AKADEMIA TECHNICZNO-ROLNICZA
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

ZESZYTY NAUKOWE NR 247

TELEKOMUNIKACJA I ELEKTRONIKA 9

WYDZIAŁ TELEKOMUNIKACJI I ELEKTROTECHNIKI

BYDGOSZCZ – 2006

AKADEMIA TECHNICZNO-ROLNICZA
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

ZESZYTY NAUKOWE NR 247

TELEKOMUNIKACJA I ELEKTRONIKA

9

BYDGOSZCZ – 2006

ISSN 0209-0589

Wydawnictwa Uczelniane Akademii Techniczno-Rolniczej
ul. Ks. A. Kordeckiego 20, 85-225 Bydgoszcz, tel. (052) 3749482, 3749426
e-mail: wydawucz@atr.bydgoszcz.pl http://www.atr.bydgoszcz.pl/~wyd

---

Wyd. I. Nakład 500 egz. Ark. aut. 3,2. Ark. druk. 4,0. Zamówienie nr 13/2006
Oddano do druku i druk ukończono w sierpniu 2006 r.
Uczelniany Zakład Małej Poligrafii ATR Bydgoszcz, ul. Ks. A. Kordeckiego 20

Contents

FOREWORD

This year, the Faculty of Telecommunications and Electrical Engineering celebrates 45th jubilee of existence. I am even more pleased that, after long break, the Scientific Book of Telecommunication and Electronics could be published. I express my hopes for publishing more numbers in the future.

Last year academical society of University of Technology and Agriculture (ATR) paid respects to departed Tadeusz Wysocki - first dean of our faculty. He was a man who has made a remarkable contribution to the Faculty of Telecommunications and Electrical Engineering, other faculties of ATR and the society of Bydgoszcz city. The memory about Tadeusz Wysocki was written by his daughter Felicja Wysocka-Schillak.

There were five articles published in the issue. Three of them were sent from abroad. Therefore the matter of two fields used in electronics and telecommunications: signal processing and analysis of nonlinear electronic systems was brought up.

The first article by Irwin W. Sandberg (The University of Texas at Austin, USA) includes compilation of results his works dealing with linear maps, which are models of multidimensional, linear and shift-invariant systems used in signal processing.

Luca Lucchese and Sanjit K. Mitra (Oregon State University, Corvallis and University of California, Santa Barbara, USA, respectively) present the article which deals with the color images processing. Authors are concentrated on determining the color mapping between two images of the same object or scene taken under different illumination conditions.

In A. H. Zemanian's (State University of New York at Stony Brook, New York, USA) article, non-standard versions of Eulerian graphs, Hamiltonian graphs, and a coloring theorem are established for non-standard graphs. In this paper the author proposes to extend the notion of well-known Eulerian and Hamiltonian graphs as well as the coloring theorem, all originally specified for standard graphs, to be valid for non-standard graphs.

Andrzej Borys (The University of Technology and Agriculture at Bydgoszcz, Poland) presents questions on analysis of nonlinear electronic systems with the use of Volterra series in his two articles.

I encourage reading this issue, and also invite specialists from this department to publish scientific articles in columns of our Scientific Book of Telecommunication and Electronics.

With friendly greetings

Editor of Telecommunication and Electronics
Sławomir Cieślik

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
TELEKOMUNIKACJA I ELEKTRONIKA 9 (2006), 7-8

MEMORY ABOUT TADEUSZ WYSOCKI THE FIRST DEAN OF FACULTY OF TELECOMMUNICATIONS AND ELECTRICAL ENGINEERING

Felicja Wysocka-Schillak

Institute for Telecommunications
Faculty of Telecommunications and Electrical Engineering
University of Technology and Agriculture
ul. Kaliskiego 7, 85-791 Bydgoszcz, Poland

Last year, we paid the final tribute to Tadeusz Wysocki, the founding dean of our faculty, the Faculty of Telecommunications Engineering at the Engineering Evening College (Tertiary Engineering College from 1964), Bydgoszcz.

Tadeusz Wysocki was born on the 8th of June 1921 in Piotrkow Trybunalski. From 1931 until 1937, he attended the Leszno Classical High School. Except during WW2, he lived in Bydgoszsz since 1937, and in 1939 completed his Final High School Exams at the Copernicus Mathematics-Physics High School. He undertook his undergraduate studies in Electrical Engineering at the Gdansk Technical University from 1945-1950, where he received the Master of Engineering Science degree. In 1951, he commenced employment at the Mechanical-Electrical Technical College in Bydgoszcz. From September 1951 until 1993, he worked at the Engineering Evening College, which evolved into the Tertiary Engineering College in 1964 and in 1974, into the Academy of Technology and Agriculture, Bydgoszcz.

He was one of the founders of the Faculty of Mechanical Engineering at the Engineering Evening College. From 1951-1953, he created from scratch, a physics and a general electrical technology laboratory at the faculty. From 1954 to 1955, he was the associate – dean and from 1962, the dean of the Faculty of Mechanical Engineering. In 1957, together with the associate – dean Zdzislaw Arnold, he developed a program to expand the faculty, which was realized in 1958 when the Faculty was relocated to the current building 'C' on Kordecki Street. In 1961, he developed a curriculum for working students, the "stacjonarno-zaoczny" system, whereby students were coming to the university from remote areas only couple of times during each semester for intensive studies and were doing their project assignments at home.

In 1962, he co-established the Faculty of Telecommunications Engineering, and became its first dean until 1967. During that time, he was also a chair of the Department of Electronics Fundamentals. From 1973-1975, he headed the Teaching Innovations Laboratory, where he introduced the Teaching Techniques course for the academics. He was a successful and respected academic teacher. He devoted his whole professional life to educating future engineers, and was inspiration for many generations of graduates. They could always count on his answers no matter if related to technical challenges or to common life problems. His personality was so strong, that his children, the daughter in-law as well as two grandsons have also chosen their careers in Electrical or Telecommunications Engineering.

Tadeusz Wysocki was also involved in numerous areas outside the Academia. From 1954-1967, he was a member of the inter-university subjects committee, specifically examining and developing curricula and teaching programs. From 1967 until 1984, he was a member of the Electronics team at the Institute of Methodological Tertiary Studies for Working Students, in Gliwice. There, he developed a study entitled "Fundamentals of supervising engineering thesis projects for working students (major in electronics engineering)", published in 1974. This was the first ever comprehensive study into this issue in Poland.

In 1958, he was nominated to the National Science & Economics Council, Bydgoszcz Division and from 1965-1967, he served as a president of that Council. From 1951, Tadeusz Wysocki was an active member of the Association of Polish Electrical Engineers (SEP). There, he worked on the establishment of a Faculty of Electrical Engineering at the Tertiary Engineering College in Bydgoszcz. For three terms, since 1984, he served as a Deputy of the Bydgoszcz Division of SEP. From 1960-1964, he was the deputy head of the Regional Arbitration Committee of the Polish Federation of Engineering Associations (NOT) for Bydgoszcz Region. He was a founding member of the Bydgoszcz Scientific Society, and was the secretary of its Faculty of Technical Studies for over forty years.

Tadeusz Wysocki was an author or co-author of 2 monographs, 8 textbooks, and 52 research papers. He presented the outcomes of his research at many scientific conferences or symposia. He was the inventor or co-inventor of 36 patents and 3 registered designs. After his retirement in 1986, Tadeusz remained an active researcher, involved in both fundamental and applied research projects.

Tadeusz Wysocki received numerous rewards for his accomplishments, including a host of national and district awards, as well as from various societies, including the Cavalry Cross of the Poland Rebirth Order, Golden Cross of Achievements, The National Education Commission Medal, the Golden Insignia of Honor from NOT, Golden Insignia of Honor from SEP, the Golden Insignia "Honored by Telecommunication Sector" and The Honor Insignia for Services to the City of Bydgoszcz. For his achievement in education, he was awarded the Individual Education Minister's Prize in 1986. For his efforts in establishing the Academy of Technology and Agriculture (ATR), he was awarded the ATR Medal of Service, the 50th Anniversary Medal of Service to the Faculty of Mechanical Engineering and the 40th Anniversary Medal of Service to the Faculty of Telecommunications and Electrical Engineering. In 2005, he received the 85th Independence Anniversary Medal of Bydgoszcz.

Tadeusz Wysocki had far-reaching interests and incredible talents, as well as a caring nature and a healthy sense of humor.

Tadeusz Wysocki passed away on the 12th of June 2005 in Bydgoszcz. His funeral was held on the 17th of June at the Sacred Heart of Our Lord parish cemetery, on Ludwikowo Street in Bydgoszcz.

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
ELEKTRONIKA I TELEKOMUNIKACJA 9 (2006), 9-24

A SHORT SURVEY OF RECENT REPRESENTATION RESULTS FOR LINEAR SYSTEM MAPS

Irwin W. Sandberg
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712, USA

Summary: We give an expression for the most general input-output map associated with the members of a certain important large family of multidimensional linear shift-invariant systems with bounded Lebesgue-measurable inputs. The expression given is an iterated function-space limit of a convolution. We also give a necessary and sufficient condition under which the limit can be written as a convolution with an integrable impulse-response function. A key role is played by a certain family of weighting operators. It is observed that for the large family of inputs and maps addressed, the Dirac impulse-response concept is in fact not the key concept concerning the representation of H, and that instead the input-output properties of H are determined in general by a certain type of family of responses. Some related material concerning other results, engineering education, and discrete-space systems, is also given.

Keywords: linear systems, multidimensional systems, impulse responses, shift-invariant systems, bounded measurable inputs, discrete-space systems.

1. INTRODUCTION

In the signal-processing literature, $x(\alpha)$ typically denotes a function. In the following we distinguish between a function x and $x(\alpha)$, the latter meaning the value of x at the point (or time) α. Sometimes a function x is denoted by $x(\cdot)$, and also we use Hx to mean $H(x)$. This notation is often useful in studies of systems in which signals are transformed into other signals.

A recent paper [14] considers continuous-time linear time-invariant systems governed by a relation $y = Hx$ in which x is an input, y is the corresponding output, and H is the system map that takes inputs into outputs. It was assumed that inputs and outputs are complex-valued functions defined on the set $\mathbb{R}$ of real numbers. As is well known, it is a widely-held belief of long standing that the input-output properties of H are completely described by its impulse response. Using a standard interpretation of what is meant by a system's impulse response, it is shown in [14] that this belief is incorrect in a simple setting in which x is drawn from the linear space of bounded uniformly-continuous complex-valued functions defined on $\mathbb{R}$. More specifically, it was shown that there is an H of the kind described above, even a causal H, whose impulse response is the zero function, but which takes certain inputs into nonzero outputs.[1] This contradicts the conclusion of a familiar engineering argument using the so-called sifting property of Dirac's impulse function (see Section 2.1). An important role in [14] is played by inputs that do not approach zero at infinity in a pointwise or certain average sense, and a similarly important role is played by such inputs in connection with related discrete-time results (see, e.g., [16]). This observation served as the motivation for the study reported on in [17], in which it is shown that the members of a certain large family of linear systems are in fact completely characterized by their (suitably defined) impulse responses – which exist as certain limits. These limits are functions in the usual sense (as opposed to generalized functions). Reference [17] addresses the case in which inputs belong to the space $L_p(\mathbb{R}^d)$ of pth-power integrable complex-valued functions x defined on $\mathbb{R}^d$, in which p satisfies $1 \leq p < \infty$, and d is an arbitrary positive integer. Outputs are taken to be elements of the space $B(\mathbb{R}^d)$ of bounded complex-valued functions y defined on $\mathbb{R}^d$, with the norm given by

$$\|y\| = \sup_{\alpha \in \mathbb{R}^d} |y(\alpha)|. \tag{1}$$

In [18] we consider the case in which outputs are elements of $B(\mathbb{R}^d)$, but inputs belong to the normed linear space $C_0(\mathbb{R}^d)$ of continuous complex-valued functions x defined on the set $\mathbb{R}^d$ of real d-vectors such that $x(\alpha) \to 0$ as $\|\alpha\|_{(d)} \to \infty$, in which $\|\cdot\|_{(d)}$ stands for the Euclidean norm on $\mathbb{R}^d$. As is widely known, such multidimensional systems are of interest in connection with, for example, image processing. The theory concerning these systems is some respects more interesting than for the $L_p(\mathbb{R}^d)$-input-$B(\mathbb{R}^d)$-output case in [17]. It is assumed that H is continuous and shown that H has a representation given by

[1] A similar result [15] holds for maps whose domain is the whole space of bounded Lebesgue-measurable signals. For related results, concerning discrete-time systems, including a representation theorem for input-output maps, see Section 3.

$$(Hx)(\alpha) = \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} (Hq_\epsilon)(\alpha - \beta)x(\beta)\, d\beta \tag{2}$$

in which the convergence is uniform with respect to α, and q_ϵ is a certain type of function that depends on the parameter ϵ. Also given is a necessary and sufficient condition under which there is a function h such that the right side of (2) can be written as a convolution

$$\int_{\mathbb{R}^d} h(\alpha - \beta)x(\beta)\, d\beta.$$

Related results are given in [19] for the case in which inputs belong to $L_p(\mathbb{R}^d) \cup L_1(\mathbb{R}^d)$, where again $1 \leq p < \infty$ (of course, the case in which $p = 2$ is of particular interest).

For reasons closely related to the material in [14] outlined above, a difficulty arises in attempts to obtain corresponding results for the important case in which the space of inputs is $L_\infty(\mathbb{R}^d)$, the normed linear space of complex-valued bounded Lebesgue-measurable functions defined on $\mathbb{R}^d$, with the norm given by (1).

Here, in the next section, we indicate in detail how this difficulty (discussed in Section 2.1) can be circumvented. More explicitly, we describe a representation theorem proved in [20] along the general lines of the theorem in [18] for the case in which both the range and domain of H is $L_\infty(\mathbb{R}^d)$, but under an additional (typically reasonable) assumption of the form

$$Hx = \lim_{\sigma \to \infty} (HW_\sigma x) \tag{3}$$

for each x, in which $\{W_\sigma : \sigma > 0\}$ is a certain set of weighting operators.[2]

A simple example of an H that satisfies the conditions assumed is given by

$$Hx = \sum_{m=1}^{n} a_m x(\cdot - \gamma_m) + \int_{\mathbb{R}^d} b(\cdot - \beta)x(\beta)\, d\beta, \quad x \in L_\infty(\mathbb{R}^d)$$

in which n is a positive integer, the a_m and γ_m respectively are real numbers and elements of $\mathbb{R}^d$, and $b \in L_1(\mathbb{R}^d)$.[3] In Section 2.1, it is noted that our representation theorem becomes false if condition (3) and its associated expressions are omitted. Also given in that section are some comments regarding engineering

[2] A similar result is given in [21] for the special (and less complex) case in which inputs are bounded functions that are continuous.

[3] This follows from material in [20]. $L_1(\mathbb{R}^d)$ has the usual meaning described in Section 2.

education and introductory courses in the area of signals and systems. Related material concerning discrete-space systems is given in Section 3.

Of particular interest is the observation in Section 2.1 that for the large family of inputs and maps H addressed, the Dirac impulse-response concept is in fact not the key concept concerning the representation of H, and that instead the input-output properties of H are determined in general by a certain type of *family* $\{Hq_\epsilon : \epsilon \in (0, \rho)\}$ of responses.

2. REPRESENTATION THEOREM

We use $L_1(\mathbb{R}^d)$ to denote the normed linear space of Lebesgue integrable complex-valued functions x defined on the set $\mathbb{R}^d$ of real d-vectors, with the usual norm given by

$$\|x\|_1 = \int_{\mathbb{R}^d} |x(\alpha)|\, d\alpha. \tag{4}$$

As usual, when $L_1(\mathbb{R}^d)$ is regarded as a metric space, the elements of $L_1(\mathbb{R}^d)$ are understood to be equivalence classes. By convergence in $L_1(\mathbb{R}^d)$, we mean convergence to an element of $L_1(\mathbb{R}^d)$ with respect to the norm in $L_1(\mathbb{R}^d)$.

As is well known, the d-dimensional extension of the concept of an impulse function as described by P. Dirac, while often useful in engineering and scientific applications, is unsatisfactory from the viewpoint of mathematics. It is unsatisfactory because according to the usual theory of integration,

$$\int_{\mathbb{R}^d} q(\alpha)\, d\alpha = 0$$

for any complex-valued function q defined on $\mathbb{R}^d$ with $q(\alpha) = 0$ for $\|\alpha\|_d > 0$, even if $q(0) = \infty$ is allowed.[4] It is also well known, at least for $d = 1$, that an alternative approach (see, for example, [7]) involves envisioning a sequence of progressively taller and narrower unit-integral functions centered at $\alpha = 0$.[5] In this spirit, but

[4]More specifically, with $q(0) = \infty$ allowed, the integral is zero as a Lebesgue integral or an improper Riemann integral. In the remainder of the paper, all integrals are meant to be interpreted as Lebesgue integrals.

[5]There is also a related theory of *distributions* [28] developed by L. Schwartz and S. Sobelov around 1948, in which the delta function is viewed as a linear functional on a certain type of space of infinitely differentiable functions of compact support. Distribution theory frees differential calculus from certain difficulties that arise because of the existence of nondifferentiable functions. No use is made of these ideas in this paper.

with no attempt to adhere strictly to the concept of a generalized function, we introduce the following definition, in which $BL_1(\mathbb{R}^d)$ denotes the linear space of bounded $L_1(\mathbb{R}^d)$ functions, with the norm defined by (4).

$\mathcal{Q}$ denotes the family of $BL_1(\mathbb{R}^d)$-valued maps q defined on $(0,1)$ such that, with $q(\epsilon)$ denoted by q_ϵ,

$$\int_{\mathbb{R}^d} q_\epsilon(\alpha)\, d\alpha = 1 \;\text{ for }\; \epsilon \in (0,1), \quad \sup_\epsilon \int_{\mathbb{R}^d} |q_\epsilon(\alpha)|\, d\alpha < \infty,$$

and

$$\lim_{\epsilon \to 0} \int_{\|\alpha\|_{(d)} > \xi} |q_\epsilon(\alpha)|\, d\alpha = 0, \quad \xi > 0.$$

Note that q given by the familiar expression

$$\begin{aligned} q_\epsilon(\alpha) &= 1/\epsilon, \quad |\alpha| \leq \epsilon/2 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

is an element of $\mathcal{Q}$ for $d = 1$.

We use $\mathcal{H}$ to stand for the family of all linear shift-invariant maps H from $L_\infty(\mathbb{R}^d)$ into itself, such that the restriction of H to $BL_1(\mathbb{R}^d)$ is a continuous map into $BL_1(\mathbb{R}^d)$.[6] Let η be a number in $[1,\infty)$. For each positive σ, let W_σ be the map from $L_\infty(\mathbb{R}^d)$ to $BL_1(\mathbb{R}^d)$ defined by $(W_\sigma x)(\alpha) = w_\sigma(\alpha)x(\alpha)$, in which $w_\sigma \in BL_1(\mathbb{R}^d)$ with $w_\sigma(\alpha) = 1$ for $\|\alpha\|_{(d)} \leq \sigma$ and $|w_\sigma(\alpha)| \leq \eta$, $\|\alpha\|_{(d)} > \sigma$. (e.g., w_σ could be taken to be equal to $2 - \sigma^{-1}\|\alpha\|_{(d)}$ for $\sigma < \|\alpha\|_{(d)} < 2\sigma$, and equal to 0 for $\|\alpha\|_{(d)} \geq 2\sigma$).

