TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Research support for this report has been provided in part by the Aerospace Corp.

Random Sampling of Random Fields: Least Squares Estimation

by J. T. Gillis

TR 91-13

Random Sampling of Random Fields: Least Squares Estimation

J. T. Gillis*

February 7, 1991

*Currently at the Systems Research Center, University of Maryland, College Park, MD. on leave from the The Aerospace Corporation, Los Angeles, CA.

Abstract

The paper begins with a discussion of deterministic sampling, where it is observed that when one can reconstruct the covariance one can also reconstruct the sample path (in quadratic mean). Then the theorem of Shapiro and Silverman, which states that Poisson based sampling allows reconstruction of the covariance at any sampling rate and a construction of an estimator of the covariance (due to Papoulis) are presented. A class of estimators for random fields using Poisson (and Poisson like) sampling is developed. The optimal estimator (minimum mean square error) is shown to exist and the error is shown to go to zero only as the sampling rate goes to infinity; Poisson sampling behaves differently from regular sampling in this respect. Poisson sampling is shown to be the best (lowest error) for a wide class of multidimensional point processes (sampling measures). One feature of the development is that it applies directly in $\mathbb{R}^n$. It is shown that the optimal estimator has many desirable properties (continuity, etc.); however, recursion in terms of the density of the sampling processes is not easily developed. A sub-optimal estimator with this desirable property is also discussed. In the case that the random field is Gaussian, the proposed estimator is seen to be the conditional mean.

Keywords: Random Sampling, Estimation, Random Fields.

Contents

1 Introduction and Definitions

1.1 Introduction

This section is a review of the origins of random sampling. The problem of reconstruction of a random process has its roots in the deterministic interpolation theorem of Whittaker and Shannon for the reconstruction of a function from regularly spaced values of the function.

Let $\widehat{f}$ ($\check{f}$) to denote the Fourier transform[1] (inverse Fourier transform) of f.

Theorem 1 (Whittaker-Shannon [1]) *If f is a function with Fourier transform: $\widehat{f}(\lambda) \in L_2(-W, W)$. Then*

$$f(t) = \sum_{n \in \mathbb{N}} f(\frac{n}{2W}) \frac{\sin 2W\pi(t - \frac{n}{2W})}{2W\pi(t - \frac{n}{2W})}$$

The first extension of this theorem for random processes is:

Theorem 2 (Balakrishnan[1]) *Let $\{X_t\}_{t \in \mathbb{R}}$, be a second order, wide sense stationary stochastic process, with covariance $R(t) = E(X_{t+s} \cdot X_s)$ such that: $\widehat{R}(\lambda) \in L_2(-W, W)$. Then*

$$X_t = l.i.m \sum_{n \in \mathbb{N}} X(\frac{n}{2W}) \frac{\sin 2W\pi(t - \frac{n}{2W})}{2W\pi(t - \frac{n}{2W})}$$

where l.i.m stands for limit in the mean square, e.g. in the norm $(E\{(f^2)\})^{\frac{1}{2}}$.

These theorems state conditions under which a function (sample path) can be pointwise interpolated (in mean square) in an error free manner. In both theorems there are restrictions on the support of the Fourier transform of the signal (restricting the covariance restricts the support of the spectral process associated with the path). In this case if one can recover the covariance one can also recover the sample path. If one can recover the sample path and the processes is ergodic then one can recover the covariance of the process via the ergodic relationship.

The *bandwidth* of a function is the support of its Fourier transform. The *bandwidth* of a stochastic process is the support of the spectral density.

A theorem which attracts our attention is the following:

Theorem 3 (Shapiro & Silverman [2]) *If $\{X_t\}_{t \in \mathbb{R}}$ is a wide sense stationary, quadratic mean continuous stochastic process with covariance $R(.)$, then $\{R(\tau_j)\}_{j \in \mathbb{N}}$ determine $R(.)$ when the times $\{\tau_j\}_{j \in \mathbb{N}}$ are chosen as Poisson random variables, regardless of the the average rate of the process $\{\tau_j\}_{j \in \mathbb{N}}$ and without bandwidth limitations on $R(\cdot)$.*

[1] The Fourier Transform is taken as: $\widehat{f}(\lambda) = \int e^{i2\pi\lambda t}\, f(t)\, dt$

Such a result is remarkable, as both of the earlier theorems relate sampling rate, $\frac{1}{2W}$, to bandwidth, $(-W, W)$; hence one might expect that the intensity of the Poisson process (which is the average sampling rate) should be related to the bandwidth of the signal or covariance.

