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

Shahar Z. Kovalsky, Guy Cohen and Joseph M. Francos

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 sub-level sets. However, the fields we consider are the sum of a deterministic component (non-random multi-dimensional 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 multi-dimensional signals in the presence of measurement noise.

Index Terms

Continuous parameter random fields, sample distribution, law of large numbers, 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 multi-dimensional

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel, Tel: +972 8 6461842, FAX: +972 8 6472949, email: shaharko,guycohen,francos@ee.bgu.ac.il

lattice. However, there are applications and problems (see an example below) 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 the continuous domain, $\mathbb{R}^m$, 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 non-random (*i.e.*, deterministic) functions. Given a measurable deterministic function $g : \mathbb{R}^m \to \mathbb{R}$, it is straightforward to define its distribution in terms of measures of the sub-level 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

$$[Tg](t) = \frac{\lambda\{\mathbf{x} \in \text{supp}\{g\} : g(\mathbf{x}) \leq t\}}{\lambda\{\text{supp}\{g\}\}}, \qquad g \in B_c(\mathbb{R}^m), \tag{1}$$

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

A. Signal Registration

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 sub-level 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*. This problem, which amounts to estimating the variation between different occurrences of a function (representing a physical entity), is an elementary problem in signal and image processing. Its explicit, or implicit, solution is an essential part of any registration or recognition algorithm.

In particular, we were interested in the case where observations on an object are assumed to simultaneously undergo an affine transformation of coordinates and a non-linear mapping of the amplitudes. More specifically, let $g : \mathbb{R}^m \to \mathbb{R}$ represent a multi-dimensional signal (*e.g.*, $m = 2$ in the case where the observed signals are images). Consider an observation f on g of the form

$$f(\mathbf{x}) = Q(g(\mathcal{A}(\mathbf{x}))) = [Q \circ g \circ \mathcal{A}](\mathbf{x}),$$

where $Q : \mathbb{R} \to \mathbb{R}$ is invertible and $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^m$ is an affine transformation. The right-hand composition of g with $\mathcal{A}$ (composition from within) can be thought of as a spatial/time deformation (*i.e.*, a deformation of the coordinate system), while its left-hand composition with Q (composition from without) can be thought of as a memoryless non-linear input/output system applied to the amplitude of the signal. Hence, in image formation terminology, the physical model corresponding to such model is that of a simultaneous deformation of both geometry and radiometry. From this point of view and in the absence of noise, given two functions (signals) g and h, the problem is then to find, if exists, a pair $(Q, \mathcal{A})$ such that $f(\mathbf{x}) = Q(g(\mathcal{A}(\mathbf{x})))$.

However, the set of possible appearances of an object is immense; for example, in images due to variations in pose, illumination and acquisition system. Thus, the task of determining the correspondence between two observations is extremely complicated. Unfortunately, straight-forward approaches for solving this problem typically lead to a high-dimensional non-convex search problem (see Section V for details). Hence, the direct approach is practically infeasible.

Nevertheless, as we show in Section II, the transformation T has two useful key features: *(i) It is invariant under right-hand affine compositions:* $T\,(g \circ \mathcal{A}) = Tg$ *for any non-singular affine transformation* $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^m$; and *(ii)* $T(Q \circ g) = [Tg] \circ Q^{-1}$ *for any strictly increasing continuous function* $Q : \mathbb{R} \to \mathbb{R}$ *such that* $Q(0) = 0$. The first property implies that Tg, the distribution of g, is invariant to any affine transformation of the domain of g (*i.e.*, "geometry"); whereas the second implies that the distribution of a signal $Q \circ g$, obtained by a non-linear mapping of the amplitudes of g, is nothing but the distribution of g, subject to "warping" of its domain by Q^{-1}.

These two properties are the key in enabling an elegant solution to the problem of jointly estimating the deformations of both the domain and range of a function (signal), described above. As we briefly discuss in Section V and in more detail in [1]–[3], in the absence of noise, the original high-dimensional non-convex search problem that needs to be solved in order to estimate the deformation can be replaced by an equivalent problem, expressed in terms of two linear systems of equations. The solution of these systems establishes an exact solution to the registration problem.

B. Noisy Observations

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

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

where $g : \mathbb{R}^m \to \mathbb{R}$ 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,

> *The question addressed in this paper is whether the "sample distribution" of a random field of the type* (2) *may be defined, such that is 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.

Considering similar problems in the case where the domain of the random field is discrete (*i.e.*, where $\mathbf{x} \in \mathbb{Z}^m$) is elementary: the case of random processes of the type (2) with discrete-parameter ($m = 1$) is straightforward and the case of multi-dimensional random fields with discrete-indices ($m > 1$) reduces to the one-dimensional case upon selection of a total order on $\mathbb{Z}^m$.

Nevertheless, there are scenarios where, due to the inherent structure of a problem, it is natural to discuss continuous-parameter multi-dimensional random fields (*i.e.*, where $\mathbf{x} \in \mathbb{R}^m$). An example of such scenario is the problem of signal registration, discussed above. Inherently, the mapping $\mathcal{A}$ of $\mathbb{R}^m$ into itself is of a continuous nature, as are the physical phenomena it represents in the various problems (*e.g.*, time warping, geometric deformation, etc.). Thus, if we impose a discrete model (*e.g.*, $\mathbf{x} \in \mathbb{Z}^m$), we find that, in general, the natural $\mathcal{A}$ to consider is incompatible (as for "almost all" $\mathbf{x} \in \mathbb{Z}^m$, $\mathcal{A}(\mathbf{x}) \notin \mathbb{Z}^m$).

