Morphological granulometric estimation of random patterns in the context of parameterized random sets

Sinan Batman, Edward R. Dougherty*

*Department of Electrical Engineering, Johns Hopkins University, USA

Received 6 April 1999; received in revised form 7 April 2000; accepted 7 April 2000

Abstract

Morphological features are used to estimate the state of a random pattern (set) governed by a multivariate probability distribution. The feature vector is composed of granulometric moments and pattern estimation involves feature-based estimation of the parameter vector governing the random set. Under such circumstances, the joint density of the features and parameters is a generalized function concentrated on a solution manifold and estimation is determined by the conditional density of the parameters given an observed feature vector. The paper explains the manner in which the joint probability mass of the parameters and features is distributed and the way the conditional densities give rise to optimal estimators according to the distribution of probability mass, whether constrained or not to the solution manifold. The estimation theory is applied using analytic representation of linear granulometric moments. The effects of random perturbations in the shape-parameter vector is discussed, and the theory is applied to random sets composed of disjoint random shapes. The generalized density framework provides a proper mathematical context for pattern estimation and gives insight, via the distribution of mass on solution manifolds, to the manner in which morphological probes discriminate random sets relative to their distributions, and the manner in which the use of additional probes can be beneficial for better estimation.

Keywords: Granulometry; Morphology; Pattern estimation; Random set

1. Introduction

Morphological granulometries were introduced to model sieving processes in the context of random set theory [1] and granulometric filters have been used in a number of contexts to restore granular images [2–7]. In the domain of image classification, numerical features induced by the rate of sieving have been used for binary texture classification and segmentation [8–11] and granulometric texture classification has been extended to gray-scale textures [12–15]. Application has generally followed the standard feature-vector approach of determining granulometric features, applying feature reduction, and placing an observed texture into a class based on the statistical nearness of its empirical feature vector to the feature distribution of the class. Regarding patterns (binary shapes), the potential for granulometric pattern recognition has long been recognized [16] and feature-based classification methods similar to those used for texture classification can be employed. Our interest here lies in the related of direction of pattern estimation, or, more precisely, estimation of random shapes. We explore parametric estimation: a random set is governed by a multivariate probability (parameter) distribution and granulometric features are used to estimate the parameters.

If we consider a pattern to be a random subset of the Euclidean plane, then, ipso facto, pattern estimation fits into the theory of random sets. In particular, if we

*Corresponding author. Tel.: +1-409-862-8154; fax: +1-409-845-6259.
E-mail address: e-dougherty@tamu.edu (E.R. Dougherty).
associate a feature vector with the pattern, then estimation of the pattern from the feature vector is an inverse problem. As a random set, the pattern possesses a distribution and, according to the manner in which the features are defined, the pattern distribution induces a multivariate distribution on the feature vector. Given a class of estimation rules, our task is to find an optimal estimator, namely, the random set formed by applying an estimation rule to the feature vector that best estimates the pattern. In this paper the features are based on granulometries. The geometric nature of granulometries makes them a natural tool for the extraction of shape information.

From a general perspective, estimation of a random closed set $S$ via a random feature vector $Y = (Y_1, Y_2, ..., Y_m)$ in which each random variable $Y_i$ is a function of $S$ involves selecting an estimation rule $\Psi$ from a class $C$ of estimation rules and forming the random set $\Psi(Y)$ as an estimator of $S$. $C$ might be chosen so that $\Psi(Y)$ belongs to a certain family of shapes, perhaps a family to which $S$ belongs, or it may be chosen to facilitate a certain form of approximation, such as $\Psi(Y)$ being a polygonal approximation. The goodness of an estimator $\Psi(Y)$ is defined in terms of a postulated error function $c[\Psi(Y), S]$. An optimal estimator, relative to $c$, is one that minimizes the error over $C$. Analysis of $c[\Psi(Y), S]$ requires consideration of its probabilistic meaning, in particular, relative to the distribution of a random closed set (see the appendix for a rigorous mathematical formulation of a random closed set).

An error measure should reflect a consequential difference between a random set and an estimator. For instance, one might consider the expected area $(\bar{\lambda})$ of the symmetric difference $E[\bar{\lambda}][\Psi(Y)\Delta S]]$ (where we implicitly assume that $Y$ and $\Psi$ are measurable functions). Without simplifying assumptions regarding the nature of $S$, the expectation $E[\bar{\lambda}][\Psi(Y)\Delta S]]$ can be problematic. Error analysis is facilitated if we assume that $S$ is parameterized by a random vector $X = (X_1, X_2, ..., X_n)$. For instance, $S$ might be an ellipse with axes of lengths $X_1$ and $X_2$, and angle of rotation $X_3$. If $S$ is parameterized by $X$, then $Y$ is a vector-valued function of $X$, say $Y = g(X)$, the area $\bar{\lambda}[\Psi(Y)\Delta S]$ is also a function of $X$, and the error is given by

$$E[\bar{\lambda}][\Psi(Y)\Delta S]] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \bar{\lambda}[\Psi(g(x_1, x_2, ..., x_n))$$

$$\Delta S(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n,$$

(1)

where $g = (g_1, g_2, ..., g_m)$. To choose an optimal estimator, one must find the area of the symmetric difference as a function of $x = (x_1, x_2, ..., x_n)$, find the integral as function of $\Psi$, and then minimize the integral over all $\Psi$ in the class of estimators being considered.

A more tractable approach is to estimate $X$ based on $Y$ and thereby obtain an estimate of $S$. To motivate this approach, consider the special case of $g$ being a one-to-one function with $n$ components and inverse $g^{-1} = (h_1, h_2, ..., h_n)$. Then we have a system of equations

$$y_1 = g_1(x_1, x_2, ..., x_n),$$
$$y_2 = g_2(x_1, x_2, ..., x_n),$$
$$\vdots$$
$$y_n = g_n(x_1, x_2, ..., x_n)$$

(2)

and inversion occurs via the system

$$x_1 = h_1(y_1, y_2, ..., y_n),$$
$$x_2 = h_2(y_1, y_2, ..., y_n),$$
$$\vdots$$
$$x_n = h_n(y_1, y_2, ..., y_n).$$

(3)

Given $Y = y$, we deterministically invert to obtain $X = x = g^{-1}(y)$. In this one-to-one setting, the distribution of $S$ plays no role.

