\langle \boldsymbol{x}, \alpha\boldsymbol{x}^+ + \beta\boldsymbol{y}^+ \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle + \beta\langle \boldsymbol{x}, \boldsymbol{y}^+ \rangle$ for every $\boldsymbol{x}, \boldsymbol{y} \in X$, $\boldsymbol{x}^+, \boldsymbol{y}^+ \in X^+$, $\alpha, \beta \in R$).

3.1. Definition. If X, X^+ are linear spaces, $\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle$ a bilinear form on $X \times X^+$ we say that the spaces X, X^+ form a *dual pair* (X, X^+) (*with respect to the bilinear form* $\langle .,. \rangle$) if

$$\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle = 0 \quad \text{for every} \quad \boldsymbol{x} \in X \quad \text{implies} \quad \boldsymbol{x}^+ = \boldsymbol{0} \in X^+$$

and

$$\langle \boldsymbol{x}, \boldsymbol{x}^+ \rangle = 0 \quad \text{for every} \quad \boldsymbol{x}^+ \in X^+ \quad \text{implies} \quad \boldsymbol{x} = \boldsymbol{0} \in X\,.$$

In Heuser [1], VI.40 the following important statement is proved.

3.2. Theorem. *Let (X, X^+) be a dual pair of linear spaces with respect to the bilinear form $\langle .,. \rangle$ defined on $X \times X^+$. Assume that $\mathbf{A} \in L(X)$ is such an operator that there is an operator $\mathbf{A}^+ \in L(X^+)$ such that*

$$\langle \mathbf{A}\mathbf{x}, \mathbf{x}^+ \rangle = \langle \mathbf{x}, \mathbf{A}^+ \mathbf{x}^+ \rangle$$

for every $\mathbf{x} \in X$, $\mathbf{x}^+ \in X^+$.

If $\operatorname{ind} \mathbf{A} = \operatorname{ind} \mathbf{A}^+ = 0$, *then*

$$\alpha(\mathbf{A}) = \alpha(\mathbf{A}^+) = \beta(\mathbf{A}) = \beta(\mathbf{A}^+) < \infty$$

and moreover

$\mathbf{A}\mathbf{x} = \mathbf{y}$ *has a solution if and only if* $\langle \mathbf{y}, \mathbf{x}^+ \rangle = 0$ *for all* $\mathbf{x}^+ \in N(\mathbf{A}^+)$,

$\mathbf{A}^+\mathbf{x}^+ = \mathbf{y}^+$ *has a solution if and only if* $\langle \mathbf{x}, \mathbf{y}^+ \rangle = 0$ *for all* $\mathbf{x} \in N(\mathbf{A})$.

In the following we assume that X and Y are Banach spaces, i.e. normed linear spaces which are complete with respect to the norm given in X, Y respectively. The norm in a normed linear space X will be denoted by $\|.\|_X$ or simply $\|.\|$ when no misunderstanding may occur.

3.3. Definition. An operator $\mathbf{A} \in L(X, Y)$ is bounded if there exists a constant $M \in R$ such that

$$\|\mathbf{A}\mathbf{x}\| \leq M \|\mathbf{x}\|$$

for all $\mathbf{x} \in X$.

The set of all bounded operators $\mathbf{A} \in L(X, Y)$ $(\mathbf{A} \in L(X))$ will be denoted by $B(X, Y)$ $(B(X))$.

It is well-known that $\mathbf{A} \in B(X, Y)$ if and only if $\mathbf{A}$ is continuous, i.e. for every sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$, $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ we have $\lim_{n\to\infty} \mathbf{A}\mathbf{x}_n = \mathbf{A}\mathbf{x}$.

For $\mathbf{A} \in B(X, Y)$ we define

$$\|\mathbf{A}\|_{B(X,Y)} = \sup_{\|\mathbf{x}\| = 1} \|\mathbf{A}\mathbf{x}\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{3,2}$$

It can be proved that by the relation (3,2) a norm on $B(X, Y)$ is given and that $B(X, Y)$ with this norm is a Banach space (see e.g. Schechter [1], Chap. III.).

3.4. Theorem (Bounded Inverse Theorem). *If $\mathbf{A} \in B(X, Y)$ is such that $R(\mathbf{A}) = Y$ and $N(\mathbf{A}) = \{\mathbf{0}\}$, then $\mathbf{A}^{-1}$ exists and $\mathbf{A}^{-1} \in B(Y, X)$.*
(See Schechter [1], III. Theorem 4.1).

3.5. Definition. We denote $X^* = B(X, R)$, where R is the Banach space of real numbers with the norm given by $\alpha \in R \to |\alpha|$. The elements of X^* are called *linear bounded functionals* on X and X^* is the *dual space* to X. Given $\mathbf{f} \in X^*$, its value at $\mathbf{x} \in X$ is denoted also by

$$\mathbf{f}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{f} \rangle_X .$$

If $f(x) = 0$ for any $x \in X$, f is said to be the zero functional on X and we write $f = 0$.

3.6. Remark. X^* equipped with the norm

$$\|f\|_{X^*} = \sup_{\|x\|_X \le 1} |f(x)| = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|_X} \qquad \text{for} \quad f \in X^*$$

(cf. (3,2)) is a Banach space. Furthermore,

$$x \in X,\ f \in X^* \to \langle x, f \rangle_X$$

is evidently a bilinear form on $X \times X^*$. Clearly, $\langle x, f \rangle_X = 0$ for any $x \in X$ if and only if f is the zero functional on X ($f = 0 \in X^*$). Moreover, it follows from the Hahn-Banach Theorem (see e.g. Schechter [1], II.3.2) that $\langle x, f \rangle_X = 0$ for any $f \in X^*$ if and only if $x = 0$. This means that the spaces X and its dual X^* form a dual pair (X, X^*) with respect to the bilinear form $\langle ., . \rangle_X$.

For some Banach spaces X there exist a Banach space E_X and a bilinear form $[., .]_X$ on $X \times E_X$ such that $f \in X^*$ if and only if there exists $g \in E_X$ such that

$$\langle x, f \rangle_X = [x, g]_X \qquad \text{for any} \quad x \in X .$$

If this correspondence between E_X and X^* is an isometrical isomorphism*), we identify E_X with X^* and put

$$\langle x, g \rangle_X = [x, g]_X .$$

3.7. Definition. Let X, Y be Banach spaces. By $X \times Y$ we denote the space of all couples (x, y), where $x \in X$ and $y \in Y$. Given $(x, y), (u, v) \in X \times Y$ and $\lambda \in R$, we put $(x, y) + (u, v) = (x + u, y + v)$, $\lambda(x, y) = (\lambda x, \lambda y)$ and

$$\|(x, y)\|_{X \times Y} = \|x\|_X + \|y\|_Y .$$

(Clearly, $\|.\|_{X \times Y}$ is a norm on $X \times Y$ and $X \times Y$ equipped with this norm is a Banach space.)

3.8. Lemma. *If (X, X^+) and (Y, Y^+) are dual pairs with respect to the bilinear forms $[., .]_X$ and $[., .]_Y$, respectively, then $(X \times Y, X^+ \times Y^+)$ is a dual pair with respect to the bilinear form*

$$(x, y) \in X \times Y,\ (x^+, y^+) \in X^+ \times Y^+ \to$$
$$[(x, y), (x^+, y^+)]_{X \times Y} = [x, x^+]_X + [y, y^+]_Y .$$

---

*) A linear operator mapping a Banach space X into a Banach space Y is called an isomorphism if it is continuous and has a continuous inverse. An isomorphism Φ: $X \to Y$ is isometrical if $\|\Phi x\|_Y = \|x\|_X$ for any $x \in X$. Banach spaces X, Y are isometrically isomorphic if there exists an isometrical isomorphism mapping X onto Y.

Proof. $[.,.]_{X\times Y}$ is clearly a bilinear form. Furthermore, let us assume that

$$[(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}^+,\boldsymbol{y}^+)]_{X\times Y}=0 \qquad \text{for all} \quad (\boldsymbol{x}^+,\boldsymbol{y}^+)\in X^+\times Y^+ . \tag{3,3}$$

In particular, we have

$$[(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}^+,\boldsymbol{y}^+)]_{X\times Y}=[\boldsymbol{x},\boldsymbol{x}^+]_X=0$$

for all $(\boldsymbol{x}^+,\boldsymbol{y}^+)\in X^+\times Y^+$ with $\boldsymbol{y}^+=\boldsymbol{0}$. Since (X,X^+) is a dual pair this implies $\boldsymbol{x}=\boldsymbol{0}$ and (3,3) reduces to $[\boldsymbol{y},\boldsymbol{y}^+]_Y=0$ for all $\boldsymbol{y}^+\in Y$, i.e. $\boldsymbol{y}=\boldsymbol{0}$. Analogously, we would show that $[(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}^+,\boldsymbol{y}^+)]_{X\times Y}=0$ for all $(\boldsymbol{x},\boldsymbol{y})\in X\times Y$ if and only if $\boldsymbol{x}^+=\boldsymbol{0}$, $\boldsymbol{y}^+=\boldsymbol{0}$.

3.9. Remark. In particular, $(X\times Y)^*=X^*\times Y^*$, where

$$\langle(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{\xi},\boldsymbol{\eta})\rangle_{X\times Y}=\langle\boldsymbol{x},\boldsymbol{\xi}\rangle_X+\langle\boldsymbol{y},\boldsymbol{\eta}\rangle_Y$$

for any $\boldsymbol{x}\in X$, $\boldsymbol{y}\in Y$, $\boldsymbol{\xi}\in X^*$ and $\boldsymbol{\eta}\in Y^*$.

3.10. Examples. (i) It is well-known (cf. Dunford, Schwartz [1]) that $\boldsymbol{A}$ is a linear operator acting from R_n into R_m if and only if there exists a real $m\times n$-matrix $\boldsymbol{B}$ such that $\boldsymbol{A}\colon \boldsymbol{x}\in R_n\to\boldsymbol{B}\boldsymbol{x}\in R_m$. Thus the space of all linear operators acting from R_n into R_m and the space of all real $m\times n$-matrices may be identified. Clearly, $B(R_n,R_m)=L(R_n,R_m)$. In particular, $R_n^*=B(R_n,R)=L(R_n,R)$ is the space of all real row n-vectors, while

$$\langle\boldsymbol{x},\boldsymbol{y}^*\rangle_{R_n}=\boldsymbol{y}^*\boldsymbol{x} \qquad \text{for any} \quad \boldsymbol{y}^*\in R_n^* \quad\text{and}\quad \boldsymbol{x}\in R_n .$$

(ii) Let $-\infty<a<b<+\infty$. The dual space to $C_n[a,b]$ is isometrically isomorphic with the space $NBV_n[a,b]$ of column n-vector valued functions of bounded variation on $[a,b]$ which are right-continuous on (a,b) and vanishes at a. Given $\boldsymbol{y}^*\in NBV_n[a,b]$, the value of the corresponding functional on $\boldsymbol{x}\in C_n[a,b]$ is

$$\langle\boldsymbol{x},\boldsymbol{y}^*\rangle_C=\int_a^b \mathrm{d}[\boldsymbol{y}^*(t)]\,\boldsymbol{x}(t) \tag{3,4}$$

and

$$\|\boldsymbol{y}^*\|_C=\sup_{\|\boldsymbol{x}\|_C=1}|\langle\boldsymbol{x},\boldsymbol{y}^*\rangle_C|=\operatorname{var}_a^b\boldsymbol{y}^*=\|\boldsymbol{y}^*\|_{BV} .$$

(The integral in (3,4) is the usual Riemann-Stieltjes integral.) This result is called the Riesz Representation Theorem (see e.g. Dunford, Schwartz [1], IV.6.3). As a consequence $\boldsymbol{K}\in B(C_n[a,b],R_m)$ if and only if there exists a function $\boldsymbol{K}\colon[a,b]\to L(R_n,R_m)$ of bounded variation on $[a,b]$ and such that

$$\boldsymbol{K}\colon \boldsymbol{x}\in C_n[a,b]\to\int_a^b \mathrm{d}[\boldsymbol{K}(t)]\,\boldsymbol{x}(t)\in R_m .$$

Let us notice that the zero functional on $C_n[a,b]$ corresponds to the function $\mathbf{y}^* \in NBV_n[a,b]$ identically vanishing on $[a,b]$.

(iii) Let $-\infty < a < b < \infty$, $1 \le p < \infty$, $q = p/(p-1)$ if $p > 1$ and $q = \infty$ if $p = 1$. The dual space to $L_n^p[a,b]$ is isometrically isomorphic with $L_n^q[a,b]$ (whose elements are row n-vector valued functions). Given $\mathbf{y}^* \in L_n^q[a,b]$, the value of the corresponding functional on $\mathbf{x} \in L_n^p[a,b]$ is

$$\langle \mathbf{x}, \mathbf{y}^* \rangle_L = \int_a^b \mathbf{y}^*(t)\, \mathbf{x}(t)\, \mathrm{d}t \tag{3,5}$$

and

$$\|\mathbf{y}^*\|_{L^*} = \sup_{\|\mathbf{x}\|_{L^p}=1} |\langle \mathbf{x}, \mathbf{y}^* \rangle_L| = \|\mathbf{y}^*\|_{L^q}$$

(see e.g. Dunford, Schwartz [1], IV.8.1). (The integral in (3,5) is the usual Lebesgue integral.) The zero functional on $L_n^p[a,b]$ corresponds to any function $\mathbf{y}^* \in L_n^q[a,b]$ such that $\mathbf{y}^*(t) = \mathbf{0}$ a.e. on $[a,b]$.

(iv) Any Hilbert space H is isometrically isomorphic with its dual space. If $\mathbf{x}, \mathbf{y} \in H \to (\mathbf{x}, \mathbf{y})_H \in R$ is an inner product on H and $\mathbf{x} \in H \to \|\mathbf{x}\|_H = (\mathbf{x}, \mathbf{x})^{1/2}$ the corresponding norm on H, then given $\mathbf{h} \in H$, the value of the corresponding functional on $\mathbf{x} \in H$ is given by

$$\langle \mathbf{x}, \mathbf{h} \rangle_H = (\mathbf{x}, \mathbf{h})_H$$

and

$$\|\mathbf{h}\|_{H^*} = \sup_{\|\mathbf{x}\|_H = 1} |\langle \mathbf{x}, \mathbf{h} \rangle_H| = \|\mathbf{h}\|_H .$$

If X, Y are Banach spaces and $\mathbf{A} \in B(X, Y)$, then for every $\mathbf{g} \in Y^*$ the mapping $\mathbf{x} \in X \to \langle \mathbf{A}\mathbf{x}, \mathbf{g} \rangle_Y$ is a linear bounded functional on X. (Given $\mathbf{x} \in X$ and $\mathbf{g} \in Y^*$, $|\langle \mathbf{A}\mathbf{x}, \mathbf{g} \rangle_Y| \le \|\mathbf{A}\mathbf{x}\|_Y \|\mathbf{g}\|_{Y^*} \le \|\mathbf{A}\|_{B(X,Y)} \|\mathbf{g}\|_{Y^*} \|\mathbf{x}\|_X$.) Thus there is an element of X^* denoted by $\mathbf{A}^*\mathbf{g}$ such that $\langle \mathbf{A}\mathbf{x}, \mathbf{g} \rangle_Y = \langle \mathbf{x}, \mathbf{A}^*\mathbf{g} \rangle_X$. This leads to the following

3.11. Definition. Given $\mathbf{A} \in B(X, Y)$, the operator $\mathbf{A}^*\colon Y^* \to X^*$ defined by

$$\langle \mathbf{A}\mathbf{x}, \mathbf{g} \rangle_Y = \langle \mathbf{x}, \mathbf{A}^*\mathbf{g} \rangle_X$$

for all $\mathbf{x} \in X$ and $\mathbf{g} \in Y^*$ is called the *adjoint operator to* $\mathbf{A}$.

Let us notice that $\mathbf{A}^* \in B(Y^*, X^*)$ and $\|\mathbf{A}^*\| = \|\mathbf{A}\|$ for any $\mathbf{A} \in B(X, Y)$. (See Schechter [1], III.2.)

3.12. Definition. For a given subset $M \subset X$ we define

$$M^\perp = \{\mathbf{f} \in X^*;\ \langle \mathbf{x}, \mathbf{f} \rangle_X = 0 \quad \text{for all} \quad \mathbf{x} \in M\}$$

and similarly for a subset $N \subset X^*$ we set

$${}^\perp N = \{\mathbf{x} \in X;\ \langle \mathbf{x}, \mathbf{f} \rangle_X = 0 \quad \text{for all} \quad \mathbf{f} \in N\} .$$

3.13. Definition. The operator $\mathbf{A} \in B(X, Y)$ is called *normally solvable* if the equation $\mathbf{Ax} = \mathbf{y}$ has a solution if and only if $\langle \mathbf{y}, \mathbf{f} \rangle_Y = 0$ for all solutions $\mathbf{f} \in Y^*$ of the adjoint equation $\mathbf{A}^*\mathbf{f} = \mathbf{0}$.

(In other words, $\mathbf{A} \in B(X, Y)$ is normally solvable if and only if the condition ${}^{\perp}N(\mathbf{A}^*) = R(\mathbf{A})$ is satisfied.)

3.14. Theorem. *If $\mathbf{A} \in B(X, Y)$, then the following statements are equivalent*

(i) *$R(\mathbf{A})$ is closed in Y,*
(ii) *$R(\mathbf{A}^*)$ is closed in X^*,*
(iii) *$\mathbf{A}$ is normally solvable $(R(\mathbf{A}) = {}^{\perp}N(\mathbf{A}^*))$,*
(iv) *$R(\mathbf{A}^*) = N(\mathbf{A})^{\perp}$.*

(See e.g. Goldberg [1], IV.1.2.)

3.15. Theorem. *Let $\mathbf{A} \in B(X, Y)$ have a closed range $R(\mathbf{A})$ in Y. Then*

$$\alpha(\mathbf{A}^*) = \beta(\mathbf{A}) \quad \text{and} \quad \alpha(\mathbf{A}) = \beta(\mathbf{A}^*) .$$

If ind $\mathbf{A}$ *is defined, then* ind $\mathbf{A}^*$ *is also defined and*

$$\operatorname{ind} \mathbf{A}^* = -\operatorname{ind} \mathbf{A} .$$

(See e.g. Goldberg [1], IV.2.3 or Schechter [1], V.4.)

3.16. Definition. If X, Y are Banach spaces then a linear operator $\mathbf{K} \in L(X, Y)$ is called *compact* (or *completely continuous*) if for every sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$, $\mathbf{x}_n \in X$ such that $\|\mathbf{x}_n\|_X \leq C = \text{const.}$ the sequence $\{\mathbf{Kx}_n\}_{n=1}^{\infty}$ in Y contains a subsequence which converges in Y.

Let the set of all compact operators in $L(X, Y)$ $(L(X))$ be denoted by $K(X, Y)$ $(K(X))$.

The set $K(X, Y) \subset L(X, Y)$ is evidently linear. Moreover every compact operator is bounded, i.e. $K(X, Y) \subset B(X, Y)$. Indeed, if $\mathbf{K} \in K(X, Y) \setminus B(X, Y)$, then there exists a sequence $\{\mathbf{x}_n\} \subset X$, $\|\mathbf{x}_n\|_X \leq C$ such that $\|\mathbf{Kx}_n\| \to \infty$ and the sequence $\{\mathbf{Kx}_n\} \subset Y$ cannot contain a subsequence which would be convergent in Y.

3.17. Theorem. *Suppose that $\mathbf{K} \in B(X, Y)$ and that there exists a sequence $\{\mathbf{K}_n\} \subset K(X, Y)$ such that $\lim_{n \to \infty} \mathbf{K}_n = \mathbf{K}$ in $B(X, Y)$. Then $\mathbf{K} \in K(X, Y)$, i.e. $K(X, Y)$ is a closed linear subspace in $B(X, Y)$.*
(See Schechter [1], IV.3.)

3.18. Proposition. *If X, Y, Z are Banach spaces, $\mathbf{A} \in B(X, Y)$, $\mathbf{K} \in K(Y, Z)$, then $\mathbf{KA} \in K(X, Z)$. Similarly $\mathbf{BL} \in K(X, Z)$ provided $\mathbf{L} \in K(X, Y)$, $\mathbf{B} \in B(Y, Z)$.*
(See Schechter [1], IV.3.)

For the adjoint of a compact operator we have

3.19. Theorem. $\mathbf{K} \in K(X, Y)$ *if and only if* $\mathbf{K}^* \in K(Y^*, X^*)$.
(See Goldberg [1], III.1.11 or Schechter [1], IV.4 for the "only if" part.)

3.20. Theorem. *Let* $\mathbf{K} \in K(X)$ *and let both the identity operator on* X *and the identity operator on* X^* *be denoted by* $\mathbf{I}$. *Then* $\mathbf{I} + \mathbf{K} \in B(X)$, $\mathbf{I} + \mathbf{K}^* \in B(X^*)$ *and*
(i) $R(\mathbf{I} + \mathbf{K})$ *is closed in* X *and* $R(\mathbf{I} + \mathbf{K}^*)$ *is closed in* X^*,
(ii) $\alpha(\mathbf{I} + \mathbf{K}) = \beta(\mathbf{I} + \mathbf{K}) = \alpha(\mathbf{I} + \mathbf{K}^*) = \beta(\mathbf{I} + \mathbf{K}^*) < \infty$.
(*In particular,* $\operatorname{ind}(\mathbf{I} + \mathbf{K}) = \operatorname{ind}(\mathbf{I} + \mathbf{K}^*) = 0$.)
(See Schechter [1], IV.3.)

3.21. Remark. It follows easily from the Bolzano-Weierstrass Theorem that any linear bounded operator with the range in a finite dimensional space is compact. ($B(X, R_n) = K(X, R_n)$ for any Banach space X.) Analogously $B(R_n, Y) = K(R_n, Y)$ for any Banach space Y.

3.22. Definition. Let E_X and E_Y be Banach spaces and let $\mathbf{J}_X \in B(X^*, E_X)$ and $\mathbf{J}_Y \in B(Y^*, E_Y)$ be isometrical isomorphisms of X^* onto E_X and Y^* onto E_Y, respectively. Let $[.,.]_X$ be a bilinear form on $X \times E_X$ such that $\langle \mathbf{x}, \xi \rangle_X = [\mathbf{x}, \mathbf{J}_X \xi]_X$ for any $\mathbf{x} \in X$ and $\xi \in X^*$ and let $[.,.]_Y$ be a bilinear form on $Y \times E_Y$ such that $\langle \mathbf{y}, \eta \rangle_Y = [\mathbf{y}, \mathbf{J}_Y \eta]_Y$ for any $\mathbf{y} \in Y$ and $\eta \in Y^*$. If $\mathbf{A} \in B(X, Y)$ and $\mathbf{B} \in L(E_Y, E_X)$ are such that

$$[\mathbf{A}\mathbf{x}, \varphi]_Y = [\mathbf{x}, \mathbf{B}\varphi]_X \qquad \text{for every} \quad \mathbf{x} \in X \quad \text{and} \quad \varphi \in E_Y,$$

then $\mathbf{B}$ is called a *representation of the adjoint operator to* $\mathbf{A}$.

3.23. Remark. If $\mathbf{A} \in B(X, Y)$ and $\mathbf{B} \in L(E_Y, E_X)$ is a representation of the adjoint operator $\mathbf{A}^* \in B(Y^*, X^*)$ to $\mathbf{A}$, then for any $\mathbf{x} \in X$ and $\varphi \in E_Y$ we have

$$[\mathbf{A}\mathbf{x}, \varphi]_Y = \langle \mathbf{A}\mathbf{x}, \mathbf{J}_Y^{-1}\varphi \rangle_Y = \langle \mathbf{x}, \mathbf{A}^*\mathbf{J}_Y^{-1}\varphi \rangle_X = [\mathbf{x}, \mathbf{J}_X \mathbf{A}^* \mathbf{J}_Y^{-1}\varphi]_X .$$

Thus $\mathbf{B} = \mathbf{J}_X \mathbf{A}^* \mathbf{J}_Y^{-1} \in B(E_Y, E_X)$. It follows easily that if we replace the dual spaces to X and Y respectively by the spaces E_X and E_Y isometrically isomorphic to them and the adjoint operators $\mathbf{A}^*$ and $\mathbf{K}^*$ to $\mathbf{A}$ and $\mathbf{K} \in B(X, Y)$, respectively, by its representations $\mathbf{B}$ and $\mathbf{C} \in B(E_Y, E_X)$ defined in 3.22, then Theorems 3.14, 3.15, 3.19 and 3.20 remain valid. This makes reasonable to use the notation $\mathbf{A}^*$ also for representations of the adjoint operator to $\mathbf{A}$.

In the rest of the section X stands for an inner product space endowed with the inner product $(.,.)_X$ and the corresponding norm $\mathbf{x} \in X \to \|\mathbf{x}\|_X = (\mathbf{x}, \mathbf{x})_X^{1/2}$. Furthermore, Y is a Hilbert space, $(.,.)_Y$ is the inner product defined on Y and $\|\mathbf{y}\|_Y = (\mathbf{y}, \mathbf{y})_Y^{1/2}$ for any $\mathbf{y} \in Y$.