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. In [4](p. 78, 102), Doob mentioned that random processes with mutually independent random variables are too irregular to discuss in the continuous-parameter case. In a sense, sample paths of such processes are too erratic to be measurable. 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 non-measurable set [5], [6]; moreover, its inner and outer measures are zero and one, respectively. Furthermore, in [6], 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.

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 [6]–[9]. For example, in [7], 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 [8] 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 Sections II-III, we redefine the sample distribution transformation by "uniformly sampling" the random field h. This transformation establishes a strong law of large numbers (in the stochastic case) and reduces to T in the deterministic case. The conditions on which this transformation may be a applied to a random field of the additive model type (2) are discussed. In Section IV we determine the sample distribution of the random field h, 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 V we apply the results to derive a solution to the foregoing registration problem in the case where the observation is subject to an additive noise.

II. The distribution transformation of a deterministic function

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

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

$$\left[T_n^{\{u_i\}} h\right](t) = \frac{1}{n} \# \left\{ i = 1, ..., n : h\left(u_i\right) \leq t \right\}, \tag{3}$$

where $\#A$ denotes the cardinality of the set A. Furthermore, whenever the limit $\lim\limits_{n\to\infty} [T_n^{\{u_i\}} h](t)$ exists for all $t \in \mathbb{R}$, we define $T^{\{u_i\}}$ by

$$T^{\{u_i\}} h = \lim_{n\to\infty} T_n^{\{u_i\}} h. \tag{4}$$

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

$$[Th](t) = \frac{\lambda\{\mathbf{x} \in \text{supp}\{h\} : h(\mathbf{x}) \leq t\}}{\lambda\{\text{supp}\{h\}\}}, \qquad h \in B_c(\mathbb{R}^m). \tag{5}$$

Notice that it also admits the following equivalent integral form

$$[Th](t) = \frac{1}{\lambda\{\text{supp}\{h\}\}} \int_{\text{supp}\{h\}} \left[\chi_{(-\infty,t]} \circ h\right](\mathbf{x}) d\lambda(\mathbf{x}), \tag{6}$$

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

Remark 1: While the transformation $T_n^{\{u_i\}}$ is well defined for any real function, the transformation $T^{\{u_i\}}$ is not necessarily defined for any selection of a sequence $\{u_i\}_{i=1}^{\infty}$.

A. *On the properties of the transformation T*

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

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

(i) The function $H(t) = [Th](t)$ is a distribution function. Furthermore, the support of the distribution $H(t)$ is bounded, in the following sense:

  a. $H(t) = 0$ for $t < \inf_{\mathbf{x}} h(\mathbf{x})$.

  b. $H(t) = 1$ for $t > \sup_{\mathbf{x}} h(\mathbf{x})$.

(ii) T is invariant under right-hand affine compositions: $T\left(h \circ \mathcal{A}\right) = Th$ for any non-singular affine transformation $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^m$.

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

Proof: Part *(i)* is immediate. Notice that $\text{supp}\{h \circ \mathcal{A}\} = \mathcal{A}^{-1}(\text{supp}\{h\})$; this is simply since $\{\mathbf{x} : h(\mathcal{A}(\mathbf{x})) = 0\} = \{\mathcal{A}^{-1}(\mathbf{y}) : h(\mathbf{y}) = 0\}$. Thus, using the properties of the Lebesgue measure, we have

$$\lambda\{\text{supp}\{h \circ \mathcal{A}\}\} = \lambda\left\{\mathcal{A}^{-1}(\text{supp}\{h\})\right\} = |\mathcal{A}^{-1}|\lambda\{\text{supp}\{h\}\},$$

where $|\mathcal{A}^{-1}|$ denotes the determinant of the transformation $\mathcal{A}^{-1}$. Next, set $\mathbf{y} = \mathcal{A}(\mathbf{x})$, thus $\mathbf{x} = \mathcal{A}^{-1}(\mathbf{y})$ and $d\lambda(\mathbf{x}) = |\mathcal{A}^{-1}|d\lambda(\mathbf{y})$. Hence, by (6), the integral form of T, and a change of variables, we have

$$\begin{aligned}
[T(h \circ \mathcal{A})](t) &= \frac{1}{\lambda\{\text{supp}\{h \circ \mathcal{A}\}\}} \int_{\text{supp}\{h \circ \mathcal{A}\}} \chi_{(-\infty,t]}(h(\mathcal{A}(\mathbf{x})))d\lambda(\mathbf{x}) \\
&= \frac{1}{|\mathcal{A}^{-1}|\lambda\{\text{supp}\{h\}\}} \int_{\text{supp}\{h\}} \chi_{(-\infty,t]}(h(\mathbf{y}))|\mathcal{A}^{-1}|d\lambda(\mathbf{y}) = [Th](t)
\end{aligned}$$

for all t, and thus *(ii)* is proved. Lastly, since W is strictly increasing and $W(0) = 0$ we have that $\text{supp}\{W \circ h\} = \text{supp}\{h\}$. Hence, for all t we have

$$\begin{aligned}
[T(W \circ h)](t) &= \frac{\lambda\left\{\mathbf{x} \in \text{supp}\{W \circ h\} : [W \circ h](\mathbf{x}) \le t\right\}}{\lambda\{\text{supp}\{W \circ h\}\}} \\
&= \frac{\lambda\left\{\mathbf{x} \in \text{supp}\{h\} : h(\mathbf{x}) \le W^{-1}(t)\right\}}{\lambda\{\text{supp}\{h\}\}} = \left[[Th] \circ W^{-1}\right](t),
\end{aligned}$$

thus *(iii)* is proved. ■

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

B. 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) [10]. For $\mathbf{a} = (a_1, \ldots, a_m)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ in $\mathbb{R}^m$, we say that $\mathbf{a} \leq \mathbf{b}$ if $a_j \leq b_j$ for all $j = 1, \ldots, m$. Define the *m-dimensional rectangle* $[\mathbf{a}, \mathbf{b}]$ as the set $\{\mathbf{x} : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$. 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 ([10]): The sequence $\{u_i\}_{i=1}^{\infty} \subseteq [\mathbf{0}, \mathbf{1}]$ is *uniformly distributed in* $[\boldsymbol{0}, \boldsymbol{1}]$ *with respect to the Lebesgue measure* λ (abbreviated λ-u.d.) if

