Strongly Consistent Estimation of the Sample Distribution of Noisy Continuous-Parameter Fields

Shahar Z. Kovalsky, Guy Cohen, and Joseph M. Francos, *Senior Member, IEEE*

***Abstract*—The general problem of defining and determining the sample distribution in the case of continuous-parameter random fields is addressed. Defining a distribution in the case of deterministic functions is straightforward, based on measures of sublevel sets. However, the fields we consider are the sum of a deterministic component (nonrandom multidimensional function) and an i.i.d. random field; an attempt to extend the same notion to the stochastic case immediately raises some fundamental difficulties. We show that by "uniformly sampling" such random fields the difficulties may be avoided and a sample distribution may be compatibly defined and determined. Not surprisingly, the obtained result resembles the known fact that the probability distribution of the sum of two independent random variables is the convolution of their distributions. Finally, we apply the results to derive a solution to the problem of deformation estimation of one- and multidimensional signals in the presence of measurement noise.**

***Index Terms*—Continuous parameter random fields, law of large numbers, sample distribution, uniformly distributed sequences.**

I. Introduction

Evaluation of the distribution function of a given function is a well known procedure when the functions, whether deterministic or random, are defined on a discrete one- or multidimensional lattice. However, there are applications and problems where the setting of the physical model and of the resulting estimation algorithm involve the evaluation of the sample distribution over some continuous domain. When the domain over which the observations are defined is some subset of $\mathbb{R}^m$ many potential difficulties arise in analyzing the properties of the sample distribution of the random process.

To clarify the notion of sample distribution considered in this paper, let us first consider the case of nonrandom (i.e., deterministic) functions. Given a measurable deterministic function $g:\mathbb{R}^m\rightarrow\mathbb{R}$, it is straightforward to define its distribution in terms of measures of the sublevel sets $\{\mathbf{x}\in\mathbb{R}^m : g(\mathbf{x})\leq t\}$, $t\in\mathbb{R}$. More specifically, let $B_c(\mathbb{R}^m)$ denote the space of bounded, compactly supported, Lebesgue measurable functions from $\mathbb{R}^m$ to $\mathbb{R}$. Let λ denote the Lebesgue measure on the m-dimensional Euclidean space $\mathbb{R}^m$. We define the transformation T on $B_c(\mathbb{R}^m)$ by

$$(Tg)(t)=\frac{\lambda\{\mathbf{x}\in\operatorname{supp}\{g\}:g(\mathbf{x})\leq t\}}{\lambda\{\operatorname{supp}\{g\}\}},\quad g\in B_c(\mathbb{R}^m) \quad (1)$$

where $\operatorname{supp}\{g\}$ denotes the support of the function g.

As shown in the next section, the transformation T plays the role of a *distribution transformation*: it maps a deterministic function $g\in B_c(\mathbb{R}^m)$ to Tg, a single variable distribution function. Tg may be thought of as the "continuous cumulative histogram" of the function g; it describes the "relative cumulative frequency" of the range of the function g, in terms of measures of its sublevel sets.

The interest in rigorously analyzing the properties of the operator T and of the resulting distribution function Tg goes beyond a mere theoretical interest. In fact, the study presented in this paper was motivated by the problem of matching (or finding the correspondence between) two related observations on the same object, that is, the problem of transformation estimation and its applications to *signal registration*, see Section IV.

Next, suppose that h takes the *additive model* form

$$h(\mathbf{x})=g(\mathbf{x})+\eta(\mathbf{x}),\qquad \mathbf{x}\in\operatorname{supp}\{g\} \quad (2)$$

where $g:\mathbb{R}^m\rightarrow\mathbb{R}$ is a *known* deterministic function and $\{\eta(\mathbf{x}):\mathbf{x}\in\mathbb{R}^m\}$ is a real-valued i.i.d. random field with a *known* distribution function F_η.

Random fields of the type (2) commonly represent noisy signals over a continuous domain, where one continuously measures some continuous physical quantity; the additive random component represents the overall measurement noise, usually due to the measurement procedure.

Fields of the type (2) are not identically distributed; moreover, their probability distribution function is location dependent, i.e., they are not, in any sense, stationary. However, one may still expect the sample distribution of h to hold information on both the deterministic and random components. Hence, the question of determining this sample distribution is an interesting problem on its own.

Intuitively, since h is the sum of two independent components, one may expect that by employing T, we can establish a law of large numbers to yield $Th=Tg*f_\eta$, where f_η is the probability density function of η. However, the transformation T *may not* be directly applied to a field of the type (2), due to inherent measurability difficulties, to be soon discussed. That being the case:

Manuscript received February 24, 2008; revised July 26, 2010; accepted August 18, 2010. Date of current version February 18, 2011.

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: shaharko@ee.bgu.ac.il; guycohen@ee.bgu.ac.il; francos@ee.bgu.ac.il).

Communicated by M. Lops, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2103770

The question addressed in this paper is whether the "sample distribution" of a random field of the type (2) may be defined, such that it has analogous properties to those introduced by the transformation T.

Of course the sample distribution of h may be defined in many ways. However, we pursue a definition that preserves the properties of T, elaborately discussed in Section II, and lets us establish a sensible law of large numbers.

However, as explained below, considering the sample distribution or, in general, laws of large numbers in the case of continuous-parameter random fields with mutually independent random variables raises severe measurability difficulties. Such i.i.d.-driven random fields are not measurable in the usual sense, and thus, the notion of sample distribution, as introduced by T, is ill-posed and has to be properly redefined. Indeed, in this case, the conditions of independence and joint measurability are incompatible with each other; in fact, the set of realizations whose corresponding sample-functions (sample-paths) are Lebesgue measurable is a nonmeasurable set [1], [2]; moreover, its inner and outer measures are zero and one, respectively. Furthermore, in [2], Judd showed that, even if the sample-measurability problem is avoided (by a proper completion of the measure), laws of large numbers may not hold; the set of realizations where the laws of large numbers hold is again not measurable. Therefore, the Lebesgue measure offers no basis for a meaningful concept of the mean or the sample-distribution of a sample function.

Let us demonstrate the above measurability problem by giving the following nonformal example (see [3] for exact details). Suppose that $\{\eta(t) : t \in [0,1]\}$ is a collection of independent and identically distributed random variables with a common finite mean μ. One would like to have $\int_0^1 \eta(t)dt = \mu$ almost surely. Assuming this is true, it is then natural to expect that $\int_a^b \eta(t)dt = (b-a)\mu$ will almost surely hold for every $[a,b] \subseteq [0,1]$. This, however, implies that η must essentially be trivial, as $\eta \equiv \mu$ almost everywhere ([3]).

Questions related to a continuum of independent and identically distributed random variables and corresponding laws of large numbers (e.g., sample-distribution) have evidently gained some interest, especially in economic theory, where various mass economic phenomena are modeled and studied, for example [2]–[5]. For example, in [3], a Riemann-like approach is invoked to integrate the sample function; then, laws of large numbers are obtained by using an L_2-norm convergence criterion. In another approach, large economies are modeled by hyperfinite processes which are measurable with respect to Loeb product spaces, and corresponding laws of large numbers are derived (see [4] and the reference therein).

