Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 342 (2008) 334–339
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

A mean value theorem for systems of integrals ☆

Slobodanka Janković [a], Milan Merkle [b,*]

[a] *Mathematical Institute SANU, Knez Mihailova 35, 11000 Belgrade, Serbia*
[b] *Faculty of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11020 Belgrade, Serbia*

Received 17 July 2007

Available online 8 December 2007

Submitted by M. Laczkovich

Abstract

More than a century ago, G. Kowalewski stated that for each n continuous functions on a compact interval $[a, b]$, there exists an n-point quadrature rule (with respect to Lebesgue measure on $[a, b]$), which is exact for given functions. Here we generalize this result to continuous functions with an arbitrary positive and finite measure on an arbitrary interval. The proof relies on a new version of Carathéodory's convex hull theorem, that we also prove in the paper. As an application, we give a discrete representation of second order characteristics for a family of continuous functions of a single random variable.

Keywords: Carathéodory's convex hull theorem; Quadrature rules; Covariance; Correlation

1. Introduction and main results

By the end of the 19th century, G. Kowalewski [3] published the following result, in a paper entitled (in translation from German language) "A mean value theorem for a system of n integrals."

Theorem 1. *Let $x_1, \dots, x_n$ be continuous functions in a variable $t \in [a, b]$. There exist real numbers $t_1, \dots, t_n$ in $[a, b]$ and non-negative numbers $\lambda_1, \dots, \lambda_n$, with $\sum_{i=1}^n \lambda_i = b - a$, such that*

$$\int_a^b x_k(t)\,\mathrm{d}t = \lambda_1 x_k(t_1) + \cdots + \lambda_n x_k(t_n) \quad \text{for each } k = 1, 2, \dots, n.$$

In [4], Kowalewski generalized Theorem 1, with $\mathrm{d}t$ replaced with $F(t)\,\mathrm{d}t$, where F is continuous and of the same sign in (a, b), and with $\sum_{i=1}^n \lambda_i = \int_a^b F(t)\,\mathrm{d}t$. It seems that these results have not found their proper place in the literature; they were simply forgotten. Except citations to Kowalewski's Theorem 1 in [1] and [7], related to Grüss' and Chebyshev's inequalities, we were not able to trace any other attempt to use, or to generalize these results. In fact,

☆ This work is supported by the Ministry of Science and Environmental Protection of Serbia, project number 144021.
* Corresponding author.
E-mail addresses: bobaj@mi.sanu.ac.yu (S. Janković), emerkle@etf.bg.ac.yu (M. Merkle).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2007.12.012

Theorem 1 and its generalization, presented in [3,4], were honestly proved there only for $n = 2$; nevertheless there is an appealing beauty, and a potential for applications in those statements.

In this paper, we offer a generalization of Theorem 1, for an arbitrary interval I (not necessarily finite), with respect to any positive finite measure, and with functions x_i that are continuous, but (if I is open or infinite) not necessarily bounded.

Our main result is the following theorem.

Theorem 2. *For an interval $I \subseteq \mathbb{R}$, let μ be a finite positive measure on the Borel sigma-field of I. Let x_k, $k = 1, \dots, n$, $n \geqslant 1$, be continuous functions on I, integrable on I with respect to the measure μ. Then there exist points $t_1, \dots, t_n$ in I, and non-negative numbers $\lambda_1, \dots, \lambda_n$, with $\sum_{i=1}^n \lambda_i = \mu(I)$, such that*

$$\int_I x_k(t)\,\mathrm{d}\mu(t) = \sum_{i=1}^n \lambda_i x_k(t_i), \quad k = 1, \dots, n.$$

In Section 2 we prove Theorem 2 as a consequence of the following version of Carathéodory's convex hull theorem, which is also of an independent interest. We show that each point in the convex hull of a continuous curve in $\mathbb{R}^n$ is a convex combination of n points of the curve, rather than of $n + 1$ points, which would follow from the classical Carathéodory's theorem.