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

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

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

Remark 2: Many constructive examples of λ-u.d. sequences in $[0, 1] \subseteq \mathbb{R}$ exist [10]. 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 [10]: a sequence $\{u_i\}_{i=1}^{\infty}$ is λ-u.d. in $[\mathbf{0}, \mathbf{1}]$ if and only if for every *Riemann integrable* function h on $[\mathbf{0}, \mathbf{1}]$

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

Remark 3: 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 non-rectangular 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 [11], [12]:

(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 [12] 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 non-rectangular 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: Let $U \subseteq \mathbb{R}^m$ be a compact, Jordan measurable subset of $\mathbb{R}^m$. A sequence $\{u_i\}_{i=1}^{\infty} \subseteq U$ is *λ-u.d. in U* if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} h(u_i) = \frac{1}{\lambda\{U\}} \int_U h(\mathbf{x}) d\lambda(\mathbf{x})$$

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

Remark 4: By using Definition 2, it is easy to see that the λ-u.d. property of a sequence is preserved under non-singular affine transformations: let $\mathcal{A}$ be a non-singular affine transformation of $\mathbb{R}^m$; $\{u_i\}_{i=1}^{\infty}$ is λ-u.d. in U if and only if $\{\mathcal{A}(u_i)\}_{i=1}^{\infty}$ is λ-u.d. in $\mathcal{A}(U)$.

Indeed, suppose $\{\mathcal{A}(u_i)\}_{i=1}^{\infty}$ is λ-u.d. in $\mathcal{A}(U)$. Let h be a Riemann integrable function with supp$\{h\} \subseteq U$. Obviously, $(h \circ \mathcal{A}^{-1})$ is a Riemann integrable function whose support is contained in $\mathcal{A}(U)$. By applying the λ-u.d. property of $\{\mathcal{A}(u_i)\}_{i=1}^{\infty}$ and substituting $\mathbf{y} = \mathcal{A}^{-1}(\mathbf{x})$, we find that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} h(u_i) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} h(\mathcal{A}^{-1}(\mathcal{A}(u_i))) \\
&= \frac{1}{\lambda\{\mathcal{A}(U)\}} \int_{\mathcal{A}(U)} h(\mathcal{A}^{-1}(\mathbf{x})) d\lambda(\mathbf{x}) \\
&= \frac{1}{|\mathcal{A}|\lambda\{U\}} \int_U h(\mathbf{y}) |\mathcal{A}| d\lambda(\mathbf{y}) = \frac{1}{\lambda\{U\}} \int_U h(\mathbf{y}) d\lambda(\mathbf{y}),
\end{aligned}$$

and therefore $\{u_i\}_{i=1}^{\infty}$ is λ-u.d. in U.

To complete the definition of λ-u.d. sequences in non-rectangular 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 $\{u_i\}_{i=1}^{\infty}$ in U.

Proof: Without loss of generality we assume that $U \subseteq [\mathbf{0}, \mathbf{1}]$; otherwise, choose $\mathcal{A}$ to be a non-singular affine transformation of $\mathbb{R}^m$ such that $\mathcal{A}(U) \subseteq [\mathbf{0}, \mathbf{1}]$ and use Remark 4.

Let $\{u_i\}_{i=1}^{\infty}$ be a λ-u.d. sequence in $[\mathbf{0}, \mathbf{1}]$. Define the subsequence $\{i_k\}_{k=1}^{\infty}$ recursively: $i_1 = \min\{i \geq 1 : u_i \in U\}$ and $i_k = \min\{i > i_{k-1} : u_i \in U\}$, $k \geq 2$. That is, $\{i_k\}_{k=1}^{\infty}$ is the maximal strictly increasing subsequence such that $u_{i_k} \in U$ 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 $\{u_{i_k}\}_{k=1}^{\infty}$ is λ-u.d. in U.

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

$$\frac{1}{n} \sum_{k=1}^{n} h(u_{i_k}) = \frac{1}{n} \sum_{i=1}^{i_n} h(u_i) = \frac{i_n}{n} \cdot \frac{1}{i_n} \sum_{i=1}^{i_n} h(u_i) \tag{7}$$

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

$$n = \sum_{i=1}^{i_n} \chi_U(u_i).$$

Notice that $n \leq i_n$ 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 $\{u_i\}_{i=1}^{\infty}$ in $[\mathbf{0}, \mathbf{1}]$ to obtain

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

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

$$\lim_{n\to\infty} \frac{1}{i_n} \sum_{i=1}^{i_n} h(u_i) = \int_{[\mathbf{0},\mathbf{1}]} h(\mathbf{x}) d\lambda(\mathbf{x}) = \int_U h(\mathbf{x}) d\lambda(\mathbf{x}).$$

Substituting into (7) yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(u_{i_k}) = \lim_{n\to\infty} \frac{i_n}{n} \cdot \frac{1}{i_n} \sum_{i=1}^{i_n} h(u_i) = \frac{1}{\lambda\{U\}} \int_U h(\mathbf{x}) d\lambda(\mathbf{x}). \tag{8}$$

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

C. On the transformation $T^{\{u_i\}}$

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

Given a function h, define $L_h(t) = \{\mathbf{x} \in \text{supp}\{h\} : h(\mathbf{x}) \le t\}$. Denote

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

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 sub-level sets, restricted to its support.