In this paper we present an approach in which the desired continuous structure of the deterministic component g is maintained while avoiding the measurability difficulties attributed to the random component η. In Section II we redefine the sample distribution transformation in terms of "uniform sampling"; the deterministic case, in which this transformation reduces to T, is discussed. In Section III the stochastic case is discussed; the sample distribution of the random field h is determined in terms of the sample distribution of the deterministic component g and of the probability distribution of the random field η. Not surprisingly, the result we obtain resembles the known fact that the probability distribution of the sum of two independent random variables is the convolution of their distributions. Finally, in Section IV we demonstrate an application of the results to derive a solution to a registration problem in the case where the observation is subject to an additive noise.

II. Distribution Transformation of a Deterministic Function

We begin by defining the three basic transformations we shall discuss.

Let $\{\mathbf{u}_i\}_{i=1}^{\infty} \subseteq \mathbb{R}^m$ be a given sequence of points in $\mathbb{R}^m$. For any function $g : \mathbb{R}^m \to \mathbb{R}$ let us define the family of transformations $\{T_n^{\{\mathbf{u}_i\}}\}_{n=1}^{\infty}$ by

$$\left(T_n^{\{\mathbf{u}_i\}} g\right)(t) = \frac{1}{n}\#\left\{i = 1,\ldots,n : g\left(\mathbf{u}_i\right) \le t\right\} \qquad (3)$$

where $\#A$ denotes the cardinality of the set A. Furthermore, whenever the limit $\lim_{n\to\infty}(T_n^{\{\mathbf{u}_i\}} g)(t)$ exists for all $t \in \mathbb{R}$, we define $T^{\{\mathbf{u}_i\}}$ by

$$T^{\{\mathbf{u}_i\}} g = \lim_{n\to\infty} T_n^{\{\mathbf{u}_i\}} g. \qquad (4)$$

Recall that the transformation T on $B_c(\mathbb{R}^m)$ has already been defined as

$$(Tg)(t) = \frac{\lambda\{\mathbf{x} \in \operatorname{supp}\{g\} : g(\mathbf{x}) \le t\}}{\lambda\{\operatorname{supp}\{g\}\}}, \quad g \in B_c(\mathbb{R}^m). \quad (5)$$

Notice that it also admits the following equivalent integral form:

$$(Tg)(t) = \frac{\int_{\operatorname{supp}\{g\}} \left(\chi_{(-\infty,t]} \circ g\right)(\mathbf{x}) d\lambda(\mathbf{x})}{\lambda\{\operatorname{supp}\{g\}\}} \qquad (6)$$

where χ_A denotes the indicator function of the set A and $\circ$ denotes the composition of functions.

The next simple lemma shows, as aforementioned, that the transformation T plays the role of a distribution transformation. It also shows some of its properties with respect to certain right- (RHS) and left-hand side (LHS) compositions.

Lemma 1 ([6]): Let $g \in B_c(\mathbb{R}^m)$ be a bounded, compactly supported, Lebesgue measurable function from $\mathbb{R}^m$ to $\mathbb{R}$. Then,

(i) The function $G(t) = (Tg)(t)$ is a distribution function. Furthermore, the support of the distribution $G(t)$ is bounded, in the following sense:
   a. $G(t) = 0$ for $t < \operatorname{ess\,inf}_{\mathbf{x}} g(\mathbf{x})$.
   b. $G(t) = 1$ for $t > \operatorname{ess\,sup}_{\mathbf{x}} g(\mathbf{x})$.

(ii) T is invariant under RHS affine compositions: $T(g \circ \mathcal{A}) = Tg$ for any nonsingular affine transformation $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^m$.

(iii) $T(W \circ g) = (Tg) \circ W^{-1}$ for any strictly increasing continuous function $W : \mathbb{R} \to \mathbb{R}$ such that $W(0) = 0$.

The above properties play an important role in the analysis of various applied problems, as will be demonstrated in Section IV.

A. Uniformly Distributed Sequences

To proceed, we introduce some basic definitions and results from the theory of uniform distribution of sequences (also known as equidistribution of sequences) [7]. For

$\mathbf{a} = (a_1, \ldots, a_m)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ in $\mathbb{R}^m$, we say that $\mathbf{a} \le \mathbf{b}$ if $a_j \le b_j$ for all $j = 1, \ldots, m$. Define the *m-dimensional rectangle* $[\mathbf{a}, \mathbf{b}]$ as the set $\{\mathbf{x} : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$. Using the notations $\mathbf{0} = (0, \ldots, 0)$ and $\mathbf{1} = (1, \ldots, 1)$, the rectangle $[\mathbf{0}, \mathbf{1}]$ is the m-dimensional unit cube.

Definition 1 ([7]): The sequence $\{\mathbf{u}_i\}_{i=1}^{\infty} \subseteq [\mathbf{0}, \mathbf{1}]$ is uniformly distributed in $[\mathbf{0}, \mathbf{1}]$ with respect to the Lebesgue measure λ (abbreviated λ-u.d.) if

$$\lim_{n\to\infty} \frac{1}{n} \# \{i = 1, \ldots, n : \mathbf{u}_i \in [\mathbf{a}, \mathbf{b}]\} = \lambda\{[\mathbf{a}, \mathbf{b}]\} = \prod_{i=1}^{n} (b_i - a_i)$$

for all $[\mathbf{a}, \mathbf{b}] \subseteq [\mathbf{0}, \mathbf{1}]$.

That is, in simple terms, the proportion of terms falling in any subrectangle is proportional to its volume.

Remark 1: Many constructive examples of λ-u.d. sequences in $[0, 1] \subseteq \mathbb{R}$ exist [7]. In fact, u.d. sequences are natural in the sense that a sequence of realizations of a uniformly distributed random variable is almost surely a λ-u.d. sequence (an immediate result of the strong law of large numbers). A generalization of the construction of u.d. sequences to $[\mathbf{0}, \mathbf{1}] \subseteq \mathbb{R}^m$ is straightforward.

The following characterization of λ-u.d. sequences is given in [7]: a sequence $\{\mathbf{u}_i\}_{i=1}^{\infty}$ is λ-u.d. in $[\mathbf{0}, \mathbf{1}]$ if and only if for every *Riemann integrable* function g on $[\mathbf{0}, \mathbf{1}]$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{u}_i) = \int_{[\mathbf{0},\mathbf{1}]} g(\mathbf{x}) d\lambda(\mathbf{x}).$$

Remark 2: This characterization *cannot* be generalized to Lebesgue measurable functions since, in general, the Lebesgue integral cannot be determined by the values of a function on any countable set of points.

We would like to expand the notion of λ-u.d. sequences to nonrectangular subsets of $\mathbb{R}^m$. In order to do so, let us briefly introduce the Jordan measure through the following characterization. Let $A \subseteq \mathbb{R}^m$ be a bounded set; the following are equivalent [8], [9]:

(i) A is *Jordan measurable*.
(ii) χ_A, the indicator function of A, is Riemann integrable.
(iii) $\lambda\{\partial A\} = 0$, that is, the boundary of A is of Lebesgue measure zero.