Suppose $g$ is not one-to-one, $m < n$, and $X$ is defined on an open subset of $\mathbb{R}^n$. Then, given $Y = y$, all we can say is that $X = x \in \mathbb{R}^{n-m}$ via the system of Eq. (2), which now has only $m$ equations, and $x \in M_g$. If a nonredundant feature is adjoined to $Y$, then $Y$ and $g$ are $(m + 1)$-dimensional, the new manifold containing $x$ has $n - m - 1$ dimensions, and the new manifold results from intersecting $M_g$ with the one-dimensional manifold determined by the equation $y_{m+1} = g_{m+1}(x)$. For $m$ equations, we were not to consider the probability distribution of $S$, then all we can say is that $x \in M_g$; however, the probability distribution of $S$ is induced by the distribution of $X$ and, via the mapping $g$, the distribution of $X$ induces a distribution on $M_g$. Indeed, this is the conditional distribution of $X$ given $Y = y$. This conditional distribution and all related probabilistic matters, including estimation of $X$, are determined by the joint distribution of $X$ and $Y$. A major portion of this paper will be devoted to the joint density for $X$ and $Y$, which exists as a generalized function on $\mathbb{R}^n$.

Generalized functions, in particular, delta functions, are commonplace in engineering when representing discrete densities; however, their use is not necessary for discrete densities. Here, delta functions occur in the representation of continuous densities, they are necessary, and their arguments are multivariate. The appendix contains some basic notions concerning generalized functions used in the paper; however, we proceed without explicit mathematical rigor to avoid losing sight of the key issues for pattern recognition among the mathematical details (noting that all operations can be rigorously justified in terms of appropriate classes of generalized functions).
functions). It is not our goal to develop new mathematical theory; the theory of generalized probability densities is well developed [17–19]. Our purpose is to explicate an appropriate context for pattern estimation, in particular, for granulometric estimation.

2. The joint parameter-feature density as a generalized function

To facilitate geometric appreciation of the joint distribution of the parameters and features, and to elucidate the role of generalized functions, we first treat the case of a single random parameter \( X \) and feature \( Y \), and we start with the simplest case in which \( Y = g(X) \) and \( g \) is a differentiable strictly monotone function (which means the inverse problem is uniquely solvable). To begin, assume \( g \) is strictly increasing. The joint probability distribution function of \( X \) and \( Y \) is defined by \( F_{X,Y}(x,y) = P(X \leq x, Y \leq y) \). For an arbitrary point \((x, y) \in \mathbb{R}^2 \), either \( g(x) \leq y \) or \( g(x) \geq y \). First suppose \( g(x) \leq y \). If \( X \leq x \), then \( Y = g(X) \leq g(x) \leq y \). Thus, the event \([X \leq x] \) is a subevent of the event \([X \leq x, Y \leq y] \), and \( P(X \leq x) \leq P(X \leq x, Y \leq y) \). The reverse inequality is trivial. Hence, \( F_{X,Y}(x,y) = F_X(x) \). Next suppose \( g(x) \geq y \). \([X \leq x, Y \leq y] \) is decomposed into a disjoint union of the events \([X \leq g^{-1}(y), Y \leq y] \) and \([g^{-1}(y) < X \leq x, Y \leq y] \). The second event has null probability because \( g^{-1}(y) < X \) implies \( Y = g(X) > y \). Thus, \( F_{X,Y}(x,y) = P(X \leq g^{-1}(y), Y \leq y) \). Since \( X \leq g^{-1}(y) \) implies \( Y = g(X) \leq y \), the event \([X \leq g^{-1}(y)] \) is a subevent of \([X \leq g^{-1}(y), Y \leq y] \), and therefore \( P(X \leq g^{-1}(y)) \leq P(X \leq g^{-1}(y), Y \leq y) \). Since the reverse inclusion always holds, \( F_{X,Y}(x,y) = F_X(g^{-1}(y)) \). Putting the two cases together yields

\[
F_{X,Y}(x,y) = \begin{cases} 
F_X(x) & \text{if } x \leq g^{-1}(y), \\
F_X(g^{-1}(y)) & \text{if } x \geq g^{-1}(y). 
\end{cases}
\] (4)

Defining \( H \) by \( H(x) = 1 \) if \( x > 0 \) and \( H(x) = 0 \) if \( x < 0 \), the partial derivative of \( F_{X,Y} \) with respect to \( x \) is

\[
\frac{\partial F_{X,Y}}{\partial x}(x,y) = f_X(x)H(g^{-1}(y) - x). 
\] (5)

Taking the generalized partial derivative with respect to \( y \) yields the generalized density

\[
f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = f_X(x) \frac{\delta(g^{-1}(y) - x)}{g'(g^{-1}(y))} = f_X(x) \frac{\delta(x - g^{-1}(y))}{g'(x)},
\] (6)

where \( \delta(x - g^{-1}(y)) \) is the delta function with argument \( x - g^{-1}(y) \). If \( g \) is decreasing, then the only change is that \( -g'(x) \) appears in the denominator. Hence, in either case the joint density is

\[
f_{X,Y}(x,y) = f_X(x) \frac{\delta(x - g^{-1}(y))}{|g'(x)|},
\] (7)

The delta function restrains \( f_{X,Y} \) to the curve \( y - g(x) = 0 \). In fact, a change of variables within the delta function shows that the joint density is also given by

\[
f_{X,Y}(x,y) = f_X(x)\delta(y - g(x)).
\] (8)

We see the equivalence of Eqs. (7) and (8) by integrating them against a test function \( \varphi \) and making the change of variable \( z = g(x) \):

\[
j_{\varphi}(x) = \int_{-\infty}^{\infty} \varphi(x)f_X(x)\delta(y - g(x)) \, dx
\]

\[
= \int_{-\infty}^{\infty} \varphi(\varphi^{-1}(z))f_X(\varphi^{-1}(z)) \frac{\delta(y - z)}{|g'(\varphi^{-1}(z))|} \, dz
\]

\[
= \varphi(g^{-1}(y))f_X(g^{-1}(y)) \frac{\delta(y - g^{-1}(y))}{|g'(g^{-1}(y))|}
\]

\[
= \int_{-\infty}^{\infty} \varphi(x)f_X(x) \frac{\delta(x - g^{-1}(y))}{|g'(x)|} \, dx.
\] (9)

The equivalence of generalized functions and changes of variable in the argument of a delta function are commented upon in the appendix.