It should be noted that the additional requirement that $L_h(t)$ is Jordan measurable for all t is not very strong. In [12] it is shown that given a Riemann integrable function h, for all except at most a countable number values of t, the subsets $L_h(t)$ are Jordan measurable. That, in turn, implies that if $L_h(t_0)$ is not Jordan measurable for some t_0 then, for arbitrarily small $\epsilon > 0$, the set $\{\mathbf{x} \in \text{supp}\{h\} : t_0 - \epsilon < h(\mathbf{x}) \le t_0\}$ 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 $\{u_i\}_{i=1}^{\infty}$ be a λ-u.d. sequence in U. For all $h \in \mathcal{R}_c(\mathbb{R}^m)$ with supp $\{h\} = U$ we have

$$Th = T^{\{u_i\}}h. \tag{9}$$

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

$$\frac{\lambda\{\mathbf{x} \in U : h(\mathbf{x}) = t\}}{\lambda\{U\}} = \lim_{n\to\infty} \frac{1}{n} \#\{i = 1, \ldots, n : h(u_i) = t\}. \tag{10}$$

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

$$[Th](t) = \frac{1}{\lambda\{U\}} \int_U \left[\chi_{(-\infty,t]} \circ h\right](\mathbf{x}) d\lambda(\mathbf{x}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[\chi_{(-\infty,t]} \circ h\right](u_i)$$

$$= \lim_{n\to\infty} \frac{1}{n} \# \{i = 1, \ldots, n : h(u_i) \le t\} = \left[T^{\{u_i\}} h\right](t).$$

Hence, the first part of the claim is proved. Denote by $\{v_1 < v_2 < \ldots < v_R\}$ the values h 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

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

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

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

We conclude this section with the following simple lemma:

Lemma 4: If $\underline{h}, \overline{h} \in B_c(\mathbb{R}^m)$ such that $\underline{h} \le \overline{h}$ then $T_n^{\{u_i\}} \overline{h} \le T_n^{\{u_i\}} \underline{h}$ for all $n \in \mathbb{N}$.

Proof: Simply follows by noticing that $\{\mathbf{x} : \overline{h}(\mathbf{x}) \le t\} \subseteq \{\mathbf{x} : \underline{h}(\mathbf{x}) \le t\}$. ■

III. Distribution transformations of random fields

So far, we have discussed the properties of a family of distribution transformations when applied to deterministic functions. We shall now discuss the results of applying the transformations $T_n^{\{u_i\}}$ and $T^{\{u_i\}}$ to a random field.

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 $\{u_i\}_{i=1}^{\infty}$ be a given sequence of *distinct* points in $\mathbb{R}^m$. The transformation $T_n^{\{u_i\}}$ can now be applied to η. Put:

$$F^{(n)}(t) = [T_n^{\{u_i\}} \eta](t) = \frac{1}{n} \# \{i = 1, ..., n : \eta(u_i) \le t\}.$$

$F^{(n)}$ is known as the *empirical distribution function* of $\{\eta(u_i)\}_{i=1}^{n}$. 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 $\eta(u_i)$.

In this context, the Glivenko-Cantelli theorem [13] can be rephrased to state that:

$\lim_{n\to\infty} F^{(n)}(t) = F_\eta(t)$ *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, $T^{\{u_i\}}\eta = \lim_{n\to\infty} T_n^{\{u_i\}}\eta = F_\eta$ with probability 1. Hence, for *any* sequence of distinct points $\{u_i\}_{i=1}^{\infty} \subseteq \mathbb{R}^m$ the transformation $T^{\{u_i\}}$ is a *strongly consistent non-parametric estimator* for the probability distribution function of the random field $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$.

We conclude this section with the following lemma:

Lemma 5: Let S be an infinite subset of $\{u_i : i = 1, 2, \ldots\}$. Then

$$\lim_{n\to\infty} \frac{\#\left\{i = 1, ..., n : u_i \in S \quad \text{and} \quad \eta\left(u_i\right) \leq t\right\}}{\#\left\{i = 1, ..., n : u_i \in S\right\}} = F_\eta(t)$$

a.s., uniformly in t.

Proof: Let

$$F_n^S(t) = \frac{\#\left\{i = 1, ..., n : u_i \in S \quad \text{and} \quad \eta\left(u_i\right) \leq t\right\}}{\#\left\{i = 1, ..., n : u_i \in S\right\}}$$

Given the subset S, let $\{i_k\}_{k=1}^{\infty}$ be a strictly increasing subsequence of indices such that $S = \{u_{i_k}\}_{k=1}^{\infty}$. We then have

$$\begin{aligned} F_{i_k}^S(t) &= \frac{\#\left\{i = 1, ..., i_k : u_i \in S \quad \text{and} \quad \eta\left(u_i\right) \leq t\right\}}{\#\left\{i = 1, ..., i_k : u_i \in S\right\}} \\ &= \frac{\#\left\{j = 1, ..., k : \eta\left(u_{i_j}\right) \leq t\right\}}{k}. \end{aligned}$$