Theorem 3. *Let $C : t \mapsto x(t)$, $t \in I$, be a continuous curve in $\mathbb{R}^n$, where $I \subset \mathbb{R}$ is an interval, and let K be the convex hull of the curve C. Then each $v \in K$ can be represented as a convex combination of n or fewer points of the curve C.*

To complete the proof of Theorem 2 we needed a general result from the theory of integration (Lemma 1), which might be well-known, but we were not able to locate it in the literature.

In Section 3 we discuss Theorem 2 in the context of quadrature rules, and in Section 4 we apply Theorem 2 to derive a representation of the second order characteristics (expectations, variances, covariances and correlation coefficients) for a family of continuous functions of a single random variable.

2. Proofs of Theorems 2 and 3

Proof of Theorem 3. According to Carathéodory's theorem, any point $v \in K$ can be represented as a convex combination of at most $n + 1$ points of the curve C. Applying a translation if necessary, we may assume that $v = 0$. Therefore, there exist real numbers $t_j \in I$ and $v_j \geqslant 0$, $0 \leqslant j \leqslant n$, such that $t_0 < t_1 < \cdots < t_n$, $v_0 + \cdots + v_n = 1$, and

$$v_0 x(t_0) + v_1 x(t_1) + \cdots + v_n x(t_n) = 0. \tag{1}$$

In the sequel, we assume that all $n+1$ points $x(t_j)$ do not belong to one hyperplane in $\mathbb{R}^n$, and that the numbers v_j are all positive; otherwise, one term from (1) can be obviously eliminated. Denote by $p_j(x)$, $0 \leqslant j \leqslant n$, the coordinates of the vector $x \in \mathbb{R}^n$ with respect to the coordinate system with the origin at 0, and with the vector base consisted of vectors $x(t_j)$, $j = 1, \dots, n$ (that is, $x = \sum_{j=1}^n p_j(x) x(t_j)$). Then from (1) we find that $p_j(x(t_0)) = -v_j/v_0 < 0$, $j = 1, \dots, n$, i.e. the coordinates of the vector $x(t_0)$ are negative. The coordinates of vectors $x(t_j)$, $j = 1, 2, \dots, n$, are non-negative: $p_j(x(t_j)) = 1$ and $p_k(x(t_j)) = 0$ for $k \neq j$. Since the functions $t \mapsto p_j(x(t))$ are continuous, the set of points $t \geqslant t_0$ at which at least one of these functions reaches zero is closed, and since it is non-empty, it has the minimum. Denoting that minimum by $\bar{t}$, we conclude that the numbers $p_j(x(\bar{t}))$, $j = 1, \dots, n$, are non-positive and at least one of them is zero. Let $p_k(x(\bar{t})) = 0$ and $p_j(x(\bar{t})) \leqslant 0$ for $j \neq k$. Then

$$x(\bar{t}) - \sum_{j=1}^{k-1} p_j\big(x(\bar{t})\big) x(t_j) - \sum_{j=k+1}^{n} p_j\big(x(\bar{t})\big) x(t_j) = 0,$$

and it follows that 0 is a convex combination of points $x(\bar{t})$ and $x(t_j)$, $j = 1, \dots, n$, $j \neq k$. □

Proof of Theorem 2. Without loss of generality, we may assume that the measure μ is a probability measure, i.e. $\mu(I) = 1$. Since the Lebesgue integral of a function $x(t) = (x_1(t), \dots, x_n(t))$ with respect to μ is in the convex hull of the set $x(I) \subset \mathbb{R}^n$ (see Lemma 1 below), the statement follows from Theorem 3. □

Lemma 1. *Let $(S, \mathcal{F}, \mu)$ be a probability space, and let $x_i : S \to \mathbb{R}$, $i = 1, \dots, n$, be μ-integrable functions. Let $x(t) = (x_1(t), \dots, x_n(t))$ for every $t \in S$. Then $\int_S x(t)\,\mathrm{d}\mu(t) \in \mathbb{R}^n$ is in the convex hull of the set $x(S) = \{x(t) \mid t \in S\} \subset \mathbb{R}^n$.*

Proof. Without loss of generality we may assume that $\int_S x_i(t)\,\mathrm{d}\mu(t) = 0$ for $i = 1, \dots, n$. If $K = K(x, S) \subset \mathbb{R}^n$ is the convex hull of the set $x(S)$, we then need to prove that $0 \in K$. We give a proof by induction.