Whenever a set is Jordan measurable, its Jordan measure (also called Jordan content) is exactly its Lebesgue measure. It should be noted that the Jordan measure is a weak notion of measure, since it is simply the restriction of the Lebesgue measure to the ring of bounded Lebesgue measurable sets having boundary of measure zero. Nevertheless, it is shown in [9] that the Riemann integral can be defined in terms of Jordan measure in about the same way that the Lebesgue integral is defined in terms of Lebesgue measure. Therefore, since λ-u.d. sequences are characterized in terms of Riemann integrable functions, the natural nonrectangular subsets of $\mathbb{R}^m$ to consider in this context are Jordan measurable sets.

Throughout, whenever we let $U \subseteq \mathbb{R}^m$ be a compact, Jordan measurable subset of $\mathbb{R}^m$, we also assume it is of a positive measure.

Definition 2 ([7]): Let $U \subseteq \mathbb{R}^m$ be a compact, Jordan measurable subset of $\mathbb{R}^m$. A sequence $\{\mathbf{u}_i\}_{i=1}^{\infty} \subseteq U$ is *λ-u.d. in U* if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{u}_i) = \frac{1}{\lambda\{U\}} \int_{U} g(\mathbf{x}) d\lambda(\mathbf{x})$$

for every Riemann integrable function g with $\operatorname{supp}\{g\} \subseteq U$.

Remark 3: By using Definition 2, it is easy to see that the λ-u.d. property of a sequence is preserved under nonsingular affine transformations: let $\mathcal{A}$ be a nonsingular affine transformation of $\mathbb{R}^m$; $\{\mathbf{u}_i\}_{i=1}^{\infty}$ is λ-u.d. in U if and only if $\{\mathcal{A}(\mathbf{u}_i)\}_{i=1}^{\infty}$ is λ-u.d. in $\mathcal{A}(U)$.

To complete the definition of λ-u.d. sequences in nonrectangular subsets of $\mathbb{R}^m$, we must validate that such sequences exist, as the next lemma asserts.

Lemma 2: Let $U \subseteq \mathbb{R}^m$ be a compact, Jordan measurable subset of $\mathbb{R}^m$. There exists a λ-u.d. sequence $\{\mathbf{u}_i\}_{i=1}^{\infty}$ in U.

Proof: Without loss of generality we assume that $U \subseteq [\mathbf{0}, \mathbf{1}]$; otherwise, choose $\mathcal{A}$ to be a nonsingular affine transformation of $\mathbb{R}^m$ such that $\mathcal{A}(U) \subseteq [\mathbf{0}, \mathbf{1}]$ and use Remark 3.

Let $\{\mathbf{u}_i\}_{i=1}^{\infty}$ be a λ-u.d. sequence in $[\mathbf{0}, \mathbf{1}]$. Define the subsequence $\{i_k\}_{k=1}^{\infty}$ recursively: $i_1 = \min\{i \ge 1 : \mathbf{u}_i \in U\}$ and $i_k = \min\{i > i_{k-1} : \mathbf{u}_i \in U\}$, $k \ge 2$. That is, $\{i_k\}_{k=1}^{\infty}$ is the maximal strictly increasing subsequence such that $\mathbf{u}_{i_k} \in U$ for all k. Notice that since U is of positive measure, i_k is finite for every k, and thus, $\{i_k\}_{k=1}^{\infty}$ is well defined. We will prove that the subsequence $\{\mathbf{u}_{i_k}\}_{k=1}^{\infty}$ is λ-u.d. in U.

Let g be a Riemann integrable function with $\operatorname{supp}\{g\} \subseteq U$. Since $\operatorname{supp}\{g\} \subseteq U$, we have $g(\mathbf{u}_i) = 0$ for $i \notin \{i_k\}_{k=1}^{\infty}$, hence

$$\frac{1}{n} \sum_{k=1}^{n} g(\mathbf{u}_{i_k}) = \frac{1}{n} \sum_{i=1}^{i_n} g(\mathbf{u}_i) = \frac{i_n}{n} \cdot \frac{1}{i_n} \sum_{i=1}^{i_n} g(\mathbf{u}_i) \tag{7}$$

for all n. By the construction of $\{i_k\}_{k=1}^{\infty}$, exactly n of the first i_n elements of $\{\mathbf{u}_{i_k}\}_{k=1}^{\infty}$ belong to U. Hence, with χ_U denoting the characteristic function of U, for all n we have

$$n = \sum_{i=1}^{i_n} \chi_U(\mathbf{u}_i).$$

Notice that $n \le i_n$ and thus $n \to \infty$ implies $i_n \to \infty$. Since U is Jordan measurable, the function χ_U is Riemann integrable so that we can use the λ-u.d. property of $\{\mathbf{u}_i\}_{i=1}^{\infty}$ in $[\mathbf{0}, \mathbf{1}]$ to obtain

$$\lim_{n\to\infty} \frac{n}{i_n} = \lim_{n\to\infty} \frac{1}{i_n} \sum_{i=1}^{i_n} \chi_U(\mathbf{u}_i) = \int_{[\mathbf{0},\mathbf{1}]} \chi_U(\mathbf{x}) d\lambda(\mathbf{x}) = \lambda\{U\}.$$

Using the same property of $\{\mathbf{u}_i\}_{i=1}^{\infty}$ again, we obtain

$$\lim_{n\to\infty} \frac{1}{i_n} \sum_{i=1}^{i_n} g(\mathbf{u}_i) = \int_{[\mathbf{0},\mathbf{1}]} g(\mathbf{x}) d\lambda(\mathbf{x}) = \int_{U} g(\mathbf{x}) d\lambda(\mathbf{x}).$$

Substituting into (7) yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}g(\mathbf{u}_{i_k}) = \lim_{n\to\infty}\frac{i_n}{n}\cdot\frac{1}{i_n}\sum_{i=1}^{i_n}g(\mathbf{u}_i) = \frac{1}{\lambda\{U\}}\int_U g(\mathbf{x})d\lambda(\mathbf{x}). \quad (8)$$

Since (8) holds for any Riemann integrable function g with $\mathrm{supp}\{g\}\subseteq U$, the sequence $\{\mathbf{u}_{i_k}\}_{k=1}^{\infty}$ is λ-u.d. in U. ■

B. *On the Transformation* $T^{\{\mathbf{u}_i\}}$

Next, we elaborate on the relationship between the transformation T and the transformation $T^{\{\mathbf{u}_i\}}$, defined in (4). In order to do so, we restrict the discussion to a better behaved class of functions.

Given a function g, define $L_g(t) = \{\mathbf{x}\in \mathrm{supp}\{g\} : g(\mathbf{x}) \le t\}$. Denote

$$\mathcal{R}_c(\mathbb{R}^m) = \{g\in B_c(\mathbb{R}^m) : g \text{ is Riemann integrable and } L_g(t) \text{ is Jordan measurable for } \textit{all } t\}.$$

That is, $\mathcal{R}_c(\mathbb{R}^m)$ is the subset of $B_c(\mathbb{R}^m)$ of Riemann integrable functions that also have Jordan measurable sublevel sets, restricted to its support.