When \( g \) is not one-to-one but is strictly monotone on a countable collection \( \{D_i\} \) of open intervals partitioning its domain \( D \), then the joint probability density is the sum of the contributions of the one-to-one constituent functions \( g_i \) on \( D_i \):

\[
f_{X,Y}(x,y) = \sum_i f_X(x) \frac{\delta(x - g_i^{-1}(y))}{|g_i'(x)|}.
\] (10)

On \( D_i \), \( \delta(x - g_i^{-1}(y)) \) concentrates the joint density of \( X \) and \( Y \) on the curve \( y - g_i(x) = 0 \). Overall, \( f_{X,Y} \) is concentrated on the curve \( y - g(x) = 0 \) on \( D \) and can be expressed over \( D \) as a whole by Eq. (8). The marginal density for \( Y \) is

\[
f_Y(y) = \int_{-\infty}^{\infty} \sum_i f_X(x) \frac{\delta(x - g_i^{-1}(y))}{|g_i'(x)|} \, dx = \sum_i \frac{f_X(g_i^{-1}(y))}{|g_i'(g_i^{-1}(y))|}
\] (11)

The conditional density for \( X \) given \( Y = y \) is

\[
f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\sum_i f_X(x)(\delta(x - g_i^{-1}(y)))/|g_i'(x)|)}{\sum_i (f_X(g_i^{-1}(y))/|g_i'(g_i^{-1}(y))|)}.
\] (12)

According to the series in the numerator, the conditional density for the parameter \( X \) given the fixed feature value \( y \) is concentrated at the points \( x_1, x_2, \ldots \) where \( g_i = g_i^{-1}(y) \in D_i \).
3. The joint density for multivariate parameters and features

Suppose there are \( n \) parameters, \( X_1, X_2, \ldots, X_n \), a single feature, \( Y \), and \( y = g(x_1, x_2, \ldots, x_n) \), with \( g \) a sufficiently well-behaved function defined on an open set \( D \subset \mathbb{R}^n \). Define the mapping

\[
\mathbf{u} = \begin{pmatrix}
    u_1 \\
    u_2 \\
    \vdots \\
    u_{n-1} \\
    u_n
\end{pmatrix} = \mathbf{T} = \begin{pmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_{n-1} \\
    g(x_1, x_2, \ldots, x_n)
\end{pmatrix}
\]

and let \( D \) be partitioned into a union of open sets \( D_i \) on each of which the restriction \( \mathbf{T}_i \) of \( \mathbf{T} \) to \( D_i \) is one-to-one and possesses nonzero Jacobian,

\[
J_{\mathbf{T}}(x_1, x_2, \ldots, x_n) = \left| \frac{\partial g(x_1, x_2, \ldots, x_n)}{\partial x_n} \right|
\]

In this case,

\[
J_{\mathbf{T}}(x_1, x_2, \ldots, x_n) = \frac{\partial g(x_1, x_2, \ldots, x_n)}{\partial x_n}
\]

On \( D_i \), there exists an inverse mapping \( \mathbf{T}^{-1} = (h_1, h_2, \ldots, h_m) \). In analogy to the single-parameter case, the joint density for \( X_1, X_2, \ldots, X_n \) and \( Y \) is given by

\[
f_{X,Y}(x_1, x_2, \ldots, x_n, y) = \sum_{i} f_{X_i}(x_1, x_2, \ldots, x_n) \frac{\delta(y - g(x_1, x_2, \ldots, x_n, y))}{\left| \frac{\partial g(x_1, x_2, \ldots, x_n)}{\partial x_n} \right|}
\]

The delta function concentrates \( f_{X,Y} \) on the manifold \( \mathbb{M}_Y \) determined by \( y = g(x_1, x_2, \ldots, x_n) \), which is concentrated on the level curve determined by \( y = g(x_1, x_2, \ldots, x_n) \). Optimal estimation of the parameters from the single feature \( Y \) depends on \( f_{X,Y} \). The choice of \( u_n = g(x_1, x_2, \ldots, x_n) \) in the definition of \( \mathbf{T} \) is arbitrary; we can let any \( u_n \) be defined by \( g(x_1, x_2, \ldots, x_n) \), with the remaining \( u \) components equal to the corresponding \( x \) components.

Various alternate forms of the joint density can be derived by changes of variables (under suitable conditions) [17–19]. In particular, it can be shown that, for a sufficiently well-behaved function \( \zeta(x_1, x_2, \ldots, x_n) \),

\[
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(y - g(x_1, x_2, \ldots, x_n)) \zeta(x_1, x_2, \ldots, x_n) \ dx_1 \ dx_2 \cdots \ dx_n = \int_{M_y} \zeta(x_1, x_2, \ldots, x_n) \frac{\varphi(x_1, x_2, \ldots, x_n)}{f_T(y)} \ ds, 
\]

where \( V \) is the gradient and \( ds \) represents surface measure on \( M_y \). Letting

\[
\zeta(x_1, x_2, \ldots, x_n) = \frac{f_K(x_1, x_2, \ldots, x_n)}{f_T(y)} \varphi(x_1, x_2, \ldots, x_n)
\]

in Eq. (19), where \( \varphi \) is a test function, and applying Eq. (18) shows that

\[
f_{X,Y}(x_1, x_2, \ldots, x_n, y) = f_{K}(x_1, x_2, \ldots, x_n) \frac{\delta(s)}{f_T(y) \left| \nabla g(x_1, x_2, \ldots, x_n) \right|}
\]

where the notation “\( \delta(s) \)” indicates that the density on the right-hand side is defined only on the manifold \( M_y \). \( \left| \nabla g(x) \right|^{-1} \) is defined as the density of \( \delta(y - g(x)) \) on \( M_y \). The occurrence of \( \left| \nabla g(x) \right|^{-1} \) in this representation is rather intuitive. It says that the conditional probability mass \( f_{X,Y}(y) \) is inversely related to the change in the feature; to wit, small change of the feature in an infinitesimal region of the parameter vector means greater probability mass for the parameter vector given the feature.