For $n = 1$, let $(S, \mathcal{F}, \mu)$ be an arbitrary probability space and let $x(t)$ be a μ-integrable function with $\int_S x(t)\,\mathrm{d}\mu(t) = 0$. Then there exists $t \in S$ such that $x(t) = 0$ or there exist $t_1, t_2 \in S$ such that $x(t_1) < 0$ and $x(t_2) > 0$. In both cases it follows that $0 \in K(x, S)$.

Let $n > 1$ and assume that the statement is true for $n - 1$ functions on any probability space. Let now $x_1, \dots, x_n$ be integrable functions on a probability space $(S, \mathcal{F}, \mu)$. Assume, contrary to what has to be proved, that $0 \notin K(x, S)$. Then there is a hyperplane that separates 0 and K, hence there are real numbers $a_1, \dots, a_n$, such that

$$L(t) := \sum_{k=1}^{n} a_k x_k(t) \geqslant 0 \quad \text{for every } t \in S. \tag{2}$$

Here we may assume that $a_n \neq 0$. By linearity, $\int_S L(t)\,\mathrm{d}\mu(t) = 0$, hence (2) yields $L(t) = 0$ for all $t \in S \setminus N$, where $\mu(N) = 0$. Therefore,

$$x_n(t) = -\sum_{k=1}^{n-1} \frac{a_k}{a_n} x_k(t) \quad \text{for every } t \in S \setminus N. \tag{3}$$

Now observe a probability space $(S \setminus N, \mathcal{F}|_{S\setminus N}, \mu|_{S\setminus N})$ and the system of functions $x_1, \dots, x_{n-1}$. By the induction assumption, 0 belongs to the convex hull of the set $\{(x_1(t), \dots, x_{n-1}(t)) \mid t \in S \setminus N\} \in \mathbb{R}^{n-1}$, so we find that

$$\sum_{i=1}^{m} \lambda_i x_k(t_i) = 0, \quad k = 1, \dots, n-1, \tag{4}$$

with some $t_1, \dots, t_m \in S \setminus N$, $0 \leqslant \lambda_i \leqslant 1$ and $\sum_{i=1}^{m} \lambda_i = 1$, and for some integer $m \geqslant 1$. Now from (3) and (4) it follows that (4) holds also for $k = n$, which implies that the statement holds for n. This ends the proof by induction. □

3. Theorem 2 from a viewpoint of quadrature rules

Theorem 2 claims that, given any set of continuous functions on I, and a finite measure μ on I, there exists an n-point quadrature rule which is exact for those functions. As it can be seen by inspection of the proofs in Section 2, this quadrature rule is not unique; a point in the interior of a convex hull can be expressed as a convex combination in infinitely many ways. This interpretation of Theorem 2 can be compared with a well-known result from [2], regarding Gaussian quadratures with respect to Chebyshev systems of functions. A brief explanation of these terms is in order.