3.24. Definition. Given $\mathbf{A} \in L(X, Y)$ and $\mathbf{y} \in Y$, $\mathbf{u} \in X$ is said to be a *least square solution to* (3,1) if

$$\|\mathbf{Au} - \mathbf{y}\|_Y \le \|\mathbf{Ax} - \mathbf{y}\|_Y \qquad \text{for all} \quad \mathbf{x} \in X\,.$$

3.25. Proposition. *If* $\mathbf{A} \in L(X, Y)$ *and* $\mathbf{u}_0 \in X$ *is such that*

$$(\mathbf{Ax}, \mathbf{Au}_0 - \mathbf{y})_Y = 0 \qquad \textit{for all} \quad \mathbf{x} \in X\,, \tag{3,6}$$

then $\mathbf{u}_0$ *is a least square solution to* (3,1). *Furthermore,* $\mathbf{x} \in X$ *is a least square solution to* (3,1) *if and only if* $\mathbf{x} - \mathbf{u}_0 \in N(\mathbf{A})$.

Proof. Given $\mathbf{x} \in X$, $\mathbf{Ax} - \mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{u}_0) + \mathbf{Au}_0 - \mathbf{y}$ and in virtue of (3,6)

$$\begin{aligned}\|\mathbf{Ax} - \mathbf{y}\|_Y^2 &= \|\mathbf{A}(\mathbf{x} - \mathbf{u}_0)\|_Y^2 + 2(\mathbf{A}(\mathbf{x} - \mathbf{u}_0), \mathbf{Au}_0 - \mathbf{y})_Y + \|\mathbf{Au}_0 - \mathbf{y}\|_Y^2 = \\ &= \|\mathbf{A}(\mathbf{x} - \mathbf{u}_0)\|_Y^2 + \|\mathbf{Au}_0 - \mathbf{y}\|_Y^2 \ge \|\mathbf{Au}_0 - \mathbf{y}\|_Y^2\,.\end{aligned}$$

Thus $\mathbf{u}_0$ is a least square solution to (3,1), while $\|\mathbf{Ax} - \mathbf{y}\|_Y = \|\mathbf{Au}_0 - \mathbf{y}\|_Y$ if and only if $\mathbf{A}(\mathbf{x} - \mathbf{u}_0) = \mathbf{0}$.

3.26. Remark. Let us notice that if $R(\mathbf{A})$ is closed in Y, then the Classical Projection Theorem (cf. e.g. Luenberger [1], p. 51) implies that the equation (3,1) possesses for any $\mathbf{y} \in Y$ a least square solution, while $\mathbf{u}_0 \in X$ is a least square solution to (3,1) if and only if (3,6) holds.

3.27. Definition. Given $\mathbf{A} \in L(X, Y)$ and $\mathbf{y} \in Y$, $\mathbf{u}_0 \in X$ is a best *approximate solution to* (3,1) if it is a least square solution to (3,1) of minimal norm (i.e. $\|\mathbf{u}_0\|_X \le \|\mathbf{u}\|_X$ for any least square solution $\mathbf{u}$ of (3,1)).

3.28. Proposition. *Let* $\mathbf{A} \in L(X, Y)$ *and let* $\mathbf{u}_0 \in X$ *fulfil* (3,6). *If besides it*

$$(\mathbf{v}, \mathbf{u}_0)_X = 0 \qquad \textit{for all} \quad \mathbf{v} \in N(\mathbf{A}) \tag{3,7}$$

holds, then $\mathbf{u}_0$ *is a best approximate solution of* (3,1).

Proof. By 3.25 $\mathbf{u}_0$ is a least square solution to (3,1) and $\mathbf{u} - \mathbf{u}_0 \in N(\mathbf{A})$ for all least square solutions $\mathbf{u}$ of (3,1). Thus assuming (3,7) we have

$$\|\mathbf{u}\|_X^2 = \|\mathbf{u} - \mathbf{u}_0\|_X^2 + 2(\mathbf{u} - \mathbf{u}_0, \mathbf{u}_0)_X + \|\mathbf{u}_0\|_X^2 = \|\mathbf{u} - \mathbf{u}_0\|_X^2 + \|\mathbf{u}_0\|_X^2 \ge \|\mathbf{u}_0\|_X^2$$

for any least square solution $\mathbf{u}$ of (3,1). Let us notice that $\|\mathbf{u}_0\|_X = \|\mathbf{u}_0\|_X$ if and only if $\mathbf{u} = \mathbf{u}_0$.

3.29. Remark. Let $\mathbf{A} \in L(X, Y)$. If $k = \dim N(\mathbf{A}) < \infty$, then applying the Gramm-Schmidt orthogonalization process we may find a orthonormal basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$

in $N(\boldsymbol{A})$, i.e. $(\boldsymbol{x}_i, \boldsymbol{x}_j)_X = 0$ if $i \neq j$ and $(\boldsymbol{x}_i, \boldsymbol{x}_i)_X = 1$. Let us put

$$\boldsymbol{P}\colon \boldsymbol{x} \in X \to \sum_{i=1}^{k} (\boldsymbol{x}, \boldsymbol{x}_i)_X \, \boldsymbol{x}_i .$$

Then $\boldsymbol{P} \in B(X)$, $R(\boldsymbol{P}) = N(\boldsymbol{A})$ and $\boldsymbol{P}^2\boldsymbol{x} = \boldsymbol{P}(\boldsymbol{P}\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$ for every $\boldsymbol{x} \in X$. Moreover,

$$(\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x}, \boldsymbol{v})_X = 0 \quad \text{for all} \quad \boldsymbol{x} \in X \quad \text{and} \quad \boldsymbol{v} \in N(\boldsymbol{A}). \tag{3,8}$$

If $R(\boldsymbol{A})$ is closed in Y, then there exists $\boldsymbol{Q} \in B(Y)$ such that $R(\boldsymbol{Q}) = R(\boldsymbol{A})$, $\boldsymbol{Q}^2 = \boldsymbol{Q}$ and

$$(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y} - \boldsymbol{y})_Y = 0 \quad \text{for all} \quad \boldsymbol{y} \in Y \quad \text{and} \quad \boldsymbol{x} \in X \tag{3,9}$$

(cf. Luenberger [1]). $\boldsymbol{P}$ is said to be a *linear bounded orthogonal projection* of X onto $N(\boldsymbol{A})$ and $\boldsymbol{Q}$ is a linear bounded orthogonal projection of Y onto $R(\boldsymbol{A})$. Let us notice that since

$$R(\boldsymbol{I} - \boldsymbol{P}) = N(\boldsymbol{P}) \quad \text{and} \quad R(\boldsymbol{I} - \boldsymbol{Q}) = N(\boldsymbol{Q}),$$

$R(\boldsymbol{I} - \boldsymbol{P})$ and $R(\boldsymbol{I} - \boldsymbol{Q})$ are closed.

As a restriction $\boldsymbol{A}|_{R(\boldsymbol{I}-\boldsymbol{P})}$ of $\boldsymbol{A}$ onto $R(\boldsymbol{I} - \boldsymbol{P})$ is a one-to-one mapping of $R(\boldsymbol{I} - \boldsymbol{P})$ onto $R(\boldsymbol{A})$, it possesses a linear inverse operator $\boldsymbol{A}^+ \in L(R(\boldsymbol{A}), R(\boldsymbol{I} - \boldsymbol{P}))$, i.e.

$$\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}. \tag{3,10}$$

As obviously $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{Q} = \boldsymbol{Q}$, it follows from (3,9) that $(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{A}^+\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{y})_Y = (\boldsymbol{A}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y} - \boldsymbol{y})_Y = 0$ for every $\boldsymbol{y} \in Y$ and $\boldsymbol{x} \in X$. Hence by 3.25 $\boldsymbol{A}^+\boldsymbol{Q}\boldsymbol{y}$ is for any $\boldsymbol{y} \in Y$ a least square solution of (3,1).

Let us put

$$\boldsymbol{A}^{\#} = (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{A}^+\boldsymbol{Q}. \tag{3,11}$$

Evidently $\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{A}$ and hence $(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{y} - \boldsymbol{y})_Y = (\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{A}^+\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{y})_Y = 0$ for every $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$. Since according to (3,8) $(\boldsymbol{v}, \boldsymbol{A}^{\#}\boldsymbol{y})_X = (\boldsymbol{v}, (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{A}^+\boldsymbol{Q}\boldsymbol{y})_X = 0$ for each $\boldsymbol{v} \in N(\boldsymbol{A})$ and $\boldsymbol{y} \in Y$, it follows from 3.28 that for every $\boldsymbol{y} \in Y$, $\boldsymbol{u}_0 = \boldsymbol{A}^{\#}\boldsymbol{y}$ is a best approximate solution to (3,1). Moreover, it is easy to verify that

$$\boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{A}, \qquad \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{A}^{\#}, \qquad \boldsymbol{A}^{\#}\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}, \qquad \boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{Q}. \tag{3,12}$$

3.30. Remark. If $\boldsymbol{A} \in B(X, Y)$, then the condition (3,6) becomes $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{u}_0 = \boldsymbol{A}^*\boldsymbol{y}$ or denoting $\boldsymbol{u}_0 = \boldsymbol{A}^{\#}\boldsymbol{y}$,

$$\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{A}^*. \tag{3,13}$$

Let us notice that if $R(\boldsymbol{A})$ is closed, then (3,12) implies (3,13). In fact, given $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$, we have by (3,9) $0 = (\boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{A}^*\boldsymbol{y})_X$, i.e. $\boldsymbol{A}^*\boldsymbol{Q} = \boldsymbol{A}^*$. This together with the relation $\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{Q}$ from (3,12) yields $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}^{\#} = \boldsymbol{A}^*\boldsymbol{Q} = \boldsymbol{A}^*$. Finally, as $\boldsymbol{A}^{\#} = \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{A}^{\#}$, $\boldsymbol{A}^{\#} = (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{A}^{\#}$ and hence by (3,8) $(\boldsymbol{v}, \boldsymbol{A}^{\#}\boldsymbol{y})_X = (\boldsymbol{v}, (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{A}^{\#}\boldsymbol{y})_X = 0$ for every $\boldsymbol{v} \in N(\boldsymbol{A})$ and $\boldsymbol{y} \in Y$. It means that (3,12) implies also (3,7).

Given $\mathbf{A} \in L(X, Y)$, any operator $\mathbf{A}^+ \in L(Y, X)$ satisfying (3,10) is called a *generalized inverse operator* to $\mathbf{A}$. If $\mathbf{A} \in B(X, Y)$, then the unique operator $\mathbf{A}^\# \in B(Y, X)$ satisfying (3,12) is called the *principal generalized inverse operator* of $\mathbf{A}$.

3.31. Remark. If $X = R_n$, $Y = R_m$ and $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{A}^\#$ its generalized inverse matrix defined by 2.2, then the vector $\mathbf{u}_0 = \mathbf{A}^\# \mathbf{b} \in R_n$ satisfies the conditions (3,13) and (3,7). In fact, as by 2.7 $\mathbf{v} \in N(\mathbf{A})$ if and only if $\mathbf{v} = (\mathbf{I} - \mathbf{A}^\# \mathbf{A})\,\mathbf{d}$ for some $\mathbf{d} \in R_n$, we have owing to (2,16) $(\mathbf{v}, \mathbf{A}^\# \mathbf{b}) = \mathbf{v}^* \mathbf{A}^\# \mathbf{b} = \mathbf{d}^*(\mathbf{I} - (\mathbf{A}^\# \mathbf{A})^*)\,\mathbf{A}^\# \mathbf{b}$ $= \mathbf{d}^*(\mathbf{A}^\# - \mathbf{A}^\# \mathbf{A} \mathbf{A}^\#)\,\mathbf{b} = 0$. Furthermore, $\mathbf{A}^* \mathbf{A} \mathbf{A}^\# = \mathbf{A}^*$ by (2,17). Thus if R_n and R_m are equipped with the Euclidean norm $|.|_e$, $\mathbf{A}^\# \mathbf{b}$ is for any $\mathbf{b} \in R_m$ a unique best approximate solution of (2,1).

4. Perron-Stieltjes integral

This section contains the definition of the Perron-Stieltjes integral based on the work of J. Kurzweil [1], [2]. Some facts concerning this integral are collected here. These facts are necessary for the subsequent study of equations and problems involving the Perron-Stieltjes integrals.

Let a fixed interval $[a, b]$, $-\infty < a < b < +\infty$ be given. We denote by $\mathscr{S} = \mathscr{S}[a, b]$ the system of sets $S \subset R_2$ having the following property:

for every $\tau \in [a, b]$ there exists such a $\delta = \delta(\tau) > 0$ that $(\tau, t) \in S$ whenever $\tau \in [a, b]$ and $t \in [\tau - \delta(\tau), \tau + \delta(\tau)]$.

Evidently any set $S \in \mathscr{S}[a, b]$ is characterized by a real function $\delta\colon [a, b] \to (0, +\infty)$.

Let $f\colon [a, b] \to R$ and $g\colon [a, b] \to R$ be real functions, $-\infty < \alpha < a < b < \beta < +\infty$. If $g(t)$ is defined only for $t \in [a, b]$ then we assume automatically that $g(t) = g(a)$ for $t < a$ and $g(t) = g(b)$ for $t > b$. It is evident that if $\operatorname{var}_a^b g < \infty$, this arrangement yields $\operatorname{var}_\alpha^\beta g \doteq \operatorname{var}_a^b g$ for any α, β such that $\alpha < a < b < \beta$.

4.1. Definition. A real valued finite function $M\colon [a, b] \to R$ is a *major function of f with respect to g* if there exists such a set $S \in \mathscr{S}[a, b]$ that

$$(\tau - \tau_0)(M(\tau) - M(\tau_0)) \geq (\tau - \tau_0)\, f(\tau_0)\,(g(\tau) - g(\tau_0))$$

for $(\tau_0, \tau) \in S$. The set of major functions of f with respect to g is denoted by $M(f, g)$.

A function $m\colon [a, b] \to R$ *is a minor function of f with respect to g* if $-m \in M(-f, g)$, i.e. if $-m$ is a major function of $-f$ with respect to g. The set of minor functions of f with respect to g is denoted by $m(f, g)$.

4.2. Definition. Let $M(f, g) \neq \emptyset$ and $m(f, g) \neq \emptyset$. The lower bound of the numbers $M(b) - M(a)$ where $M \in M(f, g)$ is called the *upper Perron-Stieltjes integral of f with respect to g from a to b* and is denoted by $\overline{\int_a^b} f\, \mathrm{d}g$. Similarly the upper bound of the numbers $m(b) - m(a)$, $m \in m(f, g)$ is called the *lower Perron-Stieltjes integral of f with respect to g from a* to b and is denoted by $\underline{\int}_a^b f \, dg$.

4.3. Lemma. *If $M(f, g) \neq \emptyset$ and $m(f, g) \neq \emptyset$, then*

$$\underline{\int_a^b} f \, dg \leq \overline{\int_a^b} f \, dg .$$

For the proof of this lemma see Kurzweil [1], Lemma 1,1,1.

4.4. Definition. If $M(f, g) \neq \emptyset$, $m(f, g) \neq \emptyset$ and the equality

$$\underline{\int_a^b} f \, dg = \overline{\int_a^b} f \, dg$$

holds, then by the relation

$$\int_a^b f \, dg = \underline{\int_a^b} f \, dg$$

the *Perron-Stieltjes integral* $\int_a^b f \, dg$ *of the function f with respect to g from a* to b is defined. In this case f is called *integrable with respect to g* on $[a, b]$. If $a = b$, then we set $\int_a^b f \, dg = 0$ and if $b < a$, then we put $\int_a^b f \, dg = -\int_b^a f \, dg$.

Now we give a different definition of the Stieltjes integral which is also included in the paper Kurzweil [1] and is equivalent to Definition 4.4. This is a definition of the integral using integral sums which is close to the Riemann-Stieltjes definition.

For the given bounded interval $[a, b] \subset R$ we consider sequences of numbers $A = \{\alpha_0, \tau_1, \alpha_1, \ldots, \tau_k, \alpha_k\}$ such that

$$a = \alpha_0 < \alpha_1 < \ldots < \alpha_k = b , \tag{4,1}$$

$$\alpha_{j-1} \leq \tau_j \leq \alpha_j , \qquad j = 1, 2, \ldots, k . \tag{4,2}$$

For a given set $S \in \mathcal{S}[a, b]$, A satisfying (4,1) and (4,2) is called a subdivision of $[a, b]$ subordinate to S if

$$(\tau_j, t) \in S \qquad \text{for} \quad t \in [\alpha_{j-1}, \alpha_j] , \quad j = 1, 2, \ldots, k . \tag{4,3}$$

The set of all subdivisions A of the interval $[a, b]$ subordinate to S is denoted by $A(S)$.

In Kurzweil [1], Lemma 1.1.1 it is proved that for every $S \in \mathcal{S}[a, b]$ we have

$$A(S) \neq \emptyset . \tag{4,4}$$

If now the real functions f: $[a, b] \to R$, g: $[a, b] \to R$ are given and $A = \{\alpha_0, \tau_1, \alpha_1, \ldots, \tau_k, \alpha_k\}$ is a subdivision of $[a, b]$ which satisfies (4,1) and (4,2), we put

$$B_{f,g}(A) = \sum_{j=1}^{k} f(\tau_j) \left(g(\alpha_j) - g(\alpha_{j-1})\right) . \tag{4,5}$$

If no misunderstanding may occur, we write simply $B(A)$ instead of $B_{f,g}(A)$.

4.5. Definition. Let $f\colon [a,b] \to R$ and $g\colon [a,b] \to R$. If there is a real number J such that to every $\varepsilon > 0$ there exists a set $S \in \mathscr{S}[a,b]$ such that

$$|B_{f,g}(A) - J| < \varepsilon \qquad \text{for any} \quad A \in A(S),$$

we define this number to be the *integral*

$$\int_a^b f \, \mathrm{d}g$$

of the function f with respect to g from a to b.

The completeness of the space R of all real numbers implies that *the integral $\int_a^b f \, \mathrm{d}g$ exists if and only if for any $\varepsilon > 0$ there exists a set $S \in \mathscr{S}[a,b]$ such that*

$$|B_{f,g}(A_1) - B_{f,g}(A_2)| < \varepsilon \qquad \textit{for all} \quad A_1, A_2 \in A(S).$$

In Kurzweil [1] (Theorem 1.2.1), the following statement is proved.

4.6. Theorem. *The integral $\int_a^b f \, \mathrm{d}g$ exists in the sense of Definition 4.4 if and only if $\int_a^b f \, \mathrm{d}g$ exists in the sense of Definition 4.5. If these integrals exist, then their values are equal.*

4.7. Remark. In Schwabik [3] it is shown that the integral introduced in 4.4 and 4.5 is equivalent to the usual Perron-Stieltjes integral defined e.g. in Saks [1]. Consequently the Riemann-Stieltjes, Lebesgue and Perron integrals are special cases of our integral. In particular, if one of the functions f, g is continuous and the other one is of bounded variation on $[a,b]$, then the integral $\int_a^b f \, \mathrm{d}g$ exists and is equal to the ordinary Riemann-Stieltjes integral of f with respect to g from a to b.

The σ-Young integral described in Hildebrandt [1] (II.19.3) is not included in the Perron-Stieltjes integral (see Example 2,1 in Schwabik [3]). However, if $f\colon [a,b] \to R$ is bounded and $g \in BV[a,b]$, then the existence of the σ-Young integral $Y\int_a^b f \, \mathrm{d}g$ implies the existence of the Perron-Stieltjes integral $\int_a^b f \, \mathrm{d}g$ and both integrals are then equal to one another (Schwabik [3], Theorem 3,2).

Now we give a survey of some fundamental properties of the Perron-Stieltjes integral. The proofs of Theorems 4.8 and 4.9 follow directly from Definition 4.5.

4.8. Theorem. *If $f\colon [a,b] \to R$, $g\colon [a,b] \to R$, $\lambda \in R$ and the integral $\int_a^b f \, \mathrm{d}g$ exists, then the integrals $\int_a^b \lambda f \, \mathrm{d}g$ and $\int_a^b f \, \mathrm{d}[\lambda g]$ exist and*

$$\int_a^b \lambda f \, \mathrm{d}g = \lambda \int_a^b f \, \mathrm{d}g, \qquad \int_a^b f \, \mathrm{d}[\lambda g] = \lambda \int_a^b f \, \mathrm{d}g.$$

4.9. Theorem. *If $f_1, f_2\colon [a,b] \to R$, $g\colon [a,b] \to R$ and the integrals $\int_a^b f_1 \, \mathrm{d}g$ and*

$\int_a^b f_2 \, \mathrm{d}g$ *exist, then the integral* $\int_a^b (f_1 + f_2) \, \mathrm{d}g$ *exists and*

$$\int_a^b (f_1 + f_2) \, \mathrm{d}g = \int_a^b f_1 \, \mathrm{d}g + \int_a^b f_2 \, \mathrm{d}g \, .$$

If $f\colon [a,b] \to R$, $g_1, g_2\colon [a,b] \to R$ *and the integrals* $\int_a^b f \, \mathrm{d}g_1$ *and* $\int_a^b f \, \mathrm{d}g_2$ *exist, then the integral* $\int_a^b f \, \mathrm{d}[g_1 + g_2]$ *exists and*

$$\int_a^b f \, \mathrm{d}[g_1 + g_2] = \int_a^b f \, \mathrm{d}g_1 + \int_a^b f \, \mathrm{d}g_2 \, .$$

4.10. Theorem. *If* $f\colon [a,b] \to R$, $g\colon [a,b] \to R$ *and* $\int_a^b f \, \mathrm{d}g$ *exists, then for any* $c, d \in R$, $a \le c < d \le b$ *the integral* $\int_c^d f \, \mathrm{d}g$ *exists.*

4.11. Theorem. *If* $f\colon [a,b] \to R$, $g\colon [a,b] \to R$, $c \in [a,b]$ *and the integrals* $\int_a^c f \, \mathrm{d}g$, $\int_c^b f \, \mathrm{d}g$ *exist, then also the integral* $\int_a^b f \, \mathrm{d}g$ *exists and the equality*

$$\int_a^b f \, \mathrm{d}g = \int_a^c f \, \mathrm{d}g + \int_c^b f \, \mathrm{d}g$$

holds.

The statement 4.10 can be proved easily if 4.6 is taken into account. The proof of 4.11 is given in Kurzweil [1] (Theorem 1.3.4).

4.12. Theorem. *Let* $f\colon [a,b] \to R$, $g\colon [a,b] \to R$ *be given and let the integral* $\int_a^b f \, \mathrm{d}g$ *exist. If* $c \in [a,b]$, *then*

$$\lim_{\substack{t \to c \\ t \in [a,b]}} \left[\int_a^t f \, \mathrm{d}g - f(c)\,(g(t) - g(c)) \right] = \int_a^c f \, \mathrm{d}g \, .$$

(See Kurzweil [1], Theorem 1.3.5.)

4.13. Corollary. *If the assumptions of* 4.12 *are satisfied, then*

$$\lim_{\substack{t \to c \\ t \in [a,b]}} \int_a^t f \, \mathrm{d}g = \int_a^c f \, \mathrm{d}g$$

if and only if $\lim_{\substack{t \to c \\ t \in [a,b]}} g(t) = g(c)$ *or* $f(c) = 0$.

If $g\colon [a,b] \to R$ *possesses the onesided limits* $g(c+)$, $g(c-)$ *at* $c \in [a,b]$ (*e.g. if* $g \in BV[a,b]$), *then*

$$\lim_{\substack{t \to c+ \\ t \in [a,b]}} \int_a^t f \, \mathrm{d}g = \int_a^c f \, \mathrm{d}g + f(c)\,(g(c+) - g(c)) = \int_a^c f \, \mathrm{d}g + f(c)\,\Delta^+ g(c) \tag{4,6}$$

and

$$(4,7)\qquad \lim_{\substack{t\to c-\\ t\in[a,b]}} \int_a^t f\,\mathrm{d}g = \int_a^c f\,\mathrm{d}g - f(c)\,(g(c) - g(c-)) = \int_a^c f\,\mathrm{d}g - f(c)\,\Delta^- g(c)$$

$$\text{for} \quad c \in (a, b]$$

where we have used the notation $\Delta^+ g(c) = g(c+) - g(c)$, $\Delta^- g(c) = g(c) - g(c-)$.

4.14. Lemma. *If* $f_i\colon [a,b] \to R$, $i = 1, 2$, $g \in BV[a,b]$ *and* $A = \{\alpha_0, \tau_1, \ldots, \tau_k, \alpha_k\}$ *is an arbitrary subdivision of* $[a,b]$ *satisfying* (4,1) *and* (4,2), *then*

$$(4,8)\qquad |B_{f_1,g}(A) - B_{f_2,g}(A)| \le \sup_{t\in[a,b]} |f_1(t) - f_2(t)| \operatorname{var}_a^b g\,.$$

Proof. Evidently

$$|B_{f_1,g}(A) - B_{f_2,g}(A)| = \left|\sum_{j=1}^{k} (f_1(\tau_j) - f_2(\tau_j))\,(g(\alpha_j) - g(\alpha_{j-1}))\right|$$

$$\le \sum_{j=1}^{k} |f_1(\tau_j) - f_2(\tau_j)|\, |g(\alpha_j) - g(\alpha_{j-1})|$$

$$\le \sup_{t\in[a,b]} |f_1(t) - f_2(t)| \sum_{j=1}^{k} |g(\alpha_j) - g(\alpha_{j-1})|$$

and (4,8) holds.

In the same trivial way the following lemma can be proved.