Now, by the Glivenko-Cantelli theorem we conclude that $\lim_{k\to\infty} F_{i_k}^S(t) = F_\eta(t)$ a.s., uniformly in t. In particular, $\{F_n^S(t)\}_{n=1}^{\infty}$ has a convergent subsequence. Furthermore, since $\{i_k\}$ is strictly increasing, $u_{i_k+1}, u_{i_k+2}, \ldots, u_{i_{(k+1)}-1} \notin S$. Thus, $F_n^S(t)$ is constant for $n = i_k, i_k+1, \ldots, i_{k+1}-1$, that is, $F_n^S(t) = F_{i_k}^S(t)$ for $i_k \leq n < i_{k+1}$. Since

$$|F_n^S(t) - F_\eta(t)| \leq |F_n^S(t) - F_{i_k}^S(t)| + |F_{i_k}^S(t) - F_\eta(t)|$$

the choice $i_k \leq n < i_{k+1}$ eliminates $|F_n^S(t) - F_{i_k}^S(t)|$, hence we have

$$\lim_{n\to\infty} F_n^S(t) = \lim_{k\to\infty} F_{i_k}^S(t) = F_\eta(t)$$

a.s., uniformly in t. ■

IV. Distribution transformation of the additive model

In this section, we return to discuss the problem stated in the beginning of the paper and derive our main results. Suppose that h takes the form

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

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

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

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

Proof: By the definition of $T_n^{\{u_i\}}$

$$[T_n^{\{u_i\}}h](t) = \frac{1}{n}\#\{i = 1, ..., n : g(u_i) + \eta(u_i) \leq t\} \tag{12}$$

for all n and all t. Since g assumes finitely many values, the right-hand side of (12) decomposes into a finite sum:

$$\frac{1}{n}\#\{i = 1, ..., n : g(u_i) + \eta(u_i) \leq t\}$$
$$= \sum_{r=1}^{R} \frac{1}{n}\#\{i = 1, ..., n : g(u_i) = v_r \quad \text{and} \quad \eta(u_i) \leq t - v_r\}. \tag{13}$$

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

$$\frac{1}{n}\#\{i = 1, ..., n : g(u_i) = v_r \quad \text{and} \quad \eta(u_i) \leq t - v_r\} = D_n^{(r)} \cdot F_n^{(r)}(t - v_r),$$

where we denote

$$D_n^{(r)} = \frac{\#\{i = 1, ..., n : g(u_i) = v_r\}}{n}$$

and

$$F_n^{(r)}(t) = \frac{\#\{i = 1, ..., n : g(u_i) = v_r \quad \text{and} \quad \eta(u_i) \leq t\}}{\#\{i = 1, ..., n : g(u_i) = v_r\}}.$$

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,

$$[T_n^{\{u_i\}}h](t) = \sum_{r=1}^{R} D_n^{(r)} \cdot F_n^{(r)}(t - v_r)$$

Now, since the conditions of Lemma 3 are satisfied,

$$\lim_{n\to\infty} D_n^{(r)} = \lim_{n\to\infty} \frac{\#\{i = 1, ..., n : g(u_i) = v_r\}}{n} = \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}$$

for $r = 1, \ldots, R$. Denote the limit $D^{(r)} = \lim_{n\to\infty} D_n^{(r)}$. Further notice that since the discrete-parameter random process $\{\eta(u_i) : i = 1, 2, \ldots\}$ satisfies the conditions of the Glivenko-Cantelli theorem, Lemma 5 may be used to obtain

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

uniformly in t. 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

$$\begin{aligned}[T^{\{u_i\}}h](t) &= \lim_{n\to\infty} [T_n^{\{u_i\}}h](t) = \sum_{r=1}^{R} \lim_{n\to\infty} \{D_n^{(r)} \cdot F_n^{(r)}(t - v_r)\} \\ &= \sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} F_\eta(t - v_r) \qquad \text{a.s.}\end{aligned} \tag{14}$$

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 $\lim\limits_{n\to\infty} T_n^{\{u_i\}}h$ exists, and

$$T^{\{u_i\}}h = [T^{\{u_i\}}g] * f_\eta \qquad \text{a.s.}$$

Furthermore, this equality also holds in $L_p(\Omega, P)$-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 $g \in \mathcal{R}_c(\mathbb{R}^m)$. We shall use the same notations as in the proof of Lemma 1.

Notice that F_η can be represented as the convolution of f_η with the unit step function $\chi_{[0,\infty)}$,

$$F_\eta(t) = \int_{-\infty}^{t} f_\eta(\tau)d\tau = \int_{-\infty}^{\infty} f_\eta(\tau)\chi_{[0,\infty)}(t-\tau)d\tau$$

for all $t \in \mathbb{R}$. Substituting into (14) yields

$$[T^{\{u_i\}}h](t) = \sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} \int_{-\infty}^{\infty} f_\eta(\tau)\chi_{[0,\infty)}(t - v_r - \tau)d\tau \tag{15}$$

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

$$[T^{\{u_i\}}h](t) = \int_{-\infty}^{\infty} f_\eta(\tau) \left[\sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}\chi_{[v_r,\infty)}\right](t-\tau)d\tau. \tag{16}$$

Clearly,

$$\begin{aligned}\left[\sum_{r=1}^{R} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}}\chi_{[v_r,\infty)}\right](t) &= \sum_{\{v_r \le t\}} \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) = v_r\}}{\lambda\{U\}} \\ &= \frac{\lambda\{\mathbf{x} \in U : g(\mathbf{x}) \le t\}}{\lambda\{U\}} = [Tg]\,(t). \end{aligned}\tag{17}$$

Substituting (17) into (16), we obtain

$$[T^{\{u_i\}}h](t) = \int_{-\infty}^{\infty} f_\eta(\tau)[Tg](t-\tau)d\tau = [[Tg] * f_\eta]\,(t) \qquad \text{a.s.} \tag{18}$$