It should be noted that the additional requirement that $L_g(t)$ is Jordan measurable for all t is not very strong. In [9] it is shown that given a Riemann integrable function g, for all except at most a countable number values of t, the subsets $L_g(t)$ are Jordan measurable. That, in turn, implies that if $L_g(t_0)$ is not Jordan measurable for some t_0 then, for arbitrarily small $\epsilon>0$, the set $\{\mathbf{x}\in\mathrm{supp}\{g\} : t_0-\epsilon < g(\mathbf{x}) \le t_0\}$ has a boundary of a positive measure. Hence, Riemann integrable functions that do not comply with the above requirement are, roughly speaking, irregular.

Moreover, from an applied point of view, restricting the discussion to $\mathcal{R}_c(\mathbb{R}^m)$ imposes no significant practical limitations being "rich" enough to describe any sampled physical signal.

Lemma 3: Let $U\subseteq\mathbb{R}^m$ be a compact, Jordan measurable subset of $\mathbb{R}^m$ and $\{\mathbf{u}_i\}_{i=1}^{\infty}$ be a λ-u.d. sequence in U. For all $g\in\mathcal{R}_c(\mathbb{R}^m)$ with $\mathrm{supp}\,\{g\}=U$ we have

$$Tg = T^{\{\mathbf{u}_i\}}g. \quad (9)$$

If, in addition, g assumes only finitely many values, then for all t we have

$$\frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x})=t\}}{\lambda\{U\}} = \lim_{n\to\infty}\frac{1}{n}\#\{i=1,\ldots,n : g(\mathbf{u}_i)=t\}. \quad (10)$$

Proof: Since $g\in\mathcal{R}_c(\mathbb{R}^m)$, the set $L_g(t)$ is Jordan measurable for all t. Equivalently, the function $\chi_{(-\infty,t]}\circ g$ is Riemann integrable on U for all t, as $\chi_{(-\infty,t]}\circ g = \chi_{L_g(t)}$ on U. Therefore, the λ-u.d. property of the sequence $\{\mathbf{u}_i\}_{i=1}^{\infty}$ may be applied to obtain

$$\begin{aligned}(Tg)(t) &= \frac{1}{\lambda\{U\}}\int_U \left(\chi_{(-\infty,t]}\circ g\right)(\mathbf{x})d\lambda(\mathbf{x}) \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\chi_{(-\infty,t]}\circ g\right)(\mathbf{u}_i) \\ &= \lim_{n\to\infty}\frac{1}{n}\#\{i=1,\ldots,n : g(\mathbf{u}_i)\le t\} = \left(T^{\{\mathbf{u}_i\}}g\right)(t).\end{aligned}$$

Hence, the first part of the claim is proved. Denote by $\{v_1 < v_2 < \cdots < v_R\}$ the values g assumes under the finite range assumption. Obviously, (10) holds for $t\notin\{v_1,v_2,\ldots,v_R\}$. Using (9), for $t=v_r$, $r=1,\ldots,R$, we find that

$$\begin{aligned}\frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x})=v_r\}}{\lambda\{U\}} &= (Tg)(v_r) - (Tg)\left(\frac{v_r+v_{r-1}}{2}\right) \\ &= \left(T^{\{\mathbf{u}_i\}}g\right)(v_r) - \left(T^{\{\mathbf{u}_i\}}g\right)\left(\frac{v_r+v_{r-1}}{2}\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\#\{i=1,\ldots,n : g(\mathbf{u}_i)=v_r\}\end{aligned}$$

where v_0 is arbitrarily set to be less than v_1, which completes the proof. ■

Thus, for a proper selection of $\{\mathbf{u}_i\}_{i=1}^{\infty}$, the transformation T can be calculated by means of $\{T_n^{\{\mathbf{u}_i\}}\}_{n=1}^{\infty}$ on the well-behaved class of functions $\mathcal{R}_c(\mathbb{R}^m)$.

III. Distribution Transformation of the Additive Model

So far, we have discussed the properties of a family of distribution transformations when applied to deterministic functions. In this section, we discuss the random case: we begin with the results of applying the transformations $T_n^{\{\mathbf{u}_i\}}$ and $T^{\{\mathbf{u}_i\}}$ to a random field; then, we return to discuss the problem of the additive model stated in the beginning of the paper and derive our main results.

Let $\{\eta(\mathbf{x}) : \mathbf{x}\in\mathbb{R}^m\}$ be a real-valued i.i.d. random field on $(\Omega,\mathcal{F},P)$ with a *known* probability distribution function F_η. Let $\{\mathbf{u}_i\}_{i=1}^{\infty}$ be a given sequence of *distinct* points in $\mathbb{R}^m$. The transformation $T_n^{\{\mathbf{u}_i\}}$ can now be applied to η. Put

$$F^{(n)}(t) = (T_n^{\{\mathbf{u}_i\}}\eta)(t) = \frac{1}{n}\#\{i=1,\ldots,n : \eta(\mathbf{u}_i)\le t\}.$$

$F^{(n)}$ is known as the empirical distribution function of $\{\eta(\mathbf{u}_i)\}_{i=1}^{n}$. For fixed t, $F^{(n)}(t)$ is a random variable (of the implicit variable ω). For a realization of the random field (i.e., fixed ω) the function $F^{(n)}(t)$ is a distribution function as it is an increasing step function jumping by $1/n$ at each point $\eta(\mathbf{u}_i)$.

In this context, the Glivenko-Cantelli theorem [10] can be rephrased to state what follows.

$\lim_{n\to\infty}F^{(n)}(t) = F_\eta(t)$ a.s., uniformly in t, that is, $\lim_{n\to\infty}\left\|F^{(n)}-F_\eta\right\|_\infty = 0$ with probability 1.

Therefore, in terms of the transformations we have previously defined, $T^{\{\mathbf{u}_i\}}\eta = \lim_{n\to\infty}T_n^{\{\mathbf{u}_i\}}\eta = F_\eta$ with probability 1. Hence, for *any* sequence of distinct points $\{\mathbf{u}_i\}_{i=1}^{\infty}\subseteq\mathbb{R}^m$ the transformation $T^{\{\mathbf{u}_i\}}$ is a strongly consistent nonparametric estimator for the probability distribution function of the random field $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$.

Now, suppose that h takes the form

$$h(\mathbf{x}) = g(\mathbf{x}) + \eta(\mathbf{x}), \qquad \mathbf{x} \in \operatorname{supp}\{g\} \tag{11}$$

where $g \in \mathcal{R}_c(\mathbb{R}^m)$ is a deterministic function and $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ is a real-valued i.i.d. random field with distribution function F_η.

Let $U = \operatorname{supp}\{g\}$ and $\{\mathbf{u}_i\}_{i=1}^\infty$ be a λ-u.d. sequence of distinct points in U (such sequence exists, according to Lemma 2).

Proposition 1: If g assumes only finitely many values $\{v_1, \ldots, v_R\}$, then

$$(T^{\{\mathbf{u}_i\}}h)(t) = \sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} F_\eta(t - v_r) \quad \text{a.s.}$$

uniformly in t. Moreover, $T^{\{\mathbf{u}_i\}}h$ is almost surely independent of the choice of $\{\mathbf{u}_i\}$ as the RHS of the equation is.