4.15. Lemma. *Let* $f\colon [a,b] \to R$, $|f(t)| \le M$ *for all* $t \in [a,b]$, $g_i \in BV[a,b]$, $i = 1, 2$. *Then for any subdivision* $A = \{\alpha_0, \tau_1, \ldots, \tau_k, \alpha_k\}$ *of the interval* $[a,b]$ *satisfying* (4,1) *and* (4,2) *we have*

$$(4,9)\qquad |B_{f,g_1}(A) - B_{f,g_2}(A)| \le M \operatorname{var}_a^b (g_1 - g_2)\,.$$

4.16. Lemma. *If* $f\colon [a,b] \to R$, $g \in BV[a,b]$ *and the integral* $\int_a^b f\,\mathrm{d}g$ *exists, then the inequality*

$$\left|\int_a^b f\,\mathrm{d}g\right| \le \sup_{t\in[a,b]} |f(t)| \operatorname{var}_a^b g$$

holds.

Proof. Since the integral $\int_a^b f\,\mathrm{d}g$ exists, for every $\varepsilon > 0$ there exists $S \in \mathscr{S}[a,b]$ such that for any $A \in A(S)$ we have

$$\left|B_{f,g}(A) - \int_a^b f\,\mathrm{d}g\right| < \varepsilon\,.$$

Let us set $f_1(t) = f(t)$, $f_2(t) = 0$ for $t \in [a,b]$. Then by 4.14 we have for any $A \in A(S)$

$$\left|\int_a^b f\,\mathrm{d}g\right| \le \left|B_{f,g}(A) - \int_a^b f\,\mathrm{d}g\right| + |B_{f,g}(A)| < \varepsilon + |B_{f_1,g}(A) - B_{f_2,g}(A)|$$

$$\le \varepsilon + \sup_{t\in[a,b]} |f(t)| \operatorname{var}_a^b g\,.$$

Hence the inequality is proved because $\varepsilon > 0$ is arbitrary.

4.17. Theorem. *If* $f_n\colon [a,b] \to R$, $n = 1,2,\ldots$, $\lim_{n\to\infty} f_n = f$ *uniformly on* $[a,b]$, $g \in BV[a,b]$ *and* $\int_a^b f_n \, \mathrm{d}g$ *exists for all* $n = 1,2,\ldots$, *then the limit* $\lim_{n\to\infty} \int_a^b f_n \, \mathrm{d}g$ *as well as the integral* $\int_a^b f \, \mathrm{d}g$ *exist and the equality*

$$\lim_{n\to\infty} \int_a^b f_n \, \mathrm{d}g = \int_a^b f \, \mathrm{d}g \tag{4,10}$$

holds.

The proof of the existence of the limit $\lim_{n\to\infty} \int_a^b f_n \, \mathrm{d}g$ and of the integral $\int_a^b f \, \mathrm{d}g$ follows from 4.14. The equality of these quantities is an immediate consequence of 4.16.

4.18. Theorem. *Let* $g_n, g \in BV[a,b]$, $n = 1,2,\ldots$ *and* $\lim_{n\to\infty} \mathrm{var}_a^b(g_n - g) = 0$. *Assume that* $f\colon [a,b] \to R$ *is bounded and* $\int_a^b f \, \mathrm{d}g_n$ *exists for all* $n = 1,2,\ldots$. *Then the limit* $\lim_{n\to\infty} \int_a^b f \, \mathrm{d}g_n$ *as well as the integral* $\int_a^b f \, \mathrm{d}g$ *exist and*

$$\lim_{n\to\infty} \int_a^b f \, \mathrm{d}g_n = \int_a^b f \, \mathrm{d}g \,. \tag{4,11}$$

(The proof follows from 4.15; cf. Schwabik [3], Proposition 2,3.)

If $f, g \in BV[a,b]$, then by Hildebrandt [1] (II.19.3.11) the σ-Young integral $Y\int_a^b f \, \mathrm{d}g$ exists. Taking into account the relationship of the σ-Young and the Perron-Stieltjes integrals (cf. 4.7) we obtain immediately the following.

4.19. Theorem. *If* $f, g \in BV[a,b]$, *then the integral* $\int_a^b f \, \mathrm{d}g$ *exists.*

4.20. Remark. For a given $\alpha \in [a,b]$ and for $t \in [a,b]$ we define

$$\psi_\alpha^+(t) = 0 \quad \text{if} \quad t \le \alpha\,, \qquad \psi_\alpha^+(t) = 1 \quad \text{if} \quad \alpha < t \tag{4,12}$$

and

$$\psi_\alpha^-(t) = 0 \quad \text{if} \quad t < \alpha\,, \qquad \psi_\alpha^-(t) = 1 \quad \text{if} \quad \alpha \le t\,. \tag{4,13}$$

The functions ψ_α^+, ψ_α^- are called *simple jump functions.*

A real function $f\colon [a,b] \to R$ is said to be a *finite-step function* on the interval $[a,b]$ if there is a finite sequence $a = d_0 < d_1 < \ldots < d_k = b$ of points in $[a,b]$ such that in every open interval (d_{i-1}, d_i) $(i = 1,2,\ldots,k)$ the function f equals identically a constant $c_i \in R$. Let us put for $t \in [a,b]$ and $i = 1,2,\ldots,k$

$$g_i(t) = c_i(\psi_{d_{i-1}}^+(t) - \psi_{d_i}^-(t)) + f(d_{i-1})(\psi_{d_{i-1}}^-(t) - \psi_{d_{i-1}}^+(t))\,.$$

It is easy to see that $g_i(t) = f(t)$ if $t \in [d_{i-1}, d_i)$ and $g_i(t) = 0$ if $t \in [a,b] \setminus [d_{i-1}, d_i)$.

Hence, for any $t \in [a, b]$ we have

$$f(t) = \sum_{i=1}^{k} g_i(t) + f(b)\,\psi_b^-(t)$$

$$= \sum_{i=1}^{k} c_i(\psi_{d_{i-1}}^+(t) - \psi_{d_i}^-(t)) + f(d_{i-1})\,(\psi_{d_{i-1}}^-(t) - \psi_{d_{i-1}}^+(t)) + f(b)\,\psi_b^-(t),$$

i.e. any finite-step function can be expressed in the form of a finite linear combination of functions of the type ψ_α^+ and ψ_α^-.

Since any function $f\colon [a, b] \to R$ which possesses the onesided limits $f(c+)$ for any $c \in [a, b)$ and $f(c-)$ for any $c \in (a, b]$ can be approximated uniformly on $[a, b]$ by a sequence of finite-step functions (see e.g. assertion 7.3.2.1 (3) in Aumann [1]), it follows from 4.17 that to prove 4.19 it is sufficient to show that the integrals $\int_a^b \psi_\alpha^+ \,\mathrm{d}g$ and $\int_a^b \psi_\alpha^- \,\mathrm{d}g$ exist for any $g \in BV[a, b]$ and any $\alpha \in [a, b]$.

4.21. Lemma. *Let $\alpha \in [a, b]$ and let $\psi_\alpha^+\colon [a, b] \to R$ and $\psi_\alpha^-\colon [a, b] \to R$ be the simple jump functions defined by (4,12) and (4,13) in 4.20.*

(a) *The integrals $\int_a^b g \,\mathrm{d}\psi_\alpha^+$ and $\int_a^b g \,\mathrm{d}\psi_\alpha^-$ exist for an arbitrary function $g\colon [a, b] \to R$ and*

$$\int_a^b g \,\mathrm{d}\psi_\alpha^+ = \begin{cases} g(\alpha) & \textit{if} \ \ \alpha < b\,, \\ 0 & \textit{if} \ \ \alpha = b\,, \end{cases} \tag{4,14}$$

$$\int_a^b g \,\mathrm{d}\psi_\alpha^- = \begin{cases} g(\alpha) & \textit{if} \ \ \alpha > a\,, \\ 0 & \textit{if} \ \ \alpha = a\,. \end{cases} \tag{4,15}$$

(b) *If $f \in BV[a, b]$ then the integrals $\int_a^b \psi_\alpha^+ \,\mathrm{d}f$, $\int_a^b \psi_\alpha^- \,\mathrm{d}f$ exist and*

$$\int_a^b \psi_\alpha^+ \,\mathrm{d}f = \begin{cases} f(b) - f(\alpha+) & \textit{if} \ \ \alpha < b\,, \\ 0 & \textit{if} \ \ \alpha = b\,, \end{cases} \tag{4,16}$$

$$\int_a^b \psi_\alpha^- \,\mathrm{d}f = \begin{cases} f(b) - f(\alpha-) & \textit{if} \ \ a < \alpha\,, \\ 0 & \textit{if} \ \ a = \alpha\,. \end{cases} \tag{4,17}$$

Proof. (a) If $\alpha = b$ then by definition $\psi_\alpha^+(t) = 0$ for every $t \in [a, b]$ and for any subdivision A: $a = \alpha_0 \le \tau_1 \le \alpha_1 \le \ldots \le \tau_k \le \alpha_k = b$ we have $B_{g,\psi_\alpha^+}(A) = 0$. Hence $\int_a^b g \,\mathrm{d}\psi_\alpha^+ = 0$. If $\alpha < b$ let us define $\delta(t) = \frac{1}{4}|t - \alpha|$ for $t \in [a, b]$, $t \neq \alpha$, $\delta(\alpha) = 1$. Evidently $\delta\colon [a, b] \to (0, +\infty)$. We define

$$S = \{(\tau, t) \in R_2;\ \tau \in [a, b],\ t \in [\tau - \delta(\tau),\ \tau + \delta(\tau)]\}\,,$$

by definition we have $S \in \mathscr{S}[a, b]$. For every subdivision $A \in A(S)$ we have $[\alpha_{j-1}, \alpha_j] \subset [\tau_j - \delta(\tau_j), \tau_j + \delta(\tau_j)]$, i.e.

$$0 < \alpha_j - \alpha_{j-1} \le 2\delta(\tau_j)$$

for any $j = 1, 2, \ldots, k$ (see (4,1), (4,2)). Moreover, there exists an index i, $1 \le i \le k$, such that $\alpha \in [\alpha_{i-1}, \alpha_i)$. If $\tau_i \neq \alpha$ then we obtain a contradictory inequality

$$0 < \alpha_i - \alpha_{i-1} \leq 2\delta(\tau_j) = \tfrac{1}{2}|\tau_i - \alpha| \leq \tfrac{1}{2}(\alpha_i - \alpha_{i-1}).$$

Hence $\tau_i = \alpha$. For every subdivision $A \in A(S)$ we have

$$B_{g,\psi_\alpha^+}(A) = \sum_{j=1}^{k} g(\tau_j)\,[\psi_\alpha^+(\alpha_j) - \psi_\alpha^+(\alpha_{j-1})] = g(\tau_i)\,[\psi_\alpha^+(\alpha_i) - \psi_\alpha^+(\alpha_{i-1})] = g(\tau_i) = g(\alpha).$$

Hence the integral $\int_a^b g\,\mathrm{d}\psi_\alpha^+$ exists and equals $g(\alpha)$ by Definition 4.5. The result for the integral $\int_a^b g\,\mathrm{d}\psi_\alpha^-$ can be proved similarly.

(b) The existence of the integrals $\int_a^b \psi_\alpha^+\,\mathrm{d}f$, $\int_a^b \psi_\alpha^-\,\mathrm{d}f$ follows imeediately from 4.19. It is not difficult to compute their values using 4.11 and 4.13. See also Schwabik [2], Proposition 2.1.

4.22. Lemma. *For* $\alpha \in [a, b]$ *define* $\psi_\alpha(t) = 0$ *if* $t \in [a, b]$, $t \neq \alpha$, $\psi_\alpha(\alpha) = 1$. *Then for any* $g \in BV[a, b]$ *the integrals* $\int_a^b \psi_\alpha\,\mathrm{d}g$, $\int_a^b g\,\mathrm{d}\psi_\alpha$ *exist and*

$$\int_a^b \psi_\alpha\,\mathrm{d}g = g(\alpha+) - g(\alpha-) = \Delta g(\alpha), \tag{4,18}$$

(recall that $g(a-) = g(a)$ *and* $g(b+) = g(b)$*)*,

$$\int_a^b g\,\mathrm{d}\psi_\alpha = 0 \qquad \textit{if} \quad \alpha \in (a, b), \tag{4,19}$$

$$\int_a^b g\,\mathrm{d}\psi_a = -g(a), \qquad \int_a^b g\,\mathrm{d}\psi_b = g(b).$$

Proof. It is easy to see that $\psi_\alpha(t) = \psi_\alpha^-(t) - \psi_\alpha^+(t)$ where ψ_α^+, ψ_α^- are given by (4,12) (4,13) respectively. The existence of the integrals is clear by 4.19, the relations (4,18), (4,19) follow immediately from 4.21.

4.23. Lemma. *Let* $g_B \in BV[a, b]$ *be a break function,* $f \in BV[a, b]$. *Then the integral* $\int_a^b f\,\mathrm{d}g_B$ *exists and*

$$\int_a^b f\,\mathrm{d}g_B = f(a)\,\Delta^+ g_B(a) + \sum_{a<\tau<b} f(\tau)\,\Delta g_B(\tau) + f(b)\,\Delta^- g_B(b)$$

where $\Delta^+ g_B(t) = g_B(t+) - g_B(t)$, $\Delta^- g_B(t) = g_B(t) - g_B(t-)$, $\Delta g_B(t) = g_B(t+) - g_B(t-)$.

Proof. Since g_B is a break function, there exists an at most countable set $(t_1, t_2, \ldots)$ of points in $[a, b]$ and two sequences c_i^+, c_i^-, $i = 1, 2, \ldots$ such that

$$g_B(t) = \sum_{a \leq t_i < t} c_i^+ + \sum_{a < t_i \leq t} c_i^-$$

where $\operatorname{var}_a^b g_B = \sum_{a<t_i\leq b} |c_i^-| + \sum_{a\leq t_i<b} |c_i^+| < +\infty$. By definition it is $c_i^+ = \Delta^+ g_B(t_i)$,

$c_i^- = \Delta^- g_B(t_i)$. Using the functions ψ_α^+, ψ_α^- defined by (4,12), (4,13) we can write

$$g_B(t) = \sum_{i=1}^{\infty} [c_i^+ \psi_{t_i}^+(t) + c_i^- \psi_{t_i}^-(t)]$$

$$= \sum_{i=1}^{\infty} [\Delta^+ g_B(t_i) \psi_{t_i}^+(t) + \Delta^- g_B(t_i) \psi_{t_i}^-(t)] .$$

Let us define

$$g_B^N(t) = \sum_{i=1}^{N} [\Delta^+ g_B(t_i) \psi_{t_i}^+(t) + \Delta^- g_B(t_i) \psi_{t_i}^-(t)] ,$$

we have

$$\operatorname{var}_a^b (g_B - g_B^N) = \operatorname{var}_a^b \left(\sum_{i=N+1}^{\infty} [\Delta^+ g_B(t_i) \psi_{t_i}^+(t) + \Delta^- g_B(t_i) \psi_{t_i}^-(t)] \right) =$$

$$= \sum_{i=N+1}^{\infty} [|\Delta^+ g_B(t_i)| + |\Delta^- g_B(t_i)|] .$$

This yields

$$\lim_{N \to \infty} \operatorname{var}_a^b (g_B - g_B^N) = 0$$

since the series $\sum_{i=1}^{\infty} [|\Delta^+ g_B(t_i)| + |\Delta^- g_B(t_i)|] = \operatorname{var}_a^b g_B$ converges by the asumption.

Evaluating $\int_a^b f \, dg_B^N$ we obtain by the results of 4.21

$$\int_a^b f \, dg_B^N = \sum_{i=1}^{N} \left[\Delta^+ g_B(t_i) \int_a^b f \, d\psi_{t_i}^+ + \Delta^- g_B(t_i) \int_a^b f \, d\psi_{t_i}^- \right] =$$

$$= \sum_{i=1}^{N} [\Delta^+ g_B(t_i) f(t_i) + \Delta^- g_B(t_i) f(t_i)] .$$

Recall that we assume $g(a-) = g(a)$, $g(b) = g(b+)$. By 4.18 we have

$$\int_a^b f \, dg_B = \lim_{N \to \infty} \int_a^b f \, dg_B^N = \sum_{i=1}^{\infty} (\Delta^+ g_B(t_i) + \Delta^- g_B(t_i)) f(t_i)$$

and the proof is complete.

In Hildebrandt [1] (II. 19.3.14) the following result is proved for the Young integrals.

Osgood Convergence Theorem. *If f_n: $[a, b] \to R$, $n = 1, 2, \ldots$ are uniformly bounded on $[a, b]$, i.e. $|f_n(t)| \le M$ for all $t \in [a, b]$ and $n = 1, 2, \ldots$, $g \in BV[a, b]$, $\lim_{n \to \infty} f_n(t) = f(t)$ for all $t \in [a, b]$, and if $Y \int_a^b f_n \, dg$ and $Y \int_a^b f \, dg$ exist, then $\lim_{n \to \infty} Y \int_a^b f_n \, dg = Y \int_a^b f \, dg$.*

In virtue of the relations between the Young integral and the Perron-Stieltjes integral mentioned in 4.7 the following statement can be deduced.

4.24. Theorem. *If $f, g, f_n \in BV[a,b]$, $|f_n(t)| \leq M$ for all $t \in [a,b]$, $n = 1,2,\dots$ and $\lim_{n\to\infty} f_n(t) = f(t)$ for all $t \in [a,b]$, then the integrals $\int_a^b f_n \,\mathrm{d}g$, $\int_a^b f \,\mathrm{d}g$ exist and $\lim_{n\to\infty} \int_a^b f_n \,\mathrm{d}g = \int_a^b f \,\mathrm{d}g$.*

This statement follows from the above quoted Osgood Convergence Theorem in the following way: Since all functions in question belong to $BV[a,b]$, the integrals $\int_a^b f_n \,\mathrm{d}g$, $\int_a^b f \,\mathrm{d}g$, $Y\int_a^b f_n \,\mathrm{d}g$ and $Y\int_a^b f \,\mathrm{d}g$ exist and $\int_a^b f_n \,\mathrm{d}g = Y\int_a^b f_n \,\mathrm{d}g$, $\int_a^b f \,\mathrm{d}g = Y\int_a^b f \,\mathrm{d}g$ (see 4.7). Hence all the assumptions of the Osgood theorem are satisfied and our statement holds.

4.25. Theorem (Substitution Theorem). *If $h \in BV[a,b]$, $g\colon [a,b] \to R$ and $f\colon [a,b] \to R$, the integral $\int_a^b g \,\mathrm{d}h$ exists and f is bounded on $[a,b]$, then the integral $\int_a^b f(t) \,\mathrm{d}(\int_a^t g(\tau) \,\mathrm{d}h(\tau))$ exists if and only if the integral $\int_a^b f(t) g(t) \,\mathrm{d}h(t)$ exists and in this case the two integrals are equal.*

Proof. Let us show that the following statement holds. If $\int_a^b g \,\mathrm{d}h$ exists then for every $\eta > 0$ there is an $S_1 \in \mathscr{S}[a,b]$ such that for every $A \in A(S_1)$, $A\colon a = \alpha_0 \leq \tau_1 \leq \dots \leq \tau_k \leq \alpha_k = b$ we have

$$\sum_{j=1}^{k} \left| g(\tau_j)\,[h(\alpha_j) - h(\alpha_{j-1})] - \int_{\alpha_{j-1}}^{\alpha_j} g \,\mathrm{d}h \right| < \eta\,. \tag{4,20}$$

Let $\eta > 0$ be given. By definition there exists $S_1 \in \mathscr{S}[a,b]$ such that if $A \in A(S_1)$ then

$$\left| B_{g,h}(A) - \int_a^b g \,\mathrm{d}h \right| = \left| \sum_{j=1}^{k} \left\{ g(\tau_j)\,[h(\alpha_j) - h(\alpha_{j-1})] - \int_{\alpha_{j-1}}^{\alpha_j} g \,\mathrm{d}h \right\} \right| < \frac{\eta}{8}$$

and if also $A' \in A(S_1)$ then

$$|B_{g,h}(A) - B_{g,h}(A')| < \frac{\eta}{4}.$$

Let $A\colon a = \alpha_0 \leq \tau_1 \leq \dots \leq \tau_k \leq \alpha_k = b$, $A \in A(S_1)$ be fixed. Assume that $U_1 = \{j_1, j_2, \dots, j_m\}$, $m \leq l$ is an arbitrary set of integers such that $1 \leq j_1 < j_2 < \dots < j_m \leq k$. Since by 4.10 the integrals $\int_{\alpha_{j_i-1}}^{\alpha_{j_i}} g \,\mathrm{d}h$, $i = 1,2,\dots,m$ exist there is an $S_2 \in \mathscr{S}[a,b]$, $S_2 \subset S_1$ such that for any subdivision A_i of the interval $[\alpha_{j_i-1}, \alpha_{j_i}]$ which is subordinate to S_1 we have

$$\left| B_{g,h}(A_i) - \int_{\alpha_{j_i-1}}^{\alpha_{j_i}} g \,\mathrm{d}h \right| < \frac{\eta}{4m}. \tag{4,21}$$

Let us refine the subdivision A in such a way that for $i = 1,\dots,m$ the points $\alpha_{j_i-1} \leq \tau_{j_i} \leq \alpha_{j_i}$ are replaced by the points of A_i and the points $\alpha_{j-1} \leq \tau_j \leq \alpha_j$, $j \notin U_1$

remain unchanged. Let us denote this refinement by A'; evidently $A' \in A(S_1)$.

We have

$$\left|\sum_{j\in U_1}\left(g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h\right)\right|$$

$$\le \left|\sum_{i=1}^{m}\left(g(\tau_{j_i})\left[h(\alpha_{j_i})-h(\alpha_{j_i-1})\right]-B_{g,h}(A_i)\right)\right| + \left|\sum_{i=1}^{m}\left(B_{g,h}(A_i)-\int_{\alpha_{j_i-1}}^{\alpha_{j_i}} g\,\mathrm{d}h\right)\right|$$

$$= \left|\sum_{j=1}^{k} g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\sum_{\substack{j=1\\ j\notin U_1}}^{m} g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\sum_{i=1}^{m} B_{g,h}(A_i)\right|$$

$$+\sum_{i=1}^{m}\left|B_{g,h}(A_i)-\int_{\alpha_{j_i-1}}^{\alpha_{j_i}} g\,\mathrm{d}h\right| < \left|B_{g,h}(A)-B_{g,h}(A')\right| + m\frac{\eta}{4m} < \frac{\eta}{4}+\frac{\eta}{4}=\frac{\eta}{2}$$

because $A, A' \in A(S_1)$ and (4,21) holds.

Since the set $U_1 \subset \{1,\dots,k\}$ of indices was arbitrary, we obtain that for a given $\eta > 0$ there exists $S_1 \in \mathscr{S}[a,b]$ such that for any $A \in A(S_1)$ and $U_1 \subset \{1,2,\dots,k\}$ the inequality

$$\left|\sum_{j=U_1} g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h\right| < \frac{\eta}{2}$$

holds. Let us set

$$d_j = g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h$$

and assume that U_1 is the set of all $j \in \{1,\dots,k\}$ for which $d_j \ge 0$, $U_2 = \{1,\dots,k\}\setminus U_1$. Then we have

$$\sum_{j=1}^{k}|d_j| = \sum_{j\in U_1} d_j - \sum_{j\in U_2} d_j \le \left|\sum_{j\in U_1} d_j\right| + \left|\sum_{j\in U_2} d_j\right| \le \eta,$$

i.e. (4,20) holds.

Now, let us prove the theorem. Assume that $\varepsilon > 0$ is given. If the integral $\int_a^b fg\,\mathrm{d}h$ exists then by definition there exists $S_1 \in \mathscr{S}[a,b]$ such that for all $A \in A(S_1)$

(i) $$\left|\sum_{j=1}^{k} f(\tau_j)g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\int_a^b fg\,\mathrm{d}h\right| < \frac{\varepsilon}{2}.$$

Since the integral $\int_a^b g\,\mathrm{d}h$ exists, by the above statement there is $S_2 \in \mathscr{S}[a,b]$ such that for any $A \in A(S_2)$ we have

(ii) $$\sum_{j=1}^{k}\left|g(\tau_j)\left[h(\alpha_j)-h(\alpha_{j-1})\right]-\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h\right| < \frac{\varepsilon}{2C}$$

where $C > 0$ is the bound for f, i.e. $|f(t)| \le C$ for all $t \in [a,b]$. If we set $S = S_1 \cap S_2$ then $S \in \mathscr{S}[a,b]$ and for any $A \in A(S)$ the inequalities (i), (ii) are satisfied. Let us set $k(t) = \int_a^t g(\tau)\,\mathrm{d}h(\tau)$, $t \in [a,b]$. Then for $A \in A(S)$ we have by (i) and (ii)

$$\left|B_{f,k}(A) - \int_a^b fg\,\mathrm{d}h\right| = \left|\sum_{j=1}^k f(\tau_j)\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h - \int_a^b fg\,\mathrm{d}h\right|$$

$$\leq \left|\sum_{j=1}^k f(\tau_j)\int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h - f(\tau_j)\,g(\tau_j)\,[h(\alpha_j) - h(\alpha_{j-1})]\right.$$

$$\left. + \sum_{j=1}^k f(\tau_j)\,g(\tau_j)\,[h(\alpha_j) - h(\alpha_{j-1})] - \int_a^b fg\,\mathrm{d}h\right|$$

$$< C\sum_{j=1}^k \left|g(\tau_j)\,[h(\alpha_j) - h(\alpha_{j-1})] - \int_{\alpha_{j-1}}^{\alpha_j} g\,\mathrm{d}h\right| + \frac{\varepsilon}{2} < \frac{C\varepsilon}{2C} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Hence according to Definition 4.5 the integral $\int_a^b f\,\mathrm{d}k = \int_a^b f(t)\,\mathrm{d}(\int_a^t g\,\mathrm{d}h)$ exists and equals $\int_a^b fg\,\mathrm{d}h$. Using the same technique the second implication can be also proved.