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

$$[T^{\{u_i\}}h](t) = \left([T^{\{u_i\}}g] * f_\eta\right)(t) \qquad \text{a.s.} \tag{19}$$

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

$$(\chi_{(-\infty,t]} \circ \overline{g}_k) \to (\chi_{(-\infty,t]} \circ g) \tag{20}$$

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 (20) is not guaranteed. Namely, fix $t \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^m$, and examine $[\chi_{(-\infty,t]} \circ \tilde{\underline{g}}_k](\mathbf{x})$ as $k \to \infty$; three possible cases arise: *(i)* if $g(\mathbf{x}) > t$ then $[\chi_{(-\infty,t]} \circ \tilde{\underline{g}}_k](\mathbf{x}) = [\chi_{(-\infty,t]} \circ g](\mathbf{x}) = 0$ for all k; *(ii)* if $g(\mathbf{x}) < t$ then there exists some $k_0 \in \mathbb{N}$ such that $[\chi_{(-\infty,t]} \circ \tilde{\underline{g}}_k](\mathbf{x}) = [\chi_{(-\infty,t]} \circ g](\mathbf{x}) = 1$ for all $k > k_0$; *(iii)* if $g(\mathbf{x}) = t$ then $[\chi_{(-\infty,t]} \circ \tilde{\underline{g}}_k](\mathbf{x}) \leq [\chi_{(-\infty,t]} \circ g](\mathbf{x}) = 1$ 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

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

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

$$(\chi_{(-\infty,t]} \circ \underline{g}_k) \to (\chi_{(-\infty,t]} \circ g) \qquad \text{a.e.} \tag{21}$$

for all t.

Recall that $h = g + \eta$, and similarly denote

$$\underline{h}_k = \underline{g}_k + \eta, \qquad \overline{h}_k = \overline{g}_k + \eta.$$

Since $\underline{g}_k$ and $\overline{g}_k$ assume only finitely many values, (19) implies that there exist subsets $\underline{\Omega}_0^{(k)}, \overline{\Omega}_0^{(k)} \subseteq \Omega$, $k = 1, 2, \ldots$, of measure one such that

$$T^{\{u_i\}} \underline{h}_k = [T^{\{u_i\}} \underline{g}_k] * f_\eta, \qquad \text{on} \;\; \underline{\Omega}_0^{(k)} \tag{22}$$

and

$$T^{\{u_i\}} \overline{h}_k = [T^{\{u_i\}} \overline{g}_k] * f_\eta, \qquad \text{on} \;\; \overline{\Omega}_0^{(k)}. \tag{23}$$

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 $\overline{g}_k \leq g \leq \underline{g}_k$, we also have $\overline{h}_k \leq h \leq \underline{h}_k$. According to Lemma 4, for all $n \in \mathbb{N}$ we have

$$T_n^{\{u_i\}} \underline{h}_k \leq T_n^{\{u_i\}} h \leq T_n^{\{u_i\}} \overline{h}_k.$$

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

$$\begin{aligned} T^{\{u_i\}}\underline{h}_k &= \lim_{n\to\infty} T_n^{\{u_i\}}\underline{h}_k \le \liminf_{n\to\infty} T_n^{\{u_i\}}h \\ &\le \limsup_{n\to\infty} T_n^{\{u_i\}}h \le \lim_{n\to\infty} T_n^{\{u_i\}}\overline{h}_k = T^{\{u_i\}}\overline{h}_k. \end{aligned} \tag{24}$$

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

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

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

and

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

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

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

for all t. That is, we have

$$\lim_{k\to\infty} T^{\{u_i\}}\underline{g}_k = T^{\{u_i\}}g.$$

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

$$\begin{aligned} \lim_{k\to\infty} [[T^{\{u_i\}}\underline{g}_k] * f_\eta](t) &= \lim_{k\to\infty} \int_{-\infty}^{\infty} [T^{\{u_i\}}\underline{g}_k](t-\tau) \cdot f_\eta(\tau) d\tau \\ &= \int_{-\infty}^{\infty} [T^{\{u_i\}}g](t-\tau) \cdot f_\eta(\tau) d\tau = [[T^{\{u_i\}}g] * f_\eta](t) \end{aligned}$$

for all t.

Lastly, we evaluate (22) as $k \to \infty$ to conclude that

$$\lim_{k\to\infty} T^{\{u_i\}}\underline{h}_k = \lim_{k\to\infty} [T^{\{u_i\}}\underline{g}_k] * f_\eta = [T^{\{u_i\}}g] * f_\eta. \tag{25}$$

Similar derivations show that

$$\lim_{k\to\infty} T^{\{u_i\}}\overline{h}_k = \lim_{k\to\infty} [T^{\{u_i\}}\overline{g}_k] * f_\eta = [T^{\{u_i\}}g] * f_\eta. \tag{26}$$

Thus, by taking the limit $k \to \infty$ in (24), we can conclude that the limit $T^{\{u_i\}}h = \lim\limits_{n\to\infty} T_n^{\{u_i\}}h$

exists and

$$[T^{\{u_i\}}h](t) = \left([T^{\{u_i\}}g] * f_\eta\right)(t) \qquad \text{a.s.}$$

for all t. Moreover, notice that since both $T^{\{u_i\}}h$ and $[T^{\{u_i\}}g] * f_\eta$ are distribution functions bounded by 1, we have that for all t,

$$\left|[T^{\{u_i\}}h](t) - \left([T^{\{u_i\}}g] * f_\eta\right)(t)\right| \leq 2.$$

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

$$[T^{\{u_i\}}h](t) = \left([T^{\{u_i\}}g] * f_\eta\right)(t) \qquad \text{in } L_p(\Omega)\text{-norm}, 1 \leq p < \infty,$$

for all t, which completes the proof. ■

V. 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, *e.g.*, computer vision, medical data processing, speech recognition, and many more. Different, however, are the assumptions made on the relation between the observations.