Proof: By the definition of $T_n^{\{\mathbf{u}_i\}}$

$$(T_n^{\{\mathbf{u}_i\}}h)(t) = \frac{1}{n}\#\{i = 1, \ldots, n : g(\mathbf{u}_i) + \eta(\mathbf{u}_i) \le t\} \tag{12}$$

for all n and all t. Since g assumes finitely many values, the RHS of (12) decomposes into a finite sum

$$\begin{aligned}&\frac{1}{n}\#\{i = 1, \ldots, n : g(\mathbf{u}_i) + \eta(\mathbf{u}_i) \le t\}\\ &= \sum_{r=1}^{R} \frac{1}{n}\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r \text{ and } \eta(\mathbf{u}_i) \le t - v_r\}.\end{aligned} \tag{13}$$

Without loss of generality, we may assume there exists an n_0 such that the sets $\{i = 1, \ldots, n_0 : g(\mathbf{u}_i) = v_r\}$ are nonempty for $r = 1, \ldots, R$; otherwise, the empty terms in (13) may be excluded. Hence, for $n \ge n_0$, each term of the sum on the RHS of (13) may be written as a product of two factors

$$\begin{aligned}\frac{1}{n}\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r \quad \text{and} \quad \eta(\mathbf{u}_i) \le t - v_r\}\\ = D_n^{(r)} \cdot F_n^{(r)}(t - v_r)\end{aligned}$$

where we denote

$$D_n^{(r)} = \frac{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r\}}{n}$$

and

$$F_n^{(r)}(t) = \frac{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r \quad \text{and} \quad \eta(\mathbf{u}_i) \le t\}}{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r\}}.$$

Notice that $\{D_n^{(r)}\}_{n=1}^\infty$ is a deterministic sequence, while $\{F_n^{(r)}(\cdot)\}_{n=1}^\infty$ is a sequence of random processes. With these notations

$$(T_n^{\{\mathbf{u}_i\}}h)(t) = \sum_{r=1}^{R} D_n^{(r)} \cdot F_n^{(r)}(t - v_r).$$

Now, since the conditions of Lemma 3 are satisfied

$$\begin{aligned}\lim_{n\to\infty} D_n^{(r)} &= \lim_{n\to\infty} \frac{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r\}}{n}\\ &= \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}\end{aligned}$$

for $r = 1, \ldots, R$. Denote the limit $D^{(r)} = \lim_{n\to\infty} D_n^{(r)}$. Also, notice that

$$\begin{aligned}F_n^{(r)}(t) &= \frac{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r \quad \text{and} \quad \eta(\mathbf{u}_i) \le t\}}{\#\{i = 1, \ldots, n : g(\mathbf{u}_i) = v_r\}}\\ &= \frac{\#\left\{j = 1, \ldots, k : \eta\left(\mathbf{u}_{i_j}\right) \le t\right\}}{k}\end{aligned}$$

where $\{i_j\}_{j=1}^\infty$ is a strictly increasing subsequence of indices such that $g(\mathbf{u}_{i_j}) = v_r$ for every j. Since the discrete-parameter random process $\{\eta(\mathbf{u}_i) : i = 1, 2, \ldots\}$ and any of its subsequences satisfy the conditions of the Glivenko-Cantelli theorem, it can be invoked to show that

$$\lim_{n\to\infty} F_n^{(r)}(t) = F_\eta(t) \qquad \text{a.s.}$$

uniformly in t, for every $r = 1, \ldots, R$. Finally, since with probability 1 we have

$$\begin{aligned}&\left\|D_n^{(r)} F_n^{(r)} - D^{(r)} F_\eta\right\|_\infty\\ &= \left\|D_n^{(r)} F_n^{(r)} - D^{(r)} F_n^{(r)} + D^{(r)} F_n^{(r)} - D^{(r)} F_\eta\right\|_\infty\\ &\le \left|D_n^{(r)} - D^{(r)}\right| \left\|F_n^{(r)}\right\|_\infty + D^{(r)} \left\|F_n^{(r)} - F_\eta\right\|_\infty\end{aligned}$$

the limit $\lim_{n\to\infty}(D_n^{(r)} \cdot F_n^{(r)}(t - v_r))$ exists almost surely for all r, and we find that

$$\begin{aligned}(T^{\{\mathbf{u}_i\}}h)(t) &= \lim_{n\to\infty} (T_n^{\{\mathbf{u}_i\}}h)(t)\\ &= \sum_{r=1}^{R} \lim_{n\to\infty} \left(D_n^{(r)} \cdot F_n^{(r)}(t - v_r)\right)\\ &= \sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} F_\eta(t - v_r) \quad \text{a.s.}\end{aligned}$$

uniformly in t, which concludes the proof. ■

In the special case, where the random field $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ has an absolutely continuous probability distribution, we have the following result.

Theorem 1: Let f_η be the probability density function of the random field $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$. Then, the limit $\lim_{n\to\infty} T_n^{\{\mathbf{u}_i\}}h$ exists, and

$$T^{\{\mathbf{u}_i\}}h = (T^{\{\mathbf{u}_i\}}g) * f_\eta \qquad \text{a.s.}$$

Furthermore, this equality also holds in $L_p(\Omega)$-norm, $1 \le p < \infty$.

Proof: We split the proof into two steps. First, we prove the assertion for $g \in \mathcal{R}_c(\mathbb{R}^m)$ that only assumes finitely many values. We then extend the result to an arbitrary $g \in \mathcal{R}_c(\mathbb{R}^m)$.

Notice that $F_\eta(t) = \int_{\mathbb{R}} f_\eta(\tau)\chi_{[0,\infty)}(t-\tau)d\tau$ for all $t \in \mathbb{R}$, so that, from Proposition 1

$$(T^{\{\mathbf{u}_i\}}h)(t) = \sum_{r=1}^{R} \frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} \int_{-\infty}^{\infty} f_\eta(\tau)\chi_{[0,\infty)}(t - v_r - \tau)d\tau \tag{14}$$

with probability 1. Since $\chi_{[0,\infty)}(t - v_r - \tau) = \chi_{[v_r,\infty)}(t-\tau)$, we have

$$(T^{\{\mathbf{u}_i\}}h)(t) = \int_{-\infty}^{\infty} f_\eta(\tau)\left(\sum_{r=1}^{R} \frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}\chi_{[v_r,\infty)}\right)(t-\tau)d\tau. \tag{15}$$

Clearly

$$\begin{aligned}\left(\sum_{r=1}^{R}\frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}\chi_{[v_r,\infty)}\right)(t) &= \sum_{\{v_r\le t\}} \frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} \\ &= \frac{\lambda\{\mathbf{x}\in U : g(\mathbf{x}) \le t\}}{\lambda\{U\}} = (Tg)(t). \end{aligned} \tag{16}$$

Substituting (16) into (15), we obtain

$$(T^{\{\mathbf{u}_i\}}h)(t) = \int_{-\infty}^{\infty} f_\eta(\tau)(Tg)(t-\tau)d\tau = ((Tg) * f_\eta)(t)$$

almost surely. Finally, since $\{\mathbf{u}_i\}_{i=1}^{\infty}$ is a λ-u.d. sequence in U, Lemma 3 implies that $Tg = T^{\{\mathbf{u}_i\}}g$, and therefore

$$(T^{\{\mathbf{u}_i\}}h)(t) = \left((T^{\{\mathbf{u}_i\}}g) * f_\eta\right)(t) \qquad \text{a.s.} \tag{17}$$

Thus, the assertion is proved, given that $g \in \mathcal{R}_c(\mathbb{R}^m)$ is also a *simple* function, that is, g only assumes finitely many values.