Turning to the case of \( m \) features, \( Y_1, Y_2, \ldots, Y_m, \) with \( y_k = g_k(x_1, x_2, \ldots, x_n), \) and \( g = (g_1, g_2, \ldots, g_m) \), a mapping \( \mathbf{T} \) can be formed with \( u_{n-m+k} = g_k(x_1, x_2, \ldots, x_n) \) for \( k = 1, 2, \ldots, m \), and

\[
J_{\mathbf{T}}(x_1, x_2, \ldots, x_n) = \begin{vmatrix}
    \frac{\partial g_1}{\partial x_{n-m+1}} & \frac{\partial g_1}{\partial x_{n-m+2}} & \cdots & \frac{\partial g_1}{\partial x_n} \\
    \frac{\partial g_2}{\partial x_{n-m+1}} & \frac{\partial g_2}{\partial x_{n-m+2}} & \cdots & \frac{\partial g_2}{\partial x_n} \\
    \vdots & \vdots & & \vdots \\
    \frac{\partial g_m}{\partial x_{n-m+1}} & \frac{\partial g_m}{\partial x_{n-m+2}} & \cdots & \frac{\partial g_m}{\partial x_n}
\end{vmatrix}
\]

If \( D \) is partitioned into a union of open sets \( D_i \) on each of which the restriction \( \mathbf{T}_i \) of \( \mathbf{T} \) to \( D_i \) is one-to-one and possesses nonzero Jacobian, then the joint density
\[ f_{X,Y}(x,y) = f_X(x) \prod_{k=1}^{m} \delta(y_k - g_k(x)), \]

\[ f_{X,Y}(x,y) = \sum_{i} f_X(x) \times \prod_{k=1}^{m} \delta(x_{n-m+k} - h_{i,n-m+k}(x_1,x_2,...,x_{n-m},y_1,y_2,...,y_m)) \]

\[ f_{X,Y} \text{ is concentrated on the intersection, } M_k = M_1 \cap M_2 \cap \cdots \cap M_m, \text{ where } M_k \text{ is determined by the equation } y_k = g_k(x_1,x_2,...,x_n). \]

From Eq. (23), the conditional density is given on \( M_k \) by

\[ f_{X|Y}(x|y) = \frac{f_X(x)}{f_Y(y)} \prod_{k=1}^{m} \delta(y_k - g_k(x)). \]

It can also be obtained from Eq. (24). Letting \( x^m = (x_1,x_2,...,x_{n-m}) \) and integrating Eq. (24) with respect to \( x \) yields

\[ f_Y(y) = \sum_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \]

\[ \times \frac{f_X(x_1,x_2,...,x_{n-m},h_{i,n-m+1}(x^m,y),...,h_{i,n}(x^m,y))}{J_T(x_1,x_2,...,x_{n-m},h_{i,n-m+1}(x^m,y),...,h_{i,n}(x^m,y))} \times dx_1 dx_2 \cdots dx_{n-m}. \]

For \( m.n \), we get the following special form of the conditional density from Eq. (24)

\[ f_{X|Y}(x|y) \]

\[ = \frac{\sum f_X(x) \prod_{k=1}^{m} \delta(x_k - h_{i,k}(y))/J_T(h_{i,1}(y),...,h_{i,n}(y))}{\sum f_X(h_{i,1}(y),...,h_{i,n}(y))/J_T(h_{i,1}(y),...,h_{i,n}(y))}, \]

where we have replaced \( J_T(x) \) in the numerator by \( J_T(h_{i,1}(y),...,h_{i,n}(y)) \) because \( f_{X|Y}(x|y) \) is concentrated on \( M_k = M_1 \cap M_2 \cap \cdots \cap M_n \). For fixed \( y \), the conditional density is concentrated at the points \( x_1,x_2,....,x_n \). If \( M_k \) is a single point, then there is a unique solution for \( x \) given \( y \). If the solution is not unique, then the probability mass is divided up among the points of intersection.

Uniqueness and nonuniqueness are illustrated in parts a and b of Fig. 1, respectively. In each, there are two curves \( M_1 \) and \( M_2 \), defined by \( y_1 - g_1(x_1,x_2) = 0 \) and \( y_2 - g_2(x_1,x_2) = 0 \), respectively. If we consider feature \( Y_1 \) alone, then the conditional density, \( f_{X|Y_1}(x_1,x_2|y_1) \), is given by Eqs. (18) and (21), and is illustrated by the density over the manifold \( M_1 \). Estimation of \( (X_1,X_2) \) from \( Y_1 \) is relative to this conditional density. A corresponding statement applies to the conditional density \( f_{X,Y}(x_1,x_2|y_1,x_2) \) over \( M_2 \) given \( Y_2 \) alone. If we consider \( Y_1 \) and \( Y_2 \) together, then in part a of Fig. 1 the curves intersect at a single point and there is a unique solution given by the intersection point \( (x_1,x_2) \in M_1 \cap M_2 \); however, in part b there are two points, \( (x_{11},x_{12}) \) and \( (x_{21},x_{22}) \) in \( M_1 \cap M_2 \), the conditional density for \( (x_{11},x_{12}) \) is given by Eq. (27), and estimation of \( (X_1,X_2) \) from \( Y_1,Y_2 \) is relative to this density.

Having found the conditional density for the parameter vector given the feature vector, an error measure can be defined directly in terms of \( X \) and \( Y \). For instance, we can use the mean-square error (MSE) between \( X \) and an estimator \( \Psi(Y) \):

\[ \text{MSE}[X,\Psi(Y)] = \sum_{k=1}^{n} E[|X_k - \psi_k(Y)|^2], \]

where \( \Psi = (\psi_1,\psi_2,...,\psi_n) \). The MSE is minimized by the conditional expectation, \( E[X|Y] \), which is determined componentwise by

\[ \psi_k(Y) = E[X_k|Y] = \int_{-\infty}^{\infty} x_k f(x_k|y_1,y_2,...,y_m) \, dx_k. \]
A drawback with using the conditional expectation is that, given an observed feature vector \( y_0, x_0 = E[X|y_0] \) will not likely lie on the solution manifold \( M_d \) determined by \( y_0 \). In such a case, the predicted value, \( x_n \), of \( X \) would not yield the observed features. To insure that the estimated parameter vector \( x_n \) is consistent with the feature data, meaning \( g(x_n) = y_0 \), we must constrain error minimization to points on \( M_d \). An optimal estimate constrained to \( M_d \) can be implemented by choosing a point on \( M_d \) that has minimum MSE with respect to the expected value of some distance metric. Since such an expected value will be computed via surface integration, a natural choice of distance metric is the geodesic distance. Calculation of geodesic distance is often time consuming because it requires sophisticated optimization techniques. It is possible to approximate geodesic distance, for instance, by partitioning the manifold into submanifolds of small curvature and approximating distance within each patch by Euclidean distance; nevertheless, computations can still be prohibitive. Since the features do not fully determine the parameter state, one may or may not be concerned with an estimate consistent with the observed features.

4. Optimal granulometric estimation

Binary granulometries are parameterized families of set operators that satisfy certain properties [1, 20–22]. In this paper we restrict our attention to the simplest form of granulometry, a family of morphological openings by a single-parameterized homothetic structuring element. Of granulometry, a family of morphological openings by this paper we restrict our attention to the simplest form of set operators that satisfy certain properties [1, 20–22].