4.26. Theorem. *Assume that for the functions $g, h \in BV[a,b]$, $f\colon [a,b] \to R$, $\varphi\colon [a,b] \to R$ the integrals $\int_a^b f\,\mathrm{d}g$, $\int_a^b \varphi\,\mathrm{d}h$ exist. If to every $\tau \in [a,b]$ there is a $\delta^*(\tau) > 0$ such that*

$$|t - \tau|\,|f(\tau)\,(g(t) - g(\tau))| \leq (t - \tau)\,\varphi(\tau)\,(h(t) - h(\tau)) \tag{4,22}$$

holds for every $\tau \in [a,b]$, $t \in [a,b] \cap [\tau - \delta^*(\tau), \tau + \delta^*(\tau)]$, then

$$\left|\int_a^b f\,\mathrm{d}g\right| \leq \int_a^b \varphi\,\mathrm{d}h\,.$$

This statement is proved in Kurzweil [2].

4.27. Corollary. *Assume that $g \in BV[a,b]$. If $f\colon [a,b] \to R$, $|f(t)| \leq M =$ const. for all $t \in [a,b]$ and $\int_a^b f\,\mathrm{d}g$ exists then for every $[c,d] \subset [a,b]$ we have*

$$\left|\int_c^d f\,\mathrm{d}g\right| \leq M\,\mathrm{var}_c^d\,g$$

and consequently $\mathrm{var}_a^b\,(\int_a^t f\,\mathrm{d}g) \leq M\,\mathrm{var}_a^b\,g < \infty$.

If $f \in BV[a,b]$ then $\int_a^b f\,\mathrm{d}g$ exists and

$$\left|\int_a^b f\,\mathrm{d}g\right| \leq \int_a^b |f(t)|\,\mathrm{d}(\mathrm{var}_a^t\,g) \leq \sup_{t\in[a,b]} |f(t)|\,\mathrm{var}_a^b\,g\,.$$

Proof. In the first case we have

$$|t - \tau|\,|f(\tau)\,(g(t) - g(\tau))| \leq (t - \tau)\,|f(\tau)|\,(\mathrm{var}_a^t\,g - \mathrm{var}_a^\tau\,g)$$
$$\leq (t - \tau)\,M(\mathrm{var}_a^t\,g - \mathrm{var}_a^\tau\,g)$$

for every $\tau \in [a,b]$, $t \in [a,b]$. Since the integral $\int_c^d M\,\mathrm{d}(\mathrm{var}_a^t g)$ exists and equals $M\,\mathrm{var}_c^d g$ we obtain the result by 4.26. The second statement can be derived in a similar way, when 4.19 and the fact that $|f| \in BV[a,b]$ are taken into account.

4.28. Theorem. *Let us assume that $g\colon [a,b] \to R$ is nondecreasing, $f_1, f_2\colon [a,b] \to R$, $f_1(t) \le f_2(t)$ for all $t \in [a,b]$ and $\int_a^b f_i\,\mathrm{d}g$ exists for $i = 1, 2$. Then*

$$\int_a^b f_1\,\mathrm{d}g \le \int_a^b f_2\,\mathrm{d}g\,.$$

This statement follows from 4.26.

4.29. Theorem. *If $h\colon [a,b] \to R$ is nonnegative, nondecreasing and continuous from the left in $[a,b]$ (i.e. $h(t-) = h(t)$ for every $t \in (a,b]$), then*

$$\int_a^b h^k(t)\,\mathrm{d}h(t) \le \frac{1}{k+1}\,[h^{k+1}(b) - h^{k+1}(a)] \tag{4,23}$$

for any $k = 0, 1, 2, \ldots$. If $h\colon [a,b] \to R$ is assumed to be nonnegative, nonincreasing and continuous from the right (i.e. $(h(t+) = h(t)$ for every $t \in [a,b))$, then

$$\int_a^b h^k(t)\,\mathrm{d}h(t) \ge \frac{1}{k+1}\,[h^{k+1}(b) - h^{k+1}(a)] \tag{4,24}$$

for any $k = 0, 1, 2, \ldots$

The proof of the first part is given in Kurzweil [2]. The second part can be proved similarly.

4.30. Theorem. *Assume that $g\colon [a,b] \to R$ is a nonnegative nondecreasing function, $\varphi\colon [a,b] \to R$ nonnegative and bounded, i.e. $\varphi(t) \le C =$ const. for all $t \in [a,b]$.*

(a) *If g is continuous from the right on $[a,b)$ and if there exist nonnegative constants K_1, K_2 such that*

$$\varphi(\xi) \le K_1 + K_2 \int_\xi^b \varphi(\tau)\,\mathrm{d}g(\tau) \tag{4,25}$$

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

$$\varphi(\tau) \le K_1 \mathrm{e}^{K_2(g(b) - g(\tau))} \tag{4,26}$$

for any $\tau \in [a,b]$.

(b) *If g is continuous from the left on $(a,b]$ and if there exist nonnegative constants K_1, K_2 such that*

$$\varphi(\xi) \le K_1 + K_2 \int_a^\xi \varphi(\tau)\,\mathrm{d}g(\tau) \tag{4,27}$$

for every $\xi \in [a, b]$, *then*

(4,28) $$\varphi(\tau) \leq K_1 e^{K_2(g(\tau) - g(a))}$$

for any $\tau \in [a, b]$.

Proof. We prove only (a). The statement (b) can be proved in the same way. Let us define

$$w(t) = L e^{K_2(g(b) - g(t))}, \qquad t \in [a, b]$$

where $L \geq 0$ is a constant.

For any $\xi \in [a, b]$ we have

$$L + K_2 \int_\xi^b w(\tau)\, dg(\tau) = L + K_2 L \int_\xi^b e^{K_2(g(b) - g(\tau))}\, dg(\tau)$$

$$= L\left(1 + K_2 \int_\xi^b \sum_{i=1}^{\infty} \frac{K_2}{i!} (g(b) - g(\tau))^i\, dg(\tau)\right).$$

Since the series $\sum_{i=0}^{\infty} K_2(g(b) - g(\tau))^i/i!$ evidently converges uniformly on $[a, b]$, 4.17 ensures that in the last term the integration and summation are interchangeable. Hence by (4,24) from 4.29 we obtain

$$L + K_2 \int_\xi^b w(\tau)\, dg(\tau) = L\left(1 + K_2 \sum_{i=0}^{\infty} \frac{K_2^i}{i!} \int_\xi^b (g(b) - g(\tau))^i\, dg(\tau)\right) =$$

$$= L\left(1 - K_2 \sum_{i=0}^{\infty} \frac{K_2^i}{i!} \int_\xi^b (g(b) - g(\tau))^i\, d(g(b) - g(\tau))\right) \leq$$

$$\leq L\left(1 + \sum_{i=0}^{\infty} \frac{K_2^i}{i!} (g(b) - g(\xi))^i\right) = L e^{K_2(g(b) - g(\xi))}.$$

Let $\varepsilon > 0$ be arbitrary. We set

$$w_\varepsilon(t) = (K_1 + \varepsilon)\, e^{K_2(g(b) - g(t))}, \qquad t \in [a, b].$$

Then

(4,29) $$K_1 + \varepsilon + K_2 \int_\xi^b w_\varepsilon(\tau)\, dg(\tau) \leq w_\varepsilon(\xi), \qquad \xi \in [a, b].$$

For the difference $m_\varepsilon(\xi) = \varphi(\xi) - w_\varepsilon(\xi)$ we have by (4,25), (4,29)

(4,30) $$m_\varepsilon(\xi) \leq -\varepsilon + K_2 \int_\xi^b m_\varepsilon(\tau)\, dg(\tau), \qquad \xi \in [a, b]$$

and, in particular, $m_\varepsilon(b) \leq -\varepsilon < 0$. Moreover, it is easy to see that $|m_\varepsilon(\xi)| \leq C_1 = \text{const.}$ for $\xi \in [a,b]$. By 4.12 we have

$$m_\varepsilon(\xi) \leq -\varepsilon + K_2\, m_\varepsilon(b)\,[g(b) - g(b-)] + \lim_{\delta\to 0+} K_2 \int_\xi^{b-\delta} m_\varepsilon(\tau)\,\mathrm{d}g(\tau)$$
$$\leq -\varepsilon + K_2\, m_\varepsilon(b)\,[g(b) - g(b-)] + C_2[g(b-) - g(\xi)], \qquad C_2 = K_2 C_1 .$$

Since $g \in BV$, there exists $\eta > 0$ such that if $0 \leq b - \xi \leq \eta$ then $C_2(g(b-) - g(\xi)) < \varepsilon/2$. Hence for $\xi \in [b-\eta, b]$ we have $m_\varepsilon(\xi) < 0$. Let us set

$$(4{,}31) \qquad T = \inf\{t \in [a,b];\ m_\varepsilon(\xi) < 0 \quad \text{for} \quad \xi \in [t,b]\} .$$

We have shown that $T < b$ and we have evidently $m_\varepsilon(t) < 0$ for $t \in (T,b]$. Further by (4,30) and 4.12

$$m_\varepsilon(T) \leq -\varepsilon + K_2 \int_T^b m_\varepsilon(\tau)\,\mathrm{d}g(\tau)$$
$$= -\varepsilon + K_2\, m_\varepsilon(T)\,(g(T+) - g(T)) + \lim_{\delta\to 0+} K_2 \int_{T+\delta}^b m_\varepsilon(\tau)\,\mathrm{d}g(\tau) \leq -\varepsilon < 0$$

since $g(T+) - g(T) = 0$ and $\int_{T+\delta}^b m_\varepsilon(\tau)\,\mathrm{d}g(\tau) \leq 0$ for every $\delta > 0$.

If $T > a$ then we repeat the above procedure and show in the same way that there exists an $\eta > 0$ such that $m_\varepsilon(\xi) < 0$ for all $\xi \in [T-\eta, T]$. This contradicts (4,31). Hence $T = a$ and $m_\varepsilon(\xi) < 0$ for all $\xi \in [a,b]$, i.e.

$$\varphi(\xi) < K_1 \mathrm{e}^{K_2(g(b) - g(\xi))} + \varepsilon \mathrm{e}^{K_2(g(b) - g(a))}$$

for all $\xi \in [a,b]$ and $\varepsilon > 0$. This yields (4,26).

4.31. Theorem. *Let $h\colon [a,b] \times [c,d] \to R$ be such that $|h(s,t)| \leq M < \infty$ and $\operatorname{var}_a^b h(.,t) + \operatorname{var}_c^d h(s,.) < \infty$ for every $(t,s) \in [a,b] \times [c,d]$. Then for any $f \in BV[a,b]$ and any $g \in BV[c,d]$ both the iterated integrals*

$$\int_a^b \mathrm{d}f(s)\left(\int_c^d h(s,t)\,\mathrm{d}g(t)\right) \quad \text{and} \quad \int_c^d \left(\int_a^b \mathrm{d}f(s)\, h(s,t)\right)\mathrm{d}g(t)$$

exist and are equal.

(See Hildebrandt [2], p. 356 and [1], II.19.)

4.32. Theorem (Dirichlet formula). *If $h\colon [a,b] \times [a,b] \to R$ is bounded on $[a,b] \times [a,b]$ and $\operatorname{var}_a^b h(s,.) < \infty$ for every $s \in [a,b]$, $\operatorname{var}_a^b h(.,t) < \infty$ for every $t \in [a,b]$, then for any $f, g \in BV[a,b]$ we have*

$$(4,32)\qquad \int_a^b \mathrm{d}g(t)\left(\int_a^t h(s,t)\,\mathrm{d}f(s)\right)$$

$$= \int_a^b \left(\int_s^b \mathrm{d}g(t)\,h(s,t)\right)\mathrm{d}f(s) + \sum_{t\in(a,b]} \Delta^- g(t)\,h(t,t)\,\Delta^- f(t) - \sum_{t\in[a,b)} \Delta^+ g(t)\,h(t,t)\,\Delta^+ f(t)$$

where $\Delta^- g(t) = g(t) - g(t-)$, $\Delta^+ g(t) = g(t+) - g(t)$.

Proof. Let us define $k(s,t) = h(s,t)$ for $a \le s \le t \le b$ and $k(s,t) = 0$ for $a \le t < s \le b$. Then $k\colon [a,b] \times [a,b] \to R$ evidently satisfies the assumptions of 4.31 and this theorem gives

$$(4,33)\qquad \int_a^b \mathrm{d}g(t)\left(\int_a^b k(s,t)\,\mathrm{d}f(s)\right) = \int_a^b \left(\int_a^b \mathrm{d}g(t)\,k(s,t)\right)\mathrm{d}f(s)\,.$$

Moreover for $t \in [a,b)$ it is

$$\int_a^b k(s,t)\,\mathrm{d}f(s) = \int_a^t h(s,t)\,\mathrm{d}f(s) + \int_t^b k(s,t)\,\mathrm{d}f(s)$$

$$= \int_a^t h(s,t)\,\mathrm{d}f(s) + h(t,t)\,\Delta^+ f(t)\,,$$

since from 4.13 and from the definition of $k(s,t)$ we have by (4,6)

$$\int_t^b k(s,t)\,\mathrm{d}f(s) = \lim_{\tau\to t+}\left[\int_\tau^b k(s,t)\,\mathrm{d}f(s) + k(t,t)\,(f(t+) - f(t))\right]$$

$$= k(t,t)\,\Delta^+ f(t) = h(t,t)\,\Delta^+ f(t)\,.$$

If $t = b$, then $\int_a^b k(s,b)\,\mathrm{d}f(s) = \int_a^b h(s,b)\,\mathrm{d}f(s)$. Hence for an arbitrary $t \in [a,b]$ we can write

$$(4,34)\qquad \int_a^b k(s,t)\,\mathrm{d}f(s) = \int_a^t h(s,t)\,\mathrm{d}f(s) + h(t,t)\,\Delta^+ f(t)$$

if we set $\Delta^+ f(b) = 0$.

A similar argument gives

$$(4,35)\qquad \int_a^b \mathrm{d}g(t)\,k(s,t) = \int_s^b \mathrm{d}g(t)\,h(s,t) + \Delta^- g(s)\,h(s,s)$$

for every $s \in [a,b]$ if the convention $\Delta^- g(a) = 0$ is used. Setting (4,34) and (4,35) into (4,33) we obtain

$$(4,36)\qquad \int_a^b \mathrm{d}g(t)\left(\int_a^t h(s,t)\,\mathrm{d}f(s)\right)$$

$$= \int_a^b \mathrm{d}f(s)\left(\int_s^b h(s,t)\,\mathrm{d}g(t)\right) + \int_a^b \Delta^- g(s)\,h(s,s)\,\mathrm{d}f(s) - \int_a^b \mathrm{d}g(t)\,h(t,t)\,\Delta^+ f(t)\,.$$

Since $g \in BV[a,b]$, there is an at most countable set of points $\alpha_1, \alpha_2, \dots$ in $[a,b]$ such that $\Delta^- g(s) = 0$ for all $s \in [a,b]$, $s \neq \alpha_i$ and $\sum_{i=1}^{\infty} |\Delta^- g(\alpha_i)| \le \operatorname{var}_a^b g < +\infty$. Let us set $H(s) = \Delta^- g(s)\, h(s,s)$ for any $s \in [a,b]$. Then $H(s) = 0$ for all $s \in [a,b]$, $s \neq \alpha_i$, $i = 1, 2, \dots$ and

$$\int_a^b \Delta^- g(s)\, h(s,s)\, \mathrm{d}f(s) = \int_a^b H(s)\, \mathrm{d}f(s).$$

Let us define for $N = 1, 2, \dots$ and $s \in [a,b]$

$$H_N(s) = \sum_{i=1}^{N} \Delta^- g(\alpha_i)\, h(\alpha_i, \alpha_i)\, \psi_{\alpha_i}(s)$$

where $\psi_\alpha(s) = 0$ if $s \neq \alpha$ and $\psi_\alpha(\alpha) = 1$.

Evidently $H_N(s) = 0$ for all $s \in [a,b]$, $s \neq \alpha_1, \alpha_2, \dots, \alpha_N$ and $H_N(\alpha_i) = H(\alpha_i)$ for $i = 1, 2, \dots, N$. For $s \in [a,b]$, $s \notin \alpha_1, \alpha_2, \dots, \alpha_N$ we have

$$\begin{aligned}|H_N(s) - H(s)| = |H(s)| \le \sup_{i=N+1,\dots} |H(\alpha_i)| &< \sum_{i=N+1}^{\infty} |\Delta^- g(\alpha_i)\, h(\alpha_i, \alpha_i)| \\ &\le M \sum_{i=N+1}^{\infty} |\Delta^- g(\alpha_i)|\end{aligned}$$

where M is the bound of $|h(s,t)|$.

Since the series $\sum_{i=1}^{\infty} |\Delta^- g(\alpha_i)|$ is convergent, we obtain that for any $\varepsilon > 0$ there is a natural N such that $M \sum_{i=N+1}^{\infty} |\Delta^- g(\alpha_i)| < \varepsilon$ and also

$$|H_N(s) - H(s)| < \varepsilon$$

for all $s \in [a,b]$, i.e. $\lim_{N\to\infty} H_N(s) = H(s)$ uniformly in $[a,b]$. Using (4,18) we conclude

$$\int_a^b H_N(s)\, \mathrm{d}f(s) = \sum_{i=1}^{N} \Delta^- g(\alpha_i)\, h(\alpha_i, \alpha_i)\, \Delta f(\alpha_i)$$

and by 4.17 we obtain

$$\begin{aligned}\int_a^b \Delta^- g(s)\, h(s,s)\, \mathrm{d}f(s) &= \int_a^b H(s)\, \mathrm{d}f(s) = \lim_{N\to\infty} \int_a^b H_N(s)\, \mathrm{d}f(s) = \\ &= \sum_{i=1}^{\infty} \Delta^- g(\alpha_i)\, h(\alpha_i, \alpha_i)\, \Delta f(\alpha_i) = \sum_{s\in(a,b]} \Delta^- g(s)\, h(s,s)\, \Delta f(s).\end{aligned}$$

Similarly it can be proved that

$$\int_a^b \mathrm{d}g(t)\,(h(t,t)\, \Delta^+ f(t)) = \sum_{t\in[a,b)} \Delta g(t)\, h(t,t)\, \Delta^+ f(t).$$

If we set these expressions into (4,36) we obtain

$$\int_a^b \mathrm{d}g(t)\left(\int_a^t h(s,t)\,\mathrm{d}f(s)\right)$$

$$= \int_a^b \left(\int_s^b \mathrm{d}g(t)\,h(s,t)\right)\mathrm{d}f(s) + \sum_{s\in(a,b)} [\Delta^- g(s)\,h(s,s)\,\Delta f(s) - \Delta g(s)\,h(s,s)\,\Delta^+ f(s)]$$
$$+ \Delta^- g(b)\,h(b,b)\,\Delta f(b) - \Delta g(a)\,h(a,a)\,\Delta^+ f(a)$$

and this yields the result.

4.33. Theorem (integration-by-parts). *Let $f, g \in BV[a,b]$; then for any interval $[c,d] \subset [a,b]$ we have*

$$\int_c^d f\,\mathrm{d}g + \int_c^d g\,\mathrm{d}f = f(d)\,g(d) - f(c)\,g(c) - \sum_{c\le\tau<d} \Delta^+ f(\tau)\,\Delta^+ g(\tau) + \sum_{c<\tau\le d} \Delta^- f(\tau)\,\Delta^- g(\tau)$$

where $\Delta^+ f(\tau) = f(\tau+) - f(\tau)$, $\Delta^- f(\tau) = f(\tau) - f(\tau-)$ and similarly for $\Delta^+ g(\tau)$, $\Delta^- g(\tau)$.

Proof. If we set $h(s,t) \equiv 1$ on $[a,b] \times [a,b]$ then for every $f, g \in BV[a,b]$ we have by 4.32

(4,37)
$$\int_c^d \left(\int_c^t \mathrm{d}f(s)\right)\mathrm{d}g(t)$$

$$= \int_c^d \left(\int_s^d \mathrm{d}g(t)\right)\mathrm{d}f(s) + \sum_{t\in(c,d]} \Delta^- g(t)\,\Delta^- f(t) - \sum_{t\in[c,d)} \Delta^+ g(t)\,\Delta^+ f(t)\,.$$

Moreover,

$$\int_c^d \left(\int_c^t \mathrm{d}f(s)\right)\mathrm{d}g(t) = \int_c^d (f(t) - f(c))\,\mathrm{d}g(t) = \int_c^d f(t)\,\mathrm{d}g(t) - f(c)\,(g(d) - g(c))$$

and similarly

$$\int_c^d \left(\int_s^d \mathrm{d}g(t)\right)\mathrm{d}f(s) = -\int_c^d g(t)\,\mathrm{d}f(t) + g(d)\,(f(d) - f(c))\,.$$

Inserting this into (4,37) we obtain the result. (A direct proof of the integration-by-parts theorem 4.33 is given in Kurzweil [3].)

The Lebesgue-Stieltjes integral has been defined and studied in many monographs on integration theory. (See e.g. Saks [1], Hildebrandt [1], Dunford, Schwartz [1] etc.) In the next theorem its relationship with the Perron-Stieltjes integral is cleared up. The proof follows e.g. from Theorem VI (8.1) in Saks [1].

4.34. Theorem. *Let $g \in BV[a,b]$ and $f\colon [a,b] \to R$ be such that the Lebesgue-Stieltjes integral (L-S) $\int_{(a,b)} f\,\mathrm{d}g$ over the open interval (a,b) exists. Then the Perron-*

Stieltjes integral $\int_a^b f\,\mathrm{d}g$ *also exists and*

$$\int_a^b f\,\mathrm{d}g = (L - S)\int_{(a,b)} f\,\mathrm{d}g + f(a)\,\Delta^+ g(a) + f(b)\,\Delta^- g(b)\,.$$

4.35. Remark. If f: $[a,b] \to R$ is bounded, h: $[a,b] \to R$ is Lebesgue integrable on $[a,b]$ $(h \in L^1[a,b])$ and $g(t) = g(a) + \int_a^t h(\tau)\,\mathrm{d}\tau$ on $[a,b]$ $(g \in AC[a,b])$, then in virtue of 4.25 and 4.34

$$\int_a^b f(t)\,\mathrm{d}g(t) = \int_a^b f(t)\,h(t)\,\mathrm{d}t\,,$$

where the right-hand side integral is the Lebesgue one.

For the proof of the following assertion see e.g. Natanson [1] (Corollary of Theorem XII.4.2). It is also included as a special case in the "symmetrical Fubini theorem" for Lebesgue-Stieltjes integrals (cf. Hildebrandt [1], X.3.2).

4.36. Theorem (Tonelli, Hobson). *If* h: $D = [a,b] \times [c,d] \to R$ *is measurable and if any one of the three Lebesgue integrals*

$$\iint_D |h(t,s)|\,\mathrm{d}t\,\mathrm{d}s\,, \qquad \int_a^b\left(\int_c^d |h(t,s)|\,\mathrm{d}s\right)\mathrm{d}t\,, \qquad \int_c^d\left(\int_a^b |h(t,s)|\,\mathrm{d}t\right)\mathrm{d}s$$

exists, then the Lebesgue integrals

$$\iint_D h(t,s)\,\mathrm{d}t\,\mathrm{d}s\,, \qquad \int_a^b\left(\int_c^d h(t,s)\,\mathrm{d}s\right)\mathrm{d}t\,, \qquad \int_c^d\left(\int_a^b h(t,s)\,\mathrm{d}t\right)\mathrm{d}s$$

all exist and are equal to one another.

One of the most helpful tools for the investigation of integro-differential and functional-differential equations is the "*unsymmetrical Fubini theorem*" 4.38. For its proof the following lemma is needed.

4.37. Lemma. *Let* h: $[a,b] \times [c,d] \to R$ *be such that* $h(.,s)$ *is measurable on* $[a,b]$ *for any* $s \in [c,d]$, $\chi(t) = |h(t,c)| + \mathrm{var}_c^d\, h(t,.) < \infty$ *for a.e.* $t \in [a,b]$ *and* $\chi \in L^p[a,b]$, $1 \le p < \infty$. *Then*

(a) *given* $f \in L^q[a,b]$ *with* $q = p/(p-1)$ *if* $p > 1$ *and* $q = \infty$ *if* $p = 1$, *the function*

$$\varphi\colon\ s \in [c,d] \to \int_a^b f(t)\,h(t,s)\,\mathrm{d}s$$

is defined for any $s \in [c,d]$, *belongs to* $BV[c,d]$ *and*

$$\varphi(s+) = \int_a^b f(t)\,h(t,s+)\,\mathrm{d}t \qquad \textit{for any}\quad s \in [c,d)\,, \tag{4,38}$$

$$\varphi(s-) = \int_a^b f(t)\,h(t,s-)\,\mathrm{d}t \qquad \textit{for any}\quad s \in (c,d]\,;$$

(b) *given* $g \in C[c,d]$ (*or* $g \in BV[c,d]$), *the function*

$$\eta\colon\ t \in [a,b] \to \int_c^d \mathrm{d}_s[h(t,s)]\, g(s)$$

is defined a.e. on $[a,b]$ *and belongs to* $L^p[a,b]$.