The most fundamental case is where observations are related strictly by transformations of the domain. In this case, an abstract noiseless signal registration problem may be formulated as follows:

> Let $\phi : \mathbb{R}^m \to \mathbb{R}^m$ be an *unknown* transformation of $\mathbb{R}^m$, where ϕ belongs to a predetermined class. Given a *known* function $s : \mathbb{R}^m \to \mathbb{R}$, representing a signal, and a measurement (observation) o of the form
>
> $$o(\mathbf{x}) = s(\phi(\mathbf{x})) = [s \circ \phi](\mathbf{x}), \qquad \mathbf{x} \in \text{supp}\{s \circ \phi\}, \tag{27}$$
>
> *solve for the transformation ϕ.*

While this simple formulation is common to many problems, there are significant differences in the assumptions made on the transformation ϕ. For example [14]–[18]:

- Speech recognition – the function s ($m = 1$) represents audio and (27) is typically referred to as "time warping". The time warping function, ϕ, is typically assumed to be piecewise linear (with additional constraints). Roughly speaking, an object is a phoneme (word), and solving this problem is equivalent to synchronizing two repetitions on the same phoneme (word), spoken at different paces.

- Image registration – the function s ($m = 2$) represents an image. Here, ϕ represents a geometric deformation of the coordinate system (rigid, affine, elastic, etc.).
- Volume registration – the function s ($m = 3$) represents pointwise measurements about a 3D object (*e.g.*, X-Ray absorption in CT scan). When modeling an elastic geometric phenomena (*e.g.*, heart beats), ϕ is assumed to be a homeomorphism (or diffeomorphism) of $\mathbb{R}^3$.

This problem has gained much interest and many solutions have been proposed, separately, in each of the research fields mentioned above. However, despite its simple formulation as a functional equation in (27), the problem of solving for ϕ is known to be extremely difficult [19]. Even the seemingly simple case, where $m = 2$ and ϕ is affine (*i.e.*, $o(\mathbf{x}) = s(\mathbf{A}\mathbf{x} + \mathbf{b})$), typically leads to a continuous-parameter, multi-dimensional, non-linear and non-convex optimization problem [14]–[18], [20]–[22].

While most of the solution methods are either approximate or optimization-based, a few explicit linear methods have been proposed [23]–[25]. Namely, the latter two show that given that ϕ admits a finite order parameterization (of certain types), it is possible to construct *linear* systems of equations in the unknown parameters; the linear systems are obtained by using non-linear moment-like functionals; sufficient conditions for these linear systems to be equivalent to the original registration problem are described and, when satisfied, the solution is explicit, unique and exact (*i.e.*, not approximated); the method extends for all m.

In practice, however, the measuring device is not ideal and introduces deviations from the ideal model. These must be taken into account in order to avoid severe model mismatches (see [1]–[3], [26] and the reference therein). 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):

> 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* non-singular 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 $s : \mathbb{R}^m \to \mathbb{R}$, representing a signal,

and a single measurement (observation) o of the form

$$o(\mathbf{x}) = [Q \circ s \circ \mathcal{A}](\mathbf{x}) + \eta(\mathbf{x}), \qquad \mathbf{x} \in \text{supp}\{s \circ \mathcal{A}\}, \tag{28}$$

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

In this formulation, the function Q represents the overall global amplitude non-linearities 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. An important special case is the additive white Gaussian noise (AWGN) model, where $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ is also assumed to be zero mean Gaussian with a *known* variance σ_η^2.

For example, in a simplistic radar model (28) becomes $o(x) = q(s(ax + b)) + \eta(x)$, where s 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 non-linearity of the receiver, and η is the measurement noise. Alternatively, in image formation terminology, the model (28) 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 non-linear mapping of the intensities (*e.g.*, due to the camera's CCD).

Hence, (28) is the complicated problem of jointly estimating the, seemingly strongly coupled, left- and right-hand compositions Q and $\mathcal{A}$ [2], [3].

To demonstrate the usability of the distribution transformation, T, let us first consider the

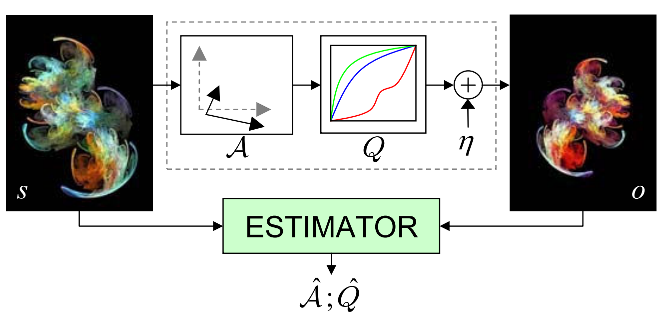

Fig. 1. Illustration of the problem description (28), where different non-linear mappings are associated with each of the color channels of the image.

noiseless case, that is, where (28) holds with $\eta \equiv 0$. Let us also assume that $s \in B_c(\mathbb{R}^m)$. In this case, the transformation T may be applied to (28). Using Lemma 1 we immediately find that

$$To = T(Q \circ s \circ \mathcal{A}) = T(Q \circ s) = [Ts] \circ Q^{-1}. \tag{29}$$