Much is known about the random sampling of random processes, e.g. asymptotically unbiased estimators and statistically efficient algorithms for spectral reconstruction. A large portion of what is known relies on the fact that $\mathbb{R}$, the parameter space for a stochastic process, has a natural ordering — which is not generally true for a random field. Additionally, many of the early theorems (this is true of the Shapiro and Silverman theorem) show that the *probability distribution of sampled points* determine the covariance (or distribution) of the underlying processes. This points out that there are several types of sampling theorems possible in stochastic sampling. The first division is based on what is determined: the second order properties, the law, the path; the second division is what determines the the properties: the second order properties of the samples, the law of the samples or the realization (path) of the samples. This gives nine possible types for theorems — of these, five types of sampling results are found in the literature. This is discussed in [3].

This paper will show that although Poisson sampling is alias free (in several senses) for the reconstruction of the covariance of a stochastic process, it is not alias free for the reconstruction of sample paths of random fields in quadratic mean. We shall show this by construction explicitly the minimum variance estimator of the field given (generalized) Poisson samples. In addition, it is shown that Poisson sampling is the best possible sampling measure in a large class. The discussion includes stochastic processes as a special case.

1.2 A Short Introduction to Random Fields

This introduction is included to establish the terminology used in this discussion, since there is not full agreement on nomenclature. Briefly, a random field is a family of random variables which are parameterized (indexed) in a general way. If the parameterization (index set) is an interval of the real line, the random field is identified with a stochastic process. Hence random fields can be shown to represent one generalization of stochastic processes.

One might take the ground wind speed as a random field, with the Borel subsets of $\mathbb{R}^2$ as the parameter sets. Similarly, the number of trees in a forest can be modeled as the selection of random points in the plane. Here the parameter space is the Borel set in $\mathbb{R}^2$ and the random variable takes on values in $\mathbb{N}^+$. Such a model is a special kind of random field called a multidimensional point process (m.p.p.).

In what follows, the parameter space (index set) is Euclidean space, usually $\mathbb{R}^2$; however definitions are framed in a general way. Relevant references for random fields are Rozanov [4], Yadrenko [5], Vanmarke [6], and Gel'fand [7]. Relevant references for m.p.p.'s are Rozanov [4], Gel'fand [7], Lewis [8], and Karr [9]. Discussions of spectral representation for random fields can be found in Yadrenko [5], and in Hannan [10].

Definition 4 *1. Let T be a locally compact space and (E, E, μ) be a measure space. A* filtration *on (E, E, μ) is a family of σ-subalgebras of E indexed by subsets of T, such that $\forall$ open $t,\ t' \subset T$, $t \subset t' \Rightarrow \mathsf{E}(t) \subset \mathsf{E}(t')$.*

2. Let (Ω, Σ, P) be a probability triple, T a locally compact space, T a distinguished ring of subsets of T. Let (E, E, μ) be a measure space. An E-valued random field *is a function X: $\Omega \times \mathsf{T} \to E$, such that $X(t,.)$ is Σ measurable $\forall t \in \mathsf{T}$ and the filtration generated by the semi-ring*

$$\begin{aligned} \Sigma(A) &= \{\omega : X(s', \omega) \in B \in \mathsf{E},\ s' \in A\} \\ & \quad \textit{is additive}: \\ \Sigma(A \cup A') &= \Sigma(A) \vee \Sigma(A')\ \forall\ A,\ A' \subset \mathsf{T} \\ & \quad (\textit{the smallest } \sigma - \textit{field containing } \Sigma(A) \textit{ and } \Sigma(A')) \end{aligned}$$

The triple (E, E, μ) is called the sample space, *T is called the* parameter space *(or the time space). By abuse of nomenclature T is occasionally referred to as the parameter space; this is especially true when T is the Borel sets of T.*

3. The function $X(\omega, .) : \mathsf{T} \to E$ is called the sample function, sample path *or* trajectory. *Additional notations used for $X(t, \omega)$ are $X_t(\omega)$, X_t; and X or $\{X_t\}$ for $\{X_t\}_{t \in T}$.*

4. The random field is homogeneous *if*

$$P_{t_1,\ t_2, \ldots\ , t_n}(\{\omega\ :\ X(t_i,\ \omega)\}\ \in\ B_i\ i = 1, 2 \ldots n)$$

is translation invariant in $t_1, \ \ldots, \ t_n \ \forall B_i \in \mathsf{E}$.

5. The random field is isotropic *if*

$$P_{t_1, t_2, \ldots, t_n}(\{\omega : X(t_i, \omega)\} \in B_i\ i = 1, 2 \ldots n)$$

is invariant in $t_1, \ \ldots, \ t_n$ under rotations $\forall B_i \in \mathsf{E}$. In the case of an ordinary stochastic process, isotropicity is equivalent to symmetry of the density function (if it exists).