Proof. Clearly, $\varphi(s)$ is defined for any $s \in [c,d]$. For an arbitrary subdivision $c = s_0 < s_1 < \ldots < s_k = d$ of $[c,d]$ we have

$$\sum_{j=1}^{k} |\varphi(s_j) - \varphi(s_{j-1})| \le \int_a^b |f(t)| \sum_{j=1}^{k} |h(t,s_j) - h(t,s_{j-1})|\, \mathrm{d}t$$
$$\le \int_a^b |f(t)|\, \chi(t)\, \mathrm{d}t < \infty\,,$$

i.e. $\varphi \in BV[c,d]$. Furthermore,

$$|f(t)\, h(t,\sigma)| \le |f(t)|\, \chi(t) \qquad \text{for a.e. } \ t \in [a,b] \ \text{ and any } \ \sigma \in [c,d]\,.$$

Applying the Lebesgue Dominated Convergence Theorem we obtain immediately (4,38).

(b) Under our assumptions $\eta(t)$ is defined a.e. on $[a,b]$. If $g\colon [c,d] \to R$ is a finite step function with jumps at $s_j \in [c,d]$ $(j = 1,2,\ldots,k)$ (cf. 4.20), then according to 4.21 $\eta(t)$ is a.e. on $[a,b]$ equal to a linear combination of the values $h(t,b)$, $h(t,a)$, $h(t,s_j+)$ and $h(t,s_j-)$ $(j = 1,2,\ldots,k)$. In particular, in this case η is measurable on $[a,b]$. Making use of the fact that any function g which is continuous on $[a,b]$ or of bounded variation on $[a,b]$ can be approximated uniformly on $[a,b]$ by finite step functions (Aumann [1]) and applying 4.17 we complete the proof of the measurability of η on $[a,b]$. By 4.16

$$|\eta(t)| \le \chi(t)\,(\sup_{s\in[c,d]} |g(s)|) \qquad \text{a.e. on} \quad [a,b]$$

and hence $\eta \in L^p[a,b]$ for any $g \in C[c,d]$ (or $g \in BV[c,d]$).

4.38. Theorem (Cameron, Martin). *Let* $h\colon [a,b] \times [c,d] \to R$ *fulfil the assumptions of* 4.37. *Then for any* $f \in L^q[a,b]$, *where* $q = p/(p-1)$ *if* $p > 1$ *and* $q = \infty$ *if* $p = 1$, *and any* $g \in C[c,d]$ (*or* $g \in BV[c,d]$) *the integrals*

$$\int_a^b f(t) \left(\int_c^d \mathrm{d}_s[h(t,s)]\, g(s) \right) \mathrm{d}t \quad \textit{and} \quad \int_c^d \mathrm{d}_s \left[\int_a^b f(t)\, h(t,s)\, \mathrm{d}t \right] g(s)$$

both exist and are equal to one another.

Proof. Let the functions $\varphi\colon [c,d] \to R$ and $\eta\colon [a,b] \to R$ be defined as in 4.37. By 4.19 and 4.37 both the integrals

$$\int_a^b f(t)\, \eta(t)\, \mathrm{d}t \quad \text{and} \quad \int_c^d \mathrm{d}[\varphi(s)]\, g(s)$$

exist. Let $g_n\colon [c,d] \to R$ $(n = 1, 2, \ldots)$ be a sequence of finite step functions such that $\lim_{n\to\infty} g_n(t) = g(t)$ uniformly on $[c,d]$. (Such a sequence exists according to 7.3.2.1 (3) in Aumann [1].) To prove the theorem it is sufficient by 4.17 and 4.20 to show that

$$\int_a^b f(t)\,\eta(t)\,\mathrm{d}t = \int_c^d \mathrm{d}[\varphi(s)]\,g(s) \tag{4,39}$$

holds for all simple jump functions $g(s) = \psi_\alpha^+(s)$ or $g(s) = \psi_\alpha^-(s)$ $(\alpha \in [c,d])$ defined by (4,12) and (4,13). Let $\alpha \in [c,d]$ and $g(s) = \psi_\alpha^+(s)$ on $[c,d]$, then in virtue of 4.21

$$\eta(t) = \begin{cases} h(t,d) - h(t,\alpha+) & \text{if } \alpha < d \\ 0 & \text{if } \alpha = d \end{cases}$$

and

$$\int_c^d \mathrm{d}[\varphi(s)]\,g(s) = \begin{cases} \varphi(d) - \varphi(\alpha+) & \text{if } \alpha < d \\ 0 & \text{if } \alpha = d \end{cases}$$

and (4,39) follows from (4,38). Analogously we can show that (4,39) holds also if $g(s) = \psi_\alpha^-(s)$ on $[c,d]$.

4.39. Integrals of matrix valued functions. If $\mathbf{F} = (f_{i,j})$, $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, r$; $\mathbf{G} = (g_{j,k})$, $j = 1, 2, \ldots, r$, $k = 1, 2, \ldots, q$ are matrix valued functions defined on the interval $[a,b]$ $(f_{i,j}\colon [a,b] \to R,\ g_{j,k}\colon [a,b] \to R)$, then we use the following symbols

$$\int_a^b \mathbf{F}\,\mathrm{d}\mathbf{G} = (\alpha_{i,k}), \qquad i = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, q,$$

and

$$\int_a^b \mathrm{d}[\mathbf{F}]\,\mathbf{G} = (\beta_{i,k}), \qquad i = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, q$$

where

$$\alpha_{i,k} = \sum_{j=1}^r \int_a^b f_{i,j}\,\mathrm{d}g_{j,k} \quad \text{and} \quad \beta_{i,k} = \sum_{j=1}^r \int_a^b g_{j,k}\,\mathrm{d}f_{i,j},$$

whenever the integrals appearing in these sums exist. In the same way it is possible to define also integrals of the type $\int_a^b \mathbf{F}\,\mathrm{d}[\mathbf{G}]\,\mathbf{H}$ etc. if the products of matrices occurring in the expressions are well defined.

Since the integral of a matrix valued function with respect to a matrix valued function is a matrix whose elements are sums of Perron-Stieltjes integrals of real scalar functions with respect to real scalar functions, all statements from this section can be used also for such integrals.

5. The space BV_n

In this section we recall some basic properties of the linear space of functions with a bounded variation from the functional analytic point of view.

Let us consider the linear set of all functions $x\colon [0,1] \to R$ with a bounded variation $\operatorname{var}_0^1 x$. Let this linear set with the norm

$$x \in BV \to \|x\|_{BV} = |x(0)| + \operatorname{var}_0^1 x \tag{5,1}$$

be denoted by $BV[0,1]$ or simply BV.

It is easy to check that (5,1) satisfies all the axioms of a norm.

If $x \in BV$, then evidently

$$|x(t)| \le |x(t) - x(0)| + |x(0)| \le |x(0)| + \operatorname{var}_0^1 x \le \|x\|_{BV} \quad \text{for any} \quad t \in [0,1]. \tag{5,2}$$

5.1. Proposition. *The normed linear space BV is a Banach space (i.e. BV is complete).*

(See Dunford, Schwartz [1] or Hildebrandt [1], II.8.6.)

Further it can be easily shown that BV is not separable. Indeed, if we set $x_\alpha(t) = 0$ for $0 \le t \le \alpha$, $x_\alpha(t) = 1$ for $\alpha < t \le 1$ for any $\alpha \in (0,1)$, then evidently $x_\alpha \in BV$ for any $\alpha \in (0,1)$ and

$$\|x_\alpha - x_\beta\|_{BV} = 2$$

provided $\alpha, \beta \in (0,1)$, $\alpha \neq \beta$. Hence BV cannot contain a countable subset which would be dense in BV. This implies that BV is not separable.

In the same way we can introduce the Banach space BV_n of all column n-vector functions $\mathbf{x} = (x_1, \dots, x_n)^*\colon [0,1] \to R_n$ of bounded variation if for the definition of $\operatorname{var}_0^1 \mathbf{x}$ some norm in R_n is used. The norm in BV_n is given by

$$\mathbf{x} \in BV_n \to \|\mathbf{x}\|_{BV_n} = |\mathbf{x}(0)| + \operatorname{var}_0^1 \mathbf{x}.$$

It is evident that $\mathbf{x}\colon [0,1] \to R_n$ belongs to BV_n if and only if any component x_i, $i = 1, 2, \dots, n$ belongs to BV. Hence it is sufficient to consider only the space BV instead of BV_n. All essential properties of BV are transferable to BV_n.

Let us consider some subspaces of BV which are of interest for the subsequent investigations.

By NBV we denote the set of all functions $\varphi \in BV$ for which $\varphi(t+) = \varphi(t)$ if $t \in (0,1)$ and $\varphi(0) = 0$.

Similarly NBV^- denotes the set of all functions $\varphi \in BV$ such that $\varphi(t-) = \varphi(t)$ for $t \in (0,1)$ and $\varphi(0) = 0$. Further we denote by S the linear set of all functions $w \in BV$ such that $w(t+) = w(t-) = c = \text{const.}$ for every $t \in (0,1)$, $w(0) = w(0+) = c$, $w(1) = w(1-) = c$.

5.2. Proposition. *The linear sets NBV, NBV^-, S are closed in BV.*

Proof. Let $\{\varphi_l\}$, $l = 1, 2, \ldots$ be a sequence with $\varphi_l \in NBV$, such that $\lim_{l\to\infty} \|\varphi_l - \varphi\|_{BV} = 0$ for some $\varphi \in BV$. For $t \in (0, 1)$ we have

$$|\varphi(t+) - \varphi(t)| = |\varphi(t+) - \varphi_l(t+) - (\varphi(t) - \varphi_l(t))| \leq \|\varphi_l - \varphi\|_{BV}$$

for any natural l since $\varphi_l \in NBV$. Hence $\varphi(t+) = \varphi(t)$. Similarly for any l we have

$$|\varphi(0)| = |\varphi_l(0) - \varphi(0)| \leq \|\varphi_l - \varphi\|_{BV}$$

and consequently $\varphi(0) = 0$ and $\varphi \in NBV$. The closedness of NBV^- and S can be proved by the same reasoning.

We denote by AC the linear set of all absolutely continuous functions on $[0, 1]$. If $x \in AC$ then by definition there exists $\delta > 0$ such that for every system $[a_i, b_i]$, $i = 1, \ldots, k$ of nonoverlapping intervals on $[0, 1]$ with

$$\sum_{i=1}^{k} (b_i - a_i) < \delta$$

we have

$$\sum_{i=1}^{k} |x(b_i) - x(a_i)| < 1 .$$

If we subdivide the interval $[0, 1]$ into m intervals by the division points $0 = c_0 < c_1 < \ldots < c_m = 1$ such that $c_i - c_{i-1} < \delta$, $i = 1, 2, \ldots, m$, then $\operatorname{var}_{c_{i-1}}^{c_i} x < 1$ for $i = 1, 2, \ldots, m$ and consequently $\operatorname{var}_0^1 x = \sum_{i=1}^{m} \operatorname{var}_{c_{i-1}}^{c_i} x < m$. Hence $x \in BV$ and the inclusion $AC \subset BV$ holds.

5.3. Proposition. *The linear set AC is closed in BV.*

Proof. Let $\lim_{k\to\infty} \|\varphi_k - \varphi\|_{BV} = 0$ for $\varphi \in BV$ and $\varphi_k \in AC$, $k = 1, 2, \ldots$. For an arbitrary system $[a_i, b_i]$, $i = 1, \ldots, k$ of nonoverlapping intervals in $[0, 1]$ we have

$$\begin{aligned}\sum_{i=1}^{k} |\varphi(b_i) - \varphi(a_i)| &\leq \sum_{i=1}^{k} |\varphi_l(b_i) - \varphi(b_i) - (\varphi_l(a_i) - \varphi(a_i))| + \sum_{i=1}^{k} |\varphi_l(b_i) - \varphi_l(a_i)| \\ &\leq \|\varphi_l - \varphi\|_{BV} + \sum_{i=1}^{k} |\varphi_l(b_i) - \varphi_l(a_i)|\end{aligned}$$

for any $l = 1, 2, \ldots$. Let $\varepsilon > 0$ be given. Let us choose an integer $l_0 \geq 1$ such that $\|\varphi_l - \varphi\|_{BV} < \varepsilon/2$ for $l > l_0$. For any fixed $l > l_0$ there is $\delta > 0$ such that if

$$\sum_{i=1}^{k} (b_i - a_i) < \delta$$

then

$$\sum_{i=1}^{k} |\varphi_l(b_i) - \varphi_l(a_i)| < \varepsilon .$$

Hence by the inequality given above we have $\sum_{i=1}^{k} |\varphi(b_i) - \varphi(a_i)| < \varepsilon$ and $\varphi \in AC$.

5.4. Remark. From 5.2 and 5.3 it is evident that if the closed linear sets NBV, NBV^-, S, AC in BV are equipped with the norm (5,1) of BV, then they are Banach spaces.

By NBV_n, NBV_n^-, S_n, AC_n we denote the closed linear subsets in BV_n which are defined similarly as NBV, NBV^-, S, AC for n-vector functions. For the same reason as above NBV_n, NBV_n^-, S_n, AC_n equipped with the norm of BV_n are Banach spaces.

Let us now assume that $x \in BV$ and define $w(0) = w(1) = x(0)$, $w(t) = x(t) - x(t+) + x(0)$ for $t \in (0, 1)$. Then evidently $w \in S$, since the difference $x(t) - x(t+)$ is nonzero only on an at most countable set $A \subset (0, 1)$ and

$$\operatorname{var}_0^1 w = 2 \sum_{t \in A} |x(t+) - x(t)| \le 2 \operatorname{var}_0^1 x < \infty .$$

Further let us set $\varphi(t) = x(t) - w(t)$ for $t \in [0, 1]$. It is $\varphi(0) = x(0) - w(0) = 0$, $\varphi(t) = x(t+) - x(0)$ for $t \in (0, 1)$, $\varphi(1) = x(1) - x(0)$, i.e. $\varphi \in NBV$.

In this way we have obtained

$$x = \varphi + w$$

for any $x \in BV$ where $\varphi \in NBV$ and $w \in S$. Since evidently $NBV \cap S = \{0\}$, this decomposition is unique. Hence the Banach space BV can be written in the form of the direct sum of closed subspaces NBV and S, i.e.

$$BV = NBV \oplus S . \tag{5,3}$$

Similarly it can be shown that also the decomposition

$$BV = NBV^- \oplus S$$

holds.

For any $x \in BV$ and $\psi \in BV$ we can define the expression

$$f(x) = \int_0^1 x(t) \, d\psi(t) . \tag{5,4}$$

By 4.19 the integral on the right-hand side in (5,4) exists. The functional f is evidently linear. Further it is

$$|f(x)| = \left| \int_0^1 x(t) \, d\psi(t) \right| \le \sup_{t \in [0,1]} |x(t)| \operatorname{var}_0^1 \psi \le \|x\|_{BV} \|\psi\|_{BV}$$

(see 4.27). Hence if f is given by (5,4) with $\psi \in BV$, then $f \in BV^*$.

5.5. Proposition. *Assume that $w \in BV$. Then*

$$\int_0^1 x(t) \, dw(t) = 0 \tag{5,5}$$

for any $x \in BV$ if and only if $w \in S$.

Proof. Let us suppose that $\int_0^1 x(t)\,\mathrm{d}w(t) = 0$ for any $x \in BV$. For a given $\alpha \in [0, 1]$ we define $x_\alpha(t) = 0$ if $t \in [0, 1] \setminus \{\alpha\}$, $x_\alpha(\alpha) = 1$. Then evidently $x_\alpha \in BV$ and we obtain by the assumption

$$\int_0^1 x_\alpha(t)\,\mathrm{d}w(t) = w(\alpha+) - w(\alpha-) = 0\,,$$

i.e. $w(\alpha+) = w(\alpha-)$ for any $\alpha \in (0, 1)$ and $\int_0^1 x_1(t)\,\mathrm{d}w(t) = w(1) - w(1-) = 0$, $\int_0^1 x_0(t)\,\mathrm{d}w(t) = w(0+) - w(0) = 0$ (cf. 4.22). This means that w differs from a continuous function only on an at most countable subset in $(0, 1)$.

Assume that $w \notin S$. Then there exist two points $\alpha, \beta \in [0, 1]$, $\alpha < \beta$ such that α, β are points of continuity for w and $w(\alpha) \neq w(\beta)$. We define $x_{[\alpha,\beta]}(t) = 1$ for $t \in [\alpha, \beta]$ and $x_{[\alpha,\beta]}(t) = 0$ for $t \in [0, 1] \setminus [\alpha, \beta]$. Evidently $x_{[\alpha,\beta]} \in BV$. Using the properties of the integral we obtain the relation

$$\int_0^1 x_{[\alpha,\beta]}(t)\,\mathrm{d}w(t) = w(\alpha) - w(\alpha-) + \int_\alpha^\beta \mathrm{d}w(t) + w(\beta+) - w(\beta)$$

$$= \int_\alpha^\beta \mathrm{d}w(t) = w(\beta) - w(\alpha) \neq 0$$

which contradicts the assumption. Hence $w \in S$. Let us assume that $w \in S$; w is evidently a break function with $\Delta w(t) = w(t+) - w(t-) = 0$ for every $t \in (0, 1)$ and $\Delta^+ w(0) = w(0+) - w(0) = 0$, $\Delta^- w(1) = w(1) - w(1-) = 0$. Hence by 4.23 we have $\int_0^1 x(t)\,\mathrm{d}w(t) = 0$ for every $x \in BV$.

5.6. Corollary. *Let $\psi \in BV$ be given. Using (5,3) ψ can be uniquely written in the form $\psi = \varphi + w$ where $\varphi \in NBV$, $w \in S$ and*

$$\int_0^1 x(t)\,\mathrm{d}\psi(t) = \int_0^1 x(t)\,\mathrm{d}\varphi(t)$$

for every $x \in BV$.

Let us define for $x \in BV$, $\varphi \in NBV$ the relation

$$\langle x, \varphi \rangle = \int_0^1 x(t)\,\mathrm{d}\varphi(t)\,. \tag{5,6}$$

This relation evidently defines a bilinear form on $BV \times NBV$.

5.7. Lemma. *Let $\varphi \in NBV$. If $\langle x, \varphi \rangle = 0$ for every $x \in BV$, then $\varphi = 0$.*
Let $x \in BV$. If $\langle x, \varphi \rangle = 0$ for every $\varphi \in NBV$, then $x = 0$.

Proof. (1) If $\langle x, \varphi \rangle = 0$ for every $x \in BV$, then $\varphi \in S$ by 5.5. Hence $\varphi \in NBV \cap S$ and by (5,3) we obtain $\varphi = 0$.

(2) Assume that $\langle x, \varphi \rangle = 0$ for every $\varphi \in NBV$ but $x \neq 0$. Then either there exists $a \in (0, 1]$ such that $x(a) \neq 0$ or $x(t) = 0$ for all $t \in (0, 1]$ and $x(0) \neq 0$. In the

first case we set $\varphi(t) = 0$ for $t \in [0, a)$, $\varphi(t) = 1$ for $t \in [a, 1]$. Evidently $\varphi \in NBV$ and φ is a simple jump function (see 4.20). By 4.21 we have $\langle x, \varphi \rangle = \int_0^1 x(t)\, d\varphi(t) = x(a) \neq 0$ and this contradicts the assumption. For the second case we set $\varphi(t) = 1$ for $t \in (0, 1]$, $\varphi(0) = 0$, then $\varphi \in NBV$ is also a simple jump function ($\varphi = \psi_0$) and by 4.21 we have $\langle x, \varphi \rangle = \int_0^1 x(t)\, d\varphi(t) = x(0) \neq 0$. Again we have obtained a contradiction and our lemma is proved.

5.8. Proposition. *The pair of spaces BV, NBV forms a dual pair (BV, NBV) with respect to the bilinear form $\langle ., . \rangle$ given by* (5,6).

Proof follows immediately from 5.7 and from the definition of a dual pair given in 3.1.

5.9. Remark. It follows easily from 5.8 that (BV_n, NBV_n) is a dual pair with respect to the bilinear form

$$\mathbf{x} \in BV_n,\ \boldsymbol{\varphi} \in NBV_n \to \langle \mathbf{x}, \boldsymbol{\varphi} \rangle = \int_0^1 \mathbf{x}^*(t)\, d\boldsymbol{\varphi}(t) = \sum_{j=1}^n \int_0^1 x_j(t)\, d\varphi_j(t).$$

Let us mention that for every fixed $\boldsymbol{\varphi} \in NBV_n$ by $\langle \mathbf{x}, \boldsymbol{\varphi} \rangle$ a bounded linear functional on BV_n is defined. In fact, we have by 4.27

$$|\langle \mathbf{x}, \boldsymbol{\varphi} \rangle| \leq \left| \int_0^1 \mathbf{x}^*(t)\, d\boldsymbol{\varphi}(t) \right| \leq \big(\sup_{t \in [0,1]} |\mathbf{x}(t)| \big) (\operatorname{var}_0^1 \boldsymbol{\varphi}) = (\operatorname{var}_0^1 \boldsymbol{\varphi})\, \|\mathbf{x}\|_{BV_n}$$

for every $\mathbf{x} \in BV_n$ and $\boldsymbol{\varphi} \in NBV_n$.

The space BV_n has important subspaces called the *Sobolev spaces* W_n^p $(1 \leq p < \infty)$ including in particular the space AC_n of absolutely continuous functions on $[0,1]$.

5.10. Definition. Given a real number p, $1 \leq p < \infty$, W_n^p denotes the space of all absolutely continuous functions $\mathbf{x}: [0, 1] \to R_n$ whose derivatives $\mathbf{x}'$ are L^p-integrable on $[0, 1]$. Furthermore,

$$\|\mathbf{x}\|_{W_n^p} = |\mathbf{x}(0)| + \left(\int_0^1 |\mathbf{x}'(t)|^p\, dt \right)^{1/p} = |\mathbf{x}(0)| + \|\mathbf{x}'\|_{L^p} \qquad \text{for any} \quad \mathbf{x} \in W_n^p.$$

($W_1^p = W^p$ and instead of $\|.\|_{W_n^p}$ we write $\|.\|_{W^p}$.)

5.11. Remark. Evidently, any W_n^p $(p \in R,\ p \geq 1)$ equipped with the norm $\|.\|_{W^p}$ is a linear normed space.

5.12. Remark. It is well-known that any $\mathbf{x} \in BV_n$ possesses a.e. on $[0, 1]$ a derivative $\mathbf{x}'(t)$ which is L-integrable on $[0, 1]$ $(\mathbf{x}' \in L_n^1)$. Furthermore, $\mathbf{x} \in AC_n$ if and only if there is $\mathbf{z} \in L_n^1$ such that

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \mathbf{z}(\tau)\, d\tau \qquad \text{on } [0, 1],$$

i.e. $W_n^1 = AC_n$. Given $\mathbf{x} \in AC_n$, we have $\operatorname{var}_0^1 \mathbf{x} = \|\mathbf{x}'\|_{L^1}$ and therefore also the norms $\|.\|_{AC}$ and $\|.\|_{W^1}$ are identical (cf. e.g. Natanson [1]).

5.13. Proposition. *Given $p \in R$, $p \geq 1$, the space W_n^p is isometrically isomorphic with the product space $L_n^q \times R_n$ and its dual space is isometrically isomorphic with $L_n^q \times R_n^*$, where $q = p/(p-1)$ if $p > 1$ and $q = \infty$ if $p = 1$.*

Proof. (a) The mapping $\mathbf{x} \in W_n^p \to (\mathbf{x}', \mathbf{x}(0)) \in L_n^p \times R_n$ and its inverse $(\mathbf{z}, \mathbf{c}) \in L_n^p \times R_n \to \mathbf{x}(t) = \mathbf{c} + \int_0^t \mathbf{z}(\tau)\,d\tau \in W_n^p$ establish an isometrical isomorphism between W_n^p and $L_n^p \times R_n$.