It is reasonable to say, roughly speaking, that an element H of $\mathcal{H}$ has an impulse response (or what might more accurately be called a "q-response limit") if for every $q \in \mathcal{Q}$ we have Hq_ϵ well defined for each $\epsilon \in (0,1)$, with $\lim_{\epsilon \to 0} Hq_\epsilon$ existing in a meaningful sense and not dependent on q. Our theorem is the following result. For the type of H addressed in the theorem, H has an impulse response h in the precise sense that statement (a) of the theorem holds. (The engineering literature says little about the existence of impulse responses for general systems, which typically are simply assumed to exist.) Finally, given $a \in L_\infty(\mathbb{R}^d)$, we say that b_σ assumed

[6] By H is shift invariant we mean as usual that $H[x(\cdot - \zeta)] = (Hx)(\cdot - \zeta)$ for all x and all $\zeta \in \mathbb{R}^d$.

equal almost everywhere to an element of $L_\infty(I\!R^d)$ for $\sigma \in (0,\infty)$ converges in $M_\infty(I\!R^d)$ to a as $\sigma \to \infty$ if we have

$$\int_A |b_\sigma(\beta) - a(\beta)|\,d\beta \to 0 \quad \text{as} \quad \sigma \to \infty$$

for every bounded Lebesgue measurable subset A of $I\!R^d$, in which case we write $a = \lim_{\sigma \to \infty} b_\sigma$ (with the sense of convergence understood). We do not distinguish between $M_\infty(I\!R^d)$ limits that agree almost everywhere.

Theorem 1 [20]: Let H be an element of $\mathcal{H}$ with the property that

$$Hx = \lim_{\sigma \to \infty}(HW_\sigma x), \quad x \in L_\infty(I\!R^d) \tag{5}$$

in the sense of convergence in $M_\infty(I\!R^d)$, and let q be an element of Q. Then, for any $x \in L_\infty(I\!R^d)$ and any $\alpha \in I\!R^d$,

$$\int_{I\!R^d} (Hq_\epsilon)(\alpha - \beta) w_\sigma(\beta) x(\beta)\,d\beta \tag{6}$$

exists as a complex number for each ϵ and each σ,

$$\lim_{\epsilon \to 0} \int_{I\!R^d} (Hq_\epsilon)(\cdot - \beta) w_\sigma(\beta) x(\beta)\,d\beta$$

(in which the integral belongs to $L_1(I\!R^d)$ for each ϵ) exists in $L_1(I\!R^d)$, is essentially an element of $L_\infty(I\!R^d)$, and we have

$$Hx = \lim_{\sigma \to \infty} \lim_{\epsilon \to 0} \int_{I\!R^d} (Hq_\epsilon)(\cdot - \beta) w_\sigma(\beta) x(\beta)\,d\beta \tag{7}$$

where the convergence with respect to σ is in $M_\infty(I\!R^d)$. In addition, the following four statements are equivalent.

(a) Hq_ϵ converges in $L_1(I\!R^d)$, and the limit is independent of q.

(b) Hq_ϵ converges in $L_1(I\!R^d)$.

(c) There is an element h of $L_1(I\!R^d)$ such that

$$(Hx)(\gamma) = \int_{I\!R^d} h(\gamma - \beta) x(\beta)\,d\beta \tag{8}$$

for almost all $\gamma \in I\!R^d$ and every $x \in L_\infty(I\!R^d)$.

(d) There is a complex-valued h defined on $I\!R^d$ such that, for any $x \in L_\infty(I\!R^d)$, g defined for any fixed $\gamma \in I\!R^d$ and all β in $I\!R^d$ by

$$g(\beta) = h(\gamma - \beta)x(\beta)$$

belongs to $L_1(\mathbb{R}^d)$, and

$$(Hx)(\gamma) = \int_{\mathbb{R}^d} h(\gamma - \beta)x(\beta)\, d\beta \tag{9}$$

for almost all $\gamma \in \mathbb{R}^d$.

2.1. Comments

The proof in [20] makes use of the following two lemmas proved in [17] and [19], respectively.

Lemma 1: Let q and f belong to Q and $L_1(\mathbb{R}^d)$, respectively. Then

$$\int_{\mathbb{R}^d} q_\epsilon(\cdot - \beta)f(\beta)\, d\beta \tag{10}$$

is an element of $L_1(\mathbb{R}^d)$ for each ϵ, and it converges in $L_1(\mathbb{R}^d)$ to f as $\epsilon \to 0$.[7]

Lemma 2: Suppose that linear shift-invariant H is a continuous map of $L_1(\mathbb{R}^d)$ into itself, and let q be an element of Q.

Then

$$Hx = \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} (Hq_\epsilon)(\cdot - \beta)x(\beta)\, d\beta, \quad x \in L_1(\mathbb{R}^d) \tag{11}$$

where the convergence is in $L_1(\mathbb{R}^d)$.

Lemma 2 is of interest in its own right. We will refer to Lemma 1 below.

The proof in [20] makes clear that h in (c) is unique in $L_1(\mathbb{R}^d)$, and that it is the limit described in (b). Also, it can be shown, using material in [5, p. 569], that for $d = 1$ there is a function K of bounded variation on $\mathbb{R}$ such that the iterated limit on the right side of (7) can be written as the limit as $\sigma \to \infty$ of a Lebesgue-Stieltjes integral

$$\int_{\mathbb{R}} z_\sigma(\cdot - \beta)\, dK(\beta)$$

in which $z_\sigma(\beta) = w_\sigma(\beta)x(\beta)$. However, this simplification obscures the role played by the Hq_ϵ.

Parts (a)–(d) of our theorem provide conditions under which the conclusion of a familiar engineering argument of long standing can be justified. That argument,

[7] Lemma 1 is related to results in [3, p. 149] and [10, p. 72], and [3] gives references concerning other related results.

which is often taught to students and which concerns the representation of linear shift-invariant systems, proceeds as follows (using our notation). Let H be the input-output map of such a system, and let X_0 be the family of all possible input functions – assumed only to be real-valued, or complex-valued, and defined on $\mathbb{R}^d$. Typically, $d = 1$. One writes

$$x(\alpha) = \int_{\mathbb{R}^d} \delta(\alpha - \beta)x(\beta)\, d\beta \tag{12}$$

for any input x, in which δ is said to be the Dirac delta function – a function on $\mathbb{R}^d$ of unit integral that vanishes except at the origin. Noting (12), which is the description of the "sifting-function" property of δ, and appealing to the linearity and shift-invariance of H, one is said to have

$$Hx = H \int_{\mathbb{R}^d} \delta(\cdot - \beta)x(\beta)\, d\beta = \int_{\mathbb{R}^d} H\delta(\cdot - \beta)x(\beta)\, d\beta \tag{13}$$

in which $H\delta(\cdot - \beta)$ is $h(\cdot - \beta)$, where h is the system's impulse response $H\delta$. Thus, one concludes that

$$(Hx)(\alpha) = \int_{\mathbb{R}^d} h(\alpha - \beta)x(\beta)\, d\beta \tag{14}$$

for all $x \in X_0$. In particular, one concludes that the input-output properties of H are completely defined by its impulse response. As indicated in the Introduction (see also [22]), these conclusions are now known to be incorrect. This fact contradicts a widely-held engineering belief typically taught to students, to the effect that the main textbook conclusions concerning continuous-time linear systems obtained using Dirac delta-function arguments can be shown to be valid using the mathematical theory of distributions.[8]

There are at least three difficulties with the engineering argument just described: (a) As noted near the beginning of Section 2, there is no function δ that has the properties described,[9] (b) even if (a) were not true, the interchange of the

[8]These remarks should in no way be interpreted as a criticism of distribution theory itself, or of the delta function as defined in the setting of distribution theory.

[9]This was understood at the outset by mathematicians. And, according to Marsden [8, p. 275]: "At the same time as the physicists and engineers were computing, mathematicians sat back in quiet amusement, pointing out that this δ-function business was really all nonsense because no such function can exist. The definition does not really make sense, as anyone can plainly see. To add to the mathematician's enjoyment, Dirac proceeds to differentiate this function δ."

order of integration and operation by H in (13) is in fact not justified by merely the linearity of H. Linearity (in particular, the superposition part of linearity) concerns H operating on finite sums, not infinite sums [10], and (c) the argument provides no justification that h exists even in the sense of generalized functions in which, as mentioned near the beginning of Section 2, we envision in the place of δ a sequence of progressively taller and narrower unit-integral inputs centered at $\alpha = 0$. In particular, the possibility that h when it exists depends on the specific sequence used is not ruled out.

Our theorem shows that (14) is valid for an important family of inputs, under the assumption that H satisfies certain continuity and limit conditions and the condition that the impulse response h of H exists as an $L_1(\mathbb{R}^d)$ limit, where by the impulse response of H is meant $\lim_{\epsilon \to 0} Hq_\epsilon$, in which q is any element of Q. More importantly, and for this family of inputs and maps H, the theorem shows that the impulse-response concept is in fact not the key concept concerning the representation of H, because in general that representation (7) is a limit of an integral rather than (9), an integral of a limit. In particular, we see that in general the input-output properties of H are determined by the *family* of responses $\{Hq_\epsilon : \epsilon \in (0, \rho)\}$, in which q is any member of Q and ρ is any number in $(0, 1]$.

With regard to engineering education, and introductory courses in the area of signals and systems, the proof (in [20]) of our theorem provides the following outline of a revised argument that yields (14), which has the advantage that it can be made as precise and rigorous as one wishes. With W_σ as described (an example of a w_σ was given before the statement of the theorem), and using Lemma 1,we have

$$(W_\sigma x)(\alpha) = \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} q_\epsilon(\alpha - \beta)(W_\sigma x)(\beta)\, d\beta \tag{15}$$

in which q is some simple element of Q. For example, one may take q_ϵ to be given by

$$q_\epsilon(\alpha) = \Pi_{j=1}^d r_\epsilon(\alpha_j)$$

where $\alpha_1, \ldots, \alpha_d$ stand for the components of α, and

$$r_\epsilon(\alpha_j) = (1/\epsilon) r(\alpha_j/\epsilon), \quad \epsilon \in (0, 1)$$

[10] In fact, it is known [23] that what might be called "infinite superposition" can fail.

in which $r \in L_1(\mathbb{R})$ is bounded and has unit integral. Therefore, under the assumption that H is continuous [on $BL_1(\mathbb{R}^d)$],

$$HW_\sigma x = \lim_{\epsilon \to 0} H \int_{\mathbb{R}^d} (q_\epsilon)(\cdot - \beta)(W_\sigma x)(\beta)\, d\beta.$$

Using the linearity of H, and under the assumptions that the domain of H is $L_\infty(\mathbb{R}^d)$, and that H is continuous, the order of performing the integration and operating by H can be shown to be able to be interchanged, yielding

$$HW_\sigma x = \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} (Hq_\epsilon)(\cdot - \beta)(W_\sigma x)(\beta)\, d\beta.$$

Using condition (5) gives

$$Hx = \lim_{\sigma \to \infty} \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} (Hq_\epsilon)(\cdot - \beta)(W_\sigma x)(\beta)\, d\beta. \tag{16}$$

And the right side of (16) is

$$\lim_{\sigma \to \infty} \int_{\mathbb{R}^d} h(\cdot - \beta)(W_\sigma x)(\beta)\, d\beta$$

when $h = \lim_{\epsilon \to 0} (Hq_\epsilon)$ exists in $L_1(\mathbb{R}^d)$, which yields (14) upon taking the limit with respect to σ. (As is indicated in [19], similar outlines hold in the setting of the representation theorems in [18] and [19].[11] In those cases, the outline is simpler because no family $\{W_\sigma : \sigma > 0\}$ of weighting operators is needed. As mentioned earlier, a similar outline holds in the setting of [21] which addresses the case of bounded continuous inputs.) Also, a result for shift-varying systems along the lines of Theorem 1 is given in [24].

The following result, which is along the lines of the theorem in [22], is of importance in the context of our theorem, because it shows that Theorem 1 becomes false if condition (5) and the w_σ terms and associated limits are omitted.

Theorem 2 [20]: There is an $H \in \mathcal{H}$ such that

(a) Hq_ϵ is the zero function for all $\epsilon \in (0,1)$ and every $q \in Q$.

(b) There are elements x of $L_\infty(\mathbb{R}^d)$ such that Hx is not the zero function.[12] [Therefore, for such H's, (7) without the w_σ term and the corresponding

[11] In this connection, and referring to the theorem in [18], it is easy to check that the statement "Hq_ϵ converges in $L_1(\mathbb{R}^d)$." can be added to the three equivalent statements.

[12] In this theorem, "the zero function" means the essentially zero function.

limit cannot hold, because the right side is the zero function for all ϵ and every $x \in L_\infty(\mathbb{R}^d)$.]

The result proved in [20] is stronger than is stated in the theorem, in that H maps $L_\infty(\mathbb{R}^d)$ into itself continuously, and that the range $H[L_\infty(\mathbb{R}^d)]$ of H can be taken to be contained in (the relatively simple space) $\mathcal{C}$. There are several variations of Theorem 2. For example, for $d = 1$, it has been shown, using the result in [14], that H can be taken to be causal (see [15]).

Assuming only that we agree to define the impulse response of H as the $L_1(\mathbb{R}^d)$ limit of Hq_ϵ for some $q \in \mathcal{Q}$, Theorem 2 directs attention to the existence of another type of map H for which the impulse response is the zero function, but which takes some inputs into nonzero outputs.

3. RELATED DISCRETE-SPACE THEOREM

Here we describe a representation result for the case of discrete time (and discrete-space) systems.

The cornerstone of the theory of discrete-time single-input single-output linear systems is the idea that every such system has an input-output map H that can be represented by an expression of the form

$$(Hx)(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(m) \tag{17}$$

in which x is the input and h is the system function associated with H in a certain familiar way. It is widely known that this, and a corresponding representation for time-invariant systems in which $h(n,m)$ is replaced with $h(n-m)$, are discussed in many books (see, for example, [9, p. 23]). Almost always it is emphasized that these representations hold *for all* linear input-output maps H. However, it is known that this claim is unwarranted. In fact, it appears that as early as 1932 Banach was aware of the lack of existence of generalized convolution-sum representations for certain linear system H's (see [4, pp. 158, 159]).[13] This writer

[13] For related material concerning discrete-time systems, in the context of the theory of conjugate spaces, see e.g., [1, p. 228, Table 1 and p. 229, Exercise 9]. For material related in a general sense concerning discrete-time systems, see [2] and [6].

does not claim that these H's are necessarily of importance in applications, but he does feel that their existence shows that the analytical ideas in the books are flawed.[14]

One of the main purposes of this section is to describe a result showing that, under some mild conditions concerning the set of inputs and H, (17) becomes correct if an additional term is added to the right side. More specifically, we have

$$(Hx)(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(m) + \lim_{k\to\infty} (HE_k x)(n)$$

for each n, in which h has the same meaning as in (1), and $E_k x$ denotes the function given by $(E_k x)(p) = x(p)$ for $|p| > k$ and $(E_k x)(p) = 0$ otherwise. This holds whenever the input set is the set of bounded functions, the outputs are bounded, and H is continuous in a certain standard sense. In particular, we see that in this important setting, an H has a representation of the form given by (17) if and only if

$$\lim_{k\to\infty} (HE_k x)(n) = 0$$

for all x and n. Since this is typically a very reasonable condition for a system map H to satisfy, it is clear that the H's that cannot be represented using just (1) are rather special.

Our results concerning H are given in the following subsection, in which the setting is more general in that we address H's for which inputs and outputs depend on an arbitrary finite number of variables. This case is of interest in connection with, for example, image processing. We also consider H's for which inputs and outputs are defined on just the nonnegative integers because that case too arises often in applications. In that setting the situation with regard to the need for an additional term in the representation is different: no additional term is needed for causal maps H.

3.1. Preliminaries

Here too let d be a positive integer, let Z be the set of all integers, and let Z_+ denote the set of nonnegative integers. Let D stand for either Z^d or Z_+^d. Let F

[14] The oversight in the books is due to the lack of validity of the interchange of the order of performing a certain infinite sum and then applying $(H\cdot)(n)$. The infinite sum at issue clearly converges pointwise, but that is not enough to justify the interchange.

be either the set of real numbers or the set of complex numbers, and let $\ell_\infty(D)$ denote the normed linear space of bounded F-valued functions x defined on D, over the field F, with the norm $\|\cdot\|$ given by $\|x\| = \sup_{\alpha \in D} |x(\alpha)|$.

For each positive integer k, let c_k stand for the discrete hypercube $\{\alpha \in D : |\alpha_j| \leq k \quad \forall j\}$ (α_j is the jth component of α), and let $\ell_1(D)$ denote the set of F-valued maps g on D such that

$$\sup_k \sum_{\beta \in c_k} |g(\beta)| < \infty.$$

For each $g \in \ell_1(D)$ the sum $\sum_{\beta \in c_k} g(\beta)$ converges to a finite limit as $k \to \infty$, and we denote this limit by $\sum_{\beta \in D} g(\beta)$.

Define maps Q_k and E_k from $\ell_\infty(D)$ into itself by $(Q_k x)(\alpha) = x(\alpha)$, $\alpha \in c_k$ and $(Q_k x)(\alpha) = 0$ otherwise, and $(E_k x)(\alpha) = x(\alpha)$, $\alpha \notin c_k$ and $(E_k x)(\alpha) = 0$ otherwise.

In the next subsection H stands for any linear map from $\ell_\infty(D)$ into itself that satisfies the condition that

$$\sup_k |(HQ_k u)(\alpha)| < \infty \tag{18}$$

for each $u \in \ell_\infty(D)$ and each $\alpha \in D$. This condition is met whenever H is continuous because then $|(HQ_k u)(\alpha)| \leq \|HQ_k u\| \leq \|H\| \cdot \|Q_k u\| \leq \|H\| \cdot \|u\|$.[15]

3.2. Representation Result

In the following theorem, $h(\cdot, \beta)$ for each $\beta \in D$ is defined by $h(\cdot, \beta) = H\delta_\beta$, where $(\delta_\beta)(\alpha) = 1$ for $\alpha = \beta$ and $(\delta_\beta)(\alpha)$ is zero otherwise. Of course $h(\cdot, \beta)$ is the response of H to a unit "impulse" occurring at $\alpha = \beta$.

Theorem 3: For any H as described, and for each $\alpha \in D$ and each $x \in \ell_\infty(D)$,

(i) g defined on D by $g(\beta) = h(\alpha, \beta)x(\beta)$ belongs to $\ell_1(D)$.