If \( t > 0 \) and \( E \) is compact and convex, then the family of openings \( \{S \odot tE\} \) is a granulometry. For the moment, we suppose \( S \) is not random. If \( t_1 < t_2 \), then \( S \odot t_2E \subset S \odot t_1E \). Hence, as \( t \) increases, the area of \( S \odot tE \) decreases. If we let \( \Omega (t) \) be the area \((\lambda)\) of \( S - S \odot tE \), with \( \Omega (0) = 0 \), then \( \Omega (t) \) is an increasing function of \( t \) known as a size distribution. If \( S \) is bounded, then \( \Omega (t) = \lambda (S) \) for sufficiently large \( t \). The normalized size distribution \( \Phi (t) = \frac{\Omega (t)}{\lambda (S)} \) increases from 0 to 1 and is continuous from the left [1], so that it is a probability distribution function and its (generalized) derivative \( \Phi (t) = d\Phi (t)/dt \) is a probability density. \( \Phi (t) \) and \( \Phi (t) \) are known as the pattern spectrum of \( S \) relative to the generator \( E \). Since \( \Phi (t) \) is a probability density, it possesses moments, \( \mu_1, \mu_2, \ldots \). These are employed as features for texture classification, and in this paper for parametric estimation. For a random set \( S \), the pattern spectrum is a random function and its moments are random variables, \( M_1, M_2, \ldots \). Each realization of the random set yields a corresponding realization of the pattern-spectrum process. Moments of an observed pattern-spectrum are realizations of moments of the pattern-spectrum. Granulometric estimation depends on the distribution of the moments \( M_1, M_2, \ldots \), in particular, the conditional expectations of the parameters given some set of granulometric moments. Consequently, we are concerned with the conditional density \( f(x|m) \), where \( X \) and \( M \) are the pattern-parameter and granulometric-moment random vectors. Here we assume that \( M \) is composed of moments resulting from linear structuring elements. These are commonly employed because they can be efficiently computed via run-lengths and a small number of moments can provide good discrimination, especially if more than one direction is used.

For a linear generator \( E \), there exists an analytic representation of the moments. First suppose that \( E \) is a unit-length vertical line segment and \( S \) is vertically convex, meaning that \( S \) can be decomposed into a disjoint union of connected linear segments, \( S = \cup (S \cap L_k) \), where \( L_k \) is the vertical line through the point \( x \) on the horizontal axis. Note that a convex set is vertically convex; indeed, a convex set can be decomposed into connected linear segments in any direction. Define the length function \( \gamma (x) \) to be the length of \( S \cap L_k \). Because \( S \) is compact, \( \gamma \) is bounded, possesses compact support, and is integrable (assuming it is measurable). We have the following representations for the pattern spectrum and its moments [23]. If the length function has the (common) form

\[
\gamma (x) = \begin{cases} \gamma_j f(x) & \text{if } a_j < x \leq b_j \text{ for } j = 1, 2, \ldots, l_1, \\ t_i & \text{if } c_i < x < d_i \text{ for } i = 1, 2, \ldots, l_2, \end{cases}
\]

where \( \gamma_j \) is one-to-one and differentiable on its domain (having one-sided derivatives at domain endpoints), and \( t_i \) is a constant, then each \( \gamma_j \) possesses a differentiable inverse function \( \gamma_j^{-1} \) with domain \( V_j \), and

\[
\Phi (t) = \frac{t}{\lambda (S)} \sum_{j=1}^{l_1} \int_{V_j} \frac{d}{dt} \gamma_j^{-1}(t) \ dt + \sum_{i=1}^{l_2} \frac{\lambda(t \cap V_i)}{\lambda (S)} t_i (t - t_i),
\]

where the expression is meant to yield \( \Phi (t) = 0 \) if both sums are null at \( t \), and

\[
\mu_k = \frac{1}{\lambda (S)} \sum_{j=1}^{l_1} \int_{V_j} \frac{d}{dt} \gamma_j^{-1}(t) \ dt + \sum_{i=1}^{l_2} \frac{\lambda(t \cap V_i)}{\lambda (S)} t_i (t - t_i). \]

If the \( \gamma_j \) are differentiable except at finite numbers of points, then the representation of \( \Phi \) holds everywhere except at a finite number of points and, since the granulometric moments are derived via integration, these points do not affect the moments and the representation of \( \mu_k \) remains valid.

There are several points concerning the representations that are germane to the present paper. First, vertical convexity generalizes to linear convexity in the direction
\( \theta \), and the representations generalize by simply rotating \( S \) through \( -\theta \). If \( S \) is not vertically convex, then it may be possible to decompose \( S \) into a disjoint union of vertically convex sets and apply the representations via the decomposition. Finally, when \( S \) is convex, there exists a simplified representation for \( \mu_k \). For convex \( S \), \( \gamma \) is continuous and the following conditions apply: (1) there exists \( x_1 \) and \( x_2 \) such that the support of \( \gamma \) equals \([x_1, x_2]\); (2) \( \gamma \) attains its maximum \( \Gamma \) over \([x_1, x_2]\) and the maximum set, \([x: \gamma(x) = \Gamma]\), is an interval, say \([z_1, z_2]\); (3) \( \gamma \) is strictly increasing over \([x_1, z_1]\) and strictly decreasing over \([z_2, x_2]\); (4) \( \gamma \) possesses the decomposition

\[
\gamma(x) = \begin{cases} 
\gamma_1(x), & x_1 \leq x \leq z_1, \\
\Gamma, & z_1 \leq x \leq z_2, \\
\gamma_2(x), & z_2 \leq x \leq x_2,
\end{cases}
\]

where \( \gamma_1 \) is strictly increasing, \( \gamma_2 \) is strictly decreasing, and, without loss of generality, we can assume that \( \gamma_1(x_1) \leq \gamma_2(x_2) \); and (5) \( \mu_k \) has the representation

\[
\mu_k = \frac{1}{\lambda(S)} \left[ \int_0^\Gamma t \zeta(t) dt + \Gamma \xi^{-1}(z_2 - z_1) \right],
\]

where

\[
\zeta(t) = \begin{cases} 
0, & t < \gamma(t_1) \text{ or } t > \Gamma, \\
\gamma_1^{-1}(t), & \gamma(t_1) \leq t \leq \gamma_2(t_2), \\
\gamma_2^{-1}(t) - \gamma_1^{-1}(t), & \gamma_2(t_2) \leq t \leq \Gamma.
\end{cases}
\]

Further reductions are possible.