(b) Let f be an arbitrary linear bounded functional on W_n^p and let us put for any $\mathbf{c} \in R_n$ and $\mathbf{z} \in L_n^p$ $f_1(\mathbf{z}) = f(\boldsymbol{\Psi}\mathbf{z})$ and $f_2(\mathbf{c}) = f(\boldsymbol{\Phi}\mathbf{c})$, where

$$\boldsymbol{\Psi}: \mathbf{z} \in L_n^p \to \int_0^t \mathbf{z}(\tau)\,d\tau \in W_n^p, \qquad \boldsymbol{\Phi}: \mathbf{c} \in R_n \to \mathbf{u}(t) \equiv \mathbf{c} \in W_n^p.$$

Then f_1 and f_2 are linear bounded functionals on L_n^p and R_n, respectively, while $f(\mathbf{x}) = f(\boldsymbol{\Psi}\mathbf{x}' + \boldsymbol{\Phi}\mathbf{x}(0)) = f_1(\mathbf{x}') + f_2(\mathbf{x}(0))$ for any $\mathbf{x} \in W_n^p$. Consequently, given $f \in (W_n^p)^*$, there exist uniquely determined $\mathbf{y}^* \in L_n^q$ ($q = p/(p-1)$ if $p > 1$, $q = \infty$ if $p = 1$) and $\boldsymbol{\lambda}^* \in R_n^*$ such that (cf. 3.10)

$$f(\mathbf{x}) = \int_0^1 \mathbf{y}^*(t)\,\mathbf{x}'(t)\,dt + \boldsymbol{\lambda}^*\,\mathbf{x}(0) \qquad \text{for any} \quad \mathbf{x} \in W_n^p.$$

Furthermore,

$$\|f_1\| = \sup_{\|\mathbf{z}\|_{L^p}=1} |f_1(\mathbf{z})| = \|\mathbf{y}^*\|_{L^q}, \qquad \|f_2\| = \sup_{|\mathbf{c}|=1} |f_2(\mathbf{c})| = |\boldsymbol{\lambda}^*|$$

and hence

$$\|f\| = \sup_{\|\mathbf{x}\|_{W^p}=1} |f(\mathbf{x})| = \|(\mathbf{y}^*, \boldsymbol{\lambda}^*)\|_{L^q \times R} = \|\mathbf{y}^*\|_{L^q} + |\boldsymbol{\lambda}^*|.$$

5.14. Remark. In accordance with 3.6 we denote for $\mathbf{x} \in W_n^p$, $\mathbf{y}^* \in L_n^q$ and $\boldsymbol{\lambda}^* \in R_n^*$

$$\langle \mathbf{x}, (\mathbf{y}^*, \boldsymbol{\lambda}^*) \rangle_W = \langle \mathbf{x}', \mathbf{y}^* \rangle_L + \boldsymbol{\lambda}^*\,\mathbf{x}(0) = \int_0^1 \mathbf{y}^*(t)\,\mathbf{x}'(t)\,dt + \boldsymbol{\lambda}^*\,\mathbf{x}(0).$$

Let us notice that $\mathbf{x} \in W_n^p \to \langle \mathbf{x}, (\mathbf{y}^*, \boldsymbol{\lambda}^*) \rangle_W$ is the zero functional on W_n^p if and only if $\mathbf{y}^*(t) = \mathbf{0}$ a.e. on $[0,1]$ and $\boldsymbol{\lambda}^* = \mathbf{0}$. As a consequence we have

5.15. Proposition. *If $\mathbf{y}^* \in L_n^q$ and $\boldsymbol{\lambda}^* \in R_n^*$, then*

$$\int_0^1 \mathbf{y}^*(t)\,\mathbf{x}'(t)\,dt + \boldsymbol{\lambda}^*\,\mathbf{x}(0) = 0 \qquad \textit{for any} \quad \mathbf{x} \in W_n^p$$

or

$$\int_0^1 \mathbf{y}^*(t)\,\mathbf{z}(t)\,dt + \boldsymbol{\lambda}^*\mathbf{c} = 0 \qquad \textit{for any} \quad \mathbf{z} \in L_n^p \quad \textit{and} \quad \mathbf{c} \in R_n$$

if and only if $\mathbf{y}^*(t) = \mathbf{0}$ a.e. on $[0,1]$ and $\boldsymbol{\lambda}^* = \mathbf{0}$.

5.16. Proposition. $\mathbf{S} \in B(W_n^p, R_m)$ *if and only if there exist an* $m \times n$*-matrix* $\mathbf{M}$ *and an* $m \times n$*-matrix valued function* $\mathbf{K}$ *with* $\|\mathbf{K}\|_{L^q} < \infty$ $(q = p/(p-1)$ *if* $p > 1$, $q = \infty$ *if* $p = 1)$ *such that*

$$\mathbf{S}\mathbf{x} = \mathbf{M}\,\mathbf{x}(0) + \int_0^1 \mathbf{K}(t)\,\mathbf{x}'(t)\,\mathrm{d}t \qquad \textit{for any} \quad \mathbf{x} \in W_n^p.$$

5.17 Lemma. *Let* $f \in BV$ *be right-continuous on* $[0, 1)$ *and left-continuous at* 1 *and* $f(1) = 0$. *Then*

$$\int_0^1 x(s)\,\mathrm{d}f(s) = 0 \qquad \text{for any} \quad x \in W_n^p \quad \text{with} \quad x(0) = x(1) = 0$$

if and only if $f(t) \equiv 0$ *on* $[0, 1]$.

Proof. Let us assume that $f(t) \not\equiv 0$ on $[0,1]$, e.g. let $f(t_0) \neq 0$. Then $\operatorname{var}_0^1 f \geq |f(1) - f(t_0)| = |f(t_0)| > 0$. Let $\varepsilon > 0$ be such that $\alpha = \operatorname{var}_0^1 f \geq 3\varepsilon > 0$. By the definition of a variation there exists a subdivision $\{0 = t_0 < t_1 < \ldots < t_m = 1\}$ of $[0, 1]$ such that

$$\sum_\sigma |\Delta f| = \sum_{j=1}^q |f(s_j) - f(s_{j-1})| > \alpha - \varepsilon$$

for any of its refinements $\sigma = \{0 = s_0 < s_1 < \ldots < s_q = 1\}$. In virtue of the one-sided continuity of f there exist $\tau_j \in (0, 1)$ $(j = 1, 2, \ldots, m)$ such that $0 < \tau_0 < < \tau_1 < \ldots < \tau_m < 1$, $t_{j-1} < \tau_{j-1} < t_j$ $(j = 1, 2, \ldots, m-1)$, $t_{m-1} < \tau_{m-1} < \tau_m < t_m = 1$ and

$$\sum_{j=0}^{m-1} |f(t_j) - f(\tau_j)| + |f(1) - f(\tau_m)| \leq \sum_{j=0}^{m-1} \operatorname{var}_{t_j}^{\tau_j} f + \operatorname{var}_{\tau_m}^1 f < \varepsilon .$$

Putting $x(0) = 0$, $x(t) = \operatorname{sign}(f(t_j) - f(\tau_{j-1}))$ for $t \in [\tau_{j-1}, t_j]$ $(j = 1, 2, \ldots, m-1)$, $x(t) = \operatorname{sign}(f(\tau_m) - f\ (\tau_{m-1}))$ for $t \in [\tau_{m-1}, \tau_m]$, $x(1) = 0$ and extending the definition of x to the whole $[0, 1]$ in such a way that x is linear on the rest of $[0, 1]$, we obtain

$$\left| \sum_{j=0}^{m-1} \int_{t_j}^{\tau_j} x(s)\,\mathrm{d}[f(s)] + \int_{\tau_m}^1 x(s)\,\mathrm{d}[f(s)] \right| \leq \sum_{j=0}^{m-1} \operatorname{var}_{t_j}^{\tau_j} f + \operatorname{var}_{\tau_m}^1 f < \varepsilon .$$

Hence

$$\left| \int_0^1 x(s)\,\mathrm{d}[f(s)] \right|$$

$$= \left| \sum_{j=1}^{m-1} |f(t_j) - f(\tau_{j-1})| + |f(\tau_m) - f(\tau_{m-1})| + \sum_{j=0}^{m-1} \int_{t_j}^{\tau_j} x(s)\,\mathrm{d}[f(s)] + \int_{\tau_m}^1 x(s)\,\mathrm{d}[f(s)] \right|$$

$$> \sum_{j=1}^{m-1} |f(t_j) - f(\tau_{j-1})| + |f(\tau_m) - f(\tau_{m-1})| - \varepsilon > \sum_\sigma |\Delta f| - 2\varepsilon > \alpha - 3\varepsilon > 0 ,$$

where $\sigma = \{0 = t_0 < \tau_0 < t_1 < \tau_1 < \ldots < t_{m-1} < \tau_{m-1} < \tau_m < t_m = 1\}$. Since obviously $x \in W^p$ and $x(0) = x(1) = 0$, this completes the proof.

6. Variation of functions of two variables

Various definitions of the variation of functions of two or more variables are known. In our considerations we use one of them, the so called *Vitali variation*. This section is devoted to the definition of this sort of variation for functions of two variables and to the fundamental properties of functions with finite variation in this sense.

Let a nondegenerate interval $I = [a, b] \times [c, d] \subset R_2$ be given. We consider a real function $k: I \to R$ defined on I.

For a given subinterval $J = [a', b'] \times [c', d'] \subset I$, $a \le a' \le b' \le b$, $c \le c' \le d' \le d$ we set

$$m_k(J) = k(b', d') - k(b', c') - k(a', d') + k(a', c'). \tag{6,1}$$

Let us define

$$\mathrm{v}_I(k) = \sup \sum_i |m_k(J_i)|, \tag{6,2}$$

where the supremum is taken over all finite systems of nonoverlapping intervals $J_i \subset I$ (i.e. for the interiors J_i^0 of the intervals J_i we assume that $J_i^0 \cap J_j^0 = \emptyset$ whenever $i \neq j$).

6.1. Definition (Vitali). The real function $k: I \to R$ is of bounded variation on I if $\mathrm{v}_I(k) < +\infty$.

6.2. Remark. If on the interval $I = [a, b] \times [c, d]$ an $n \times n$-matrix $\mathbf{K}(s, t) = (k_{ij}(s, t))$ $(i, j = 1, \ldots, r)$ is given, i.e. $\mathbf{K}: I \to L(R_n)$, then we can set

$$m_K(J) = \mathbf{K}(b', d') - \mathbf{K}(b', c') - \mathbf{K}(a', d') + \mathbf{K}(a', c')$$

as above and define the number $\mathrm{v}_I(\mathbf{K}) = \sup \sum_i |m_K(J_i)|$ in the same way as in (6,2) where the norm in the sum on the right-hand side is some norm of an $n \times n$-matrix (cf. 1.1). For the case of the norm defined in 1.1 we have evidently $\mathrm{v}_I(k_{ij}) \le \mathrm{v}_I(\mathbf{K})$ for all $i, j = 1, 2, \ldots, n$.

6.3. Remark. Assume that $a = \alpha_0 < \alpha_1 < \ldots < \alpha_k = b$, $c = \gamma_0 < \gamma_1 < \ldots < \gamma_l = d$ are some finite subdivisions of the intervals $[a, b]$, $[c, d]$ respectively. The finite system of subintervals

$$J_{ij} = [\alpha_{i-1}, \alpha_i] \times [\gamma_{j-1}, \gamma_j], \qquad i = 1, \ldots, k, \quad j = 1, \ldots, l$$

is called a net-type subdivision of the interval $I = [a, b] \times [c, d]$. Evidently every net-type subdivision of I is a finite system of nonoverlapping intervals.

It is easy to see that for every finite system of nonoverlapping intervals $J_i \subset I$ there is a net-type subdivision of I such that every J_i is the union of some of its elements. Using this fact it is not difficult to show that for the definition of $\mathrm{v}_I(k)$

from (6,2) the supremum can be taken over all finite net-type subdivisions and the number $v_I(k)$ remains unchanged.

6.4. Examples. Assume that $f \in BV[a,b]$, $g \in BV[c,d]$. Then for $k(s,t) = f(s)\,g(t)$: $[a,b] \times [c,d] \to R$ we have by definition

$$v_{[a,b]\times[c,d]}(k) = \operatorname{var}_a^b f \operatorname{var}_c^d g < \infty\,.$$

Let us set

$$h(s,t) = 0 \quad \text{for} \quad 0 \le t < s \le 1\,, \qquad h(s,t) = 1 \quad \text{for} \quad 0 \le s \le t \le 1\,.$$

Then for every net-type subdivision $J_{ij} = [\alpha_{i-1}, \alpha_i] \times [\alpha_{j-1}, \alpha_j]$, $i, j = 1, \ldots, k$, $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_k = 1$ we have

$$\sum_{i,j=1}^{k} |m_h(J_{i,j})| = \sum_{i=1}^{k} |m_h(J_{i,i})| + \sum_{i=2}^{k} |m_h(J_{i,j-1})| = 2k - 1$$

since $m_h(J_{i,i}) = 1$, $m_h(J_{i,i-1}) = 1$ and $m_h(J_{i,j}) = 0$ if $j \neq i, i-1$. Hence $v_{[0,1]\times[0,1]}(h)$ cannot be finite.

The following lemma can be easily verified.

6.5. Lemma. *If $I_j \subset I \subset R_2$, $j = 1, \ldots, m$ is a finite system of nonoverlapping intervals in I and k: $I \to R$, then*

$$\sum_{j=1}^{m} v_{I_j}(k) \le v_I(k)\,. \tag{6,3}$$

6.6. Lemma. *Let k: $I = [a,b] \times [c,d] \to R$ be given such that $v_I(k) < \infty$, $\operatorname{var}_a^b k(.,\gamma_0) < \infty$ for some $\gamma_0 \in [c,d]$, i.e. $k(.,\gamma_0) \in BV[a,b]$ for some $\gamma_0 \in [c,d]$. Then $k(.,\gamma) \in BV[a,b]$ for all $\gamma \in [c,d]$ and*

$$\operatorname{var}_a^b k(.,\gamma) \le v_I(k) + \operatorname{var}_a^b k(.,\gamma_0)\,. \tag{6,4}$$

If k: $[a,b] \times [c,d] \to R$ and $\gamma \in [c,d]$ is fixed, then we denote the usual variation of the function $k(s,\gamma)$ in the interval $[a,b]$ by $\operatorname{var}_a^b k(.,\gamma)$. Similarly for $\operatorname{var}_c^d k(\alpha, .)$ where $\alpha \in [a,b]$ is fixed.

Proof. For any $\gamma, \gamma_0 \in [c,d]$, $\alpha_{j-1}, \alpha_j \in [a,b]$ we have

$$|k(\alpha_j,\gamma) - k(\alpha_{j-1},\gamma)| \le |m_{J_j}(k)| + |k(\alpha_j,\gamma_0) - k(\alpha_{j-1},\gamma_0)|$$

where $J_j = [\alpha_{j-1}, \alpha_j] \times [\gamma_0, \gamma]$. Hence for each finite decomposition $a = \alpha_0 < \alpha_1 < \ldots < \alpha_k = b$ we have

$$\sum_{j=1}^{k} |k(\alpha_j,\gamma) - k(\alpha_{j-1},\gamma)|$$

$$\le \sum_{j=1}^{k} |m_{J_j}(k)| + \sum_{j=1}^{k} |k(\alpha_j,\gamma_0) - k(\alpha_{j-1},\gamma_0)| \le v_I(k) + \operatorname{var}_a^b k(.,\gamma_0)$$

and this inequality implies (6,4).

For a given $k\colon I \to R$, $I = [a,b] \times [c,d]$ we put

$$\omega_1(a) = 0\,, \qquad \omega_1(\sigma) = \mathrm{v}_{[a,\sigma]\times[c,d]}(k) \qquad \text{for} \quad \sigma \in (a,b] \tag{6,5}$$

and similarly

$$\omega_2(c) = 0\,, \qquad \omega_2(\tau) = \mathrm{v}_{[a,b]\times[c,\tau]}(k) \qquad \text{for} \quad \tau \in (c,d]\,. \tag{6,6}$$

6.7. Lemma. *The function $\omega_1\colon [a,b] \to R$ from (6,5) is nondecreasing on $[a,b]$, $\omega_1(b) = \mathrm{v}_I(k)$; hence $\omega_1 \in BV[a,b]$ if $\mathrm{v}_I(k) < +\infty$. Similarly for the function $\omega_2\colon [c,d] \to R$ from* (6,6).

The proof follows easily from the definitions.

6.8. Lemma. *If $k\colon I \to R$, $I = [a,b] \times [c,d]$, $\mathrm{v}_I(k) < \infty$ and $\operatorname{var}_a^b k(.,c) < \infty$, then the set of discontinuity points of k in the first variable s lies on a denumerable system of lines in I, which are parallel to the t-axis.*

Proof. For any $s, s_0 \in [a,b]$, $t \in [c,d]$ we have

$$\begin{aligned} |k(s,t) - k(s_0,t)| &\le |k(s,t) - k(s,c) - k(s_0,t) + k(s_0,c)| + |k(s,c) - k(s_0,c)| \\ &\le |\omega_1(s) - \omega_1(s_0)| + |\operatorname{var}_a^s k(.,c) - \operatorname{var}_a^{s_0} k(.,c)| \end{aligned}$$

where $\omega_1\colon [a,b] \to R$ is given by (6,5). Since $\omega_1 \in BV[a,b]$ by 6.7 and the function $\operatorname{var}_a^s k(.,c)$ is also of bounded variation on $[a,b]$, the above inequality gives that there exists an at most denumerable set of points $M \subset [a,b]$ such that $\lim_{s\to s_0} k(s,t) = k(s_0,t)$ whenever $s_0 \in [a,b]\setminus M$ and $t \in [c,d]$ are arbitrary. This yields our proposition.

6.9. Lemma. *If $k\colon I \to R$, $\mathrm{v}_I(k) < \infty$, $\operatorname{var}_a^b k(.,c) < \infty$, $\operatorname{var}_c^d k(a,.) < \infty$, then the set of discontinuities of k in $I = [a,b] \times [c,d]$ lies on a denumerable set of lines in I parallel to the coordinate axes.*

This proposition is proved in Hildebrandt [1], III.5.4. If $k(s,t)$ satisfies the assumptions of 6.8 then $h(s,t) = k(s,t) - k(a,t)$ satisfies the assumptions of 6.9 and 6.8 is a corollary of 6.9.

6.10. Lemma. *If $k\colon I \to R$, $I = [a,b] \times [c,d]$, $\mathrm{v}_I(k) < +\infty$, then for an arbitrary subdivision $c = \gamma_0 < \gamma_1 < \ldots < \gamma_l = d$ and any two points $s_1, s_2 \in [a,b]$ we have*

$$\left|\sum_{j=1}^{l} [\operatorname{var}_a^{s_2}(k(.,\gamma_j) - k(.,\gamma_{j-1})) - \operatorname{var}_a^{s_1}(k(.,\gamma_j) - k(.,\gamma_{j-1}))]\right| \le |\omega_1(s_2) - \omega_1(s_1)|$$

where $\omega_1\colon [a,b] \to R$ is defined by (6,5).

Proof. Let us set $h(s,t) = k(s,t) - k(s,c)$ for $(s,t) \in I$. Then $h(s,c) = 0$ for any $s \in [a,b]$ and by 6.6 $\operatorname{var}_a^b h(.,t) < \infty$ for any $t \in [c,d]$ because evidently $\mathrm{v}_I(h) < \infty$. Hence $\operatorname{var}_a^s h(.,t)$ is finite for any $s \in [a,b]$, $t \in [c,d]$. For any $j = 1, \ldots, l$ we have

$h(s,\gamma_j) - h(s,\gamma_{j-1}) = k(s,\gamma_j) - k(s,\gamma_{j-1})$ and $\operatorname{var}_a^s(k(.,\gamma_j) - k(.,\gamma_{j-1}))$ is also finite for every $s \in [0,1]$. This implies that for any $j = 1,\dots,l$ we have

$$\begin{aligned}&|\operatorname{var}_a^{s_2}(k(.,\gamma_j) - k(.,\gamma_{j-1})) - \operatorname{var}_a^{s_1}(k(.,\gamma_j) - k(.,\gamma_{j-1}))|\\ &\quad\le |\operatorname{var}_{s_1}^{s_2}(k(.,\gamma_j) - k(.,\gamma_{j-1}))| \le \mathrm{v}_{[s_1,s_2]\times[\gamma_{j-1},\gamma_j]}(k).\end{aligned}$$

By 6.5 we obtain the inequality

$$\begin{aligned}&\sum_{j=1}^{l} \mathrm{v}_{[s_1,s_2]\times[\gamma_{j-1},\gamma_j]}(k) \le \mathrm{v}_{[s_1,s_2]\times[c,d]}(k)\\ &\le |\mathrm{v}_{[a,s_2]\times[c,d]}(k) - \mathrm{v}_{[a,s_1]\times[c,d]}(k)| = |\omega_1(s_2) - \omega_1(s_1)|\end{aligned}$$

which yields our result.

6.11. Lemma. *If $k\colon I \to R$, $I = [a,b] \times [c,d]$, $\mathrm{v}_I(k) < \infty$ and for some $s_0 \in [a,b]$ the relation*

(6,7) $$\lim_{s\to s_0\pm} |k(s,t) - k(s_0,t)| = 0$$

holds for all $t \in [c,d]$, then

(6,8) $$\lim_{s\to s_0\pm} \omega_1(s) = \omega_1(s_0)$$

where $\omega_1\colon [a,b] \to R$ is defined by (6.5).

This is proved in Schwabik [2], Lemma 2.1.

6.12. Remark. If for $k\colon I \to R$ we have $\mathrm{v}_I(k) < \infty$ and $\operatorname{var}_a^b k(.,c) < \infty$, then by 6.8 the relation (6,7) is satisfied for all $s_0 \in [a,b]$ except for a denumerable set of points in $[a,b]$. Moreover, in this case $k(.,t) \in BV[a,b]$ for every $t \in [c,d]$ (cf. 6.6). Hence by the elementary properties of functions of bounded variation the onesided limits $\lim_{\sigma\to s_0+} k(\sigma,t) = k(s_0+,t)$, $\lim_{\sigma\to s_0-} k(\sigma,t) = k(s_0-,t)$ exist for every $s_0 \in [a,b)$, $s_0 \in (a,b]$, respectively, and for every $t \in [c,d]$.

6.13. Lemma. *If $k\colon I \to R$ $(I = [a,b] \times [c,d])$ is given, then for every $s_1, s_2 \in [a,b]$ we have*

(6,9) $$\operatorname{var}_c^d(k(s_2,.) - k(s_1,.)) \le |\omega_1(s_2) - \omega_1(s_1)|$$

where $\omega_1\colon [a,b] \to R$ is defined in (6,5).

Proof. For an arbitrary subdivision $c = \gamma_0 < \gamma_1 < \dots < \gamma_l = d$ we have by 6.5

$$\begin{aligned}&\sum_{j=1}^{l} |k(s_2,\gamma_j) - k(s_1,\gamma_j) - k(s_2,\gamma_{j-1}) + k(s_1,\gamma_{j-1})|\\ &\le \mathrm{v}_{[s_1,s_2]\times[c,d]}(k) \le |\mathrm{v}_{[a,s_2]\times[c,d]}(k) - \mathrm{v}_{[a,s_1]\times[c,d]}(k)| = |\omega_1(s_2) - \omega_1(s_1)|\end{aligned}$$

and proceeding to the supremum for all finite subdivisions of $[c,d]$ we obtain (6,9).

6.14. Lemma. *Assume that* $k\colon I \to R$ $(I = [a,b] \times [c,d])$ *is given with* $\mathrm{v}_I(k) < \infty$ *and for some* $s_0 \in [a,b)$ *the limit*

$$\lim_{s \to s_0+} k(s,t) = k(s_0+,t) \tag{6,10}$$

exists for every $t \in [c,d]$. *Then*

$$\lim_{\delta \to 0+} \operatorname{var}_c^d (k(s_0+\delta,.) - k(s_0+,.)) = 0\,.$$

Proof. Define $k^0\colon I \to R$ such that $k^0(s,t) = k(s,t)$ if $(s,t) \in I$, $s \neq s_0$ and $k^0(s_0,t) = k(s_0+,t)$. Since $\operatorname{var}_c^d (k(s_0+,.) - k(s_0,.)) < \infty$ we obtain $\mathrm{v}_I(k^0) < \infty$. Let $\omega_1^0\colon [a,b] \to R$, $\omega_1^0(a) = 0$, $\omega_1^0(\sigma) = \mathrm{v}_{[a,\sigma]\times[c,d]}(k^0)$ for $\sigma \in (a,b]$. Since $\lim_{s\to s_0+} (k^0(s,t) - k^0(s_0,t)) = 0$ for every $t \in [c,d]$, we have by 6.11 $\lim_{s\to s_0+} \omega_1^0(s) = \omega_1^0(s_0)$. For every $\delta > 0$ such that $s_0 + \delta \in [a,b]$ we have by 6.13

$$\begin{aligned}\operatorname{var}_c^d (k^0(s_0+\delta,.) - k^0(s_0,.)) &= \operatorname{var}_c^d (k(s_0+\delta,.) - k(s_0+,.))\\ &\leq |\omega_1^0(s_0+\delta) - \omega_1^0(s_0)|\,.\end{aligned}$$

The limitation process $\delta \to 0+$ yields our result.

6.15. Corollary. *If* $k\colon I \to R$ $(I = [a,b] \times [c,d])$ *is such that* $\mathrm{v}_I(k) < \infty$ *and* $\operatorname{var}_a^b k(.,c) < \infty$, *then for any* $s_0 \in [a,b)$ *we have*

$$\operatorname{var}_c^d (k(s_0+,.) - k(s_0,.)) \leq \omega_1(s_0+) - \omega_1(s_0)$$

where $\omega_1\colon [a,b] \to R$ *is given by* (6,5).

Proof. The assumptions assure by 6.6 that $\operatorname{var}_a^b k(.,t) < \infty$ for every $t \in [c,d]$ and consequently the limit $\lim_{s\to s_0+} k(s,t) = k(s_0+,t)$ exist for every $t \in [c,d]$. The statement follows immediately from 6.13.