(ii) $\lim_{k \to \infty} (HE_k x)(\alpha)$ exists and is finite.

(iii) We have

$$(Hx)(\alpha) = \sum_{\beta \in D} h(\alpha, \beta)x(\beta) + \lim_{k \to \infty} (HE_k x)(\alpha).$$

[15] Here we have used the well-known fact that boundedness and continuity are equivalent for a linear operator between normed linear spaces.

Theorem 3 is proved in [11] (alternatively, see [12, Appendix G] or [13, pp. 68–69]). For $D = \mathcal{Z}_+$ and H causal in the usual sense, the term $\lim_{k\to\infty}(HE_k x)(\alpha)$ is always zero. However, for $D = \mathcal{Z}^d$, and also for $D = \mathcal{Z}_+$ with H not causal, there are maps H for which the additional term is not always zero (see [11] or [13]). A related result addressing systems with stochastic inputs is given in [25].

There are some other results related to Theorem 3 that are of direct interest. For example, it is a widespread engineering belief that linear *shift-invariant* input-output operators that take a set of functions (closed under translation) into itself commute in the sense that $H_1 H_2 = H_2 H_1$ for any two such operators. But this is based on the belief that such operators, for which $h(\alpha, \beta)$ can be written in the form $h(\alpha - \beta)$, have convolution representations. In [26] theorems are given to the effect that, in the discrete-space setting described above, it is *not true* that shift-invariant operators commute (e.g., always commute), even though $H_1 H_2 = H_2 H_1$ holds on certain interesting subsets of inputs. A result showing the lack of cummutativity for some continuous-space systems is also given.[16]

BIBLIOGRAPHY

[1] G. Bachman and L. Narici, *Functional Analysis*, New York: Academic Press, 1966.

[2] W. J. Borodziewicz, K. J. Jaszczak, and M. A. Kowalski, "A Note on Mathematical Formulation of Discrete-Time Linear Systems," *Signal Processing*, vol. 5, pp. 369–375, 1983.

[3] W. Cheney and W. Light, *A Course in Approximation Theory*, Pacific Grove: Brooks/Cole, 2000.

[4] R. E. Edwards, *Functional Analysis*, New York: Dover, 1995.

[5] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, Providence: American Mathematical Society, 1957.

[6] P. Hughett, "Representation Theorems for Semilocal and Bounded Linear Shift-Invariant Operators on Sequences," *Signal Processing*, vol. 67, pp. 199–209, 1998.

[7] M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions*, Cambridge: Cambridge University Press, 1960.

[16] And it was found [27] that not all linear shift-invariant continuous-space systems are characterized by their frequency responses (even when they exist).

[8] J. E. Marsden, *Elementary Classical Analysis*, New York: W. H. Freeman and Company, 1974.

[9] A. V. Oppenheim, R. W. Shafer, and J. R. Buck, *Discrete-Time Signal Processing (Second Edition)*, Upper Saddle River: Prentice Hall, 1999.

[10] B. E. Petersen, *Introduction to the Fourier Transform and Pseudo-Differential Operators*, Pitman Publishing Inc., Marshfield, Massachusetts: 1983.

[11] I. W. Sandberg, "A Representation Theorem for Linear Discrete-Space Systems," *Mathematical Problems in Engineering*, vol. 4, pp. 369–375, 1998.

[12] I. W. Sandberg, "Multidimensional Nonlinear Myopic Maps, Volterra Series, and Uniform Neural-Network Approximations," pp. 99–128 in D. Docampo, A. Figueiras-Vidal, and F. Perez-Gonzalez (eds.), *Intelligent Methods in Signal Processing and Communications*, Boston: Birkhauser, 1997.

[13] I. W. Sandberg, J. T. Lo, C. Francourt, J. Principe, S. Katagiri, and S. Haykin *Nonlinear Dynamical Systems: Feedforward Neural Network Perspectives*, New York: John Wiley, 2001.

[14] I. W. Sandberg, "Causality and the Impulse Response Scandal," *IEEE Transactions on Circuits and Systems I*, vol. 50, no. 6., pp. 810–811, 2003.

[15] ______________, "On Causality and the Impulse Response Scandal," *WSEAS Transactions on Circuits and Systems*, vol. 3, no. 9, pp. 1741–1744, 2004.

[16] ______________, "Notes on Representation Theorems for Linear Discrete-Space Systems," *Proceedings of the International Symposium on Circuits and Systems*, Orlando, Florida, May 30–June 2, 1999 (four pages on CD).

[17] ______________, "Notes on Linear Systems and Impulse Responses," *Circuits, Systems and Signal Processing*, vol. 23, no. 5, pp. 339–350, 2004.

[18] ______________, "Notes on Multidimensional Linear System Representations," *38th Annual Conference on Information Sciences and Systems*, Department of Electrical Engineering, Princeton University, Princeton, NJ, pp. 801–806, March 2004.

[19] ______________, "On the Representation of Linear System Maps: Inputs that Need Not be Continuous," *Proceedings of the 6th WSEAS International Conference on Applied Mathematics*, Corfu Island, Greece, August 17–19, 2004 (five pages on CD).

[20] ______________, "Bounded Inputs and the Representation of Linear System Maps," *Circuits, Systems, and Signal Processing*, vol. 24, no. 1, pp. 103-115, 2005.

[21] ______________, "On the Representation of Linear System Maps: Bounded Inputs," *Proceedings of the 6th WSEAS International Conference on Applied Mathematics*," Corfu Island, Greece, August 17–19, 2004 (six pages on CD).

[22] ___________, "Continuous Multidimensional Systems and the Impulse Response Scandal," *Multidimensional Systems and Signal Processing*, 15, pp. 295–299, 2004.

[23] ___________, "The Superposition Scandal," *Circuits, Systems, and Signal Processing*, vol. 17, no. 6, pp. 733–735, 1998.

[24] ___________, "On the Representation of Shift-Varying Linear System Maps," *Circuits, Systems, and Signal Processing*, 2006 (to appear).

[25] ___________, "A Note on Representation Theorems for Linear Discrete-Space Systems with Stochastic Inputs," *International Journal of Circuit Theory and Applications*, vol. 29, pp. 505–509, 2001.

[26] ___________, "Linear Shift-Invariant Input-Output Maps Do Not Necessarily Commute," *International Journal of Circuit Theory and Applications*, vol. 28, pp. 513–518, 2000.

[27] ___________, "Notes on Linear Systems and Frequency Responses," *International Journal of Circuit Theory and Applications*, vol. 33, pp. 175–181, 2005.

[28] A. H. Zemanian, *Distribution Theory and Transform Analysis*, Mineola, NY: Dover, 1987. Originally published: New York, McGraw-Hill, 1965 (International Series in Pure and Applied Mathematics).

PRZEGLĄD NAJNOWSZYCH WYNIKÓW OTRZYMANYCH W DZIEDZINIE OPISU SYSTEMÓW LINIOWYCH

Streszczenie

W artykule zostało przedstawione ogólne wyrażenie opisujące odwzorowanie zbioru sygnałów wejściowych na zbiór sygnałów wyjściowych, związane z elementami pewnej bardzo ważnej, dużej rodziny wielowymiarowych, liniowych oraz stacjonarnych systemów określonych dla sygnałów wejściowych ograniczonych i mierzalnych w sensie Lebesgue'a. Wyrażenie to stanowi granicę splotu zdefiniowanego w iterowanej przestrzeni funkcyjnej. W pracy został również podany warunek konieczny i dostateczny, pod którym ww. granica może być przedstawiona jako splot z odpowiedzią impulsową w postaci funkcji całkowalnej. Kluczową rolę pełni w tym przypadku pewna rodzina operatorów wagowych. Pokazano też, że dla dużej rodziny przyjętych zbiorów sygnałów wejściowych i odwzorowań, koncepcja odpowiedzi impulsowej otrzymywanej w wyniku pobudzenia systemu tzw. impulsem Diraca nie jest kluczowa dla opisu systemu H. To znaczy, że zamiast tej koncepcji właściwości systemu H w konwencji wejścia-wyjścia mogą być, w ogólności, określone za pomocą pewnej rodziny odpowiedzi. W artykule przedstawiono również inne pokrewne wyniki, dotyczące systemów dyskretnych oraz aspektów edukacyjnych.

Słowa kluczowe: systemy liniowe, systemy wielowymiarowe, odpowiedzi impulsowe, systemy stacjonarne, sygnały ograniczone oraz mierzalne, systemy dyskretne

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
TELEKOMUNIKACJA I ELEKTRONIKA 9 (2006), 25-35

Combined Achromatic and Chromatic Correction of Color Images

Luca Lucchese[§] and Sanjit K. Mitra[‡]

[§]School of Electrical Engineering and Computer Science
Oregon State University, Corvallis, OR 97331
[‡]Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Abstract

This paper advances a new algorithm for determining the color mapping between two images of the same object or scene taken under different illumination conditions. The proposed algorithm compensates for differences in colors by separately equalizing their achromatic and chromatic components. The equalization of the 1-D achromatic channel is carried out with a standard technique for gray-level images whereas the equalization of the 2-D chromatic channel is considered as a problem of image warping. This method can also be used to color calibrate a trichromatic sensing device, provided that a color chart is available.

Keywords: Color correction, 2-D warping, chromatic coordinates, luminance, color chart, color constancy.

1 Introduction

Color is an important feature in many computer vision systems for object recognition. Such feature is subject to a great variability which depends on various factors, such as the spectral characteristics of the color sensor, the chromatic aberration of its lens system, and the illuminant(s) casting light on the object or scene of interest. The simple mathematical formulation of the image acquisition process taking place in a color imaging sensor shows these dependencies and help understand their importance in the appearance of color in digital images. The pixels of a color image can be represented with triplets $(\Theta_R, \Theta_G, \Theta_B)$ whose components are given by

$$\Theta_k = \int_S \rho_k(\lambda)\mathcal{R}(\lambda)\mathcal{I}(\lambda)d\lambda, \quad k = R, G, B, \tag{1}$$

where $\rho_k(\lambda)$'s are the spectral sensitivities, in the red, green, and blue bands, of the color filters of the camera which acquired the image, $\mathcal{R}(\lambda)$ is the reflectance of the object to which the pixel belongs, $\mathcal{I}(\lambda)$ is the spectral power distribution of the illuminant, and S is the visible spectrum, usually comprised between 400 and 700 *nm* [1].

Let us suppose that two color images $\mathcal{C}_1$ and $\mathcal{C}_2$ portray the same scene, recorded by the same imaging sensor, in different illumination conditions; then, only $\mathcal{I}(\lambda)$ varies in Eq. (1). If we knew the analytical description of this function for the illuminants of interest and estimates of the reflectances in the scene were also at disposal along with the sensor sensitivities, we would have a complete description of the mechanism which generated the two images. We could then easily render the colors of $\mathcal{C}_1$ as though they were acquired under the illuminant(s) of $\mathcal{C}_2$ and *vice versa*. However, unless one may resort to sophisticated spectral measuring systems, the information about reflectances and illuminants in Eq. (1) is seldom known and the rendering problem stated above cannot be easily solved from an analytical point of view. Such problem is also related to another well-known problem in the computer vision literature which is referred to as the *color constancy problem* [1].

A different interpretation of Eq. (1) may be adopted in order to perform the color calibration of a sensing device which establishes correspondences between the true colors in the scene, encoded by $\mathcal{R}(\lambda)$, and the colors in the image of the scene, encoded by $(\Theta_R, \Theta_G, \Theta_B)$. This process has a major relevance in any computer vision task where the 'true color' of objects has to be retrieved, *e.g.*, for recognition purposes. Color calibration is usually performed by imaging a color chart with different patches of known reflectance and under controlled lighting conditions. The calibration process leads to three sets of transformations relating the true colors in the scene to colors in the image of the scene which can then be used to correct the color of subsequently acquired images.

Various approaches have been advanced for correcting the color in digital images. These methods can be broadly divided into:

1) white-point mapping,

2) illumination estimation methods, and

3) methods based on lookup tables.

White-point mapping methods force the white point of the recorded image to an ideal white point. Usually, a white reference surface is imaged and the value of each band is corrected such that all three bands have the same value [2]. Illumination estimation methods first estimate the chromaticity of the scene illumination, either directly or indirectly, and then use it to correct the image color [3, 4]. Methods based on lookup tables generate lists of correspondences between a number of image colors and relative true colors. If a color that is being reproduced does not belong to the lookup table, it is estimated through interpolation of the colors in the table [5, 6, 7].

The method advanced in this paper can be classified as a lookup table method, even though the lookup table is not explicitly created [8]. With reference to the formulation presented above, our method transforms the colors of $\mathcal{C}_1$ into those of $\mathcal{C}_2$ without using Eq. (1) and, therefore, without any knowledge of the spectral descriptions of object reflectances and illuminants. If one represents the colors of the two images in a certain color space, the change of illuminant(s) can be regarded as a geometric warping within such a space. A similar approach has been adopted by Pappas and Pitas [9] for digital color restoration of old paintings; they search for the best transformation in the $L^*a^*b^*$ color space by trying, among others, affine and roto-translational models. Instead, we decompose the colors of $\mathcal{C}_1$ and $\mathcal{C}_2$ into achromatic and chromatic components, and separately search for warpings in the two channels. The 1-D achromatic component is represented by the CIE luminance Y whereas the 2-D chromatic component is represented by the CIE chromaticities u' and v' [10]. The first is equalized with a standard technique for gray-level images [11, 12]; the second by resorting to image warping [13]. Two examples of application of our method are reported in this contribution which confirm its good performance.

This paper has four sections. Section 2 presents the representation of color we have adopted. Section 3 discusses the equalization algorithm. Section 4 finally draws the conclusions.

2 Color Representation

Our algorithm operates on color signals decomposed into a 1-D achromatic channel – the CIE luminance Y – and a 2-D chromatic channel – expressed by the two CIE chromaticities u' and v'. We briefly describe such a decomposition. The two color images $\mathcal{C}_1$ and $\mathcal{C}_2$ are acquired as RGB bands, each one normalized between 0 and 1; these are then converted into CIE XYZ coordinates through the linear transformation[1] [10]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49000 & 0.31000 & 0.20000 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00000 & 0.01000 & 0.99000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{2}$$

We will loosely refer to Y as the luminance signal, even if the luminance $\mathcal{L}$ is properly defined as $\mathcal{L} = R + 4.5907G + 0.0601B$ [10]. This is motivated by the fact that the numbers 1, 4.5907, and 0.0601 are in the same ratios as the coefficients 0.17697, 0.81240, and 0.01063 in the definition of Y in Eq. (2). From Eq. (2), the two chromatic channels u' and v' are derived, respectively, as $u' = 4X/(X+15Y+3Z)$ and $v' = 9Y/(X+15Y+3Z)$. We have chosen the space $u'v'$ because of its characteristics of uniformity.

As an example, Figs. 1 (a) and (c) show two test images[2] used in this paper; they are the portions cropped from a larger color checkerboard which was acquired, respectively, with outdoor and indoor illumination. Figs. 1 (b) and (d) show the chromaticity diagrams $u'v'$ relative, respectively, to the images of Figs. 1 (a) and (c). The curved line represents the locus of spectral colors, the straight line the

[1] This transformation implicitly assumes that the red, green, and blue components of the image are expressed in terms of CIE RGB components [10], even though this is not usually the case. This formulation was however adopted for convenience in the algorithmic implementation.

[2] An electronic version of this paper with color images is available at `http://web.engr.oregonstate.edu/~luca/publications.htm`.

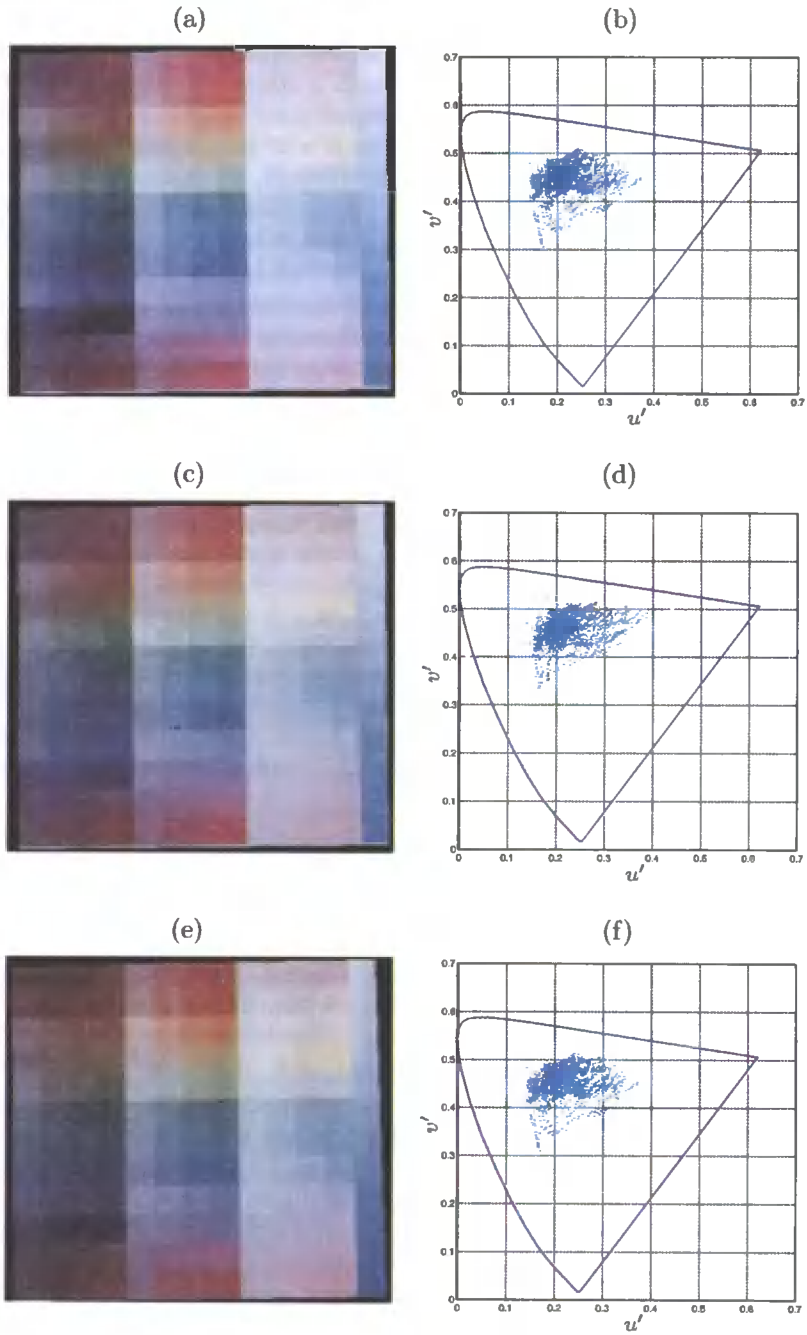

Figure 1: (a) Image $\mathcal{C}_1$ of a checkerboard taken under outdoor illumination; (c) Image $\mathcal{C}_2$ of the same target taken under indoor illumination; (e) Image $warp\{\mathcal{C}_1\}$, obtained by equalizing $\mathcal{C}_1$ against $\mathcal{C}_2$ with the proposed algorithm. (b), (d), and (f) Chromaticity diagrams relative, in the order, to the images of Fig. 1 (a), (c), and (e).