Hence, T has converted the joint problem (28), 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 (28), using (29) and since $Q(0) = 0$ we have

$$\begin{aligned}[Ro] &= [To] \circ o - [To](0) = ([Ts] \circ Q^{-1}) \circ (Q \circ s \circ \mathcal{A}) - [Ts](Q^{-1}(0)) \\ &= [Ts] \circ s \circ \mathcal{A} - [Ts](0) = ([Ts] \circ s - [Ts](0)) \circ \mathcal{A} = [Rs] \circ \mathcal{A}\end{aligned}$$

where the before last equality holds since $[Ts](0)$ is constant over all of $\mathbb{R}^m$. Hence, R (which has been defined in terms of T) has converted the joint problem (28), in the unknowns Q and $\mathcal{A}$, to a "new" problem in a single unknown, $\mathcal{A}$.

We conclude the results above in the following symmetric corollary:

Corollary 1: With the notations of (28) we have

$$To = [Ts] \circ Q^{-1} \tag{30}$$

and

$$[Ro] = [Rs] \circ \mathcal{A}. \tag{31}$$

Thus, using the distribution transformation T we have successfully decoupled the problem (28) into two "new" registration problems, (30) and (31). Each of these problems is exactly of the type (27), that is, where observations are strictly related by transformations of the domain. In particular, one may solve for the unknowns Q^{-1} and $\mathcal{A}$ by solving linear systems of equations, as previously mentioned [24], [25].

In fact, these properties of the transformation T led to the investigation of the properties of the transformation T in the random case. However, as mentioned in the introduction, To is not properly defined in the case where η does not vanish in (28). We were therefore interested in determining whether the sample distribution of o may be defined, such that is has analogous properties to those introduced by the transformation T.

This question is answered by Theorem 1; under the assumptions that $s \in \mathcal{R}_c(\mathbb{R}^m)$ and that $\{\eta(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}$ admits a probability density function f_η, we may conclude the following [3]:

Corollary 2: Let $\{u_i\}_{i=1}^{\infty}$ and $\{\tilde{u}_i\}_{i=1}^{\infty}$ be λ-u.d. sequences of distinct points in supp$\{s\}$ and supp$\{s \circ \mathcal{A}\}$, respectively, then

$$T^{\{\tilde{u}_i\}}o = \left([T^{\{u_i\}}s] \circ Q^{-1}\right) * f_\eta \qquad \text{a.s.} \tag{32}$$

Notice that (32) is the stochastic-case analog to (30), 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 (32) to the form (30) derived for the noise-free case. Having estimated Q, (28) 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.

For practical application examples of the derived methodology for the problem of joint radiometric-geometric image registration, we refer the interested reader to [1]–[3].

Acknowledgement

The authors would like to thank Daniel Berend and Haya Kaspi for helpful discussions, guidance and comments.

References

[1] S. Z. Kovalsky, G. Cohen, and J. M. Francos, "Registration of joint geometric and radiometric image deformations in the presence of noise," *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, pp. 561–565, 2007.

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

[3] S. Z. Kovalsky, R. Hagege, G. Cohen, and J. M. Francos, "Estimation of joint geometric and radiometric image deformations," 2008, submitted for publication.

[4] J. L. Doob, *Stochastic processes*. New York: John Wiley & Sons Inc., 1990, reprint of the 1953 original, A Wiley-Interscience Publication.

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

[6] 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.

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

[8] 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.

[9] 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.

[10] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*. New York: Wiley-Interscience [John Wiley & Sons], 1974.

[11] M. Spivak, *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*. W. A. Benjamin, Inc., New York-Amsterdam, 1965.

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

[13] P. Billingsley, *Probability and measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., 1995, a Wiley-Interscience Publication.

[14] Y. Amit, *2d Object Detection and Recognition: Models, Algorithms, and Networks*. Cambridge, MA, USA: MIT Press, 2002.

[15] K. Wang and T. Gasser, "Alignment of curves by dynamic time warping," *The Annals of Statistics*, vol. 25, no. 3, pp. 1251–1276, 1997.

[16] C. S. Myers and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition," *Bell System Technical Journal*, vol. 60, no. 7, pp. 1389–1409, 1981.

[17] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[18] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[19] D. Keysers and W. Unger, "Elastic image matching is np-complete," *Pattern Recognition Letters*, vol. 24, pp. 445–453, 2003.

[20] U. Grenander, *General pattern theory*, ser. Oxford Mathematical Monographs. New York: The Clarendon Press Oxford University Press, 1993, a mathematical study of regular structures, Oxford Science Publications.

[21] A. Trouvé, "Diffeomorphisms groups and pattern matching in image analysis," *Int. J. Comput. Vision*, vol. 28, no. 3, pp. 213–221, 1998.

[22] M. I. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *Int. J. Comput. Vision*, vol. 41, no. 1-2, pp. 61–84, 2001.

[23] J. Kannala, E. Rahtu, J. Heikkilä, and M. Salo, "A new method for affine registration of images and point sets." *Image Analysis, SCIA 2005 Proceedings, Lecture Notes in Computer Science 3540*, vol. 3540, pp. 224–234, 2005.

[24] R. Hagege and J. M. Francos, "Parametric estimation of two-dimensional affine transformations [object recognition applications]," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 3, pp. iii–305–308, 2004.

[25] J. M. Francos, R. Hagege, and B. Friedlander, "Estimation of multidimensional homeomorphisms for object recognition in noisy environments," *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1615–1619, 2003.

[26] J. O. Ramsay and X. Li, "Curve registration," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 60, no. 2, pp. 351–363, 1998.