To elucidate the method, consider a square of random side \( R \) and angle \( \Theta \) with respect to the horizontal axis, and assume that \( R \) and \( \Theta \) are uniformly distributed on \([a, b]\) and \([0, \pi/2]\), respectively. From Eq. (34), the \( k \)th granulometric moment for the vertical linear granulometry is

\[
\mu_k = g_k(r, \theta) = \frac{r}{\cos \theta} \left( 1 - \frac{k}{k + 2} \tan \theta \right).
\]

From Eqs. (18) and (21), the conditional density for \((R, \Theta)\) given \( M_k = \mu_k \) is given by

\[
f(r, \theta | \mu_k) = \frac{f_{R, \Theta}(r, \theta)}{f_{R}(\mu_k)} \delta(\mu_k - g_k(r, \theta))
\]

\[
= \frac{f_{R, \Theta}(r, \theta)}{f_{R}(\mu_k)} \delta(s)
\]

where \( \delta(s) \) restricts the latter expression to the curve \( M_k \) defined by

\[
r = \left( 1 - \frac{k}{k + 2} \tan \theta \right)^{-1/k} \mu_k^{-k/2} \cos \theta = 0
\]

for \( \theta \in [0, \pi/4] \) and \( r \in [a, b] \) [24].

Given the first two moments, \( \mu_1 \) and \( \mu_2 \), the conditional densities \( f(r, \theta | \mu_1) \) and \( f(r, \theta | \mu_2) \) are concentrated on the curves \( M_1 \) and \( M_2 \), respectively, in the \( r-\theta \)-plane. Relative to \( \mu_1 \), the conditional expectation for \( R \) and \( \Theta \) is determined by the probability mass \( f(r, \theta | \mu_1) \), and it lies off the curve \( M_1 \). If desired, error minimization can be restricted to \( M_1 \). Similar statements apply to \( \mu_2 \). The curves \( M_1 \) and \( M_2 \) intersect at a single point \((r_0, \theta_0)\) in the \( r-\theta \)-plane, so that \((r_0, \theta_0)\) uniquely solves the inverse problem for the system

\[
\begin{align*}
y_1 &= \mu_1(r, \theta), \\
y_2 &= \mu_2(r, \theta), \quad (r \in [a, b], \theta \in [0, \pi/4]).
\end{align*}
\]

Suppose now we let \( \Theta \) vary between 0 and \( \pi/2 \). On \([\pi/4, \pi/2]\),

\[
\mu_k = g_k(r, \theta) = \frac{r}{\sin \theta} \left( 1 - \frac{k}{k + 2} \cot \theta \right).
\]

There is a unique solution to the inverse problem on \([\pi/4, \pi/2]\); however, over the full interval \([0, \pi/2]\) there are two solutions, each in its corresponding domain. The probability mass for each can be calculated from Eq. (27). One might try to use more than two moments to obtain a unique solution, but this will not help here because all granulometric moments are functionally dependent on the first two. Moments of a different linear granulometry can be used in addition to the two used here to obtain a unique solution, so long as the linear structuring element is not obtained by rotating the current one by a multiple of \( \pi/4 \).

5. Random perturbations

The ideal random set may be determined by the parameter vector \( X \), but the set we observe may be determined by a vector \( Z = K(X, N) \), where \( N \) is a random perturbation in the parameter vector. In this case, the probabilistic analysis involves \( Z \) and the feature vector \( Y = g(X, N) = g(K(X, N)) \). Estimation of \( X \) results from estimation of \((X, N)\) by the preceding methods; however, the conditional density of the function of the form \( f_{X,N,Y}(x, n, y) \) and the conditional density is given by

\[
f_{X,N,Y}(x, n, y) = \frac{f_{X,N,Y}(x, n, y)}{f_Y(y)}.
\]

A special case occurs when

\[
Z = K(X, N) = (\kappa_1(X_1, N), \kappa_2(X_2, N), \ldots, \kappa_n(X_n, N)).
\]

Here, \( X \) is replaced componentwise by \( Z \). For \( k = 1, 2, \ldots, n, \kappa_k \) serves as a coupling function between the perturbed and the original parameter, with \( X_k \) being replaced by \( Z_k = \kappa_k(X_k, N) \). Whereas in the general case, the perturbed set can belong to a different class of shapes, when there is coupling the perturbed set must belong to the same shape class. A natural way to proceed is to...
estimate the states of the coupling functions by replacing $f_{X,Y}$ by $f_{Z,Y}$ in the probabilistic analysis and then, having estimated $Z$, proceed to find the optimal estimate of $X_k$ given $Z_k$ via the function $Z = K(X,N)$. This approach is geometrically sound in that it tries to determine the state of the distorted shape first in order to filter out the perturbation afterwards.

We illustrate the situation in which the perturbed set remains in the same shape class, first without using coupling and then using coupling. Again consider the random square of side $R$ and angle $\Theta$; however, now $\Theta \in [\pi/10, \pi/4 - \pi/10]$. The perturbation is embedded in the rotation by an additive noise term $N$ uniformly distributed on $[\pi/10, \pi/10]$. The perturbed shape is a randomly sized and rotated square with random angle $\Theta + N$. For $\theta + n \in [0, \pi/4]$, the vertical linear granulometry has $k$th moment $\mu_k = g_k^v(r, \theta, n)$ given by Eq. (36) with $\theta + n$ in place of $\theta$. From Eqs. (18) and (21), the conditional density for $(R, \Theta, N)$ given $M_k = \mu_k$ is

$$f(r, \theta, n|\mu_k) = \frac{f_{R,\Theta,N}(r, \theta, n)}{f_{M_k}(\mu_k)} \frac{\delta(\mu_k - g_k^v(r, \theta, n))}{\delta(s)} = \frac{f_{R,\Theta,N}(r, \theta, n)}{f_{M_k}(\mu_k)} \frac{\delta(s)}{|\nabla g_k^v(r, \theta, n)|}$$

(43)

where $\delta(s)$ restricts the latter expression to the surface $M_k$ described in parametric form by Eq. (38) with $\theta + n$ in place of $\theta$. Fig. 2 shows the manifolds corresponding to particular values of the pattern-spectrum mean, $\mu_1 = 140$, and the second moment, $\mu_2 = 20,246$. Also shown is the conditional-expectation vector, $(E[R|\mu_1], E[\Theta|\mu_1], E[N|\mu_1])$, which lies on the manifold determined by $\mu_1$ in $r-\theta-n$ space. The perturbation has spread the probability masses from one-dimensional curves to two-dimensional surfaces. Thus, the intersection of the manifolds is one dimensional, the linear segment described by $(r = 147, \theta + n = \pi/16)$, as opposed to a point (or points) in the unperturbed case. Rather than a unique solution for the inverse problem using two moments, we only obtain a solution curve in $r-\theta-n$ space, together with a conditional density $f(r, \theta, n|\mu_1, \mu_2)$. The noise has inhibited our ability to estimate the pattern. Owing to functional dependence among the moments, using additional moments from the same granulometry will not reduce the solution space.