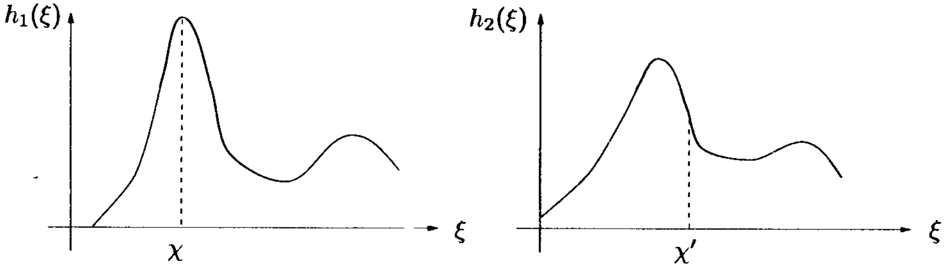

Figure 2: Histograms $h_1(\xi)$ and $h_2(\xi)$ considered as probability density functions.

locus of purples, and the small circle the reference white chosen as the chromatic point corresponding to the standard illuminant D_{65} ($u'_N = 0.1978$ and $v'_N = 0.4683$) [10].

The two images $\mathcal{C}_1$ and $\mathcal{C}_2$, having size $L_x \times L_y$ are represented in a palettized format[3] as $\mathcal{C}_m = \{\mathcal{Q}_m, \mathcal{P}_m, \boldsymbol{w}_m\}$, $m = 1, 2$, where $\mathcal{Q}_m \in \mathbb{N}^{L_x \times L_y}$ are matrices of pointers to the palettes $\mathcal{P}_m \in \mathbb{R}^{N_m \times 3}$ arranged as $\mathcal{P}_m = [\boldsymbol{Y}_m\ \boldsymbol{u}'_m\ \boldsymbol{v}'_m]$, where $\boldsymbol{Y}_m \in \mathbb{R}^{N_m}$ contains the luminance values of $\mathcal{C}_m$, $\boldsymbol{u}'_m \in \mathbb{R}^{N_m}$ and $\boldsymbol{v}'_m \in \mathbb{R}^{N_m}$ carry the chromaticity components, and $\boldsymbol{w}_m \in \mathbb{N}^{N_m}$ contains the number of pixels associated with the entries of the palette $\mathcal{P}_m$; N_m is the number of distinct colors within each image (in general, $N_1 \neq N_2$).

3 Color Correction Algorithm

3.1 Correction of the achromatic channel

From vectors $\boldsymbol{Y}_1$, $\boldsymbol{w}_1$, $\boldsymbol{Y}_2$, and $\boldsymbol{w}_2$, we build the two histograms $h_1(\xi)$ and $h_2(\xi)$ where $0 \leq \xi \leq 1$, because of the definition of Y in Eq. (2), and $h_1(\bar{\xi})$ and $h_2(\bar{\xi})$ represent the number of pixels having values of luminance comprised between $\bar{\xi} - \delta\bar{\xi}/2$ and $\bar{\xi} + \delta\bar{\xi}/2$, $\delta\bar{\xi}$ being the width of the histogram bins; in our implementation, a hundred bins are used for the construction of the histograms. In order to interpret the luminance histograms as *probability density functions* (pdf's), they are normalized in such a way that $\int_0^1 h_m(\xi) d\xi = 1$, $m = 1, 2$.

Enhancement of gray-level images is commonly accomplished through histogram equalization which consists in flattening the pdf's associated with the luminance histograms [11]. In our task instead, we want to render the luminance values of $\mathcal{C}_1$ as close as possible to those of $\mathcal{C}_2$; to this end, we enforce the equality of the *cumulative distributions* [12] (see Fig. 2), *i.e.*,

$$\int_0^{\chi} h_1(\xi) d\xi = \int_0^{\chi'} h_2(\xi) d\xi, \tag{3}$$

where $\chi' = \phi(\chi)$ is a monotonic function to estimate. This function is numerically computed by dividing the domain $[0, 1]$ of the two histograms into N intervals $\Delta\chi$

[3]Symbol $\mathbb{R}$ denotes the set of real numbers, symbol $\mathbb{N}$ the set of natural numbers.

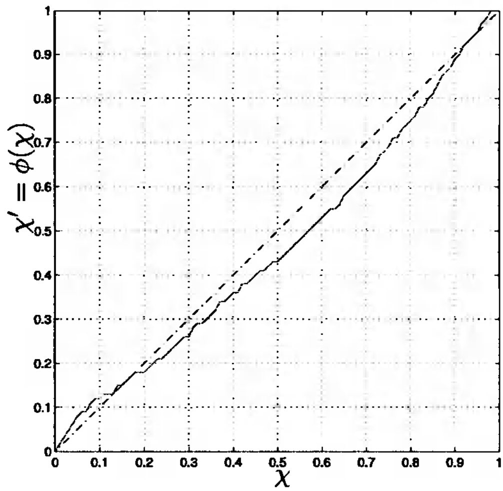

Figure 3: Plot of $\chi' = \phi(\chi)$ versus χ.

and applying the following routine

for $n = 1$ to $n = N$

$$\text{find } \chi'_n \text{ such that } \int_0^{n\Delta\chi} h_1(\xi)d\xi = \int_0^{\chi'_n} h_2(\xi)d\xi \tag{4}$$

end

Figure 3 shows with a solid line the plot of χ' versus χ for the pair of images of Figs. 1 (a) and (b); for a comparison, the identity transformation $\chi' = \chi$ is drawn with a dash-dotted line. The function $\chi' = \phi(\chi)$ allows the luminances $\boldsymbol{Y}_1$ to be transformed into $warp_Y\{\boldsymbol{Y}_1\}$, whose cumulative distribution is the same as that of $\boldsymbol{Y}_2$.

3.2 Correction of the chromatic channel

The problem of finding the geometric transformation which maps the chromaticities of Fig. 1 (b) into those of Fig. 1 (d) can be regarded as an *image warping* problem [13]. The chromaticity diagrams are 2-D plots, but they can be interpreted as 2-D histograms since any given chromaticity can be associated with a height expressing the number of pixels having that chromaticity. Instead of using coarse 2-D histograms, we prefer to use smooth surfaces obtained as follows. Let $\boldsymbol{u} = [u\ v]^T \in \mathbb{R}^2$ denote the Cartesian coordinates[4] of the two chromatic axes; the origin of this reference system is established for convenience at $(0.35, 0.35)$ since the chromaticity diagram is contained within the square $[0, 0.7] \times [0, 0.7]$. The two

[4]They should have been denoted by u' and v', but the superscripts have been dropped for notational convenience.

surfaces $\zeta_1(\boldsymbol{u})$ and $\zeta_2(\boldsymbol{u})$ are computed as

$$z_m(\boldsymbol{u}) = \sum_{n=1}^{N_m} w_m[n] \exp\left(-\tfrac{1}{2\sigma^2}\left((u - u'_m[n])^2 + (v - v'_m[n])^2\right)\right), \quad \text{(5a)}$$

$$\zeta_m(\boldsymbol{u}) \doteq z_m(\boldsymbol{u}) / \int_{\mathbb{R}^2} z_m(\boldsymbol{u}) d\boldsymbol{u}, \quad m = 1, 2. \quad \text{(5b)}$$

The aperture σ of the Gaussian functions summed up in Eq. (5a) determines the smoothness of the two surfaces; a good value for σ was found to be 0.01. Fig. 4 shows the surfaces $\zeta_1(\boldsymbol{u})$ and $\zeta_2(\boldsymbol{u})$ relative to the chromaticity diagrams of Figs. 1 (b) and (d). The normalization of Eq. (5b) allows the two surfaces to be regarded as 2-D pdf's and, at the same time, as images where 0 corresponds to black and 1 to white.

In principle, the change of illuminant(s) entails a complex transformation of $\zeta_1(\boldsymbol{u})$ into $\zeta_2(\boldsymbol{u})$ involving a warping of both domain and codomain. Suppose, for instance, that the imaged scene in $\boldsymbol{C}_1$ contains several colored surfaces illuminated by daylight; $\zeta_1(\boldsymbol{u})$ will thus exhibit various peaks. If the same surfaces are now exposed to a reddish illuminant in $\boldsymbol{C}_2$, those peaks will decrease in height and will migrate into a large peak in correspondence to the portion of the red chromaticities in the $u'v'$-diagram. This transformation appears to be difficult to capture with a mathematical model; an approximate description of the "migration of the chromaticities" can however be sought by disregarding the transformation which affects the ranges.

The geometric transformation between $\zeta_1(\boldsymbol{u})$ and $\zeta_2(\boldsymbol{u})$ is thus modeled as a twelve-parameter 2-D biquadratic warping, which can capture fairly complicated warpings [13]. Let $\boldsymbol{u}' = [u' \; v']^T \in \mathbb{R}^2$ be the coordinates of $\zeta_2(\boldsymbol{u})$ corresponding to the point $\boldsymbol{u}$ of the image, or pdf, $\zeta_1(\boldsymbol{u})$; they relate as

$$\begin{cases} u' = \kappa_{11}u^2 + \kappa_{12}uv + \kappa_{13}v^2 + \kappa_{14}u + \kappa_{15}v + \kappa_{16} \\ v' = \kappa_{21}u^2 + \kappa_{22}uv + \kappa_{23}v^2 + \kappa_{24}u + \kappa_{25}v + \kappa_{26}. \end{cases} \quad \text{(6)}$$

The estimate of the parameters κ_{ij} in Eq. (6) is formulated as the nonlinear minimization problem

$$\hat{\kappa}_{ij} = \min_{\kappa_{ij}} \|\zeta_2(\boldsymbol{u}') - \zeta_1(\boldsymbol{u})\|^2_{\mathcal{L}^2(\mathbb{R}^2)} = \min_{\kappa_{ij}} \int_{\mathbb{R}^2} \left(\zeta_2(\boldsymbol{u}') - \zeta_1(\boldsymbol{u})\right)^2 d\boldsymbol{u}, \quad \text{(7)}$$

where $\|\cdot\|^2_{\mathcal{L}^2(\mathbb{R}^2)}$ denotes the squared $\mathcal{L}^2$-norm of real functions defined over $\mathbb{R}^2$. The problem in Eq. (7) is then solved with the *Levenberg-Marquardt* (LM) *method* which is a standard minimization algorithm [15, 16].

A proper initialization is necessary in order for the LM algorithm to avoid being trapped into local minima. In our implementation, a good starting point was found to be the translational displacement provided by a phase correlation technique applied to $\zeta_1(\boldsymbol{u})$ and $\zeta_2(\boldsymbol{u})$; by denoting with $\boldsymbol{t} = [t_x \; t_y]^T \in \mathbb{R}^2$ this translational displacement, the initial values are set as $\kappa_{14} = 1$, $\kappa_{16} = t_x$, $\kappa_{25} = 1$, and $\kappa_{26} = t_y$, all the remaining parameters being set to zero. The estimates

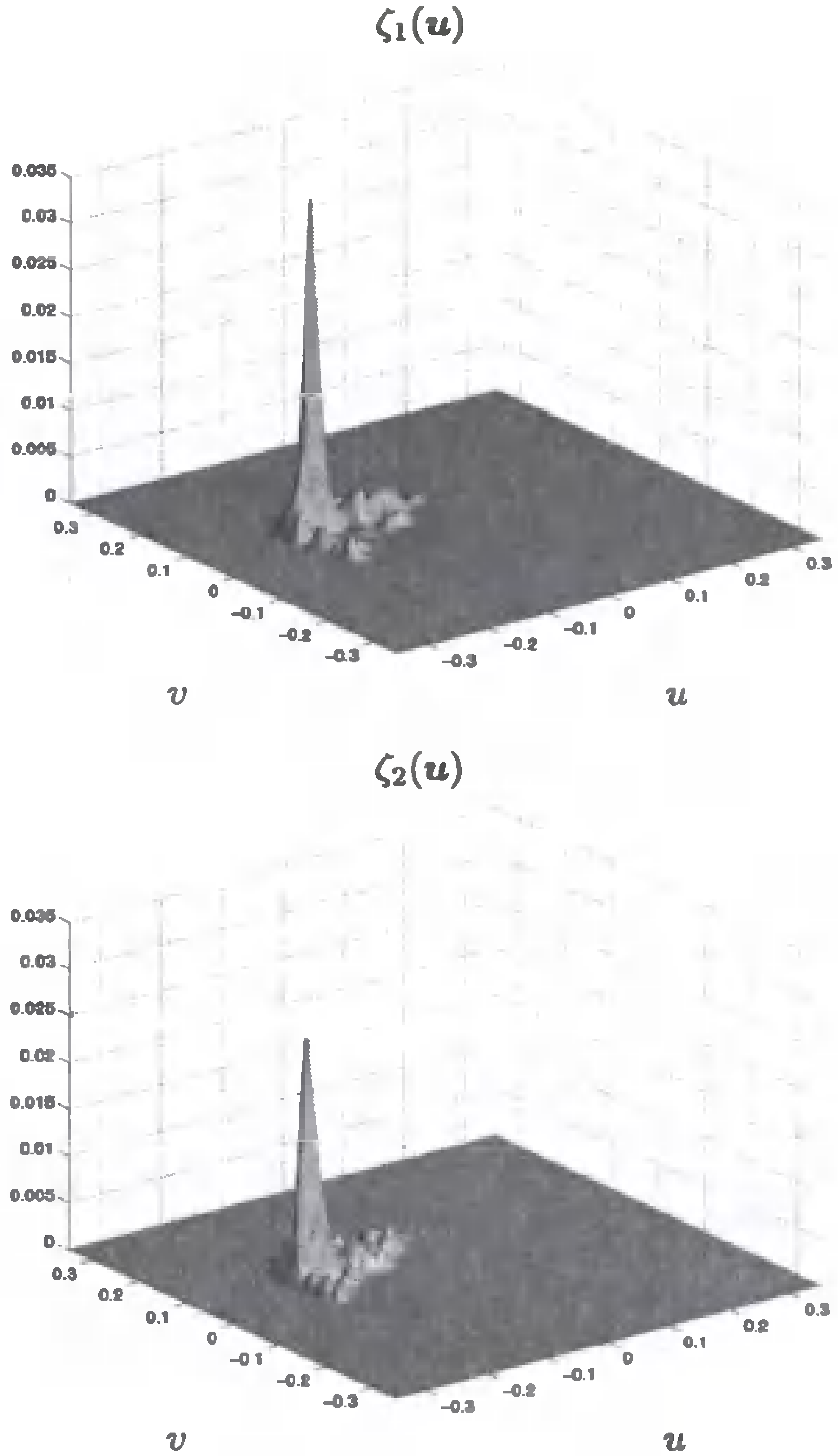

Figure 4: Surfaces $\zeta_1(\boldsymbol{u})$ and $\zeta_2(\boldsymbol{u})$ derived from the chromaticity diagrams of Figs. 1 (b) and (d), respectively.

returned by the LM algorithm were:

$$
\begin{array}{rrrr}
\kappa_{11} = & 1.59 \times 10^{-5} & \kappa_{21} = & -5.85 \times 10^{-5} \\
\kappa_{12} = & -3.13 \times 10^{-6} & \kappa_{22} = & 4.17 \times 10^{-5} \\
\kappa_{13} = & 1.71 \times 10^{-6} & \kappa_{23} = & -2.94 \times 10^{-5} \\
\kappa_{14} = & 1 & \kappa_{24} = & 4.06 \times 10^{-4} \\
\kappa_{15} = & 2.13 \times 10^{-5} & \kappa_{25} = & 1 \\
\kappa_{16} = & 4.32 \times 10^{-4} & \kappa_{26} = & 4.27 \times 10^{-3}
\end{array}
\quad (8)
$$

Based on the estimates of Eq. (8), the chromaticities $\{\boldsymbol{u}'_1, \boldsymbol{v}'_1\}$ are trans-

(a)

(b)

(c)

Figure 5: A second example of color equalization: (a) Image $\mathcal{C}_1$ taken under a bluish illuminant; (b) Image $\mathcal{C}_2$ of the same scene taken under a yellowish illuminant; (c) Image *warp*$\{\mathcal{C}_1\}$, obtained by equalizing $\mathcal{C}_1$ against $\mathcal{C}_2$ with our algorithm.

formed into $warp_{u'v'}\{u'_1, v'_1\}$; Figure 1 (f) displays such chromaticities for the reported example. From $\mathcal{P}_1$ a new palette can now be built as $warp\{\mathcal{P}_1\} \doteq [warp_Y\{Y_1\}\ warp_{u'v'}\ \{u'_1, v'_1\}]$ and a new image, in turn, can be built from $\mathcal{C}_1$ as $warp\{\mathcal{C}_1\} \doteq \{\mathcal{Q}_1, warp\ \{\mathcal{P}_1\}, w_1\}$. Figure 1 (e) shows the image $warp\{\mathcal{C}_1\}$, obtained by equalizing $\mathcal{C}_1$ against $\mathcal{C}_2$ of the same figure with our algorithm. We may notice that the appearance of the colors of $warp\{\mathcal{C}_1\}$ is very close to that of the colors of $\mathcal{C}_2$. Fig. 5 displays the results obtained with another pair of test images [14]; in this case, the change of illuminant between the two in rather severe since the first was acquired with a bluish illuminant and the second with a yellowish one. Nonetheless, the proposed color correction algorithm performs satisfactorily.

4 Conclusions

A new method for correcting color of digital images has been presented. It works separately on the achromatic and chromatic channels of a color image. The correction of the achromatic channel is carried out with a standard technique for gray-level images while the correction of the 2-D chromatic channel is considered as a problem of image warping. The effectiveness of this method has been confirmed by the two experiments reported in the paper.

Acknowledgements

This work was supported by a University of California MICRO grant with matching supports from National Instruments, NEC Corporation and Intel Corporation.

References

[1] B.A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., Sunderland, Massachussets, 1995.

[2] K. Staes,"Light Sources as an Integral Part of the Color Photographic System, *Soc. of Motion Picture and Television Eng. J.*, Vol. 86, pp. 537-543, 1997.

[3] H.-C. Lee, "Method for Computing the Scene-Illuminant Chromaticity from Specular Highlights, *Journal of the Optical Society of America*, Vol. A-3, No. 10, pp. 1,694-1,699, 1986.

[4] M.J. Vrhel and H.J. Trussell, "Color Correction Using Principal Components, *Color Research and Application*, Vol. 17, No. 5, pp. 328-338, 1992.

[5] P.-C. Hung,"Color Rendition Using Three-Dimensional Interpolation, *Proc. of the SPIE: Imaging Applications in the Work World*, pp. 111- 115, 1988.

[6] R.H. Kang and P.G. Anderson, "Neural Network Applications to the Color Scanner and Printer Calibration, *Journal of Electronic Imaging*, Vol. 1, No. 12, pp. 125-135, 1992.