1.2.1 Examples

1. Let $\{X_t\}$ be a stochastic process with probability triple (Ω, Σ, P). Then X is a random field with $T = I\!R$, T = Borel sets; (E, E, μ) is ($I\!R$, Lebesgue measurable sets, Lebesgue measure). The canonical identification is that t is identified with the interval $(-\infty, t]$. By definition $X(t, .)$ is Σ measurable and the filtration $\Sigma(t) \equiv \Sigma(X^{-1}(t', B), t' \in [-\infty, t), B \in \mathsf{E})$ is additive.

   Informally, $\Sigma(t)$ is what is known about ω by observing $X(t', \omega)\ \forall t' \in [-\infty, t)$. A processes X_t is said to be *adapted* to a filtration $\Sigma'(t)$ if it is $\Sigma'(t)$ measurable – then $\Sigma(t) \subset \Sigma'(t)$, and informally $\Sigma'(t)$ contains at least as much information about ω as $\Sigma(t)$ [11].

2. The canonical example is a two dimensional random field. Let T be $I\!R^2$, T = the Borel sets; let (E, E, μ) be ($I\!R^2$, Lebesgue measurable sets, Lebesgue measure). Then X(t,.) is a random variable $\forall t \in I\!R^2$. Here the associated filtration is: $\Sigma(A) \equiv \Sigma(X^{-1}(\tau, B),\ \tau \in A, B \in \mathsf{E})$ for Borel $A \subset I\!R^2$.

3. A most important example is that of a *multidimensional point process* (m.p.p.). Let T = $I\!R^n$ and T be the Borel measurable sets and (E, E, μ) be ($I\!N^+$, all subsets, μ = cardinality). Then $\forall B,\ B' \subset \mathsf{T}$:

   (a) $\eta(B)\ \in\ I\!N^+$

   (b) $\eta(B \cup B') = \eta(B) + \eta(B')\quad if\ B \cap B = \emptyset$

Definition 5 *Associate with a realization of a m.p.p, $N(\omega, .)$, a set of points in $I\!R^n$ in the following manner: say τ_i is* associated *with $N(A)$ if for any open system of neighborhoods of τ_i, $\{B_n\}$ with* $\lim_n B_n = \{\tau_i\}$ *has* $\lim_n N(B_n \cap A) \neq 0$. *If such a limit is either one or zero with probability one, then the process is said to be* simple. *The association of all such points in A will be written as $\{\tau_i \in A\}$.*

Assumption 6 *Unless otherwise stated, all point processes are assumed to be simple, and $N(B) < \infty\ if\ B$ is compact.*

Definition 7 *A simple m.p.p., N, is defined to be a Poisson m.p.p. with rate μ if:*

1.

$$E\{N(A)\} = \mu \int_A dt$$

2.

$$cov(N(A),\ N(B)) = \mu \int_{A \cap B} dt$$

The usual Poisson process is a special case of a Poisson m.p.p., the advantage is that the definition is not based on the increment properties of the process. Define the stochastic integral of a function, ϕ, on a set, A, with respect to an m.p.p., η, as:

$$\int_A \phi(t)\eta(dt) \equiv \sum_{\tau_i \in A} \phi(\tau_i(\omega)) \tag{1}$$

Associated with equation 1 is

$$\int_A \phi(t)\ \mu(dt) \equiv E_N\{\int_A \phi(t)\ \eta(dt)\}$$

where μ is known as the mean measure (if it exists), and

$$\begin{aligned}\int_A \int_B \phi(t,s)\ \sigma(ds \times dt) \quad &\equiv \quad E_N\{\int_A \int_B \phi(t,s)\ \eta(ds)\eta(dt)\} \\ &\quad - \int_A \int_B \phi(t,s)\ \mu(ds)\mu(ds)\end{aligned}$$

where σ is known as the covariance measure (if it exists). When ϕ is stochastic, the computation of:

$$E\{\int_A \phi(t,\omega)\eta(dt)\} = E\{\sum_{\tau_i \in A} \phi(\tau_i(\omega),\omega)\}$$

which is a generalization of Wald's equation [12], may not always be possible – as implicitly assumed here, independence will be crucial. The process formed in 1 is also known as a *compound* or *marked* process [8], [9].