6.16. Corollary. *If* $k\colon I \to R$, $\mathrm{v}_I(k) < \infty$, $\operatorname{var}_a^b k(.,c) < \infty$, *then for any* $s_0 \in [a,b)$ *we have*

$$\lim_{\delta\to 0+} \sup_{t\in[c,d]} |k(s_0+\delta,t) - k(s_0+,t)| = 0\,,$$

i.e.

$$\lim_{\delta\to 0+} k(s_0+\delta,t) = k(s_0+,t) \quad \textit{uniformly in } [c,d]\,.$$

Proof. For any $t \in [c,d]$ we have evidently

$$|k(s_0+\delta,t) - k(s_0+,t)| \leq |k(s_0+\delta,c) - k(s_0+,c)| + \operatorname{var}_c^d (k(s_0+\delta,.) - k(s_0+,.))$$

and our result follows immediately from the fact that $\lim_{s\to s_0+} k(s,c) = k(s_0+,c)$ exists and from 6.14.

6.17. Remark. It is easy to see that the statements from 6.14, 6.15 and 6.16 are also reformulable for the case of left-hand limits.

Further it is clear that 6.4–6.16 are also valid if the real function $k\colon I \to R$ is replaced by a matrix valued function $\mathbf{K}(s,t) = (k_{ij}(s,t))$. If some continuity properties are needed, then the usual norm of a matrix is used. Compare also 6.2.

6.18. Theorem. *Let* $k\colon I \to R$, $I = [a,b] \times [c,d]$ *be given. Let us suppose that* $v_I(k) < +\infty$ *and* $\operatorname{var}_c^d k(a, .) < \infty$.

If $g \in BV[c,d]$, *then the integral*

$$\int_c^d g(t)\, d_t[k(s,t)] \tag{6,12}$$

exists for every $s \in [a,b]$. *For any* $s \in [a,b]$ *the inequality*

$$\left|\int_c^d g(t)\, d_t[k(s,t)]\right| \le \int_c^d |g(t)|\, d_t[\operatorname{var}_c^t k(s, .)] \le \sup_{t\in[c,d]} |g(t)| \operatorname{var}_c^d k(s, .) \tag{6,13}$$

holds and moreover

$$\operatorname{var}_a^b\left(\int_c^d g(t)\, d_t[k(., t)]\right) \le \int_c^d |g(t)|\, d\omega_2(t) \le \sup_{t\in[c,d]} |g(t)|\, v_I(k) \tag{6,14}$$

where $\omega_2\colon [c,d] \to R$ *is defined by* (6,6). *Thus the integral* (6,12) *as a function of the variable s belongs to* $BV[a,b]$.

Proof. By 6.6 $k(s, .) \in BV[c,d]$ for every $s \in [a,b]$. Hence by 4.19 the integral (6,12) exists for every $s \in [a,b]$. The inequality (6,13) follows immediately from 4.27. In order to prove (6,14) we assume that an arbitrary subdivision $a = \alpha_0 < \alpha_1 < \ldots < \alpha_k = b$ of the interval $[a,b]$ is given. By 4.27 we have

$$\left|\int_c^d g(t)\, d_t[k(\alpha_i, t) - k(\alpha_{i-1}, t)]\right| \le \int_c^d |g(t)|\, d(\operatorname{var}_c^t (k(\alpha_i, .) - k(\alpha_{i-1}, .))).$$

Consequently

$$\sum_{i=1}^k \left|\int_c^d g(t)\, d_t[k(\alpha_i, t) - k(\alpha_{i-1}, t)]\right| \tag{6,15}$$

$$\le \int_c^d |g(t)|\, d\left(\sum_{i=1}^k \operatorname{var}_c^t (k(\alpha_i, .) - k(\alpha_{i-1}, .))\right).$$

Using 6.10 we obtain for all $t, \tau \in [a,b]$

$$|t-\tau|\, |g(\tau)| \left|\sum_{i=1}^k \operatorname{var}_c^t (k(\alpha_i, .) - k(\alpha_{i-1}, .)) - \sum_{i=1}^k \operatorname{var}_c^\tau (k(\alpha_i, .) - k(\alpha_{i-1}, .))\right|$$

$$\le (t-\tau)\, |g(\tau)|\, (\omega_2(t) - \omega_2(\tau))$$

since ω_2: $[c, d] \to R$ is nondecreasing and consequently 4.26 gives the estimate

$$\int_c^d |g(t)| \, \mathrm{d}\left(\sum_{i=1}^k \operatorname{var}_c^t (k(\alpha_i, .) - k(\alpha_{i-1}, .))\right) \le \int_c^d |g(t)| \, \mathrm{d}\omega_2(t).$$

Since this holds for every subdivision of $[a, b]$ we get by (6,15) the inequality

$$\operatorname{var}_a^b\left(\int_c^d g(t) \, \mathrm{d}_t[k(., t)]\right) \le \int_c^d |g(t)| \, \mathrm{d}\omega_2(t).$$

By 4.27 we have

$$\int_c^d |g(t)| \, \mathrm{d}\omega_2(t) \le \sup_{t \in [c,d]} |g(t)| \operatorname{var}_c^d \omega_2 = \sup_{t \in [c,d]} |g(t)| \, \mathrm{v}_I(k).$$

6.19. Corollary. *If the assumptions of* 6.18 *are satisfied, then*

$$\sup_{s \in [a,b]} \left|\int_c^d g(t) \, \mathrm{d}_t[k(s, t)]\right| \le \sup_{t \in [c,d]} |g(t)| \, (\operatorname{var}_c^d k(a, .) + \mathrm{v}_I(k)). \tag{6,16}$$

Proof. For any $s \in [a, b]$ we have by 4.27

$$\left|\int_c^d g(t) \, \mathrm{d}_t[k(s, t)]\right| \le \left|\int_c^d g(t) \, \mathrm{d}_t[k(a, t)]\right| + \operatorname{var}_a^b\left(\int_c^d g(t) \, \mathrm{d}_t[k(., t)]\right)$$
$$\le \sup_{t \in [c,d]} |g(t)| \operatorname{var}_c^d k(a, .) + \operatorname{var}_a^b\left(\int_c^d g(t) \, \mathrm{d}_t[k(., t)]\right).$$

(6,16) follows now easily from (6,14).

6.20. Theorem. *Let* k: $I = [a, b] \times [c, d] \to R$ *be given. Suppose that* $\mathrm{v}_I(k) < \infty$, $\operatorname{var}_c^d k(a, .) < \infty$ *and* $\operatorname{var}_a^b k(., c) < \infty$. *If* $f \in BV[a, b]$, $g \in BV[c, d]$, *then*

$$\int_c^d g(t) \, \mathrm{d}_t\left(\int_a^b k(s, t) \, \mathrm{d}f(s)\right) = \int_a^b \left(\int_c^d g(t) \, \mathrm{d}_t[k(s, t)]\right) \mathrm{d}f(s) \tag{6,17}$$

holds and the integrals on both sides of (6,17) *exist.*

Proof. By 6.18 $\int_c^d g(t) \, \mathrm{d}_t[k(., t)] \in BV[a, b]$ and 4.19 yields the existence of the integral on the right-hand side of (6,17). By 6.6 we obtain $k(., t) \in BV[a, b]$ for every $t \in [c, d]$ and by 4.19 also the existence of the integral $\int_a^b k(s, t) \, \mathrm{d}f(s)$ for any $t \in [c, d]$. Let $c = \gamma_0 < \gamma_1 < \ldots < \gamma_l = d$ be an arbitrary subdivision of $[c, d]$. For any $s \in [a, b]$ and $i = 1, \ldots, l$ we have

$$|k(s, \gamma_i) - k(s, \gamma_{i-1})|$$
$$\le |k(s, \gamma_i) - k(a, \gamma_i) - k(s, \gamma_{i-1}) + k(a, \gamma_{i-1})| + |k(a, \gamma_i) - k(a, \gamma_{i-1})|$$
$$\le \mathrm{v}_{[a,c] \times [\gamma_{i-1}, \gamma_i]}(k) + |k(a, \gamma_i) - k(a, \gamma_{i-1})|.$$

Hence by 4.27 and 6.5

$$\sum_{i=1}^{l} \left| \int_a^b (k(s,\gamma_i) - k(s,\gamma_{i-1}))\, \mathrm{d}f(s) \right|$$

$$\leq \sum_{i=1}^{l} \left[\mathrm{v}_{[a,b]\times[\gamma_{i-1},\gamma_i]}(k) + |k(a,\gamma_i) - k(a,\gamma_{i-1})| \right] \operatorname{var}_a^b f$$

$$\leq (\mathrm{v}_I(k) + \operatorname{var}_c^d k(a,\,.))\operatorname{var}_a^b f < \infty\,.$$

Taking the supremum over all finite subdivisions of $[c, d]$ on the left-hand side of this inequality we obtain

$$(6,18) \qquad \operatorname{var}_c^d \left(\int_a^b k(s,\,.\,)\, \mathrm{d}f(s) \right) \leq (\mathrm{v}_I(k) + \operatorname{var}_c^d k(a,\,.))\operatorname{var}_a^b f < \infty\,.$$

From 4.27 the existence of the integral on the left-hand side of (6,17) follows.

Let now $\alpha \in [c, d]$ and let $\psi_\alpha^+(t)$ be the simple jump function defined for $t \in [c, d]$ (see 4.20). By 4.21 we have

$$\int_c^d \psi_\alpha^+(t)\, \mathrm{d}_t[k(s,t)] = k(s,d) - k(s,\alpha+)$$

and

$$(6,19) \qquad \int_a^b \left(\int_c^d \psi_\alpha^+(t)\, \mathrm{d}_t[k(s,t)] \right) \mathrm{d}f(s) = \int_a^b (k(s,d) - k(s,\alpha+))\, \mathrm{d}f(s)\,.$$

On the other hand, we have by 4.21

$$(6,20) \qquad \int_c^d \psi_\alpha^+(t)\, \mathrm{d}_t \left[\int_a^b k(s,t)\, \mathrm{d}f(s) \right] = \int_a^b k(s,d)\, \mathrm{d}f(s) - \lim_{\delta\to 0+} \int_a^b k(s,\alpha+\delta)\, \mathrm{d}f(s)$$

$$= \lim_{\delta\to 0+} \int_a^b (k(s,d) - k(s,\alpha+\delta))\, \mathrm{d}f(s)\,.$$

By 4.27 we have

$$\left| \int_a^b (k(s,\alpha+) - k(s,\alpha+\delta))\, \mathrm{d}f(s) \right| \leq \sup_{s\in[a,b]} |k(s,\alpha+) - k(s,\alpha+\delta)| \operatorname{var}_a^b f$$

and by 6.16 we obtain

$$\lim_{\delta\to 0+} \int_a^b (k(s,\alpha+) - k(s,\alpha+\delta))\, \mathrm{d}f(s) = 0\,.$$

Hence by (6,20)

$$\int_c^d \psi_\alpha^+(t)\, \mathrm{d}_t \left[\int_a^b k(s,t)\, \mathrm{d}f(s) \right] = \int_a^b (k(s,d) - k(s,\alpha+))\, \mathrm{d}f(s)$$

and this together with (6,19) yields that for $g = \psi_\alpha^+$ the equality (6,17) is satisfie

In the same way it can be proved that (6,17) holds if we set $g(t) = \psi_\alpha^-(t)$, where ψ_α^- is the simple jump function given by (4,13). From these facts and from the linearity of the integral it is now clear that (6,17) holds whenever $g \in BV[c, d]$ is a finite step function (cf. 4.20).

Let now $g \in BV[c, d]$. There is a sequence $g_l \in BV[c, d]$, $l = 1, 2, \ldots$ of finite step functions such that $\lim_{l \to \infty} g_l(t) = g(t)$ uniformly on $[c, d]$ (see Aumann [1], 7.3.2.1).

Since by (6,18) it is $\int_a^b k(s, .)\, \mathrm{d}f(s) \in BV[c, d]$, we have by 4.17

$$\lim_{l \to \infty} \int_c^d g_l(t)\, \mathrm{d}_t \left(\int_a^b k(s, t)\, \mathrm{d}f(s) \right) = \int_c^d g(t)\, \mathrm{d}_t \left(\int_a^b k(s, t)\, \mathrm{d}f(s) \right). \tag{6,21}$$

Further by 6.19 we obtain

$$\sup_{s \in [a,b]} \left| \int_c^d (g(t) - g_l(t))\, \mathrm{d}_t[k(s, t)] \right| \leq \sup_{t \in [c,d]} |g(t) - g_l(t)| \, (\mathrm{var}_a^d\, k(a, .) + \mathrm{v}_I(k)).$$

Hence

$$\lim_{l \to \infty} \int_c^d g_l(t)\, \mathrm{d}_t[k(s, t)] = \int_c^d g(t)\, \mathrm{d}_t[k(s, t)]$$

uniformly on $[a, b]$ and by 4.17 the relation

$$\lim_{l \to \infty} \int_a^b \left(\int_c^d g_l(t)\, \mathrm{d}_t[k(s, t)] \right) \mathrm{d}f(s) = \int_a^b \left(\int_c^d g(t)\, \mathrm{d}_t[k(s, t)] \right) \mathrm{d}f(s) \tag{6,22}$$

holds. Since g_l are finite step functions we have for any $l = 1, 2, \ldots$

$$\int_a^b \left(\int_c^d g_l(t)\, \mathrm{d}_t[k(s, t)] \right) \mathrm{d}f(s) = \int_c^d g_l(t)\, \mathrm{d}_t \left(\int_a^b k(s, t)\, \mathrm{d}f(s) \right)$$

as was shown above. Consequently, by (6,21) and (6,22) we obtain the desired equality (6,17) and the proof is complete.

6.21. Remark. If all assumptions of 6.20 are satisfied, then it can be proved that the equality

$$\int_c^d g(t)\, \mathrm{d}_t \left(\int_a^b f(s)\, \mathrm{d}_s[k(s, t)] \right) = \int_a^b f(s)\, \mathrm{d}_s \left(\int_c^d g(t)\, \mathrm{d}_t[k(s, t)] \right) \tag{6,23}$$

also holds (see Schwabik [2]).

6.22. Theorem. *Let* $\mathbf{K}(s, t)$: $I = [a, b] \times [c, d] \to L(R_n)$ *be given,* $\mathbf{K}(s, t) = (k_{ij}(s, t))$, $i, j = 1, \ldots, n$. *Suppose that* $\mathrm{v}_I(\mathbf{K}) < \infty$, $\mathrm{var}_c^d\, \mathbf{K}(a, .) < \infty$, $\mathrm{var}_a^b\, \mathbf{K}(., c) < \infty$. *If* $\boldsymbol{x} \in BV_n[c, d]$, $\boldsymbol{y} \in BV_n[a, b]$, *then the equality*

$$\int_a^b \left(\int_c^d \mathrm{d}_t[\mathbf{K}(s, t)]\, \boldsymbol{x}(t) \right)^* \mathrm{d}\boldsymbol{y}(s) = \int_c^d \boldsymbol{x}^*(t)\, \mathrm{d}_t \left(\int_a^b \mathbf{K}^*(s, t)\, \mathrm{d}\boldsymbol{y}(s) \right) \tag{6,24}$$

holds and the integrals on both sides of (6,24) *exist.*

Proof. By definition we have

$$(6,25)\qquad \int_a^b \left(\int_c^d \mathrm{d}_t[\mathbf{K}(s,t)]\,\mathbf{x}(t)\right)^* \mathrm{d}\mathbf{y}(s)$$

$$= \sum_{i=1}^n \int_a^b \left(\sum_{j=1}^n \int_c^d x_j(t)\,\mathrm{d}_t[k_{ij}(s,t)]\right) \mathrm{d}y_i(s) = \sum_{i=1}^n \sum_{j=1}^n \int_a^b \left(\int_c^d x_j(t)\,\mathrm{d}_t[k_{ij}(s,t)]\right) \mathrm{d}y_i(s).$$

Since all x_j, y_i, k_{ij}, $i, j = 1, \dots, n$ satisfy the assumptions of 6.20 we can use this theorem for the interchanging of the order of integrations in the expression (6,25). If we do this we obtain

$$\int_a^b \left(\int_c^d \mathrm{d}_t[\mathbf{K}(s,t)]\,\mathbf{x}(t)\right)^* \mathrm{d}\mathbf{y}(s) = \sum_{i=1}^n \sum_{j=1}^n \int_c^d x_j(t)\,\mathrm{d}_t\left(\int_a^b k_{ij}(s,t)\,\mathrm{d}y_i(s)\right)$$

$$= \sum_{j=1}^n \int_c^d x_j(t)\,\mathrm{d}_t\left(\int_a^b \sum_{i=1}^n k_{ij}(s,t)\,\mathrm{d}y_i(s)\right) = \int_c^d \mathbf{x}^*(t)\,\mathrm{d}_t\left(\int_a^b \mathbf{K}^*(s,t)\,\mathrm{d}\mathbf{y}(s)\right)$$

and (6,24) is proved.

6.23. Remark. A similar formulation in terms of a matrix valued function $\mathbf{K}$ and vectors $\mathbf{x}, \mathbf{y}$ can be given for the equality (6,23) from 6.21.

6.24. Remark. In this paragraph only such results on functions of bounded variation in two variables are presented which are in some manner used in the forthcomming investigations of integral equations in the space BV_n. For the reader interested in this topic we refer to further results contained in the book Hildebrandt [1], III.4. (for example Helly's Choice Theorem, Jordan decomposition, etc.).

6.25. Remark. Let $I = [a, b] \times [c, d]$ be given. Let us denote by $SBV(I)$ the set of all functions $k\colon I \to R$ such that $v_I(K) < \infty$, $\operatorname{var}_a^b k(., c) < \infty$, $\operatorname{var}_c^d k(a, .) < \infty$. $SBV(I)$ is evidently a linear set. $SBV(I)$ can be normed by setting

$$\|k\| = |k(a, c)| + \operatorname{var}_a^b k(., c) + \operatorname{var}_c^d k(a, .) + v_I(k).$$

Evidently

$$|k(s, t)| \le \|k\| \qquad \text{for every} \quad (s, t) \in I.$$

The same holds even if the functions on I are matrix valued.

7. Nonlinear operators and nonlinear operator equations in Banach spaces

This section provides the basic tools for the investigation of nonlinear boundary value problems for ordinary differential equations contained in Chapter V. The reader interested in more details concerning differential and integral calculus on

Banach spaces is referred to the monographs on functional analysis (e.g. Kantorovič, Akilov [1]).

Throughout the paragraph, X, Y and Z are Banach spaces.

7.1. Preliminaries. Given a Banach space X with the norm $\|.\|_X$, $\varrho_0 > 0$ and $\boldsymbol{x}_0 \in X$, $\mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X)$ denotes the set of all $\boldsymbol{x} \in X$ such that $\|\boldsymbol{x} - \boldsymbol{x}_0\|_X \leq \varrho_0$.

Let $\boldsymbol{F}$ be an operator acting from X into Y and defined on $D \subset X$ ($\boldsymbol{F}$: $D \to Y$). $\boldsymbol{F}$ is *lipschitzian* on $D_0 \subset D$ if there exists a real number λ, $0 \leq \lambda < \infty$, such that

$$\|\boldsymbol{F}(\boldsymbol{x}') - \boldsymbol{F}(\boldsymbol{x}'')\|_Y \leq \lambda \|\boldsymbol{x}' - \boldsymbol{x}''\|_X$$

for all $\boldsymbol{x}', \boldsymbol{x}'' \in D_0$. If $\lambda < 1$, $\boldsymbol{F}$ is said to be *contractive* on D_0.

The operator $\boldsymbol{F}$: $D \subset X \times Z \to Y$ is said to be *locally lipschitzian* on $D_0 \subset D$ near $\boldsymbol{z} = \boldsymbol{z}_0$ if for any $\boldsymbol{x}_0 \in D_0$ there exist $\varrho_0 > 0$, $\sigma_0 > 0$ and $\lambda \geq 0$ such that $\boldsymbol{x}', \boldsymbol{x}'' \in \mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X)$ and $\boldsymbol{z} \in \mathfrak{B}(\boldsymbol{z}_0, \sigma_0; Z)$ implies $(\boldsymbol{x}', \boldsymbol{z}) \in D$, $(\boldsymbol{x}'', \boldsymbol{z}) \in D$ and

$$\|\boldsymbol{F}(\boldsymbol{x}', \boldsymbol{z}) - \boldsymbol{F}(\boldsymbol{x}'', \boldsymbol{z})\|_Y \leq \lambda \|\boldsymbol{x}' - \boldsymbol{x}''\|_X .$$

7.2. Gâteaux derivative. The operator $\boldsymbol{F}$ acting from X into Y and defined on $D \subset X$ is *Gâteaux differentiable at* $\boldsymbol{x}_0 \in D$ if there exists a bounded linear operator $\boldsymbol{G} \in B(X, Y)$ such that for any $\xi \in X$

$$\lim_{\vartheta \to 0} \left\| \frac{\boldsymbol{F}(\boldsymbol{x}_0 + \vartheta\xi) - \boldsymbol{F}(\boldsymbol{x}_0)}{\vartheta} - \boldsymbol{G}\xi \right\|_Y = 0 .$$

$\boldsymbol{G}$ is the *Gâteaux derivative* of $\boldsymbol{F}$ at $\boldsymbol{x} = \boldsymbol{x}_0$ and is denoted by $\boldsymbol{G} = \boldsymbol{F}'(\boldsymbol{x}_0)$. If $\boldsymbol{F}'(\boldsymbol{x})$ exists for all $\boldsymbol{x} \in D'$, where $D' \subset D$ is an open subset in X, and the mapping

$$\boldsymbol{F}': \boldsymbol{x} \in D_0 \to \boldsymbol{F}'(\boldsymbol{x}) \in B(X, Y)$$

possesses the Gâteaux derivative $\boldsymbol{H} \in B(X, B(X, Y))$ at $\boldsymbol{x} = \boldsymbol{x}_0 \in D_0$, $\boldsymbol{H}$ is said to be the *second order Gâteaux derivative* of $\boldsymbol{F}$ at $\boldsymbol{x} = \boldsymbol{x}_0$ and $\boldsymbol{H} = \boldsymbol{F}''(\boldsymbol{x}_0)$.

In general, if $\boldsymbol{H}$ is the *k-th order Gâteaux derivative* of $\boldsymbol{F}$ on $D_0 \subset D \subset X$ and $\boldsymbol{L}$ is the Gâteaux derivative of $\boldsymbol{H}$ at $\boldsymbol{x} = \boldsymbol{x}_0 \in D_0$, then $\boldsymbol{L}$ is the $(k+1)$-th order Gâteaux derivative of $\boldsymbol{F}$ at $\boldsymbol{x}_0$ and $\boldsymbol{L} = \boldsymbol{F}^{(k+1)}(\boldsymbol{x}_0)$.

Let $X_1, X_2, \ldots, X_n$ be Banach spaces. Let $\boldsymbol{F}$: $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \to \boldsymbol{F}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \in Y$ be an operator from the product space $\Xi = X_1 \times X_2 \times \ldots \times X_n$ into a Banach space Y. The derivative of $\boldsymbol{F}$ at a point $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ with respect to the j-th variable (i.e. if we fix the other variables and $\boldsymbol{F}$ is considered as an operator from X_j into Y) is denoted by $\boldsymbol{F}_j'(\boldsymbol{x})$ or $\boldsymbol{F}_{x_j}'(\boldsymbol{x})$. ($\boldsymbol{F}'(\boldsymbol{x})$ is defined and continuous on the open subset $D \subset \Xi$ if and only if $\boldsymbol{F}_j'(\boldsymbol{x})$ $(j = 1, 2, \ldots, n)$ are defined and continuous on D. Then for any $\boldsymbol{x} \in \Xi$ and $\xi = (\xi_1, \xi_2, \ldots, \xi_n) \in \Xi$

$$[\boldsymbol{F}'(\boldsymbol{x})]\,\xi = \sum_{j=1}^{n} [\boldsymbol{F}_j'(\boldsymbol{x})]\,\xi_j .) .$$

If on $D \subset \Xi$ $\boldsymbol{F}$ possesses all the derivatives $\boldsymbol{F}_j^{(p_j)}(\boldsymbol{x})$ $(j = 1, 2, \dots, n)$ which are continuous in $\boldsymbol{x}$ on D, we shall write $\boldsymbol{F} \in C^{p_1, p_2, \dots, p_n}(D)$. If $\boldsymbol{F}$ is continuous on D, we shall write $\boldsymbol{F} \in C(D)$.

Let us summarize some basic properties of the Gâteaux derivative.

(i) *Any linear mapping* $\boldsymbol{A} \in B(X, Y)$ *is Gâteaux differentiable on* X *and* $\boldsymbol{A}'(\boldsymbol{x}) = \boldsymbol{A}$ *for any* $\boldsymbol{x} \in X$.