[7] K. Kanamori, H. Kawakami, and H. Kotera, "A Novel Color Transformation Algorithm and Its Applications, *Proc. of the SPIE: Image Processing Algorithms and Techniques*, Vol. 1,244, pp. 272-281, 1990.

[8] L. Lucchese and S.K. Mitra, "A New Method for Color Image Equalization," *Proc. of 2001 Int'l Conference on Image Processing (ICIP 2001)*, Thessaloniki, Greece, Sept. 2001, Vol. I, pp. 133-136.

[9] M. Pappas and I. Pitas, "Digital Color Restoration of Old Paintings," *IEEE Transactions on Image Processing*, Vol. 9, No. 2, pp. 291-294, Feb. 2000.

[10] R.W.G. Hunt, *Measuring Colour*, 2^{nd} Ed., Ellis Horwood Ltd. Publ., Chichester, UK, 1987.

[11] R.C. Gonzales and R.E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[12] M. Petrou and P. Bosdogianni, *Image Processing – The Fundamentals*, John Wiley & Sons, Ltd., Chichester, UK, 2000.

[13] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, California, 1992.

[14] D. H. Brainard, Hyperspectral Image Data, `http://color.psych.ucsb.edu/hyperspectral`.

[15] J.E. Dennis, Jr., and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[16] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2^{nd} Ed., Cambridge University Press, Cambridge, UK, 1992.

KOMBINACYJNA ACHROMATYCZNA I CHROMATYCZNA KOREKCJA OBRAZÓW KOLOROWYCH

Streszczenie

W artykule przedstawiono nowy algorytm służący do określenia odwzorowania kolorów pomiędzy dwoma obrazami tego samego obiektu lub sceny, uzyskanymi w różnych warunkach oświetleniowych. Działanie zaproponowanego algorytmu oparte jest na zasadzie kompensacji różnic kolorów, co wykonuje się niwelując różnice oddzielnie dla składowej achromatycznej i chromatycznej. Wyrównanie (niwelacja różnic) w achromatycznym kanale 1-D jest dokonywane poprzez wykorzystanie standardowej techniki używanej w przypadku obrazów czarno-białych, natomiast operacja wyrównania w chromatycznym kanale 2-D jest rozpatrywana jako problem deformacji obrazu. Przedstawiona metoda może być również zastosowana do kalibracji kolorów czujników kolorów, przy założeniu, że referencyjny rozkład kolorów jest znany.

Słowa kluczowe: korekcja kolorów, deformacje 2-D, składniki: achromatyczny i chromatyczny, luminancja, rozkłady kolorów, stałość kolorów

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
TELEKOMUNIKACJA I ELEKTRONIKA 9 (2006), 37-44

NONSTANDARD EULERIAN AND HAMILTONIAN GRAPHS, AND A COLORING THEOREM FOR NONSTANDARD GRAPHS

Armen H. Zemanian

Electrical Engineering Department
State University of New York at Stony Brook
Stony Brook, New York 11794-2350, USA

From any given sequence of finite or infinite graphs, a nonstandard graph can be constructed. The procedure is similar to an ultrapower construction of an internal set from a sequence of subsets of the real line, but now the individual entities are the nodes of the graphs instead of real numbers. The transfer principle can then be invoked to extend several graph-theoretic results to the nonstandard case. In this work, nonstandard versions of Eulerian graphs, Hamiltonian graphs, and a coloring theorem are established for nonstandard graphs

Keywords: Nonstandard graphs, transfer principle, ultrapower constructions, Eulerian graphs, Hamiltonian graphs, a graph-coloring theorem.

1. INTRODUCTION

A nonstandard graph can be defined by applying the transfer principle to a set X of nodes, with the set B of branches defined by a symmetric irreflexive binary relationship on X [3], [4]. Thus, the conventional graph $G=\{X,B\}$ is transferred to the nonstandard graph $^*G=\{^*X,{}^*E\}$. This result can be used to establish in a nonstandard way the standard theorem that, if every finite subgraph of a conventional infinite graph G has a coloring with finitely many colors, namely k colors, then G itself has such a finite coloring with k colors.

A different and more general construction of a nonstandard graph [6] starts with an arbitrary sequence of conventional (finite or infinite) graphs and constructs from that a nonstandard graph in much the same way as an internal set in the hyperreal line $^*\mathbb{R}$ is constructed from a given sequence of subsets of the real line $\mathbb{R}$, that is, by means of an ultrapower construction. In this case, the resulting nonstandard graph has nonstandard branches and nonstandard nodes. This construction of a nonstandard graph is presented in [6]. It is more general than the prior approach based on a symmetric irreflexive binary relation in that the graphs of the sequence need not be node-induced subgraphs of a given graph.

After setting up our nonstandard graphs using an ultrapower approach, we invoke the transfer principle to lift several standard graph-theoretic results into a nonstandard setting. In particular, standard theorems concerning Eulerian graphs, Hamiltonian graphs, and a coloring theorem are extended to the nonstandard setting. By virtue of the transfer principle, this only requires that the standard theorems be stated as sentences in symbolic logic, which are then transferred to appropriate sentences for nonstandard graphs.

In the following, a set A is denoted with braces, as for example by $A = \{a, b.c, ...\}$, where the terms within the braces are the elements of the set A. A may have any cardinality, the latter being denoted by $|A|$. $\mathbb{N}$ denotes the set of natural numbers: $\mathbb{N} = \{0, 1, 2, ...\}$. Thus, a sequence is a function from $\mathbb{N}$ into a set A and is denoted by $\langle a_n : n \in \mathbb{N} \rangle$ or simply by $\langle a_n \rangle$ where $n \in \mathbb{N}$ is understood. $\mathbb{R}$ denotes the set of real numbers. Hence, ${}^*\mathbb{N}$ is the set of hypernaturals, and ${}^*\mathbb{R}$ is the set of hyperreals.

2. NONSTANDARD GRAPHS

A standard graph G is a conventional (finite or infinite) graph $G = \{X, B\}$, where X is the set of its nodes and B is the set of its branches. Each branch $b \in B$ is a two-element set $b = \{x, y\}$ with $x, y \in X$ and $x \neq y$; b and x are said to be incident and so, too, are b and y. Also, x and y are said to be adjacent through b. By this definition, there are no multiedges (i.e., no parallel branches) and no self-loops (i.e., no branches consisting of a single node). Paths, loops (i.e., closed paths), trails, and closed trails in a standard finite graph have their usual definitions.

Next, let $\langle G_n : n \in \mathbb{N} \rangle$ be a given sequence of graphs. For each n, we have $G_n = \{X_n, B_n\}$, where X_n is the set of branches and B_n is the set of nodes. We allow $X_n \cap X_m \neq \varnothing$ so that G_n and G_m may be subgraphs of a larger graph. In fact, we may have $X_n = X_m$ and $B_n = B_m$ for all $n, m \in \mathbb{N}$ so that G_n may be the same graph for all $n \in \mathbb{N}$. Furthermore, let F be a chosen nonprincipal ultrafilter on $\mathbb{N}$ [3].

In the following, $\langle x_n \rangle = \langle x_n : n \in \mathbb{N} \rangle$ will denote a sequence of nodes with $x_n \in X_n$ for all $n \in \mathbb{N}$. A nonstandard node ${}^*x$ is an equivalence class of such sequences of nodes, where two such sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ are taken to be equivalent if $\{n: x_n = y_n\} \in F$, in which case we write "$\langle x_n \rangle = \langle y_n \rangle$ a.e." or say that $x_n = y_n$ "for almost all n." We also write ${}^*x = [x_n]$, where it is understood that the x_n are the members of any one sequence in the equivalence class. We let ${}^*X$ denote the set of nonstandard nodes.

Next, we define the nonstandard branches: Let ${}^*x = [x_n]$ and ${}^*y = [y_n]$ be two nonstandard nodes. Let $N_{xy} = \{n: \{x_n, y_n\} \in B_n\}$ and $N_{xy}^c = \{n: \{x_n, y_n\} \notin B_n\}$. Since F is an ultrafilter, exactly one of N_{xy} and N_{xy}^c is a member of F. If it is N_{xy}, then

${}^*b = [\{x_n, y_n\}]$ is defined to be a nonstandard branch; that is, ${}^*b$ is an equivalence class of sequences $\langle b_n \rangle$ of branches $b_n = \{x_n, y_n\} \in B_n$, $n = 0,1,2,\ldots$. In this case, we also write ${}^*x, {}^*y \in {}^*b$ and ${}^*b = \{{}^*x, {}^*y\}$. We let ${}^*B$ denote the set of nonstandard branches. On the other hand, if $N^c_{xy} \in F$, then $[\{x_n, y_n\}]$ is not a nonstandard branch.

Finally, we define a nonstandard graph ${}^*G$ to be the pair ${}^*G = \{{}^*X, {}^*B\}$.

That all this properly defines the nonstandard nodes, the nonstandard branches, and a nonstandard graph is shown in [6, Section 8.1].

A special case arises when all the G_n are the same standard graph $G = \{X, B\}$. When this is so, we call ${}^*G = \{{}^*X, {}^*B\}$ an *enlargement* of G, in conformity with an "enlargement" ${}^*A$ of a subset A of $\mathbb{R}$ [3]. If moreover G is a finite graph, each node $x \in {}^*X$ can be identified with a node of X because the enlargement of a finite set equals the set itself. Similarly, every branch $b \in {}^*B$ can be identified with a branch in B. Thus, ${}^*G = G$. On the other hand, if G is a conventionally infinite graph, X is an infinite set, and its enlargement ${}^*X$ has more elements, namely, nonstandard nodes that are not equal to standard nodes, (i.e., ${}^*X \setminus X$ is not empty). Similarly, ${}^*B \setminus B$ is not empty too. In short, ${}^*G$ is a proper enlargement of G.

For an example of a nonstandard enlargement, consider a standard one-way infinite path $P = \{X, B\}$; thus, its nodes form a sequence such that $b = \{x, y\} \in B$ if and only if x and y are adjacent in that sequence. Its enlargement ${}^*P = \{{}^*X, {}^*B\}$ has hypernodes that form a hypersequence (i.e., its hypernodes are totally ordered according to the total order of ${}^*\mathbb{N}$, and ${}^*b = \{{}^*x, {}^*y\}$, where ${}^*x = [x_n]$ and ${}^*y = [y_n]$, is a member of ${}^*B$ if and only if $\{n\colon x_n$ and y_n are adjacent$\} \in F$.

More generally, any infinite graph G has an enlargement ${}^*G$ that contains G as a nonstandard subgraph. In fact, just as the set $\mathbb{N}$ of natural numbers enlarges into the set ${}^*\mathbb{N}$ of hypernatural numbers with infinitely many galaxies and with $\mathbb{N}$ being its principal galaxy [3], here too the graphical galaxies can be defined as nonstandard subgraphs of ${}^*G$. But now, ${}^*G$ either has just one galaxy, namely G, or ${}^*G$ has infinitely many galaxies with G being its principal galaxy. Moreover, in the latter case, the galaxies of ${}^*G$ are partially ordered according to their closeness to the principal galaxy G, but there need not be a total ordering of those galaxies. These ideas about graphical galaxies are discussed in another paper [7].

A different special case arises when almost all the G_n are (possibly different) finite graphs. We will refer to the resulting nonstandard graph ${}^*G$ as a hyperfinite graph. We can lift many theorems concerning finite graphs to hyperfinite graphs. It is just a matter of writing the standard theorems in an appropriate form using symbolic logic and then applying the transfer principle.

What needs to be known about nonstandard analysis and nonstandard graphs for a comprehension of this paper is presented in [6]. Incidences and adjacencies between hypernodes and hyperbranches are defined and examined in [6]. Similarly, nonstandard hyperfinite paths and loops appear in [6], and connectedness for nonstandard graphs is discussed in [6]. These ideas are used in the following. Some knowledge of symbolic logic and the transfer principle are also used in the following. The needed information is given in [6, Appendix A].

Sections 2 through 4 in this paper present new results that have not been published elsewhere.

3. EULERIAN GRAPHS

A finite trail is defined much as a finite path is defined except that the condition that all the nodes be distinct is relaxed; however, branches are still required to be distinct. Thus, the truth of the following symbolic sentence defines a *trail* T in a finite graph $G=\{X,B\}$, with T having two or more branches. This time we use the notation $b_m=\{x_m,y_m\}$ to display the nodes x_m and y_m that are incident to b_m.

$$(\exists k\in \mathbb{N}\setminus\{0\})\,(\exists b_0,b_1,\dots,b_k\in B)\,(\forall m\in\{0,\dots,k-1\})\,(y_m=x_{m+1})$$

That B is a set insures that the branches $b_0,b_1,\dots,b_k$ are all distinct. On the other hand, this sentence allows nodes to repeat in a trail.

For a *closed trail*, we have the truth of the following symbolic sentence as its definition.

$$(\exists k\in \mathbb{N}\setminus\{0\})\,(\exists b_0,b_1,\dots,b_k\in B)\,(\forall m\in\{0,\dots,k-1\})\,(y_m=x_{m+1})\wedge(y_k=x_0)$$

With Q denoting a trail, we denote the set of branches in Q by $B(Q)$. Also, we let $\mathcal{Q}(G)$ denote the set of closed trails in a given graph $G=\{X,B\}$.

By attaching asterisks (as usual when invoking the transfer principle), we obtain by transfer the corresponding sentence for trails in a given nonstandard graph ${}^*G=\{{}^*X,{}^*B\}$. Thus, a *nonstandard trail* ${}^*Q$ is defined by the truth of the following symbolic sentence; now, $b_m=\{x_m,y_m\}$ is a nonstandard branch with the nonstandard nodes x_m and y_m.

$$(\exists k\in {}^*\mathbb{N}\setminus\{0\})\,(\exists b_0,b_1,\dots,b_k\in {}^*B)\,(\forall m\in\{0,\dots,k-1\})\,(y_m=x_{m+1}).$$

A similar expression holds for a *nonstandard closed trail* (just append $\wedge(y_k=x_0)$). With ${}^*Q$ denoting a nonstandard trail, we denote the set of nonstandard branches in ${}^*Q$ by ${}^*B({}^*Q)$. Also, we let ${}^*\mathcal{Q}({}^*G)$ denote the set of nonstandard closed trails in a given nonstandard graph ${}^*G=\{{}^*X,{}^*B\}$.

Let $\mathcal{C}_f$ denote the set of finite connected graphs. Thus, ${}^*\mathcal{C}_f$ denotes the set of hyperfinite connected nonstandard graphs [6]. A finite connected graph $G=\{X,B\}\in\mathcal{C}_f$ is called *Eulerian* if it contains a closed trail that meets every node of X. The degree d_x of $x\in X$ is the natural number $d_x=|\{b\in B\colon x\in b\}|$. The nonstandard version of this definition is as follows: Let ${}^*G=\{{}^*X,{}^*B\}\in{}^*\mathcal{C}_f$. For any $x\in{}^*X$, the *degree* of x is $d_x=|\{b\in{}^*B\colon x\in b\}|$. In this case, d_x may be an unlimited hypernatural number when ${}^*G$ is a hyperfinite graph. However, ${}^*G$ might happen to be a finite graph

$G \in C_f$, which from the point of view of an ultrapower construction can occur if all the G_n for ${}^*G = [G_n]$ are the same finite graph $G \in \mathcal{G}_f$; in this case, d_x will be a natural number for all $x \in {}^*X$.

Let $\mathcal{E}_u$ (resp. ${}^*\mathcal{E}_u$) denote the set of all standard Eulerian graphs (resp. nonstandard Eulerian graphs). Then, Eulerian graphs can be defined by asserting the truth of the following symbolic sentence to the right of $\leftrightarrow$, where as usual $G = \{X, B\}$.

$$G \in \mathcal{E}_u \leftrightarrow (\exists\, Q \in \mathcal{Q}(G))\,((\forall\, b \in B)\,(b \in B(Q))$$

By transfer, the truth of the following right-hand side defines *nonstandard Eulerian graphs*. Now, ${}^*G = \{{}^*X, {}^*B\}$.

$${}^*G \in {}^*\mathcal{E}_u \leftrightarrow (\exists\, {}^*Q \in {}^*\mathcal{Q}(G))\,((\forall\, b \in {}^*B)\,(b \in {}^*B({}^*Q)))$$

Now an ancient theorem of Euler asserts that a graph G is Eulerian if and only if the degree of every node of $G = \{X, B\}$ is an even natural number. Symbolically, this can be stated as follows.

$$G \in \mathcal{E}_u \leftrightarrow (\forall\, x \in X)\,(d_x/2 \in \mathbb{N})$$

Transferring this, we get the nonstandard version of this theorem of Euler:

$${}^*G \in {}^*\mathcal{E}_u \leftrightarrow (\forall\, x \in {}^*X)\,(d_x/2 \in {}^*\mathbb{N})$$

4. HAMILTONIAN GRAPHS

In this and the next section, the symbol $x \Diamond y$ denotes the condition that the standard nodes x and y are adjacent (i.e., are incident to the same branch). The same symbol is used when x and y are adjacent hypernodes (i.e., are incident to the same hyperbranch).