Real functions $x_1, \dots, x_m$ defined on an interval $[a, b]$ are said (see [2]) to constitute a Chebyshev system on $[a, b]$ if all functions are continuous on $[a, b]$ and

$$\begin{vmatrix} x_1(t_1) & x_1(t_2) & \dots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \dots & x_2(t_m) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(t_1) & x_m(t_2) & \dots & x_m(t_m) \end{vmatrix} \neq 0 \tag{5}$$

for any choice of points $t_1, \dots, t_m \in [a, b]$ with $t_i \neq t_j$ whenever $i \neq j$. A classical example of a Chebyshev system on any interval $[a, b]$ is furnished with functions $x_i(t) = t^{i-1}$, $i = 1, \dots, m$. The condition (5) is equivalent to the requirement that no m points of the curve parametrized by $x_1 = x_1(t)$, …, $x_m = x_m(t)$, $t \in [a, b]$, belong to a hyperplane which contains the origin. Another way to express (5) is to require that any function of the form $g(t) = c_1 x_1(t) + \dots + c_m x_m(t)$, $c_i \in \mathbb{R}$, $\sum_{i=1}^{m} c_i^2 > 0$, must not have more than $m - 1$ different zeros on $[a, b]$.

According to [5], for a positive and finite measure μ on $[a, b]$, a quadrature rule of the form

$$\int_{[a,b]} f(s)\,\mathrm{d}\mu(s) = \sum_{k=1}^{n} A_k f(t_k) + R_n(f), \quad A_k \in \mathbb{R},\ t_k \in [a, b], \tag{6}$$

is called Gaussian with respect to a collection of functions $x_1, \dots, x_{2n}$ if (6) is exact for all functions x_i in place of f, i.e. if $R_n(x_i) = 0$ for $i = 1, \dots, 2n$. A quadrature rule of the form (6) is determined by a choice of coefficients A_k and points t_k, $k = 1, \dots, n$.

The next theorem, which can be derived from [2, Chapter 2], claims the existence and uniqueness of a Gaussian quadrature rule with respect to a Chebyshev system of continuous functions $x_1, \dots, x_{2n}$ on $[a, b]$.

Theorem 4. *(See [2].) There exists a unique n-point Gaussian quadrature rule* (6) *with respect to any Chebyshev system of continuous functions $x_1, \dots, x_{2n}$ on a finite interval $[a, b]$. Moreover, all coefficients $A_1, \dots, A_n$ are positive.*

There are variations and generalizations of Theorem 4 in various directions, see, for example, [5] or recent paper [6].

Clearly, Theorem 4 yields a particular case of Theorem 2 if functions $x_1, \dots, x_n$ can be complemented with suitably chosen functions (for example $1, t, t^2, \dots, t^n$) to make a Chebyshev system of $2n$ functions on interval $[a, b]$. However, Theorem 2 is much more general, it is not limited to compact intervals, it allows unbounded functions, and above all, it does not require the condition (5), which is very restrictive and difficult to check.

Let us finally note that Theorem 2 can be applied to a system of n integrals with respect to different measures which are absolutely continuous with respect to one measure, with continuous Radon–Nikodym derivatives. Indeed, suppose that $\mu_1, \dots, \mu_n$ are positive finite measures on an interval I, with continuous Radon–Nikodym derivatives $\varphi_1, \dots, \varphi_n$ with respect to a measure μ. Then applying Theorem 2 to the system of $2n$ integrals with respect to μ, we get, for $k = 1, \dots, n$:

$$\int_I x_k(t)\,\mathrm{d}\mu_k(t) = \int_I x_k(t)\varphi_k(t)\,\mathrm{d}\mu(t) = \sum_{i=1}^{2n} \lambda'_i x_k(t_i)\varphi_k(t_i), \tag{7}$$

$$\mu_k(I) = \int_I \varphi_k(t)\,\mathrm{d}\mu(t) = \sum_{i=1}^{2n} \lambda'_i \varphi_k(t_i), \tag{8}$$

for some λ'_i and t_i, $i = 1, \dots, 2n$. Introducing $\lambda_{ki} = \lambda'_i \varphi_k(t_i)$, we have that

$$\int_I x_k(t)\,\mathrm{d}\mu_k(t) = \sum_{i=1}^{2n} \lambda_{ki} x_k(t_i), \quad \text{where } \sum_{i=1}^{2n} \lambda_{ki} = \mu_k(I), \quad k = 1, \dots, n.$$

Hence, the points $t_1, \dots, t_{2n}$ are common for all n quadratures, whereas the weights are specified for each one.