If we use coupling, then $Z = (R, \Omega)$, where $\Omega = \Theta + N$, and we obtain the conditional densities $f(r,\omega|\mu_1)$ and $f(r,\omega|\mu_2)$. The curves upon which they are concentrated intersect in a single point $(r_0, \omega_0)$ giving the state of the distorted square. One can now make an estimate of $\Theta$ based on $\omega_0$, namely, $E[\Theta|\Omega = \omega_0]$. Granulometric estimation yields a unique solution using coupling; however, we are left with estimating $\Theta$ via the distribution of $\Omega$.

We now treat a perturbation in which the perturbed shape can belong to a different shape class than the unperturbed pattern. Consider a random ellipse with horizontal and vertical axes described by the equation $(x/A)^2 + (y/B)^2 = 1$, where $A$ and $B$ are random. Distortion is introduced by using the random power term $Q$, so that the equation becomes $(x/A)^2 + (y/B)^2 = 1$. In accordance with the distribution of $Q$, there is continuous deformation through various shapes. For a vertical linear structuring element, some computation yields

$$\mu_k = \frac{(2B)^k}{1 + k + Q} \frac{\Gamma(1/Q)(2 + (1 + k)/Q)\Gamma(1 + 2/Q)}{\Gamma(1 + 1/Q)(1 + (2 + k)/Q)},$$

(44)

where $\Gamma$ denotes the gamma function. Since $A$ does not appear in the expression, given $\mu_k$, the conditional probability density, $f(b,q|\mu_k)$, is a generalized function concentrated on the line in the $B-Q$ plane defined by Eq. (44).

6. Disjoint grain unions

Thus far, we have focused our attention on shapes (single-component random sets); however, as mentioned earlier, the analytic representation of linear granulometric moments extends to random sets that possess a certain type of decomposition (applicable to most typical patterns), and for these the representation can be applied to each component of the decomposition [23]. In particular, it applies to a disjoint union

$$S = \bigcup_{i=1}^t S_i + z_i$$

(45)

of convex, compact sets $S_1, S_2, \ldots, S_t$. Suppose $S_i$ has length function $\gamma_i$, the maximum $\Gamma_i$ of $\gamma_i$ is attained over an interval of length $L_i$, $\Gamma_0 = 0$, and, without loss of generality, $\Gamma_0 \leq \Gamma_1 \leq \cdots \leq \Gamma_t$. Then

$$\mu_k = \frac{\sum_{i=1}^t \sum_{l=1}^t \int_{\Gamma_i}^{\Gamma_{i+1}} \gamma_i(t) \, dt + \sum_{i=1}^t \Gamma_i^{k+1} L_i}{\sum_{i=1}^t \int_{\Gamma_i}^{\Gamma_{i+1}} Q(S_i)}.$$
For instance, if the random set is composed of a disjoint union of $I$ squares possessing random edge lengths $R_1, R_2, \ldots, R_I$ and random angles $\Theta_1, \Theta_2, \ldots, \Theta_I \in [0, \pi/4]$, then Eq. (46) yields

$$
\mu_k = \sum_{i=1}^{I} \sum_{j=1}^{I} \frac{2 \sin \theta_r \cos \theta_j}{r_{ij}^2} \left( (r_i \cos \theta_i)^{k+2} - (r_i \cos \theta_i)^{k+1} \right) = \sum_{i=1}^{I} r_i^{k+2},
$$

where we define $r_0 = \theta_0 = 0$. The problem with this expression is that the densities are too complicated for practical numerical implementation unless the number of grains is small.

There is a special, more tractable case of the union model of Eq. (45). According to the granulometric mixing theorem, [23] if $S$ is composed of disjoint homothetic translates arising from $m$ compact sets $A_1, A_2, \ldots, A_m$, namely,

$$
S = \bigcup_{i=1}^{m} \bigcup_{j=1}^{m} (r_{ij} A_i + z_{ij}) \quad (48)
$$

then, for the granulometry generated by a convex, compact (not necessarily linear) set $E$,

$$
\mu_k = \sum_{i=1}^{m} \sum_{j=1}^{m} (A_i) \mu_k (A_i) r_{ij}^{k+2}, \quad (49)
$$

where $\mu_k (A_i)$ is the $k$th granulometric moment of $A_i$. If grain sizing is independent, then the distribution of the parameter vector $R = (R_{11}, R_{12}, \ldots, R_{m,n})$ is a product of the grain sizing distribution with itself. Moreover, according to the asymptotic granulometric mixing theorem [24,25], if the grain types $A_1, A_2, \ldots, A_m$ occur in fixed proportions as the number of grains goes to infinity and grain sizing is independent, then, under mild conditions, $M_k$ possesses an asymptotically normal distribution and there exist analytic expressions for the moments of $M_k$.

If there is only a single grain primitive $A$, then $S = \bigcup_{i=1}^{I} (r_i A + z_i)$ and Eq. (49) reduces to

$$
\mu_k = \mu_k (A) \sum_{i=1}^{I} r_i^{k+2} \sum_{i=1}^{I} r_i^2.
$$

If we apply this to a union of randomly sized squares at fixed angle of rotation $\theta \in [0, \pi/4]$, each square being of the form $R_i \Lambda$, where $\Lambda$ is the unit square, then

$$
\mu_k = g_k (r_1, r_2, \ldots, r_I) = \left( \frac{1}{\cos \theta} \right)^4 \left( 1 - \frac{k}{k + 2} \tan \theta \right) \sum_{i=1}^{I} r_i^{k+2} \sum_{j=1}^{I} r_j^2.
$$

From Eq. (21), the conditional density for $R$ given $M_k = \mu_k$ is given by

$$
f_{R|M_k}(r | \mu_k) = \frac{f_R(r)}{f_{M_k}(\mu_k)} \delta(s) / |\nabla g_k(r)|,
$$

where $\delta(s)$ concentrates the conditional density to the $I$-dimensional manifold determined by the equation $\mu_k = g_k (r_1, r_2, \ldots, r_I) = 0$. Under independent sizing, $f_{M_k}$ is a simple product, and for a large number of grains we can use the asymptotic normal approximation of $f_{M_k}$.