In this section, it is assumed that each graph $G = \{X, B\}$ is connected and finite and has at least three nodes (i.e., $|X| \geq 3$). A graph G is called *Hamiltonian* if it contains a loop (i.e., a closed path) that meets every node in the graph. Let $\mathcal{L}(G)$ denote the set of all loops in G. Also, for any loop $L \in \mathcal{L}(G)$, let $X(L)$ denote the node set of L. Then, a Hamiltonian graph $G = \{X, B\} \in C_f$ is also defined by the truth of the following symbolic sentence to the right of $\leftrightarrow$. The set of Hamiltonian graphs will be denoted by $\mathcal{H}$, where $\mathcal{H} \subset C_f$.

$$G \in \mathcal{H} \leftrightarrow (\exists\, L \in \mathcal{L}(G))\,((\forall\, x \in X)\,(x \in X(L)))$$

By transfer of this symbolic sentence, we define a *nonstandard Hamiltonian graph* as follows, where now ${}^*\mathcal{H}$ is the set of nonstandard hyperfinite Hamiltonian graphs, $\mathcal{L}({}^*G)$ is the set of all nonstandard loops in ${}^*G$, ${}^*X(L)$ is the set of nonstandard nodes in $L \in \mathcal{L}({}^*G)$, and ${}^*G = \{{}^*X, {}^*B\}$.

$$^*G \in {}^*\mathcal{H} \leftrightarrow (\exists\, L \in \mathcal{L}(^*G))\,((\forall\, x \in {}^*X)\,(x \in {}^*X(L))).$$

A sufficient condition for a graph $G = \{X, B\}$ to be Hamiltonian is that the degree d_x of each of its nodes x be no less than one half of $|X|$ [1], [2]. Symbolically, this condition is expressed as follows:

$$((\forall\, x \in X)\,(d_x \geq |X|/2)) \rightarrow \mathcal{G} \in \mathcal{H}.$$

By transfer, we get the following criterion for a nonstandard Hamiltonian graph.

$$((\forall\, x \in {}^*X)\,(d_x \geq |{}^*X|/2)) \rightarrow {}^*\mathcal{G} \in {}^*\mathcal{H}.$$

A more general criterion due to Ore asserts that $G = \{X, B\}$ is Hamiltonian if, for every pair of nonadjacent nodes x and y, $d_x + d_y \geq |X|$ [1], [2]. Symbolically, we have

$$((\forall\, x, y \in X)\,((\neg(x \Diamond y)) \rightarrow (d_x + d_y \geq |X|))) \rightarrow G \in \mathcal{H},$$

which by transfer becomes

$$((\forall\, x, y \in {}^*X)\,((\neg(x \Diamond y)) \rightarrow (d_x + d_y \geq |{}^*X|))) \rightarrow {}^*G \in {}^*\mathcal{H}.$$

Still more general is Posa's theorem [1], [2]: If, for every $j \in \mathcal{N}$ satisfying $1 \leq j < |X|/2$, the number of nodes of degree no larger than j is less than j, then the graph $G = \{X, B\}$ is Hamiltonian. The following symbolic sentence states this criterion.

$$((\forall\, j \in \mathcal{N})\,(\forall\, x \in X)\,((1 \leq j < |X|/2) \rightarrow (|\{x \in X : d_x \leq j\}| < j))) \rightarrow G \in \mathcal{H}$$

By transfer the following criterion holds for nonstandard graphs ${}^*G = \{{}^*X, {}^*B\}$.

$$((\forall\, j \in {}^*\mathcal{N})\,(\forall\, x \in {}^*X)\,((1 \leq j < |{}^*X|/2) \rightarrow (|\{x \in {}^*X : d_x \leq j\}| < j))) \rightarrow {}^*G \in {}^*\mathcal{H}$$

5. A COLORING THEOREM

At the beginning of this paper, it was mentioned that a standard coloring theorem for infinite graphs could be established in a nonstandard way. This is accomplished by first lifting that standard theorem to a nonstandard setting. In this section a different coloring theorem for finite graphs will be lifted to an assertion for nonstandard hyperfinite graphs.

A simple graph-coloring theorem that is not restricted to planar graphs asserts that, if the largest of the degrees for the nodes of a finite graph $G = \{X, B\}$ is k, then the graph is $(k+1)$-colorable [5].[1] To express this symbolically, first let $M(X, \mathcal{N}_{k+1})$ denote the set of all functions that map a set X into the set $\mathcal{N}_{k+1}$ of those natural numbers j satisfying $1 \leq j \leq k+1$. Also, let $\mathcal{G}_f$ (resp. ${}^*\mathcal{G}_f$) denote the set of finite

[1] There exists a function f that assigns to each node one of $k+1$ colors such that no two adjacent nodes have the same color.

graphs (resp. hyperfinite graphs). Then, the following restates this theorem for any graph $G = \{X, B\}$.

$$(\forall G \in \mathcal{G}_f)(\exists\, k \in \mathbb{N})\,(\forall\, x, y \in X)$$

$$((d_x \le k) \to ((\exists\, f \in M(X, \mathbb{N}_{k+1}))\,((x \lozenge y) \to (f(x) \ne f(y)))))$$

To transfer this to any hyper finite graph ${}^{*}G = \{{}^{*}X, {}^{*}B\}$, we first let ${}^{*}M({}^{*}X, \mathbb{N}_{k+1})$ be the set of all internal functions mapping the enlargement ${}^{*}X$ into $\mathbb{N}_{k+1}$. Then, this theorem is transferred to nonstandard graphs simply by appending asterisks, as usual:

$$(\forall^{*}G \in {}^{*}\mathcal{G}_f)(\exists\, k \in \mathbb{N})\,(\forall\, x, y \in {}^{*}X)$$

$$((d_x \le k) \to ((\exists\, {}^{*}f \in {}^{*}M({}^{*}X, \mathbb{N}_{k+1}))\,((x \lozenge y) \to ({}^{*}f(x) \ne {}^{*}f(y)))))$$

Note here that the assumption of a natural-number bound k on the degrees of all the nonstandard nodes has been maintained. This conforms to the fact that the enlargement of the finite set $\mathbb{N}_{k+1}$ is $\mathbb{N}_{k+1}$. As a consequence, the conclusion remains strong.

On the other hand, we could generalize this transferred theorem as follows: In terms of an ultrapower construction, we could replace $\mathbb{N}_{k+1}$ by an internal set ${}^{*}\mathbb{N}_{k+1}$ obtained from a sequence $\langle \mathbb{N}_{k_n+1} : n \in \mathbb{N} \rangle$ of finite sets $\mathbb{N}_{k_n+1}$, one set for each G_n with regard to ${}^{*}G = [G_n]$ and with k_n being the maximum node degree in G_n. But then, our conclusion would be weakened to a coloring with a hypernatural number ${}^{*}k = [k_n]$ of colors.

6. A FINAL COMMENT

Undoubtedly, other standard results for graphs can be lifted in this way to nonstandard settings.

BIBLIOGRAPHY

[1] Behzad M., Chartrand G., 1971. Introduction to the Theory of Graphs. Allyn and Bacon Inc. Boston.

[2] Buckley F., Harary F., 1990. Distances in Graphs. Addison-Wesley Publishing Co. New York.

[3] Goldblatt R., 1998. Lectures on the Hyperreals. Springer New York.

[4] Mendelson E., 1992 Introduction to Mathematical Logic, Fourth Edition. Chapman Hall/CRC, Boca Raton Florida.

[5] Wilson R.J., 1972 Introduction to Graph Theory. Academic Press New York.

[6] Zemanian A.H., 2004. Graphs and Networks: Transfinite and Nonstandard. Birkhauser Boston.

[7] Zemanian A.H., 2005. The Galaxies of Nonstandard Enlargements of Infinite and Transfinite Graphs: II. CEAS Technical Report 819, University at Stony Brook April.

EULEROWSKIE I HAMILTONOWSKIE GRAFY NIESTANDARDOWE ORAZ TWIERDZENIE O KOLOROWANIU DLA GRAFÓW NIESTANDARDOWYCH

Streszczenie

Z dowolnej sekwencji grafów skończonych lub nieskończonych można zbudować graf niestandardowy. Procedura w tym przypadku jest podobna do konstrukcji tzw. ultramocy zbioru wewnętrznego z sekwencji podzbiorów zbioru liczb rzeczywistych, gdzie teraz poszczególne wielkości są węzłami grafów, a nie liczbami rzeczywistymi. Wtedy zasada transferu może być zastosowana w celu rozszerzenia zakresu stosowalności wielu teoretycznych wyników dotyczących grafów standardowych na grafy niestandardowe. W tej pracy są rozpatrywane niestandardowe przypadki grafów Eulerowskich i Hamiltonowskich oraz twierdzenie o kolorowaniu dla przypadku grafów niestandardowych.

Słowa kluczowe: grafy niestandardowe, zasada transferu, konstrukcje ultramocy, grafy Eulerowskie i Hamiltonowskie, twierdzenie o kolorowaniu grafów.

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
TELEKOMUNIKACJA I ELEKTRONIKA 9 (2006), 45-52

PRODUCTS OF DIRAC IMPULSES IN NONLINEAR ANALYSIS WITH THE USE OF VOLTERRA SERIES

Andrzej Borys

Institute for Telecommunications
Faculty of Telecommunications and Electrical Engineering
University of Technology and Agriculture
ul. Kaliskiego 7, 85-791 Bydgoszcz, Poland

It is shown in this paper that the products of Dirac impulses can occur in the nonlinear analysis with the use of Volterra series. Then these products must be however treated as the products of Dirac impulses of different arguments. Moreover, they can be assumed to be multi-dimensional Dirac impulses satisfying similar conditions as those regarding the ordinary one-dimensional Dirac impulse, but now in the corresponding multi-dimensional time or frequency domain. The defining relations for these multi-dimensional Dirac impulses are derived. Also the expressions for their Fourier transforms are given.

Keywords: nonlinear analysis, Volterra series, Dirac impulse, multi-dimensional Dirac impulses.

1. INTRODUCTION

There are researchers who believe that the products of Dirac impulses, which can occur in nonlinear analysis with the use of Volterra series of continuous time, are not allowed. They say that such products are forbidden by mathematics, or more precisely, by the theory of distributions [1]. Expressing the above statement, they however forget that in the nonlinear analysis we do not have to do with the products of one variable but with the products of more variables. That is we deal in the nonlinear analysis with the products like $\delta(t_1)\delta(t_2)$, $\delta(t_1)\delta(t_2)\delta(t_3)$, and so on, not with the products of the form $\delta(t)\delta(t)$, $\delta(t)\delta(t)\delta(t)$, and so on, where $\delta(t)$ means a Dirac impulse of a continuous time variable t.

The products $\delta(t)\delta(t)$, $\delta(t)\delta(t)\delta(t)$, and of higher order of this type do not exist [8], of course, contrary to their multi-dimensional counterparts $\delta(t_1)\delta(t_2)$, $\delta(t_1)\delta(t_2)\delta(t_3)$, and so on, which make sense.

The objective of this paper is to show, using a quite simple and understandable mathematics, that the usage of Dirac impulse products in the Volterra nonlinear analysis, as it has been done in [2, 3, 4, 5, 6, 7], is fully correct.

2. A DIRAC IMPULSE AND AN ACCOMPANYING NORMALIZING EQUATION

A Dirac impulse is considered mathematically as a distribution [1]. In the literature, it is also called a generalized function or a Dirac delta function. The inherent feature of the definition of this generalized function or distribution is the following normalizing condition:

$$\int_{-\infty}^{\infty} \delta(t)dt = 1. \tag{1}$$

It is the essential part of the Dirac delta function definition, causing that this generalized function is well defined.

Using $\delta(t)$ without the condition given by (1) has no meaning and leads therefore to errors. It is however common to write out Dirac impulses $\delta(t)$ alone, as for example in the following sentence: "let us now put the signal $\delta(t)$ at the circuit input". Writing so, we include tacitly also the normalizing equation given by (1). In other words, it is assumed per definition that any Dirac impulse $\delta(t)$ is accompanied by a normalizing condition of the form (1).

From the theory of distributions [8], we know that a simple product of two Dirac impulses $\delta(t)\delta(t)$ (of the same continuous time variable t) does not exist. However, a one-dimensional convolution of two Dirac impulses, that is $\int_{-\infty}^{\infty} \delta(\tau)\delta(t-\tau)d\tau$ does exist [8]. As a result, this convolution gives a new Dirac delta function

$$\int_{-\infty}^{\infty} \delta(\tau)\delta(t-\tau)d\tau = \delta(t). \tag{2}$$

Note that a product of two Dirac impulses of two different continuous time variables τ and t occurs in (2) under the integral symbol.

Let us apply the normalizing condition given by (1) to the Dirac impulse occurring on the right-hand side of (2). As a result, we obtain then

$$1 = \int_{-\infty}^{\infty} \delta(t)dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\tau)\delta(t-\tau)d\tau dt. \tag{3}$$

Now, by introducing a new variable $t - \tau = t_1$ with $dt = dt_1$ on the right-hand side of (3), we get

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\tau)\delta(t-\tau)d\tau dt &= \int_{-\infty}^{\infty} \delta(\tau)\int_{-\infty}^{\infty} \delta(t_1)dt_1 d\tau = \\ &= \left(\int_{-\infty}^{\infty} \delta(\tau)d\tau\right)\left(\int_{-\infty}^{\infty} \delta(t_1)dt_1\right) = 1 \cdot 1 = 1.\end{aligned} \tag{4}$$

The result given by (4) can be rewritten in the following form:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(t_1)\delta(t_2)dt_1dt_2=\left(\int_{-\infty}^{\infty}\delta(t_1)dt_1\right)\left(\int_{-\infty}^{\infty}\delta(t_2)dt_2\right)=1. \tag{5}$$

Note that (5) represents a kind of a normalizing condition put upon the product of two Dirac impulses of two different time variables t_1 and t_2. This result can be interpreted in such a way that the product of two Dirac impulses of two different time variables does exist, according to (2), because it is a subject to a constraint given by (5).

The above result can be easily extended for more dimensions. To see this, consider again (2), and write the convolution of a Dirac impulse occurring on the right-hand side of (2) (now $\delta(\tau_1)$) with another Dirac impulse ($\delta(t-\tau_1)$). This leads to

$$\begin{aligned}\delta(t)&=\int_{-\infty}^{\infty}\delta(\tau_1)\delta(t-\tau_1)d\tau_1=\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau)\delta(\tau_1-\tau)d\tau\delta(t-\tau_1)d\tau_1=\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau)\delta(\tau_1-\tau)\delta(t-\tau_1)d\tau d\tau_1.\end{aligned} \tag{6}$$

Applying then the normalizing condition given by (1) to (6) gives

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau)\delta(\tau_1-\tau)\delta(t-\tau_1)d\tau d\tau_1 dt. \tag{7}$$

Afterwards, introducing a new variable $t-\tau_1=t_1$ with $dt=dt_1$ on the right-hand side of (7), we obtain

$$\begin{aligned}1&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau)\delta(\tau_1-\tau)\delta(t_1)d\tau d\tau_1 dt_1=\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau)\delta(\tau_1-\tau)\int_{-\infty}^{\infty}\delta(t_1)dt_1 d\tau d\tau_1.\end{aligned} \tag{8}$$

Proceeding similarly by introducing $\tau_1-\tau=t_2$ with $dt_2=d\tau_1$ in (8), we get

$$\begin{aligned}1&=\int_{-\infty}^{\infty}\delta(\tau)\left(\int_{-\infty}^{\infty}\delta(t_2)dt_2\right)\left(\int_{-\infty}^{\infty}\delta(t_1)dt_1\right)d\tau=\\&=\left(\int_{-\infty}^{\infty}\delta(t_2)dt_2\right)\left(\int_{-\infty}^{\infty}\delta(t_1)dt_1\right)\left(\int_{-\infty}^{\infty}\delta(\tau)d\tau\right)=1\cdot1\cdot1.\end{aligned} \tag{9}$$

Concluding, we see that the result of (9) can be rewritten as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(t_1)\delta(t_2)\delta(t_3)dt_1dt_2dt_3 = \\ =\left(\int_{-\infty}^{\infty}\delta(t_1)dt_1\right)\left(\int_{\infty}^{\infty}\delta(t_2)dt_2\right)\left(\int_{-\infty}^{\infty}\delta(t_3)dt_3\right)=1 . \tag{10}$$

Equation (10) represents a kind of a normalizing condition put upon the product of three Dirac impulses of three different time variables t_1, t_2, and t_3. Moreover, this result can be interpreted in such a way that the product of three Dirac impulses of three different time variables does exist, similarly as $\delta(t_1)\delta(t_2)$ before, because of the fact that it is a subject to a constraint given by (10).

It is evident that the above procedure can be continued for higher dimensions.

The results derived in this section will be used to prove the correctness of the Volterra nonlinear analysis, in which the products of ordinary Dirac impulses occur.

3. A MULTI-DIMENSIONAL GENERALIZATION OF A DIRAC IMPULSE

Consider first an example of a simple linear circuit being a linear conductor possessing an i-v characteristic given by the following equation:

$$i(t) = gv(t), \tag{11}$$

where $i(t)$ and $v(t) \neq \delta(t)$ are, respectively, the current and voltage at this conductor, and g represents its conductance. Note that (11) can be rewritten using the terminology of systems theory [8] that is by means of a convolution, with the use of a Dirac impulse, as

$$i(t) = \int_{-\infty}^{\infty} g\delta(\tau)v(t-\tau)d\tau, \tag{12}$$

where $g\delta(\tau)$ represents an impulse response of a circuit considered, i.e. of a linear conductor. This impulse response is in the form of a constant multiplied by a Dirac impulse. It illustrates one of the possible applications of the one-dimensional Dirac impulse in the theory of linear circuits and systems.

Let us consider another example to illustrate now the description of nonlinear circuits in the terminology of systems theory. To this end, we consider a nonlinear conductor having a quadratic i-v characteristic given by

$$i(t) = g_2\left(v(t)\right)^2, \tag{13}$$

where $i(t)$ and $v(t) \neq \delta(t)$ are, respectively, the current and voltage at the above nonlinear conductor, and g_2 is a constant. Similarly as (11), (13) can be put into the form exploited in the theory of systems. That is we can rewrite (13) as

$$
\begin{aligned}
i(t) &= g_2\left(\int_{-\infty}^{\infty}\delta(\tau_1)v(t-\tau_1)d\tau_1\right)\left(\int_{-\infty}^{\infty}\delta(\tau_2)v(t-\tau_2)d\tau_2\right)= \\
&= g_2\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\delta(\tau_1)v(t-\tau_1)d\tau_1\right)\delta(\tau_2)v(t-\tau_2)d\tau_2\right)= \\
&= g_2\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau_1)v(t-\tau_1)d\tau_1\delta(\tau_2)v(t-\tau_2)d\tau_2\right)= \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g_2\delta(\tau_1)\delta(\tau_2)v(t-\tau_1)v(t-\tau_2)d\tau_1 d\tau_2 \ ,
\end{aligned}
\tag{14}
$$

where $g_2\delta(\tau_1)\delta(\tau_2)$ represents a nonlinear impulse response of the second order - according to the terminology widely used in the literature (see for example [9]) - of the nonlinear conductor considered.

A comparison of (14) with (12) shows that (14) can be viewed as a generalization of one-dimensional convolution occurring in (12) to two-dimensional convolution in (14). Moreover, the product $\delta(\tau_1)\delta(\tau_2)$ can be assumed to be a natural generalization of one-dimensional Dirac impulse $\delta(\tau)$ for two dimensions. And because of this reason, the following normalizing condition in two dimensions:

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau_1)\delta(\tau_2)d\tau_1 d\tau_2 = \\
&=\left(\int_{-\infty}^{\infty}\delta(\tau_1)d\tau_1\right)\left(\int_{-\infty}^{\infty}\delta(\tau_2)d\tau_2\right)=1\cdot 1=1
\end{aligned}
\tag{15}
$$

should hold in this case, and it really holds. Note that this is the same normalizing condition as (5) derived in section 2. Moreover, note that the nonlinear (quadratic) counterpart of (2) will have now the form

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau_1)\delta(\tau_2)\delta(t_1-\tau_1)\delta(t_2-\tau_2)d\tau_1 d\tau_2 = \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\delta(\tau_1)\delta(t_1-\tau_1)d\tau_1\right)\delta(\tau_2)\delta(t_2-\tau_2)d\tau_2 = \\
&=\left(\int_{-\infty}^{\infty}\delta(\tau_1)\delta(t_1-\tau_1)d\tau_1\right)\left(\int_{-\infty}^{\infty}\delta(\tau_2)\delta(t_2-\tau_2)d\tau_2\right)= \\
&=\delta(t_1)\delta(t_2)\ .
\end{aligned}
\tag{16}
$$

A comparison of (16) with (2), and (15) with (1) shows that the product $\delta(t_1)\delta(t_2)$ on the right-hand side of (16) can be assumed to be a two-dimensional Dirac impulse. The normalizing condition (15) (in the form of a two-dimensional integral) applies to the above two-dimensional Dirac impulse. Moreover, note that this two-dimensional Dirac impulse can not be written as $\delta(t_1=t)\delta(t_2=t)=\delta(t)\delta(t)$, what would suggest that it is a product of two one-dimensional

Dirac impulses. Such the description is not simply possible because the product $\delta(t)\delta(t)$ does not exist [8]. On the other hand, it follows from the above discussion that the two-dimensional Dirac impulse $\delta(t_1)\delta(t_2)$ can be regarded in calculations as something like a product of two ordinary one-dimensional Dirac impulses. Then, however, the calculations must be carried out in two dimensions (the dimension of the Dirac impulse $\delta(t_1)\delta(t_2)$ can not be forgotten, dropped).