Next, we extend this result to an arbitrary $g \in \mathcal{R}_c(\mathbb{R}^m)$ by means of approximation from below and from above.

Let $\overline{g}_k = \lfloor kg \rfloor / k$, $k \ge 1$. It is easy to see that $\{\overline{g}_k\}$ is a sequence of *simple* functions in $\mathcal{R}_c(\mathbb{R}^m)$ such that $\overline{g}_k \le g$ and $\overline{g}_k \to g$ pointwise. Importantly, this also implies that

$$\chi_{(-\infty,t]} \circ \overline{g}_k \to \chi_{(-\infty,t]} \circ g \tag{18}$$

pointwise, for all t. This important property is simply due to the *left continuity* of $\chi_{(-\infty,t]}$ and the fact that $\overline{g}_k \le g$.

Similarly, let $\underline{\tilde{g}}_k = \lceil kg \rceil / k$, $k \ge 1$. Then, $\{\underline{\tilde{g}}_k\}$ is a sequence of *simple* functions in $\mathcal{R}_c(\mathbb{R}^m)$ such that $g \le \underline{\tilde{g}}_k$ and $\underline{\tilde{g}}_k \to g$ pointwise. In this case, however, a property similar to (18) is not guaranteed. Namely, fix $t \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^m$, and examine $(\chi_{(-\infty,t]} \circ \underline{\tilde{g}}_k)(\mathbf{x})$ as $k \to \infty$; three possible cases arise: (i) if $g(\mathbf{x}) > t$ then $(\chi_{(-\infty,t]} \circ \underline{\tilde{g}}_k)(\mathbf{x}) = (\chi_{(-\infty,t]} \circ g)(\mathbf{x}) = 0$ for all k; (ii) if $g(\mathbf{x}) < t$ then there exists some $k_0 \in \mathbb{N}$ such that $(\chi_{(-\infty,t]} \circ \underline{\tilde{g}}_k)(\mathbf{x}) = (\chi_{(-\infty,t]} \circ g)(\mathbf{x}) = 1$ for all $k > k_0$; (iii) if $g(\mathbf{x}) = t$ then $(\chi_{(-\infty,t]} \circ \underline{\tilde{g}}_k)(\mathbf{x}) \le (\chi_{(-\infty,t]} \circ g)(\mathbf{x}) = 1$ for all k. Hence, a problem may occur for values of t such that $\{\mathbf{x} : g(\mathbf{x}) = t\}$ has a positive measure. This problem, however, is simple to rectify since $\{\mathbf{x} : g(\mathbf{x}) = t\}$ has zero measure for all except at most a countable number values of t. Let $\{t_k\}$ be the values of t for which $\{\mathbf{x} : g(\mathbf{x}) = t\}$ has a positive measure, and define

$$\underline{g}_k(\mathbf{x}) = \begin{cases} g(\mathbf{x}), & g(\mathbf{x}) = t_j, j = 1,\ldots,k \\ \underline{\tilde{g}}_k(\mathbf{x}), & \text{otherwise.}\end{cases}$$

Clearly, $\{\underline{g}_k\}$ is a sequence of *simple* functions in $\mathcal{R}_c(\mathbb{R}^m)$ such that $g \le \underline{g}_k$ and $\underline{g}_k \to g$ pointwise. Moreover

$$\chi_{(-\infty,t]} \circ \underline{g}_k \to \chi_{(-\infty,t]} \circ g \qquad \text{a.e.} \tag{19}$$

for all t.

Recall that $h = g + \eta$, and similarly denote $\underline{h}_k = \underline{g}_k + \eta$ and $\overline{h}_k = \overline{g}_k + \eta$. Since $\underline{g}_k$ and $\overline{g}_k$ assume only finitely many values, (17) implies that there exist subsets $\underline{\Omega}_0^{(k)}, \overline{\Omega}_0^{(k)} \subseteq \Omega$, $k = 1, 2, \ldots$, of measure one such that

$$T^{\{\mathbf{u}_i\}}\underline{h}_k = (T^{\{\mathbf{u}_i\}}\underline{g}_k) * f_\eta, \qquad \text{on } \underline{\Omega}_0^{(k)} \tag{20}$$

and

$$T^{\{\mathbf{u}_i\}}\overline{h}_k = (T^{\{\mathbf{u}_i\}}\overline{g}_k) * f_\eta, \qquad \text{on } \overline{\Omega}_0^{(k)}. \tag{21}$$

Denote

$$\Omega_0 = \bigcap_k \left(\underline{\Omega}_0^{(k)} \cap \overline{\Omega}_0^{(k)}\right)$$

so that, Ω_0 is again of measure one, being a countable intersection of sets of measure one. Now, fix $\omega \in \Omega_0$ (i.e., fix a realization of η). Since $\overline{g}_k \le g \le \underline{g}_k$, we also have $\overline{h}_k \le h \le \underline{h}_k$. This inequality implies that $\{\mathbf{x} : \underline{h}_k(\mathbf{x}) \le t\} \subseteq \{\mathbf{x} : h(\mathbf{x}) \le t\} \subseteq \{\mathbf{x} : \overline{h}_k(\mathbf{x}) \le t\}$, and therefore

$$T_n^{\{\mathbf{u}_i\}}\underline{h}_k \le T_n^{\{\mathbf{u}_i\}}h \le T_n^{\{\mathbf{u}_i\}}\overline{h}_k.$$

Taking $n \to \infty$ gives, for every k

$$\begin{aligned} T^{\{\mathbf{u}_i\}}\underline{h}_k &= \lim_{n\to\infty} T_n^{\{\mathbf{u}_i\}}\underline{h}_k \le \liminf_{n\to\infty} T_n^{\{\mathbf{u}_i\}}h \\ &\le \limsup_{n\to\infty} T_n^{\{\mathbf{u}_i\}}h \le \lim_{n\to\infty} T_n^{\{\mathbf{u}_i\}}\overline{h}_k = T^{\{\mathbf{u}_i\}}\overline{h}_k. \end{aligned} \tag{22}$$

We shall show that $T^{\{\mathbf{u}_i\}}\underline{h}_k$ and $T^{\{\mathbf{u}_i\}}\overline{h}_k$ tend to the same limit as $k \to \infty$.