4. A representation of second order characteristics for continuous functions of a single random variable

Given a measurable function f and a probability measure μ on the Borel sigma field $\mathcal{B}$ on $\mathbb{R}$, with $\mu(I) = 1$, the integral

$$\int_I f(x)\,\mathrm{d}\mu(x)$$

can be thought of as the expectation $\mathrm{E}f(X)$, where X is a random variable on some probability space $(\Omega, \mathcal{F}, P)$ with $P(X \in B) = \mu(B)$ for $B \in \mathcal{B}$, concentrated on I, i.e. $P(X \in I) = 1$. The covariance between $f(X)$ and $g(X)$ is then

$$\begin{aligned}\mathrm{Cov}\big(f(X), g(X)\big) &= \mathrm{E}\big(f(X) - \mathrm{E}f(X)\big)\big(g(X) - \mathrm{E}g(X)\big)\\ &= \int_I f(x)g(x)\,\mathrm{d}\mu(x) - \int_I f(x)\,\mathrm{d}\mu(x)\int_I g(x)\,\mathrm{d}\mu(x).\end{aligned}$$

A probabilistic statement that directly corresponds to Theorem 2 is the following one: *For any given random variable X and continuous functions $f_1, \dots, f_n$, there exists a discrete random variable T, taking at most n values, $t_1, \dots, t_n$, with probabilities $\lambda_1, \dots, \lambda_n$, such that*

$$\mathrm{E}f_i(X) = \mathrm{E}f_i(T), \quad i = 1, 2, \dots, n.$$

A similar statement holds for the second order characteristics, as in the next theorem.

Theorem 5. *Let X be a random variable concentrated on an interval $I \subseteq \mathbb{R}$. Suppose that $f_1, \dots, f_n$ are continuous functions on I, with finite $\mathrm{E} f_i(X)^2$, $i = 1, \dots, n$. Then there exist $m = n(n+3)/2$ points $t_1, \dots, t_m$ from I, and non-negative numbers $\lambda_1, \dots, \lambda_m$ with $\sum \lambda_k = 1$, such that*

$$\mathrm{E} f_i(X) = \sum_{k=1}^{m} \lambda_k f_i(t_k), \quad i = 1, \dots, n, \tag{9}$$

$$\mathrm{Cov}\big(f_i(X), f_j(X)\big) = \sum_{k=1}^{m-1} \sum_{l=k+1}^{m} \lambda_k \lambda_l \big(f_i(t_k) - f_i(t_l)\big)\big(f_j(t_k) - f_j(t_l)\big). \tag{10}$$

Proof. Let $F_i = \mathrm{E} f_i(X)$ and $C_{ij} = \mathrm{Cov}(f_i(X), f_j(X))$, $i, j = 1, \dots, n$. Since $C_{ij} = C_{ji}$, the total number of different expectations is $m = n(n+3)/2$. By Theorem 2, there are points $t_1, \dots, t_m$ and non-negative coefficients $\lambda_1, \dots, \lambda_m$, with $\sum \lambda_i = 1$, such that

$$F_i = \sum_{k=1}^{m} \lambda_k f_i(t_k), \quad i = 1, \dots, n, \quad \text{and} \tag{11}$$

$$C_{ij} = \sum_{k=1}^{m} \lambda_k \big(f_i(t_k) - F_i\big)\big(f_j(t_k) - F_j\big), \quad i, j = 1, \dots, n. \tag{12}$$

For a fixed $i \leqslant j$, expressions (11) and (12) yield

$$C_{ij} = \sum_{k=1}^{m} \lambda_k \left(f_i(t_k) - \sum_{l=1}^{m} \lambda_l f_i(t_l) \right) \big(f_j(t_k) - F_j\big),$$

which, after some organizing reduces to (10), and the theorem is proved. $\square$