Note also that the form $\delta(t_1)\delta(t_2)$ of the two-dimensional Dirac impulse shows that this impulse can be factored.

Furthermore, it follows from the above derivations, presented for the case of a circuit with quadratic nonlinearity, that the theory of the two-dimensional Dirac impulse can be generalized to more dimensions. That is the three-dimensional Dirac impulse will have the form of $\delta(t_1)\delta(t_2)\delta(t_3)$ with the normalizing condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\tau_1)\delta(\tau_2)\delta(\tau_3)d\tau_1 d\tau_2 d\tau_3 = \left(\int_{-\infty}^{\infty}\delta(\tau_1)d\tau_1\right)\cdot \cdot\left(\int_{-\infty}^{\infty}\delta(\tau_2)d\tau_2\right)\left(\int_{-\infty}^{\infty}\delta(\tau_3)d\tau_3\right) = 1\cdot 1\cdot 1 = 1 \quad . \tag{17}$$

Similar forms and conditions like (17) will hold for four dimensions, five dimensions, and so on.

4. MULTI-DIMENSIONAL FOURIER TRANSFORMS OF MULTI-DIMENSIONAL DIRAC IMPULSES

A n-dimensional Fourier transform of a function $h(t_1,...,t_n)$ of the n-dimensional time is defined as [9]

$$H(f_1,...,f_n) = \mathcal{F}\{h(t_1,...,t_n)\} = = \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} h(t_1,...,t_n)\cdot \exp(-j2\pi(f_1t_1+...+f_nt_n))dt_1\cdots dt_n \,, \tag{18}$$

where $f_1,...,f_n$ mean the frequencies from the n-dimensional frequency space.

Applying the n-dimensional Dirac impulse $\delta(\tau_1)\cdots\delta(\tau_n)$ instead of $h(t_1,...,t_n)$ in (18), we get

$$\mathcal{F}\{\delta(t_1)\cdots\delta(t_n)\} = \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\delta(t_1)\cdots\delta(t_n)\cdot\exp(-j2\pi(f_1t_1+...+f_nt_n))dt_1\cdots dt_n = = \left(\int_{-\infty}^{\infty}\delta(t_1)\exp(-j2\pi f_1t_1)\right)\cdots\left(\int_{-\infty}^{\infty}\delta(t_1)\exp(-j2\pi f_nt_n)\right) = 1\cdots 1 = 1_n \tag{19}$$

The inverse n-dimensional Fourier transform is given by [9]

$$h(t_1,\dots,t_n) = \mathcal{F}^{-1}\{H(f_1,\dots,f_n)\} = \\ = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} H(f_1,\dots,f_n)\cdot\exp(j2\pi(f_1t_1+\dots+f_nt_n))df_1\cdots df_n . \quad (20)$$

Applying (20) to the Fourier transform of the n-dimensional Dirac impulse, which is equal to 1_n (see (19)), we arrive at

$$\mathcal{F}^{-1}\{1_n\} = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} 1_n \exp(j2\pi(f_1t_1+\dots+f_nt_n))df_1\cdots df_n = \\ = \left(\int_{-\infty}^{\infty} 1\exp(j2\pi f_1t_1)\right)\cdots\left(\int_{-\infty}^{\infty} 1\exp(j2\pi f_nt_n)\right) = \delta(t_1)\cdots\delta(t_n) . \quad (21)$$

The results given by (19) and (21) are of course the correct results in view of the theory presented.

The multi-dimensional Dirac impulses can also appear in the Volterra nonlinear analysis, which is carried out in the frequency-domain (in the multi-dimensional frequency space). In such the analysis, we can have to do with such functions in the n-dimensional time domain like [6]

$$\exp(j2\pi(f_at_1+\dots+f_at_n) , \quad (22)$$

where f_a is the frequency of a harmonic signal applied to the circuit input.

To calculate the n-dimensional Fourier transform of the function (22), we must use the relation (18). So applying (22) instead of $h(t_1,\dots,t_n)$ in (18), we get

$$\mathcal{F}\{\exp(j2\pi(f_at_1+\dots+f_at_n)\} = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\exp(j2\pi(f_at_1+\dots+f_at_n)\cdot \\ \cdot\exp(-j2\pi(f_1t_1+\dots+f_nt_n))dt_1\cdots dt_n = \left(\int_{-\infty}^{\infty} 1\exp(-j2\pi(f_1-f_a)t_1)\right)\cdots \\ \cdot\left(\int_{-\infty}^{\infty} 1\exp(-j2\pi(f_n-f_a)t_n)\right) = \delta(f_1-f_a)\cdots\delta(f_n-f_a) . \quad (23)$$

Note the occurrence of the products of Dirac impulses in the frequency domain on the right-hand side of (23). This is the correct result in view of the theory presented. It is called here the n-dimensional Dirac impulse in the n-dimensional frequency space.

5. CONCLUSION

The multi-dimensional Dirac impulses have been introduced in this paper. It has been shown that these impulses are factorable, that is they can be viewed as the products of ordinary one-dimensional Dirac impulses of different arguments. And this fact is fundamental for the nonlinear analysis with the use of Volterra series because such the products can occur in this analysis. However, as shown here, carrying out this analysis makes sense and leads to the correct results.

BIBLIOGRAPHY

[1] Schwartz L., 1966. Théorie des distributions. Hermann & Cie Paris.
[2] Boyd S., Chua L. O., and Desoer C. A., 1984. Analytical foundations of Volterra series. IMA Journal of Mathematical Control and Information, vol. 1, pp. 243-282.
[3] Gilbert E. G., 1977. Functional expansions for the response of nonlinear differential systems. IEEE Trans. on Automatic Control, vol. AC-22, no. 6, pp. 909- 921.
[4] Pupkov K. A., Kapalin V. I., and Yuscenko A. S., 1986. Functional Series in the Theory of Nonlinear Systems. Science Moscow (in Russian).
[5] Roszkiewicz J. and Borys A., 1979. Computer small signal analysis of analog circuits. Part I, fundamental analysis. Rozpr. Elektrotech., z. 1 (in Polish).
[6] Borys A., 1996. On Analysis of Nonlinear Analog Circuits Using Volterra Series. Wydawnictwa Uczelniane ATR Bydgoszcz (in Polish).
[7] Borys, A., 1987. Elementary deterministic theories of frequency and amplitude stability in feedback oscillators. IEEE Trans. Circuits and Systems, vol. CAS-34, pp. 254-258.
[8] Fliege N. J., 1991. Systemtheorie. Teubner Stuttgart.
[9] Bussgang J. J., Ehrman L., and Graham J. W., 1974. Analysis of nonlinear systems with multiple inputs. Proceedings of the IEEE, vol. 62, pp. 1088-1119.

ILOCZYNY IMPULSÓW DIRACA W ANALIZIE NIELINIOWEJ Z WYKORZYSTANIEM SZEREGU VOLTERRY

Streszczenie

W tym artykule pokazano, że iloczyny impulsów Diraca mogą występować w analizie nieliniowej, w której wykorzystuje się szereg Volterry. W takiej analizie te iloczyny muszą być jednakże traktowane jako iloczyny impulsów Diraca różnych argumentów. Pokazano, że mogą one wtedy być rozpatrywane jako wielowymiarowe impulsy Diraca, spełniające podobne warunki do tych, jakie dotyczą zwykłego, jednowymiarowego impulsu Diraca, ale teraz we właściwej, wielowymiarowej dziedzinie czasu lub częstotliwości. Wyprowadzono zależności definicyjne dla tych wielowymiarowych impulsów Diraca i podano wyrażenia określające ich transformaty Fouriera.

Słowa kluczowe: analiza nieliniowa, szereg Volterry, impuls Diraca, wielowymiarowe impulsy Diraca

AKADEMIA TECHNICZNO-ROLNICZA IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY
ZESZYTY NAUKOWE NR 247
TELEKOMUNIKACJA I ELEKTRONIKA 9 (2006), 53-63

THE MODIFIED NODAL FORMULATION FOR NONLINEAR CIRCUITS WITH MULTIPLE INPUTS

Andrzej Borys

Institute for Telecommunications
Faculty of Telecommunications and Electrical Engineering
University of Technology and Agriculture
ul. Kaliskiego 7, 85-791 Bydgoszcz, Poland

The modified nodal formulation (MNF) for nonlinear circuits with multiple inputs is presented in this paper. It is shown that in principle this approach is identical for the nonlinear circuits analyzed apart from that they possess a single or multiple inputs. That is the nonlinear models of basic circuit elements derived for the MNF are identical for both of these two cases as well as is the form of the MNF matrix equations for the corresponding orders of the nonlinear analysis. However, some particular properties are different. For example, the nonlinear transfer functions of a circuit with a single input can be calculated in a simple way by solving the MNF matrix equations for a Dirac impulse excitation. The nonlinear transfer functions are then equal to the corresponding circuit nodal voltages and branch currents for such the excitation. This does not however holds when a nonlinear circuit possesses more inputs. Furthermore, it is shown that solving the MNF matrix equations leads also to the correct results even when the circuit nonlinear impulse responses are the multidimensional Dirac impulses and the input signals at this circuit are the Dirac impulse signals.

Keywords: nonlinear analysis, nonlinear systems, modified nodal formulation, nonlinear circuits with multiple inputs, Volterra series.

1. INTRODUCTION

The modified nodal formulation (MNF) of circuit equations has proven to be an efficient and useful description suitable for computer programs of the analysis of linear circuits [1]. This description has been further extended in [2] to the case of the computer analysis of nonlinear analog circuits, which can be described by the Volterra series. The algorithms developed in [2] have been used in computer calculations of the responses of nonlinear circuits to periodic excitations [3], [4].

One of the objectives of this paper is to show that the approach developed in [2] is basically the same for circuits with single as well as for circuits with multiple inputs. These two cases (of single and multiple inputs) differ however from each other with regard to some properties. And we discuss here these issues in detail. As for example, the fact that getting the nonlinear transfer functions of the second and higher orders from the Fourier (or Laplace) transforms of the nodal voltages and branch currents, which one obtains by solving the linear equations containing the modified admittance matrix (occurring in the MNF), is possible for nonlinear circuits with a single input. But for nonlinear circuits with multiple inputs this is not generally possible. Note however that the same also regards the other methods like, as for example, the classic method of Bussgang et al. [5], based on the admittance matrix.

The method using the modified nodal formulation for nonlinear circuits, which is presented here, is a generalization of the method based on the admittance matrix published by Bussgang et al. [5]. The MNF is applicable to a wider class of nonlinear circuits (not only to those having the admittance matrix description) and aims rather in calculations of the components of the corresponding orders of the nodal voltages and branch currents than in evaluation of the circuit nonlinear transfer functions.

2. MODELS OF BASIC ELEMENTS OF NONLINEAR CIRCUITS WITH SINGLE AND MULTIPLE INPUTS

The basic principle used in derivations of models of basic elements of nonlinear circuits [2] for the use in the MNF is illustrated here on an example of a nonlinear conductor. So assume that the i-v characteristic of such the nonlinear conductor is given by a power series of the form

$$i(t) = g_1 v(t) + g_2 \left(v(t)\right)^2 + g_3 \left(v(t)\right)^3 + \dots \tag{1}$$

where $i(t)$ and $v(t)$ are the current and voltage, respectively, at the conductor, t is a continuous time, and g_1, g_2, g_3, and so on, are some constants. Furthermore, assume that this nonlinear conductor is embedded somewhere in a nonlinear circuit, and that the current i and the voltage v at the conductor can be expressed by the following Volterra series [6], [7]:

$$i(t) = \sum_{n=1}^{\infty} i^{(n)}(t) \tag{2a}$$

and

$$v(t) = \sum_{n=1}^{\infty} v^{(n)}(t) \ , \tag{2b}$$

where $i^{(n)}(t)$ and $v^{(n)}(t)$ are the partial responses in the current and voltage, respectively, associated with the corresponding orders $n = 1$ (linear), 2,3,..., of nonlinearity. These responses are the corresponding terms in the Volterra series and are given by [6], [7]

$$i^{(n)}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_I^{(n)}(\tau_1, \dots, \tau_n) \prod_{k=1}^{n} x(t - \tau_k) d\tau_k \tag{3a}$$

and

$$v^{(n)}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_V^{(n)}(\tau_1, \dots, \tau_n) \prod_{k=1}^{n} x(t - \tau_k) d\tau_k \ , \tag{3b}$$

for the input signal $x(t)$ of a nonlinear circuit with a single input. In (3a), $h_I^{(n)}(\tau_1,...,\tau_n)$ means the circuit nonlinear impulse response of the n-th order regarding the current at the conductor and related to the circuit input assumed. Similarly, $h_V^{(n)}(\tau_1,...,\tau_n)$ in (3b) is the circuit nonlinear impulse response of the n-th order regarding the voltage across the conductor and related to the circuit input assumed.

In the case of nonlinear circuits with multiple inputs, the expressions for $i^{(n)}(t)$ and $v^{(n)}(t)$ are more complicated and have the following forms:

$$i^{(n)}(t)=\sum_{i_1,..,i_n\in\{1,..,N\}}^{N^n}\int_{-\infty}^{\infty}..\int_{-\infty}^{\infty}h_{I i_1..i_n}^{(n)}(\tau_{i_1},..,\tau_{i_n})\prod_{k=1}^{n}x_{i_k}(t-\tau_{i_k})d\tau_{i_k} \tag{4a}$$

and

$$v^{(n)}(t)=\sum_{i_1,..,i_n\in\{1,..,N\}}^{N^n}\int_{-\infty}^{\infty}..\int_{-\infty}^{\infty}h_{V i_1..i_n}^{(n)}(\tau_{i_1},..,\tau_{i_n})\prod_{k=1}^{n}x_{i_k}(t-\tau_{i_k})d\tau_{i_k}\,, \tag{4b}$$

where $h_{I i_1..i_n}^{(n)}(\tau_{i_1},..,\tau_{i_n})$ and $h_{V i_1..i_n}^{(n)}(\tau_{i_1},...,\tau_{i_n})$ mean the circuit nonlinear impulse responses of the n-th order associated with the current at and the voltage across the conductor, respectively, and associated also with the $i_1,..,i_n\in\{1,..,N\}$ arrangement of the input signals. N in (4a) and (4b) is the number of circuit inputs. Moreover, $x_{i_k}(t-\tau_{i_k})$ means the input signal at the input $i_k\in\{1,..,N\}$ of the circuit, and the sum $\sum_{i_1,..,i_n\in\{1,..,N\}}^{N^n}$ stands for the summation operation over all the arrangements of indices $i_1,..,i_n\in\{1,..,N\}$ (of which the total number is N^n).

Dropping for convenience the argument t in (2a) and (2b), and introducing then the expressions given by (2a) and (2b) in (1), we obtain

$$\begin{aligned}i^{(1)}+i^{(2)}+i^{(3)}+...&=g_1v^{(1)}+\left[g_1v^{(2)}+g_2v^{(1)}v^{(1)}\right]+\\&+\left[g_1v^{(3)}+g_2v^{(1)}v^{(2)}+g_2v^{(2)}v^{(1)}+g_3v^{(1)}v^{(1)}v^{(1)}\right]+...\ .\end{aligned} \tag{5}$$

Comparison of the components of the same order on both sides of (5) gives

$$i^{(1)}=g_1v^{(1)} \tag{6a}$$

$$i^{(2)}=g_1v^{(2)}+g_2v^{(1)}v^{(1)} \tag{6b}$$

$$i^{(3)}=g_1v^{(3)}+g_2v^{(1)}v^{(2)}+g_2v^{(2)}v^{(1)}+g_3v^{(1)}v^{(1)}v^{(1)} \tag{6c}$$

and so on.

To transform the expressions given by (6a), (6b), and (6c) into the multi-frequency domain, one needs first to introduce the multidimensional time variables in these expressions. And this is carried out by a standard procedure as described, for instance, in [5], [6], [8]. So one obtains, for example, for $i^{(n)} \to i^{(n)}(t_1,..,t_2)$,

$$i^{(n)}(t_1,..,t_n)=\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}h_l^{(n)}(\tau_1,..,\tau_n)\prod_{k=1}^{n}x(t_k-\tau_k)d\tau_k \tag{7}$$

from (3a), where the n-dimensional time means a set of time variables $\{t_1,..,t_n\}$.

Afterwards one uses the multidimensional Fourier transforms [5], [6], [8] in (6a), (6b), and (6c) with the multidimensional time introduced. These Fourier transforms are defined by the following formula:

$$Y^{(n)}(f_1,...,f_n)=\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}y^{(n)}(t_1,...,t_n)\cdot\exp(-j2\pi(f_1t_1+...+f_nt_n))dt_1\cdots dt_n \tag{8}$$

where $Y^{(n)}(f_1,...,f_n)$ means the n-dimensional Fourier transform of a function $y^{(n)}(t_1,...,t_n)$ of the n-dimensional time. Moreover, $f_1,...,f_n$ in (8) are the frequencies from the n-dimensional frequency space.

As a result, one gets

$$I^{(1)}(f_1)=g_1V^{(1)}(f_1) \tag{9a}$$

$$I^{(2)}(f_1,f_2)=g_1V^{(2)}(f_1,f_2)+g_2V^{(1)}(f_1)V^{(1)}(f_2) \tag{9b}$$

$$\begin{aligned}I^{(3)}(f_1,f_2,f_3)&=g_1V^{(3)}(f_1,f_2,f_3)+g_2V^{(1)}(f_1)V^{(2)}(f_2,f_3)+\\&+g_2V^{(2)}(f_1,f_2)V^{(1)}(f_3)+g_3V^{(1)}(f_1)V^{(1)}(f_2)V^{(1)}(f_3)\end{aligned} \tag{9c}$$

and so on. In (9a), (9b), and (9c), f_1, f_2, and f_3 are the frequency variables in the multidimensional frequency space. More precisely, f_1 belongs to the one-dimensional frequency space, f_1 and f_2 belong to the two-dimensional frequency space, f_1, f_2, and f_3 belong to the three-dimensional frequency space, and so on.