Notice that, by using Lemma 3 and the integral form of T, we have

$$(T^{\{\mathbf{u}_i\}}g)(t) = (Tg)(t) = \frac{1}{\lambda\{U\}}\int_U (\chi_{(-\infty,t]} \circ g)(\mathbf{x})d\lambda(\mathbf{x})$$

and

$$(T^{\{\mathbf{u}_i\}}\underline{g}_k)(t) = (T\underline{g}_k)(t) = \frac{1}{\lambda\{U\}}\int_U \chi_{(-\infty,t]} \circ \underline{g}_k(\mathbf{x})d\lambda(\mathbf{x}).$$

Recall that $\{\underline{g}_k\}$ satisfies $(\chi_{(-\infty,t]} \circ \underline{g}_k) \to (\chi_{(-\infty,t]} \circ g)$ a.e. for all t. Hence, Lebesgue's bounded convergence theorem may be employed to show that

$$\begin{aligned}\lim_{k\to\infty} (T^{\{\mathbf{u}_i\}}\underline{g}_k)(t) &= \lim_{k\to\infty} \frac{1}{\lambda\{U\}} \int_U (\chi_{(-\infty,t]} \circ \underline{g}_k)(\mathbf{x}) d\lambda(\mathbf{x}) \\ &= \frac{1}{\lambda\{U\}} \int_U (\chi_{(-\infty,t]} \circ g)(\mathbf{x}) d\lambda(\mathbf{x}) = (T^{\{\mathbf{u}_i\}}g)(t)\end{aligned}$$

for all t. That is, we have

$$\lim_{k\to\infty} T^{\{\mathbf{u}_i\}}\underline{g}_k = T^{\{\mathbf{u}_i\}}g.$$

Since $(T^{\{\mathbf{u}_i\}}\underline{g}_k)(t-\tau)\cdot f_\eta(\tau) \leq f_\eta(\tau)$ and f_η is integrable, the dominated convergence theorem may used to show that

$$\begin{aligned}\lim_{k\to\infty} &\left((T^{\{\mathbf{u}_i\}}\underline{g}_k) * f_\eta\right)(t) \\ &= \lim_{k\to\infty} \int_{-\infty}^{\infty} (T^{\{\mathbf{u}_i\}}\underline{g}_k)(t-\tau)\cdot f_\eta(\tau) d\tau \\ &= \int_{-\infty}^{\infty} (T^{\{\mathbf{u}_i\}}g)(t-\tau)\cdot f_\eta(\tau) d\tau = \left((T^{\{\mathbf{u}_i\}}g) * f_\eta\right)(t)\end{aligned}$$

for all t.

Last, we evaluate (20) as $k \to \infty$ to conclude that

$$\lim_{k\to\infty} T^{\{\mathbf{u}_i\}}\underline{h}_k = \lim_{k\to\infty} (T^{\{\mathbf{u}_i\}}\underline{g}_k) * f_\eta = (T^{\{\mathbf{u}_i\}}g) * f_\eta. \quad (23)$$

Similar derivations show that

$$\lim_{k\to\infty} T^{\{\mathbf{u}_i\}}\overline{h}_k = \lim_{k\to\infty} (T^{\{\mathbf{u}_i\}}\overline{g}_k) * f_\eta = (T^{\{\mathbf{u}_i\}}g) * f_\eta. \quad (24)$$

Thus, by taking the limit $k \to \infty$ in (22), we can conclude that the limit $T^{\{\mathbf{u}_i\}}h = \lim_{n\to\infty} T_n^{\{\mathbf{u}_i\}}h$ exists and

$$(T^{\{\mathbf{u}_i\}}h)(t) = \left((T^{\{\mathbf{u}_i\}}g) * f_\eta\right)(t) \qquad \text{a.s.}$$

for all t. Moreover, notice that since both $T^{\{\mathbf{u}_i\}}h$ and $(T^{\{\mathbf{u}_i\}}g) * f_\eta$ are distribution functions bounded by 1, we have that for all t,

$$\left|(T^{\{\mathbf{u}_i\}}h)(t) - \left((T^{\{\mathbf{u}_i\}}g) * f_\eta\right)(t)\right| \leq 2.$$

Therefore, by using Lebesgue's bounded convergence theorem, we may also conclude that

$$(T^{\{\mathbf{u}_i\}}h)(t) = \left((T^{\{\mathbf{u}_i\}}g) * f_\eta\right)(t)$$

in $L_p(\Omega)$-norm, $1 \leq p < \infty$, for all t, which completes the proof. ■

IV. A Signal Registration Application

Consider the problem of matching (or finding the correspondence between) two related observations on the same object. Throughout, objects are single physical entities represented by functions; for example, a pulse (in radar), an isolated word (in speech analysis), an isolated image (in computer vision), etc. Thus the same fundamental problem is common to various applications. We elaborate here on a special case of the general problem, where the domain is transformed by an affine transformation of $\mathbb{R}^m$; this case is basic and provides a "first-order" approximation to more complex cases. In this case, a more practical formulation of the (affine) domain registration problem is the following (see Fig. 1 for an illustration).

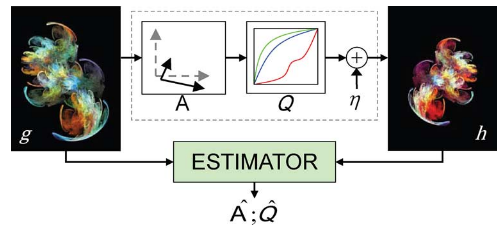

Fig. 1. Illustration of the problem description (25), where different nonlinear mappings are associated with each of the color channels of the image.

Let $Q : \mathbb{R} \to \mathbb{R}$ be an *unknown* strictly increasing continuous function that vanishes at 0; let $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^m$ be an *unknown* nonsingular affine transformation of $\mathbb{R}^m$; and let $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ be a real-valued i.i.d. random field with a *known* probability distribution function F_η. Given a *known* function $g : \mathbb{R}^m \to \mathbb{R}$, representing a signal, and a single measurement (observation) h of the form

$$h(\mathbf{x}) = (Q \circ g \circ \mathcal{A})(\mathbf{x}) + \eta(\mathbf{x}), \quad \mathbf{x} \in \operatorname{supp}\{g \circ \mathcal{A}\} \quad (25)$$

find an estimate for Q and $\mathcal{A}$.

In this formulation, the function Q represents the overall global amplitude nonlinearities in the measuring process (typically due to the nonlinear characteristics of the source, the sensing device itself, amplifiers, etc.); the random component η represents the overall measurement noise, modeled as a random field with mutually independent and identically distributed random variables.

For example, in a simplistic radar model (25) becomes $h(x) = q(g(ax+b)) + \eta(x)$, where g is the transmitted pulse signal, a and b are related to the target velocity and range (due to the Doppler effect and the propagation time), q represents the nonlinearity of the receiver, and η is the measurement noise. Alternatively, in image formation terminology, the model (25) describes the case where the global variability associated with the observation is both *geometric* and *radiometric*. Observations on an object are assumed to simultaneously undergo an affine transformation of coordinates and a nonlinear mapping of the intensities (e.g., due to the recording device). Hence, (25) is the complicated problem of jointly estimating the, seemingly strongly coupled, left- and right-hand compositions Q and $\mathcal{A}$ [6], [11].

To demonstrate the usability of the distribution transformation, T, let us first consider the noiseless case, that is, where (25) holds with $\eta \equiv 0$. Let us also assume that $g \in B_c(\mathbb{R}^m)$.

In this case, the transformation T may be applied to (25). Using Lemma 1 we immediately find that

$$Th = T(Q \circ g \circ \mathcal{A}) = T(Q \circ g) = (Tg) \circ Q^{-1}. \quad (26)$$

Hence, T has converted the joint problem (25), in the unknowns Q and $\mathcal{A}$, to a "new" problem in a single unknown, Q^{-1}.