Let us note that (10) can be represented in a symmetric form

$$\mathrm{Cov}\big(f_i(X), f_j(X)\big) = \frac{1}{2} \sum_{k=1}^{m} \sum_{l=1}^{m} \lambda_k \lambda_l \big(f_i(t_k) - f_i(t_l)\big)\big(f_j(t_k) - f_j(t_l)\big).$$

Clearly, Theorem 5 yields representations for variances, correlation coefficients and the covariance matrix in terms of m points and weights. For a single covariance $\mathrm{Cov}(f(X), g(X))$, where f, g are continuous functions, there is a simpler representation. Indeed, we can start with the representation of Theorem 2,

$$\mathrm{Cov}\big(f(X), g(X)\big) = \lambda\big(f(t_1) - F\big)\big(g(t_1) - G\big) + (1-\lambda)\big(f(t_2) - F\big)\big(g(t_2) - G\big),$$
$$F = \lambda f(t_1) + (1-\lambda) f(t_2),$$

where $F = \mathrm{E} f(X)$, $G = \mathrm{E} g(X)$, and, in the same way as in the proof of Theorem 5, we find that

$$\mathrm{Cov}\big(f(X), g(X)\big) = \lambda(1-\lambda)\big(f(t_1) - f(t_2)\big)\big(g(t_1) - g(t_2)\big), \tag{13}$$

for some $\lambda \in [0, 1]$ and $t_1, t_2 \in I$. The term $\lambda(1-\lambda)$ reaches its maximum $1/4$ for $\lambda = 1/2$, and by continuity of functions f, g, there are another two points, call them again t_1 and t_2, so that (13) holds with $\lambda = 1/2$. Hence, we have the following result.

Theorem 6. *Given a random variable X concentrated on an interval $I \subseteq \mathbb{R}$, and functions f and g that are continuous on I, there exist points $t_1, t_2 \in I$, such that*

$$\mathrm{Cov}\big(f(X), g(X)\big) = \frac{1}{4}\big(f(t_1) - f(t_2)\big)\big(g(t_1) - g(t_2)\big). \tag{14}$$

The representation (14) is a generalization of Karamata's result [1] for the Lebesgue measure on a compact interval I, i.e. the uniform distribution of X.

Acknowledgments

We are grateful to the referee for useful suggestions and comments about the presentation in Section 2, in particular concerning the formulation and the proof of Lemma 1. We also thank Gradimir V. Milovanović for a discussion related to quadrature rules and Theorem 4. Second author acknowledges his partial affiliation to Računarski fakultet, Beograd, Serbia, and Instituto de Mathemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

References

[1] J. Karamata, Sur certain inégalités relatives aux quotients et à la difference de $\int fg$ et $\int f \int g$, Publ. Inst. Math. (Beograd) 2 (1948) 131–145.
[2] S. Karlin, W. Studden, Tchebycheff Systems with Applications in Analysis and Statistics, Wiley–Interscience, New York, 1966.
[3] G. Kowalewski, Ein Mittelwertsatz für ein System von n Integralen, Z. Math. Phys. (Schlömilch Z.) 42 (1895) 153–157.
[4] G. Kowalewski, Bemerkungen zu dem Mittelwertsatze für ein System von n Integralen, Z. Math. Phys. (Schlömilch Z.) 43 (1896) 118–120.
[5] J. Ma, V. Rokhlin, S. Wandzura, Generalized Gaussian quadrature rules for systems of arbitrary functions, SIAM J. Numer. Anal. 3 (3) (1996) 971–996.
[6] G.V. Milovanović, A.S. Cvetković, Gaussian-type quadrature rules for Müntz systems, SIAM J. Sci. Comput. 27 (3) (2005) 893–913.
[7] D.S. Mitrinović, P.M. Vasić, History, variations and generalizations of the Čebyšev inequality and the question of some priorities, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. 461–497 (1974) 1–30.