Note that (9a), (9b), and (9c) can be further rewritten as

$$I^{(1)}(f_1)=g_1V^{(1)}(f_1)+S_l^{(1)}(f_1) \tag{10a}$$

$$I^{(2)}(f_1,f_2)=g_1V^{(2)}(f_1,f_2)+S_l^{(2)}(f_1,f_2) \tag{10b}$$

$$I^{(3)}(f_1,f_2,f_3)=g_1V^{(3)}(f_1,f_2,f_3)+S_l^{(3)}(f_1,f_2,f_3) \tag{10c}$$

with

$$S_I^{(1)}(f_1) = 0 \quad (11a)$$

$$S_I^{(2)}(f_1, f_2) = g_2 V^{(1)}(f_1) V^{(1)}(f_2) \quad (11b)$$

$$S_I^{(3)}(f_1, f_2, f_3) = g_2 V^{(1)}(f_1) V^{(2)}(f_2, f_3) + g_2 \cdot \\ \cdot V^{(2)}(f_1, f_2) V^{(1)}(f_3) + g_3 V^{(1)}(f_1) V^{(1)}(f_2) V^{(1)}(f_3) \quad (11c)$$

where $S_I^{(1)}(f_1)$, $S_I^{(2)}(f_1, f_2)$, $S_I^{(3)}(f_1, f_2, f_3)$, and so on, play a role of the independent current sources (independent of the voltages $V^{(1)}$, $V^{(2)}$, and $V^{(3)}$, respectively, and so on), connected in parallel with the conductor g_1. They are used in the analyses of the corresponding orders $n = 1, 2, 3, \ldots$.

From (10a-10c), it follows that the current transforms $I^{(n)}$ and the voltage transforms $V^{(n)}$ depend upon exactly n frequencies $f_1, \ldots, f_n$ of the n-dimensional frequency spaces, $n = 1, 2, 3, \ldots$. Moreover, (10a), (10b), and (10c) are similar to each other in the sense that the term of the form $g_1 V^{(n)}$ is a component of the currents $I^{(n)}$ for all n. However, there occur some additional components for $n > 1$ as in (10b) and (10c) (for $n = 1$, this additional component is per definition equal to zero). The additional components in (10b) and (10c) can be considered as the independent current sources connected parallel to the linear conductor g_1. Furthermore, note from (11b) and (11c) that the values of these sources for a given order $n = 2$ or $n = 3$ are calculated using the values of the voltages $V^{(1)}$, and $V^{(1)}$ and $V^{(2)}$, respectively, calculated before, in the analyses of lesser orders $n = 1$, and $n = 1$ and 2, accordingly.

It follows from the above observations that the model of the nonlinear conductor [2] for nonlinear analysis in the frequency domain with the use of the Volterra series consists of a linear conductor having the conductance $G = g_1$ and the independent current sources connected to it in parallel, of which values are calculated for the each order of the analysis separately, according to the relations given by (11a), (11b), (11c), and so on.

From the above derivations (5-11) of the model of the nonlinear conductor, it follows that the form of this model (see the relations given by (10) and (11)) is independent of that whether the nonlinear conductor is embedded in a nonlinear circuit having a single input (then (3a) and (3b) apply) or in a nonlinear circuit possessing multiple inputs (then (4a) and (4b) apply).

Using the scheme sketched above, the models of other basic nonlinear circuit elements like a nonlinear resistor, a nonlinear capacitor, a nonlinear inductor, a nonlinear voltage-controlled voltage source, a nonlinear current-controlled voltage source, a nonlinear voltage-controlled current source, and a nonlinear current-controlled current source can be derived similarly. These models have been derived and presented in [2], [4]. To complete, the models of the linear elements and of the independent current and voltage sources, for the use in nonlinear analysis, have been derived in [2, 4], too.

3. MODIFIED NODAL FORMULATION FOR NONLINEAR CIRCUITS WITH SINGLE AND MULTIPLE INPUTS

In the description of linear circuits in the frequency domain, the basic circuit elements can be split into two groups: one formed by elements possessing an admittance description, and the second consisting of those that do not. For details, see [1].

The elements which do not possess the admittance matrix description can be incorporated into the equations describing a circuit in such a way that they extend the classic nodal formulation of circuit equations (based on the use of the admittance matrix) by new matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{Z}$. Then one arrives at the so-called modified nodal formulation (the modified admittance matrix description), which has the following form [2]:

$$\begin{bmatrix} \boldsymbol{Y} & \boldsymbol{\alpha} \\ \boldsymbol{\beta} & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{V} \\ \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_I \\ \boldsymbol{X}_V \end{bmatrix} \tag{12}$$

where $\boldsymbol{Y}$ is an admittance matrix determined by the circuit elements possessing the admittance description, but $\boldsymbol{Z}$ is a matrix of impedance character determined by the non-admittance-type elements. Furthermore, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are dimensionless matrices, $\begin{bmatrix} V & I \end{bmatrix}^T$ is a vector of unknown nodal voltages and branch currents, but $\begin{bmatrix} X_I & X_V \end{bmatrix}^T$ is a vector of the external independent current (X_I) and voltage (X_V) sources applied to the circuit inputs. The letter T in the above vectors stands for the transposing operation. Moreover, the elements of the matrices and vectors occurring in (12) can depend upon the frequency f as being the Fourier transforms.

The modified admittance matrix occurring in (12) is formed with the use of a mnemonic technique of stamps developed for basic linear circuit elements [1] (for example, for a linear conductor the associated equation is of the form $I - G \cdot V = 0$, where I and V are the Fourier transforms of the current and voltage, respectively, at the conductor, but G means its conductance).

Taking into account the form of the models of the basic nonlinear elements, basic linear elements, and external independent sources derived for the MNF and discussed in the previous section, we see that to arrive at the MNF for nonlinear circuits one needs to modify the description given by (12) by adding to it an additional vector consisting of the values of the independent current and voltage sources arising in the models of the nonlinear elements (for instance, see (11b) and (11c)). This leads to the following description:

$$\begin{bmatrix} \boldsymbol{S}_I^{(n)} \\ \boldsymbol{S}_V^{(n)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Y}^{(n)} & \boldsymbol{\alpha}^{(n)} \\ \boldsymbol{\beta}^{(n)} & \boldsymbol{Z}^{(n)} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}^{(n)} \\ \boldsymbol{I}^{(n)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_I^{(n)} \\ \boldsymbol{X}_V^{(n)} \end{bmatrix} \tag{13}$$

where the superscript $n = 1$ (linear), 2, 3, ...means the order of the analysis, and $\boldsymbol{S}_I^{(n)}$ and $\boldsymbol{S}_V^{(n)}$ are the corresponding vectors of the internal independent current and voltage sources introduced to the description by the circuit nonlinear elements. Moreover, note that the vector $\left[X_I^{(n)}\ X_V^{(n)}\right]^T$ for the external independent sources, occurring on the right-hand side of (13), has the form

$$\begin{bmatrix} X_I^{(1)} \\ X_V^{(1)} \end{bmatrix} = \begin{bmatrix} X_I \\ X_V \end{bmatrix} \text{ and } \begin{bmatrix} X_I^{(n)} \\ X_V^{(n)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ for } n > 1 . \tag{14}$$

For more details, see [2].

4. DIRAC IMPULSE EXCITATIONS IN THE MODIFIED NODAL FORMULATION FOR NONLINEAR CIRCUITS

First, note that the MNF for nonlinear circuits given by (13) is, similarly as the models presented in section 2, independent of that whether a nonlinear circuit possesses one or multiple inputs. For both the cases, the vector of unknowns, $\left[V^{(n)}\ I^{(n)}\right]^T$, consists of the n-dimensional Fourier transforms of the components of the circuit nodal voltages and branch currents of the corresponding orders. (At this point, we remind the fact that the n-dimensional Fourier transform depends upon the frequencies $f_1,...,f_n$ from the n-dimensional frequency space.) Some of the elements of the vector $\left[V^{(n)}\ I^{(n)}\right]^T$ are the circuit output quantities (at a circuit single output or multiple outputs). Furthermore, the elements of the vectors $\left[V^{(m)}\ I^{(m)}\right]^T$, $m = 1,2,3,...,n-1$ with $n = 2,3,...$ are also used to calculate the values of the elements of the vectors $\left[S_I^{(n)}\ S_V^{(n)}\right]^T$ for a given $n \geq 2$, which describe the internal independent sources used in the analyses of higher orders $n \geq 2$ (for example, see (11b) and (11c)).

Consider now the case of a nonlinear circuit with a single input and excited by a Dirac impulse signal $\delta(t)$. Further, assume for the purpose of illustration of this case, that the output signal is a nodal voltage given by (2b) and (3b). Assuming now that none of $h_V^{(n)}(\tau_1,...,\tau_n)$, $n \geq 2$, occurring in (3b) is a multidimensional Dirac impulse [9], introducing there $x(t) = \delta(t)$, and the multidimensional time (as shown in (7)), we get from (3b)

$$v^{(n)}(t_1,..,t_n) = h_V^{(n)}(t_1,..,t_n) \tag{15}$$

according to the theory presented in [10].

Applying the n-dimensional Fourier transform defined by (8) to (15) gives

$$V^{(n)}(f_1,..,f_n) = H_V^{(n)}(f_1,..,f_n) \tag{16}$$

where $V^{(n)}(f_1,..,f_n)$ and $H_V^{(n)}(f_1,..,f_n)$ are the n-th order component of the nodal voltage at the output (more precisely, its n-dimensional Fourier transform) and the n-th order nonlinear transfer function from the circuit input to the node represented by this voltage.

Equation (16) shows that the nonlinear transfer functions of a nonlinear circuit with a single input can be calculated in a very simple way in the calculations applying the MNF. They are then obtained as the corresponding nodal voltages and/or branch currents by performing the analyses of the corresponding orders with the circuit input signal being a Dirac impulse. Note also that the nodal voltages and branch currents of the corresponding orders, which are used in the calculations of the elements of the vectors $\left[S_I^{(n)} \;\; S_V^{(n)} \right]^T$, are then equal to the corresponding nonlinear transfer functions of the circuit, too.

The computer algorithms for the calculation of the nonlinear transfer functions of nonlinear circuits with single inputs, based on the approach described above, which uses the MNF and the Dirac impulse input excitation, have been implemented and presented in [2], [3], [4]. Note that this approach is much simpler than the approach called the exponential input method [5], [6], [11], [12], which was used so far in the calculations of the nonlinear transfer functions. The exponential input method uses a more complicated input signal, that is the sum of n exponentials $\exp(j2\pi f_i t)$, $i = 1,...,n$, to calculate the nonlinear transfer function of the n-th order (the number n changes with the order of the analysis). Moreover, this method does not lead to such the direct results as, for example, presented by (16).

In the case of nonlinear circuits possessing multiple inputs, such relations like (16) do not hold. To see this, consider now a nonlinear circuit with N inputs and driven by an input signal being the Dirac impulse $\delta(t)$. Further, assume that the output signal of this circuit is a nodal voltage given by (2b) and (4b). Hence, on assuming that none of $h_{Vi_1,..,i_n}^{(n)}(\tau_1,...,\tau_n)$, $n \geq 2$, in (4b) is a multidimensional Dirac impulse [9], substituting there $x_{i_k}(t) = \delta(t)$, $i_k \in \{1,..,N\}$, and introducing the multidimensional time (as shown in (7)), we get from (4b)

$$v^{(n)}(t_1,..,t_n) = \sum_{i_1,..,i_n \in \{1,..,N\}}^{N^n} h_{Vi_1,..,i_n}^{(n)}(t_{i_1},..,t_{i_n}) \tag{17}$$

according to the theory presented in [10].

Applying the n-dimensional Fourier transform defined by (8) to (17) leads to

$$V^{(n)}(f_1,..,f_n) = \sum_{i_1,..,i_n \in \{1,..,N\}}^{N^n} H_{Vi_1,..,i_n}^{(n)}(f_{i_1},..,f_{i_n}) \tag{18}$$

where $V^{(n)}(f_1,..,f_n)$ is the n-th order component of the nodal voltage at the output (more precisely, its n-dimensional Fourier transform). Moreover, the nonlinear transfer functions $H^{(n)}_{V i_1 .. i_n}(f_{i_1},..,f_{i_n})$, $i_1...,i_n \in \{1..,N\}$, in (18) are the n-th order ones associated with the N inputs and the circuit node represented by this voltage. A number of these nonlinear transfer functions to be taken into account in (18) for the n-th order nonlinear analysis is equal to N^n.

Concluding, we see from (18) that $V^{(n)}(f_1,..,f_n)$ in the case of multiple inputs does not represent a single nonlinear transfer function, but it is a sum of the corresponding nonlinear functions. This fact lets us also to say that the problem of calculation of the nonlinear transfer functions in the nonlinear circuits with multiple inputs is generally harder than that in circuits with single inputs. Its discussion is however outside the scope of this paper.

Finally, we pay the reader's attention to that that the assumption that the nonlinear impulse responses $h^{(n)}_V(\tau_1,...,\tau_n)$, $n \geq 2$, are not the multidimensional Dirac impulses, which was used in derivation of (15), is in fact not necessary for the correctness of (15). We show that the relation given by (15) is also valid, when the $h^{(n)}_V(\tau_1,...,\tau_n)$'s are the multidimensional Dirac impulses. And similarly, the assumption that the nonlinear impulse responses $h^{(n)}_{V i_1 .. i_n}(\tau_1,...,\tau_n)$, $n \geq 2$, are not the multidimensional Dirac impulses, which was used in derivation of (17), is not necessary for the correctness of (17).

To be quite correct mathematically in allowing the above nonlinear impulse responses $h^{(n)}_V(\tau_1,...,\tau_n)$ in (15) and $h^{(n)}_{V i_1 .. i_n}(\tau_1,...,\tau_n)$ in (17) to be also the multidimensional Dirac impulses, we need only to reverse the order of operations used in derivations of (15) and (17). So we introduce then the multidimensional time in (3b) and (4b) (as shown in (7)) first, and afterwards perform the substitution $x(t) = \delta(t)$ in (3b) or $x_{i_k}(t) = \delta(t)$, $i_k \in \{1,..,N\}$, in (4b). As a result, we get then (15) and (17) valid also for the nonlinear impulse responses occurring in them being the multidimensional Dirac impulses $\delta^{(n)}(t_1,t_2,...,t_n) = \delta(t_1)\delta(t_2)\cdots\delta(t_n)$. Note that such the multidimensional Dirac impulses are well defined, according the theory presented in [9], [10].

BIBLIOGRAPHY

[1] Vlach J. and Singhal K., 1983. Computer Methods for Circuit Analysis and Design. Van Nostrand-Reinhold New York.

[2] Borys A., 1984. A simplified analysis of nonlinear distortion in analog electronic circuits using the Volterra-Wiener series. Scientia Electrica, vol. 30, no. 3, pp. 78-103.

[3] Borys A., 1987. The response of a nonlinear circuit to an excitation in the form of Fourier series. Scientia Electrica, vol. 33, no. 3, pp. 3-8.

[4] Borys A., 1996. On Analysis of Nonlinear Analog Circuits Using Volterra Series. Wydawnictwo Uczelniane ATR Bydgoszcz (in Polish).

[5] Bussgang J.J., Ehrman L., and Graham J.W., 1974. Analysis of nonlinear systems with multiple inputs. Proceedings of the IEEE, vol. 62, pp. 1088-1119.

[6] Chua L.O. and Ng C.-Y., 1979. Frequency domain analysis of nonlinear systems: general theory. IEE J. Electron. Circuits and Systems, vol. 3, pp. 165-185.
[7] Schetzen M., 1980. The Volterra and Wiener Theories of Nonlinear Systems. John Wiley & Sons New York.
[8] Chua L.O. and Ng C.-Y., 1979. Frequency domain analysis of nonlinear systems: formulation of transfer functions. IEE J. Electron. Circuits and Systems, vol. 3, pp. 257-269.
[9] Borys A., 2006. Products of Dirac impulses in nonlinear analysis with the use of Volterra series. Zeszyty Naukowe ATR w Bydgoszczy, Bydgoszcz, submitted for publication.
[10] Borys A., 2006. On the nonlinear analysis with the Dirac impulse excitations. Conference Signal Processing'2006, Poznań, submitted for presentation and publication.
[11] Kuo Y.L., 1977. Frequency-domain analysis of weakly nonlinear networks. "Canned" Volterra analysis, part 1. IEEE Circuits and Systems Magazine, vol. 11, pp. 2-8.
[12] Bedrosian E., Rice S.O., 1971. The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs. Proceedings of the IEEE, vol. 59, pp. 1688-1707.

ZMODYFIKOWANA MACIERZ ADMITANCYJNA W ANALIZIE NIELINIOWYCH UKŁADÓW ELEKTRONICZNYCH Z WIELOMA WEJŚCIAMI

Streszczenie

W artykule tym przedstawiono sposób, w jaki zmodyfikowana macierz admitancyjna (ZMA) może być wykorzystana w analizie nieliniowych układów elektronicznych z wieloma wejściami. Pokazano, że podstawowa zasada wykorzystywana w tym podejściu jest taka sama jak w analizie układów nieliniowych z pojedynczym wejściem. W wyniku zastosowania tej zasady, nieliniowe modele podstawowych elementów składowych układów, wyprowadzone dla opisu za pomocą ZMA, jak również i postać równań macierzowych opisujących układy z wykorzystaniem ZMA (dla poszczególnych rzędów nieliniowości), są takie same – niezależnie od tego, czy dany układ posiada tylko jedno wejście, czy więcej wejść. Pewne szczególne własności są jednakże różne. I tak, przykładowo, nieliniowe transmitancje układu z pojedynczym wejściem mogą być obliczone przy wykorzystaniu ZMA dla poszczególnych rzędów nieliniowości przy pobudzeniu wejścia sygnałem w postaci impulsu Diraca. W tym przypadku nieliniowe transmitancje są równe odpowiednim napięciom węzłowym i prądom gałęziowym w układzie przy powyższym pobudzeniu. Takie relacje nie zachodzą jednakże w układzie, gdy posiada on więcej niż jedno wejście. W pracy pokazano także, że rozwiązując układy równań napisanych z wykorzystaniem ZMA, otrzymuje się poprawne wyniki nawet w takich granicznych przypadkach, jak wtedy, gdy układ posiada nieliniowe odpowiedzi impulsowe w postaci wielowymiarowych impusów Diraca, a sygnały pobudzające jego wejścia mają postać jednowymiarowych impulsów Diraca.

Słowa kluczowe: analiza nieliniowa, układy nieliniowe z wieloma wejściami, zmodyfikowana macierz admitancyjna, szereg Volterry

Publikacje Wydawnictw Uczelnianych
Akademii Techniczno-Rolniczej w Bydgoszczy
można nabywać bezpośrednio
w Dziale Udostępniania Biblioteki Głównej ATR
85-796 Bydgoszcz, ul. Prof. S. Kaliskiego 7
lub poprzez Internet: www.bg.atr.bydgoszcz.pl

Zamówienia można składać telefonicznie
052-340-8072, 8078
listownie oraz elektronicznie:
wusprzedaz@atr.bydgoszcz.pl

ISSN 0209-0589