7. Conclusion

Estimation of random patterns is placed in the context of generalized probability densities concentrated on manifolds to characterize estimation of parameterized random sets. This framework has been employed to characterize granulometric parametric pattern estimation. The manner in which the joint probability mass of the parameters and features is distributed on a solution manifold has been explained, together with the way the conditional densities give rise to optimal estimators according to the distribution of probability mass, whether constrained or not to the solution manifold. The estimation theory has been applied using the analytic representation of linear granulometric moments. If chosen properly, these moments act as independent shape probes, effectively reducing the solution space. The effects of random perturbations in the parameter vector have been considered, both when they are coupled and uncoupled. The generalized density framework not only provides a proper mathematical context for pattern estimation, it also gives insight via the distribution of mass on solution manifolds, to the manner in which geometric probes discriminate random sets relative to their distributions, and the manner in which the use of additional probes can be beneficial for better estimation.

Appendix

$A.1.\textit{Distribution of a random closed set}$

To define a random closed set, let $F$ be the space of closed sets in the Euclidean plane $\mathbb{R}^n$, $\mathcal{J}$ be the hit-or-miss topology on $F$, $\Sigma_\mathcal{F}$ be the Borel $\sigma$-algebra generated by $\mathcal{J}$, and $(S, \Sigma_S, \nu)$ be a probability space with $\sigma$-algebra $\Sigma_S$ and measure $\nu$. A random closed set is a measurable function $S : (Z, \Sigma_Z, \nu) \to (F, \Sigma_{\mathcal{F}})$. This means that for every $z \in Z, S(z)$ is a closed subset in $\mathbb{R}^n$. Keeping in mind that “points” in $F$ are closed sets and Borel sets in $F$ are sets of closed sets, for any Borel set $B \in \Sigma_\mathcal{F}$, $S^{-1}(B) \in \Sigma_Z$ and a probability measure is induced on $\Sigma_\mathcal{F}$ by

$$
P_S(B) = \nu(S^{-1}(B)) = \nu(\{ z \in Z : S(z) \in B \}). \quad (A.1)$$
The induced probability $P_S(\mathcal{H})$ gives $P(S \in \mathcal{H})$. For greater detail, including an introduction to the hit-or-miss topology, see Ref. [26].

A.2. Note on generalized functions

The term “generalized function” (or “distribution”) is often used generically, there being various classes of generalized functions and corresponding test functions. Here we present some basic notions. Let $D$ be an open subset of $\mathbb{R}^n$ and $\mathcal{D}$ be the set of infinitely differentiable functions having compact support in $D$. $\mathcal{D}$ is called the space of test functions, a generalized function is a linear functional on $\mathcal{D}$ that is continuous with respect to a certain topology on $\mathcal{D}$ [27], we denote by $\mathcal{D}'$ the space of generalized functions on $\mathcal{D}$, and we write the operation of a generalized function $\alpha$ on a test function $\phi$ as $\langle \phi, \alpha \rangle$. Two generalized functions $\alpha_1$ and $\alpha_2$ are equal if $\langle \phi, \alpha_1 \rangle = \langle \phi, \alpha_2 \rangle$ for any $\phi \in \mathcal{D}$. If $f$ is a locally integrable function on $D$, then $f$ defines a generalized function by

$$\langle \phi, f \rangle = \int_D \phi(x)f(x)\,dx.$$  \hspace{1cm} (A.2)

If $F$ is a probability distribution function, then it defines a generalized function via the Lebesgue–Stieltjes integral

$$\langle \phi, F \rangle = \int_D \phi(x)\,dF(x).$$  \hspace{1cm} (A.3)

If $\alpha \in \mathcal{D}'$ and $f$ is an infinitely differentiable function, then a product generalized function, $fx$, is defined by $\langle \phi, fx \rangle = \langle \phi, \alpha \rangle f$. If $\alpha \in \mathcal{D}'$ and $D^a$ denotes the differential operator

$$D^a = \left( \frac{\partial}{\partial x_1} \right)^{a_1} \left( \frac{\partial}{\partial x_2} \right)^{a_2} \cdots \left( \frac{\partial}{\partial x_n} \right)^{a_n} \hspace{1cm} \text{(A.4)}$$

then the partial derivative $D^a\alpha$ of $\alpha$ is defined as a generalized function by

$$\langle \phi, D^a\alpha \rangle = (-1)^{|a|} \langle D^a \phi, \alpha \rangle,$$  \hspace{1cm} (A.5)

where $|a| = a_1 + a_2 + \cdots + a_n$. In particular, the density for a probability distribution function $F$ is the generalized function $f$ defined by

$$\langle \phi, f \rangle = (-1)^n \left( \frac{\partial}{\partial x_1} \right)^{a_1} \left( \frac{\partial}{\partial x_2} \right)^{a_2} \cdots \left( \frac{\partial}{\partial x_n} \right)^{a_n} \langle \phi, F \rangle.$$  \hspace{1cm} (A.6)

The delta function is defined by $\langle \phi, \delta \rangle = \phi(0)$ and written in integral form as $\int_0^1 \phi(x)\,dx = \phi(0)$. The equivalent forms of $fx, y(x, y)$ in Eqs. (7) and (8) essentially involve the change of variables

$$\delta(y - g(x)) = \frac{\delta(x - g^{-1}(y))}{|g'(x)|}.$$  \hspace{1cm} (A.7)

We justify the change of variables by integration against an arbitrary test function in Eq. (9). Changing variables is not a trivial matter and various conditions exist relative to the classes of generalized and test functions involved. Similar changes of variables for multivariate transformations involve the Jacobian of the transformation.

References


About the Author—SINAN BATMAN holds a B.S. from the Middle East Technical University in Turkey, an M.S. from Rochester Institute of Technology, and a Ph.D. from Texas A & M University (1998), all in Electrical Engineering. His Ph.D. research was in the areas of nonlinear stochastic signal processing, pattern estimation theory and random image modelling. After completing his Ph.D., he remained in the Electrical Engineering Department at Texas A & M for another year as a postdoctoral research associate and lecturer. Since then he has been an associate research scientist in the Center for Imaging Science and the Electrical and Computer Engineering Department at The Johns Hopkins University. His current research efforts are focused in the areas of biomedical image processing and automated target detection.

About the Author—EDWARD DOUGHERTY holds an M.S. in Computer Science from Stevens Institute of Technology and a Ph.D. in Mathematics from Rutgers University. He is a professor in the Electrical Engineering Department at Texas A & M University. He is the author of 11 books and editor of four others, and has written numerous papers on nonlinear image and signal processing. He is currently editor of the SPIE/IS & T Journal of Electronic Imaging and of the SPIE/IEEE Series on Imaging Science and Engineering. His current research is in the areas of automatic design of nonlinear operators, pattern recognition, and genomic signal processing.