(ii) *If the operators* $\boldsymbol{F}_1, \boldsymbol{F}_2$: $X \to Y$ *are Gâteaux differentiable at* $\boldsymbol{x}_0 \in X$ *and* $\alpha_1, \alpha_2 \in R$, *then also* $\alpha_1 \boldsymbol{F}_1 + \alpha_2 \boldsymbol{F}_2$ *is Gâteaux differentiable at* $\boldsymbol{x}_0$ *and*

$$(\alpha_1 \boldsymbol{F}_1 + \alpha_2 \boldsymbol{F}_2)'(\boldsymbol{x}_0) = \alpha_1 \boldsymbol{F}_1'(\boldsymbol{x}_0) + \alpha_2 \boldsymbol{F}_2'(\boldsymbol{x}_0).$$

(iii) *Let the operators* $\boldsymbol{F}$: $X \to Y$ *and* $\boldsymbol{G}$: $Y \to Z$ *be Gâteaux differentiable on open subsets* $D_{\boldsymbol{F}} \subset X$ *and* $D_{\boldsymbol{G}} \subset Y$ $(D_{\boldsymbol{F}} \supset \boldsymbol{F}(D_{\boldsymbol{F}}))$, *respectively. Then, if the mapping*

$$\boldsymbol{y} \in D_{\boldsymbol{G}} \subset Y \to \boldsymbol{G}'(\boldsymbol{y}) \in B(Y, B(Y, Z))$$

is continuous $(\boldsymbol{G} \in C^1(D_{\boldsymbol{G}}))$, *then the composed operator* $\boldsymbol{T} = \boldsymbol{G}\boldsymbol{F}$: $X \to Z$ *is Gâteaux differentiable on* D . *If, moreover,* $\boldsymbol{F} \in C^1(D\ )$, *then also* $\boldsymbol{T} \in C^1(D\ )$.

(iv) *If the operator* $\boldsymbol{F}$: $X \to Y$ *is Gâteaux differentiable at any point* $\boldsymbol{x}$ *of the domain* D *in* X *and* $\|\boldsymbol{F}'(\boldsymbol{x})\|_{B(X,Y)} \leq M < \infty$ *for any* $\boldsymbol{x} \in D$, *then* $\boldsymbol{F}$ *is lipschitzian on* D (*with the Lipschitz constant* M).

7.3. Abstract functions. *The operators acting from* R *into a Banach space* Y *are called abstract functions.*

The derivative $\boldsymbol{f}'$ of the abstract function $\boldsymbol{f}$: $R \to Y$ at the point $t_0 \in R$ is defined by

$$\lim_{t \to t_0} \left\| \frac{\boldsymbol{f}(t) - \boldsymbol{f}(t_0)}{t - t_0} - \boldsymbol{f}'(t_0) \right\|_Y = 0.$$

Let the abstract function $\boldsymbol{f}$: $R \to Y$ be defined and continuous on the interval $[a, b]$ $(-\infty < a < b < \infty)$. Then there exists $\boldsymbol{y} \in Y$ such that given $\varepsilon > 0$, there is a $\delta > 0$ such that for any subdivision $\sigma = \{a = t_0 < t_1 < \dots < t_{m_\sigma} = b\}$ of the interval $[a, b]$ with $(t_j - t_{j-1}) \leq \delta$ $(j = 1, 2, \dots, m_\sigma)$ and for an arbitrary choice of $t_j' \in (t_{j-1}, t_j)$ $(j = 1, 2, \dots, m_\sigma)$ it holds

$$\left\| \sum_{j=1}^{m_\sigma} \boldsymbol{f}(t_j')(t_j - t_{j-1}) - \boldsymbol{y} \right\|_Y < \varepsilon.$$

We denote

$$\boldsymbol{y} = \int_a^b \boldsymbol{f}(t)\, \mathrm{d}t$$

and $\boldsymbol{y}$ is said to be the abstract Riemann integral of $\boldsymbol{f}(t)$ over the interval $[a, b]$.

The abstract Riemann integral possesses analogous properties as the usual

Riemann integral of functions $[a, b] \to R$. In particular, if $\|\boldsymbol{f}(t)\|_Y \le M < \infty$ on $[a, b]$, then

$$\left\| \int_a^b \boldsymbol{f}(t)\,\mathrm{d}t \right\|_Y \le \int_a^b \|\boldsymbol{f}(t)\|_Y\,\mathrm{d}t \le M(b - a).$$

Furthermore, if $\boldsymbol{f}'$ exists and is continuous on $(\alpha, \beta) \supset [a, b]$, then

$$\int_a^b \boldsymbol{f}'(t)\,\mathrm{d}t = \boldsymbol{f}(b) - \boldsymbol{f}(a).$$

7.4. Lemma (Mean Value Theorem). *Let X, Y, Z be Banach spaces, and $\boldsymbol{x}_0 \in X$, $\boldsymbol{z}_0 \in Z$. Let the operator $\boldsymbol{F}$: $X \times Z \to Y$ be defined and Gâteaux differentiable on $\mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X) \times \mathfrak{B}(\boldsymbol{z}_0, \sigma_0; Z)$ $(\varrho_0 > 0,\ \sigma_0 > 0)$. Then for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X)$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathfrak{B}(\boldsymbol{z}_0, \sigma_0; Z)$*

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{x}_2, \boldsymbol{z}_2) - \boldsymbol{F}(\boldsymbol{x}_1, \boldsymbol{z}_1) = &\int_0^1 [\boldsymbol{F}'_x(\boldsymbol{x}_1 + \vartheta(\boldsymbol{x}_2 - \boldsymbol{x}_1), \boldsymbol{z}_1 + \vartheta(\boldsymbol{z}_2 - \boldsymbol{z}_1))](\boldsymbol{x}_2 - \boldsymbol{x}_1)\,\mathrm{d}\vartheta \\ &+ \int_0^1 [\boldsymbol{F}'_z(\boldsymbol{x}_1 + \vartheta(\boldsymbol{x}_2 - \boldsymbol{x}_1), \boldsymbol{z}_1 + \vartheta(\boldsymbol{z}_2 - \boldsymbol{z}_1))](\boldsymbol{z}_2 - \boldsymbol{z}_1)\,\mathrm{d}\vartheta. \end{aligned}$$

(The mapping

$$\begin{aligned} \vartheta \in [0, 1] \to &[\boldsymbol{F}'(\boldsymbol{x}_1 + \vartheta(\boldsymbol{x}_2 - \boldsymbol{x}_1), \boldsymbol{z}_1 + \vartheta(\boldsymbol{z}_2 - \boldsymbol{z}_1))][(\boldsymbol{x}_2, \boldsymbol{z}_2) - (\boldsymbol{x}_1, \boldsymbol{z}_1)] \\ = &[\boldsymbol{F}'_x(\boldsymbol{x}_1 + \vartheta(\boldsymbol{x}_2 - \boldsymbol{x}_1), \boldsymbol{z}_1 + \vartheta(\boldsymbol{z}_2 - \boldsymbol{z}_1))](\boldsymbol{x}_2 - \boldsymbol{x}_1) \\ &+ [\boldsymbol{F}'_z(\boldsymbol{x}_1 + \vartheta(\boldsymbol{x}_2 - \boldsymbol{x}_1), \boldsymbol{z}_1 + \vartheta(\boldsymbol{z}_2 - \boldsymbol{z}_1))](\boldsymbol{z}_2 - \boldsymbol{z}_1) \in Y \end{aligned}$$

is an abstract function.)

7.5. Theorem (Implicit Function Theorem). *Let X, Y and Z be Banach spaces, $\boldsymbol{x}_0 \in X$, $\boldsymbol{z}_0 \in Z$, $\varrho_0 > 0$, $\sigma_0 > 0$. Let the operator $\boldsymbol{F}$: $X \times Z \to Y$ be defined and continuous on $\mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X) \times \mathfrak{B}(\boldsymbol{z}_0, \sigma_0; Z)$, while*

(i) $\boldsymbol{F}(\boldsymbol{x}_0, \boldsymbol{z}_0) = \boldsymbol{0}$;
(ii) $\boldsymbol{F} \in C^{1,0}(\mathfrak{B}(\boldsymbol{x}_0, \varrho_0; X) \times \mathfrak{B}(\boldsymbol{z}_0, \sigma_0; Z))$ (cf. 7.2);
(iii) $\boldsymbol{F}'_x(\boldsymbol{x}_0, \boldsymbol{z}_0)$ *possesses a bounded inverse operator.*

Then there exist $\varrho > 0$ and $\sigma > 0$ such that for any $\boldsymbol{z} \in \mathfrak{B}(\boldsymbol{z}_0, \sigma; Z)$ there exists a unique solution $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{z}) \in \mathfrak{B}(\boldsymbol{x}_0, \varrho; X)$ to the equation

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{0}. \tag{7,1}$$

Moreover, the mapping $\boldsymbol{z} \in \mathfrak{B}(\boldsymbol{z}_0, \sigma; Z) \to \boldsymbol{\varphi}(\boldsymbol{z}) \in \mathfrak{B}(\boldsymbol{x}_0, \varrho; X)$ is continuous.

(Proof follows easily by applying Corollary 7.7 of the Contraction Mapping Principle 7.6 to the equation

$$\boldsymbol{x} = \boldsymbol{x} - [\boldsymbol{F}'_x(\boldsymbol{x}_0, \boldsymbol{z}_0)]^{-1}\,\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{z}).)$$

7.6. Theorem (Contraction Mapping Principle). *Let X be a Banach space and let $D \subset X$ be closed. Let the operator $\mathbf{T}$: $X \to X$ be contractive on D and $\mathbf{T}(D) \subset D$. Then there exists a unique $\mathbf{x} \in D$ such that $\mathbf{x} = \mathbf{T}(\mathbf{x})$.*
(The sought solution is the limit of successive approximations

$$\mathbf{x}_n = \mathbf{T}(\mathbf{x}_{n-1}) \quad (n = 1, 2, \dots),$$

where $\mathbf{x}_0$ may be an arbitrary element of D.)

7.7. Corollary. *Let X and Z be Banach spaces. Let $\mathbf{x}_0 \in X$, $\mathbf{z}_0 \in Z$, $\varrho_0 > 0$, $\sigma_0 > 0$, $0 \leq \lambda < 1$ and let $\mathbf{T}$ be a continuous mapping of $\mathfrak{B}(\mathbf{x}_0, \varrho_0; X) \times \mathfrak{B}(\mathbf{z}_0, \sigma_0; Z)$ into X such that*

(i) $$\|\mathbf{T}(\mathbf{x}_1, \mathbf{z}) - \mathbf{T}(\mathbf{x}_2, \mathbf{z})\|_X \leq \lambda \|\mathbf{x}_1 - \mathbf{x}_2\|_X$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{B}(\mathbf{x}_0, \varrho_0; X)$ and $\mathbf{z} \in \mathfrak{B}(\mathbf{z}_0, \sigma_0; Z)$;

(ii) $$\|\mathbf{T}(\mathbf{x}_0, \mathbf{z}) - \mathbf{x}_0\|_X < \varrho_0(1 - \lambda)$$

for all $\mathbf{z} \in \mathfrak{B}(\mathbf{z}_0, \sigma_0; Z)$.

Then, given $\mathbf{z} \in \mathfrak{B}(\mathbf{z}_0, \sigma_0; Z)$, there exists a unique element $\mathbf{x} = \boldsymbol{\varphi}(\mathbf{z}) \in \mathfrak{B}(\mathbf{x}_0, \varrho_0; X)$ such that $\mathbf{x} = \mathbf{T}(\mathbf{x}, \mathbf{z})$.

The mapping $\mathbf{z} \in \mathfrak{B}(\mathbf{z}_0, \sigma_0; Z) \to \boldsymbol{\varphi}(\mathbf{z}) \in \mathfrak{B}(\mathbf{x}_0, \varrho_0; X)$ is continuous.

Another version of the Implicit Function Theorem which is of interest for our purposes is the following theorem which also follows from the Contraction Mapping Principle.

7.8. Theorem. *Let X and Y be Banach spaces. Let $\mathbf{x}_0 \in X$, $\varrho_0 > 0$ and $\varkappa_0 > 0$. Let the operators $\mathbf{F}$: $X \to Y$ and $\mathbf{G}$: $X \times [0, \varkappa_0] \to Y$ satisfy the assumptions*

(i) $\mathbf{F}(\mathbf{x}_0) = \mathbf{0}$;
(ii) $\mathbf{F} \in C^1(\mathfrak{B}(\mathbf{x}_0, \varrho_0; X))$;
(iii) $\mathbf{F}'(\mathbf{x}_0)$ *possesses a bounded inverse operator*;
(iv) $\mathbf{G}$ *is locally lipschitzian on $\mathfrak{B}(\mathbf{x}_0, \varrho_0; X)$ near $\varepsilon = 0$.*

Then there exist $\varrho > 0$ and $\varkappa > 0$ such that for any $\varepsilon \in [0, \varkappa]$ there is a unique solution $\mathbf{x} = \boldsymbol{\varphi}(\varepsilon) \in \mathfrak{B}(\mathbf{x}_0, \varrho; X)$ of the equation

$$\mathbf{F}(\mathbf{x}) + \varepsilon\, \mathbf{G}(\mathbf{x}, \varepsilon) = \mathbf{0}\,. \tag{7,2}$$

Moreover, the mapping $\varepsilon \in [0, \varkappa] \to \boldsymbol{\varphi}(\varepsilon) \in \mathfrak{B}(\mathbf{x}_0, \varrho; X)$ is continuous.

7.9. Quasilinear equation — noncritical case. Of special interest are quasilinear (weakly nonlinear) equations of the form

$$\mathbf{L}\mathbf{x} - \varepsilon\, \mathbf{N}(\mathbf{x}, \varepsilon) = \mathbf{0}\,, \tag{7,3}$$

where $\mathbf{L}$ is a linear bounded operator acting from a Banach space X into a Banach space Y with the definition domain $D(\mathbf{L}) = X$ $(\mathbf{L} \in B(X, Y))$ and $\mathbf{N}$ is in general a nonlinear operator acting from $X \times R_1$ into Y.

The case when $\mathbf{L}$ possesses a bounded inverse operator is called *noncritical case*. In such a case the equation (7,3) is reduced to the equivalent equation

$$(7{,}4) \qquad \mathbf{x} = \varepsilon \mathbf{L}^{-1}\,\mathbf{N}(\mathbf{x}, \varepsilon)\,.$$

For $\varepsilon = 0$ (7,4) has the unique solution $\mathbf{x}_0 = \mathbf{0}$. To solve it for $\varepsilon > 0$ we may apply Theorem 7.8, where $\mathbf{F} = \mathbf{L}$ and $\mathbf{G} = -\mathbf{N}$.

7.10. Quasilinear equation — critical case. A linear bounded operator $\mathbf{L} \in B(X, Y)$ possesses a bounded inverse if and only if $N(\mathbf{L}) = \{\mathbf{0}\}$ and $R(\mathbf{L}) = Y$ (cf. Bounded Inverse Theorem 3.4).

In a general case when either $\dim N(\mathbf{L}) > 0$ or $R(\mathbf{L}) \subsetneqq Y$ the projection method may sometimes be used to consider the equation (7,3).

Let $\mathbf{L} \in B(X, Y)$ be such that

$$(7{,}5) \qquad R(\mathbf{L}) \text{ is closed}, \quad \alpha(\mathbf{L}) = \dim N(\mathbf{L}) < \infty\,, \quad \beta(\mathbf{L}) = \operatorname{codim} R(\mathbf{L}) < \infty$$

($\mathbf{L}$ is said to be *noetherian*). Then there exist linear bounded projections $\mathbf{P}$ of X onto $N(\mathbf{L})$ ($\mathbf{P} \in B(X)$, $R(\mathbf{P}) = N(\mathbf{L})$, $\mathbf{P}^2 = \mathbf{P}$) and $\mathbf{Q}$ of Y onto $R(\mathbf{L})$ ($\mathbf{Q} \in B(Y)$, $R(\mathbf{Q}) = R(\mathbf{L})$, $\mathbf{Q}^2 = \mathbf{Q}$) such that $R(\mathbf{I} - \mathbf{P})$ is closed in X, $\dim R(\mathbf{I} - \mathbf{Q}) = \beta(\mathbf{L})$ and

$$(7{,}6) \qquad X = N(\mathbf{L}) \oplus R(\mathbf{I} - \mathbf{P})\,, \qquad Y = R(\mathbf{L}) \oplus R(\mathbf{I} - \mathbf{Q})$$

(cf. Goldberg [1] II.1.14 and II.1.16). Thus $\mathbf{L}\mathbf{x} = \varepsilon\,\mathbf{N}(\mathbf{x}, \varepsilon)$ if and only if both

$$(7{,}7) \qquad \mathbf{Q}(\mathbf{L}\mathbf{x} - \varepsilon\,\mathbf{N}(\mathbf{x}, \varepsilon)) = \mathbf{L}\mathbf{x} - \varepsilon\mathbf{Q}\,\mathbf{N}(\mathbf{x}, \varepsilon) = \mathbf{0}$$

and

$$(7{,}8) \qquad (\mathbf{I} - \mathbf{Q})\,(\mathbf{L}\mathbf{x} - \varepsilon\,\mathbf{N}(\mathbf{x}, \varepsilon)) = -\varepsilon(\mathbf{I} - \mathbf{Q})\,\mathbf{N}(\mathbf{x}, \varepsilon) = \mathbf{0}\,.$$

Any $\mathbf{x} \in X$ may be written in the form $\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{I} - \mathbf{P})\,\mathbf{x}$. For $\mathbf{x} \in X$ let us denote $\mathbf{u} = (\mathbf{I} - \mathbf{P})\,\mathbf{x}$ and $\mathbf{v} = \mathbf{P}\mathbf{x}$. Then the system (7,7), (7,8) becomes

$$\mathbf{L}_1\mathbf{u} - \varepsilon\mathbf{Q}\,\mathbf{N}_1(\mathbf{u}, \mathbf{v}, \varepsilon) = \mathbf{0}\,, \qquad (\mathbf{I} - \mathbf{Q})\,\mathbf{N}_1(\mathbf{u}, \mathbf{v}, \varepsilon) = \mathbf{0}\,,$$

where

$$\mathbf{L}_1 :\ \mathbf{u} \in R(\mathbf{I} - \mathbf{P}) \to \mathbf{L}\mathbf{u} \in R(\mathbf{L}) = R(\mathbf{Q})$$

and

$$\mathbf{N}_1(\mathbf{u}, \mathbf{v}, \varepsilon) = \mathbf{N}(\mathbf{u} + \mathbf{v}, \varepsilon)$$

for $\mathbf{u} \in R(\mathbf{I} - \mathbf{P})$, $\mathbf{v} \in N(\mathbf{L})$ and $\varepsilon \in [0, \varkappa_0]$. Clearly, $\mathbf{L}_1 \in B(R(\mathbf{I} - \mathbf{P}), R(\mathbf{L}))$ is a one-to-one mapping of $R(\mathbf{I} - \mathbf{P})$ onto $R(\mathbf{L})$. ($\mathbf{L}_1\mathbf{u} = \mathbf{0}$ implies $\mathbf{u} \in R(\mathbf{P})$ and since $R(\mathbf{P}) \cap R(\mathbf{I} - \mathbf{P}) = \{\mathbf{0}\}$, $\mathbf{u} = \mathbf{0}$).

7.11. Theorem. *Let* $\mathbf{L} \in B(X, Y)$ *fulfil* (7,5) *and let* $\mathbf{P} \in B(X)$ *and* $\mathbf{Q} \in B(Y)$ *be the corresponding projections of* X *onto* $N(\mathbf{L})$ *and of* Y *onto* $R(\mathbf{L})$, *respectively. Let* $\mathbf{h} \in R(\mathbf{L})$ *and* $\mathbf{L}\mathbf{x}_0 = \mathbf{h}$.

Let $\varrho_0 > 0$, $\varkappa_0 > 0$ *and* $D = \mathfrak{B}(\mathbf{x}_0, \varrho_0; X) \times [0, \varkappa_0]$, *Let* $\mathbf{N} \in C^{1,0}(D)$, $\mathbf{N}(\mathbf{x}_0, 0) \in R(\mathbf{L})$ *and* $(\mathbf{I} - \mathbf{Q})\, \mathbf{N}'_x(\mathbf{x}_0, 0)$ *possesses a bounded inverse.*

Then there are $\varkappa > 0$ *and* $\varrho > 0$ *such that for any* $\varepsilon \in [0, \varkappa]$ *there exists a unique solution* $\mathbf{x} = \boldsymbol{\varphi}(\varepsilon) \in \mathfrak{B}(\mathbf{x}_0, \varrho; X)$ *of the equation*

$$\mathbf{L}\mathbf{x} = \mathbf{h} + \varepsilon\, \mathbf{N}(\mathbf{x}, \varepsilon). \tag{7,9}$$

The mapping $\boldsymbol{\varphi}\colon \varepsilon \in [0, \varkappa] \to \boldsymbol{\varphi}(\varepsilon) \in \mathfrak{B}(\mathbf{x}_0, \varrho; X)$ *is continuous.*

Proof. Let us denote $U = R(\mathbf{I} - \mathbf{P})$, $V = R(\mathbf{P}) = N(\mathbf{L})$. Then U and V are Banach spaces with the norms induced by $\|.\|_X$. Given $\mathbf{x} \in X$, let us put $\mathbf{u} = (\mathbf{I} - \mathbf{P})\,\mathbf{x}$ and $\mathbf{v} = \mathbf{P}\mathbf{x}$. In particular, $\mathbf{u}_0 = (\mathbf{I} - \mathbf{P})\,\mathbf{x}_0$, $\mathbf{v}_0 = \mathbf{P}\mathbf{x}_0$. Since $\mathbf{h} \in R(\mathbf{L})$, $(\mathbf{I} - \mathbf{Q})\,\mathbf{h} = \mathbf{0}$ and (7,9) becomes

$$\mathbf{L}_1\mathbf{u} - \mathbf{h} - \varepsilon\mathbf{Q}\,\mathbf{N}(\mathbf{u} + \mathbf{v}, \varepsilon) = \mathbf{0}, \qquad (\mathbf{I} - \mathbf{Q})\,\mathbf{N}(\mathbf{u} + \mathbf{v}, \varepsilon) = \mathbf{0},$$

where $\mathbf{L}_1 = \mathbf{L}|_U \in B(U, R(\mathbf{L}))$ possesses a bounded inverse. Let $D_1 \subset U \times V \times [0, \varkappa_0]$ denote the set of all $(\mathbf{u}, \mathbf{v}, \varepsilon) \in U \times V \times [0, \varkappa_0]$ such that $\|\mathbf{u} - \mathbf{u}_0\|_X \leq \frac{1}{2}\varrho_0$ and $\|\mathbf{v} - \mathbf{v}_0\|_X \leq \frac{1}{2}\varrho_0$. Given $(\mathbf{u}, \mathbf{v}, \varepsilon) \in D_1$, $(\mathbf{u} + \mathbf{v}, \varepsilon) \in D$ and we may define

$$\mathbf{T}(\mathbf{u}, \mathbf{v}, \varepsilon) = \begin{pmatrix} \mathbf{L}_1\mathbf{u} - \mathbf{h} - \varepsilon\mathbf{Q}\,\mathbf{N}(\mathbf{u} + \mathbf{v}, \varepsilon) \\ (\mathbf{I} - \mathbf{Q})\,\mathbf{N}(\mathbf{u} + \mathbf{v}, \varepsilon) \end{pmatrix} \in R(\mathbf{L}) \times R(\mathbf{I} - \mathbf{Q}).$$

Clearly, $\mathbf{T}$ is a continuous mapping of $D_1 \subset U \times V \times [0, \varkappa_0]$ into $Y \times Y$. Moreover, for any $(\mathbf{u}, \mathbf{v}, \varepsilon) \in D_1$ and $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in U \times V$

$$[\mathbf{T}'_{(u,v)}(\mathbf{u}, \mathbf{v}, \varepsilon)]\,(\boldsymbol{\xi}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{L}_1\boldsymbol{\xi} - \varepsilon\mathbf{Q}[\mathbf{N}'_x(\mathbf{u} + \mathbf{v}, \varepsilon)]\,(\boldsymbol{\xi} + \boldsymbol{\eta}) \\ (\mathbf{I} - \mathbf{Q})\,[\mathbf{N}'_x(\mathbf{u} + \mathbf{v}, \varepsilon)]\,(\boldsymbol{\xi} + \boldsymbol{\eta}) \end{pmatrix},$$

the mapping $(\mathbf{u}, \mathbf{v}, \varepsilon) \in D_1 \to \mathbf{T}'_{(u,v)}(\mathbf{u}, \mathbf{v}, \varepsilon) \in B(U \times V, Y \times Y)$ being continuous.

Since $\mathbf{N}(\mathbf{u}_0 + \mathbf{v}_0, 0) \in R(\mathbf{L})$ and $\mathbf{L}_1\mathbf{u}_0 = \mathbf{h}$, $\mathbf{T}(\mathbf{u}_0, \mathbf{v}_0, 0) = \mathbf{0}$. Moreover,

$$[\mathbf{T}'_{(u,v)}(\mathbf{u}_0, \mathbf{v}_0, 0)]\,(\boldsymbol{\xi}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{L}_1\boldsymbol{\xi} \\ (\mathbf{I} - \mathbf{Q})\,\mathbf{N}'_x(\mathbf{x}_0, 0)\,(\boldsymbol{\xi} + \boldsymbol{\eta}) \end{pmatrix}$$

for any $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in U \times V$. It is easy to see that for any $\mathbf{p} \in R(\mathbf{L})$ and $\mathbf{q} \in R(\mathbf{I} - \mathbf{Q})$

$$[\mathbf{T}'_{(u,v)}(\mathbf{u}_0, \mathbf{v}_0, 0)]\,(\boldsymbol{\xi}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}$$

if and only if $\boldsymbol{\xi} = \mathbf{L}_1^{-1}\mathbf{p}$ and $\boldsymbol{\eta} = [(\mathbf{I} - \mathbf{Q})\,\mathbf{N}'_x(\mathbf{x}_0, 0)]^{-1}\,\mathbf{q} - \boldsymbol{\xi}$. Applying the Implicit Function Theorem 7.5 we complete the proof.