In order to obtain a parallel result with respect to $\mathcal{A}$, let us define an auxiliary operator R on $B_c(\mathbb{R}^m)$ by

$$Rh = (Th) \circ h - (Th)(0).$$

By applying R to (25), using (26) and since $Q(0) = 0$, we have

$$\begin{aligned}(Rh) &= (Th) \circ h - (Th)(0)\\ &= ((Tg) \circ Q^{-1}) \circ (Q \circ g \circ \mathcal{A}) - (Tg)(Q^{-1}(0))\\ &= (Tg) \circ g \circ \mathcal{A} - (Tg)(0)\\ &= ((Tg) \circ g - (Tg)(0)) \circ \mathcal{A} = (Rg) \circ \mathcal{A} \end{aligned} \quad (27)$$

where the before last equality holds since $(Tg)(0)$ is constant over all of $\mathbb{R}^m$. Hence, R (which has been defined in terms of T) has converted the joint problem (25), in the unknowns Q and $\mathcal{A}$, to a "new" problem in a single unknown, $\mathcal{A}$. Moreover, one may solve for the unknowns Q^{-1} and $\mathcal{A}$ by solving linear systems of equations [12], [13].

As mentioned in the introduction, Th is not properly defined in the case where η does not vanish in (25). We were therefore interested in determining whether the sample distribution of h may be defined, such that it has analogous properties to those introduced by the transformation T.

This question is answered by Theorem 1; under the assumptions that $g \in \mathcal{R}_c(\mathbb{R}^m)$ and that $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ admits a probability density function f_η, we may conclude the following.

Corollary 1: Let $\{\mathbf{u}_i\}_{i=1}^{\infty}$ and $\{\tilde{\mathbf{u}}_i\}_{i=1}^{\infty}$ be λ-u.d. sequences of distinct points in $\mathrm{supp}\{g\}$ and $\mathrm{supp}\{g \circ \mathcal{A}\}$, respectively, then

$$T^{\{\tilde{\mathbf{u}}_i\}} h = \left((T^{\{\mathbf{u}_i\}} g) \circ Q^{-1} \right) * f_\eta \qquad \text{a.s.} \quad (28)$$

Notice that (28) is the stochastic-case analog to (26), and indeed reduces to it as f_η approaches the Dirac delta. Hence, in order to estimate the left-hand composition Q, the original stochastic registration problem can be replaced, with probability one, with a "new" deterministic problem. This deterministic problem has the form of a "classic" deconvolution problem. Solution of the deconvolution problem reduces (28) to the form (26) derived for the noise-free case. Having estimated Q, (25) may be reformulated and solved as a registration problem of strictly the domain (i.e., geometry). As indicated above, this problem has an explicit solution.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Berend and Prof. H. Kaspi for helpful discussions, guidance, and comments. The authors would also like to thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] J. L. Doob, "Stochastic processes depending on a continuous parameter," *Trans. Amer. Math. Soc.*, vol. 42, no. 1, pp. 107–140, 1937.

[2] K. L. Judd, "The law of large numbers with a continuum of IID random variables," *J. Econom. Theory*, vol. 35, no. 1, pp. 19–25, 1985.

[3] H. Uhlig, "A law of large numbers for large economies," *Econom. Theory*, vol. 8, no. 1, pp. 41–50, 1996.

[4] Y. Sun, "A theory of hyperfinite processes: The complete removal of individual uncertainty via exact LLN," *J. Math. Econom.*, vol. 29, no. 4, pp. 419–503, 1998.

[5] P. J. Hammond and Y. Sun, "Joint measurability and the one-way Fubini property for a continuum of independent random variables," *Proc. Amer. Math. Soc.*, vol. 134, no. 3, pp. 737–747, 2006.

[6] S. Z. Kovalsky, G. Cohen, R. Hagege, and J. M. Francos, "Decoupled linear estimation of affine geometric deformations and nonlinear intensity transformations of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 940–946, 2010.

[7] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. New York: Wiley-Intersci., 1974.

[8] M. Spivak, *Calculus on Manifolds. A Modern Approach to Classical Theorems of Advanced Calculus*. New York-Amsterdam: W. A. Benjamin, 1965.

[9] O. Frink Jr., "Jordan measure and Riemann integration," *Ann. Math. (2)*, vol. 34, no. 3, pp. 518–526, 1933.

[10] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995, Wiley Series in Probabil. Math. Statist..

[11] S. Z. Kovalsky, R. Hagege, and J. M. Francos, "Registration of joint radiometric and geometric image deformations," in *Proc. 10th IASTED Intr. Conf. Signal Image Process.*, 2007.

[12] R. Hagege and J. M. Francos, "Parametric estimation of affine transformations: An exact linear solution," *J. Math. Imag. Vision*, 2010.

[13] J. M. Francos, R. Hagege, and B. Friedlander, "Estimation of multidimensional homeomorphisms for object recognition in noisy environments," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1615–1619.

Shahar Z. Kovalsky was born on February 21, 1980, in Israel. He received the B.Sc. degree in electrical engineering and the B.Sc. degree in mathematics in 2007, both *summa cum laude*, and the M.Sc. degree in 2008, in electrical engineering (*summa cum laude*), all from the Ben-Gurion University, Beer-Sheva, Israel.

Dr. Kovalsky was awarded the Wolf foundation award for his M.Sc. studies.

Guy Cohen was born on April 12, 1971, in Beer Sheva, Israel. He received the Ph.D. degree in computer engineering in 2003 from the Ben-Gurion University, Beer-Sheva.

From 2003 to 2004, he was a Postdoctoral Research Fellow with the Mathematics Department, Ben-Gurion University. From 2004 to 2005, he was a Postdoctoral Research Fellow with the Erwing Schrödinger Institute, Vienna. From 2005 to 2006, he was a Postdoctoral Research Fellow with the Einstein Institute, HUJI. In 2006 he joined the Department of Electrical and Computer Engineering, Ben-Gurion University, where he is now a senior lecturer. His current research interests are ergodic theory, stochastic processes, and harmonic analysis.

Joseph M. Francos (S'89–M'90–SM'97) received the B.Sc. degree in computer engineering in 1982, and the D.Sc. degree in electrical engineering in 1991, both from the Technion—Israel Institute of Technology, Haifa.

From 1982 to 1987, he was with the Signal Corps Research Laboratories, Israeli Defense Forces. From 1991 to 1992, he was with the Department of Electrical Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, as a Visiting Assistant Professor. During 1993, he was with Signal Processing Technology, Palo Alto, CA. In 1993, he joined the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel, where he is now a Professor. He heads the Mathematical Imaging Group and the Signal Processing track. He also held visiting positions with the Massachusetts Institute of Technology Media Laboratory, Cambridge; the Electrical and Computer Engineering Department, University of California, Davis; the Electrical Engineering and Computer Science Department, University of Illinois, Chicago; and at the Electrical Engineering Department, University of California, Santa Cruz. His current research interests are in image registration, estimation of object deformations form images, parametric modeling and estimation of 2-D random fields, random fields theory, and texture analysis and synthesis.

Dr. Francos served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2001.