Assumption 8 *Adopt the implicit convention that the sampling schemes are always independent of the basic random fields under consideration.*

1.2.2 Random Sampling

The following problem is discussed in Papoulis [13]. Since it forms the basis of what follows, the discussion is repeated. Proceeding heuristically — consider a Poisson process with parameter μ; think of samples as being taken at the jumps of the Poisson process[2]. Observe that as required for overlapping intervals $[s_1, t_1)$ and $[s_2, t_2)$:

$$E\{N([s_1,t_1))\} \quad = \quad \mu(t_1 - s_1) = \mu \int_{s_1}^{t_1} d\tau,$$

$$\begin{aligned}Cov\{N([s_1,t_1))N((s_2,t_2])\} \quad &= \quad \mu \int_{s_1 \vee s_2}^{t_1 \wedge t_2} dt \\ &= \quad \mu \int_{s_1}^{t_1} \int_{s_2}^{t_2} \delta(\tau_1 - \tau_2)\ d\tau_1 d\tau_2.\end{aligned}$$

[2]This is sometimes modeled as the Poisson impulse process: $Z((s,t]) = \sum_{s \leq \tau_i < t} \delta(t' - t_i)$. Then $Z(t)$ is the "generalized" derivative of $N([s,t)) = \sum_{s \leq \tau_i < t}$, which is the counting measure associated with the Poisson process.

These two statments can be expressed differentially as $E\{N(dt)\} = \mu dt$ and $E\{N(dt)N(ds)\} = \{\mu\delta(t-s) + \mu^2\}dtds$. We shall adopt this notation.

Lemma 9 (Papoulis[13], p. 337) *Let $\{X_t\}_{t\in\mathbb{R}}$ be a wide sense stationary stochastic process with continuous paths and covariance $R_X \in L_1$, spectral measure $S_X(\lambda)d\lambda$ with $S(\lambda) \in L_1$. Let*

$$\begin{aligned} Y_t^a(\lambda) &= \frac{1}{\mu}\int_{-a}^{a} X_t e^{i2\pi\lambda\cdot t} \ N(dt) \\ &\equiv \frac{1}{\mu}\sum_{t_j\in(-a,a)} X_{t_j} e^{i2\pi\lambda\cdot t_j} \end{aligned}$$

Then:

$$\lim_{a\to\infty} \frac{1}{2a} E\{|Y_t^a|^2\} = S_X(\lambda) + \frac{1}{\mu} R_X(0)$$

Note that one can always estimate $R(0) \approx \frac{1}{\mu}\int_{-a}^{a} X_\tau^2 N(d\tau)$, and this allows us to estimate the covariance of the processes. This argument adapts to $\mathbb{R}^n$ transparently; evidently recovery of the spectrum is independent of the average sampling rate, μ. This line of investigation has been extensively developed, in fact we have:

Theorem 10 (Karr[9]) *Let X be a random field continuous in probability, N be a Poisson process with absolutely continuous mean, both on $\mathbb{R}^n$ and independent. Then the law of the marked process $Y(dt) = X_t N(dt)$ determines the law of X.*

The proof of this theorem and discussion of estimation techniques for the mean and covariance of X based on observations Y can be found in Karr's book [9].

In the remainder of this discussion the focus will be on sample path recovery from the sample path of the samples. Specifically we shall construct estimators of the form $Y(dt) = \int_A X_t N(dt)$, which minimize the estimation error. We shall examine the properties of such estimators, such as the dependance of the estimator on the region A, and the sampling rate, μ.

2 The Estimation Problem

Let $N(d\tau)$ be an homogeneous multidimensional point process with:

$$\begin{aligned} E\{N(dt)\} &= \mu dt \\ E\{N(dt)N(ds)\} &= \{\mu\delta(t-s) + \mu^2\{c(t-s)+1\}\}dtds \end{aligned}$$

This is not the most general form of an m.p.p., which would take $c(t-s)dtds$ as a measure $C(d(t-s))d(t+s)$; however the absolutely continuous form is sufficiently general for our purpose. We shall assume that both c and $\hat{c}$ are bounded.

Let X be an homogeneous random field with:

$$\begin{aligned} E\{X_t\} &= 0 \\ E\{X_t X_s\} &= R(t-s) \end{aligned}$$

and $Z(A) = \int_A h(\tau)X_\tau N(d\tau)$ be the compound process formed from X and N. It has second order statistics:

$$\begin{aligned} E\{Z(A)\} &= \mu \int_A h(\tau)E\{X_\tau\}d\tau \ = \ 0 \\ E\{ \int_A Z(dt) \int_B Z(ds)\} &= \int_A \int_B h(t)h(s)R(t-s)\ E\{N(dt)N(ds)\} \\ &= \int_A \int_B h(t)h(s)R(t-s) \left\{\mu\delta(t-s) + \mu^2\{c(t-s)+1\} \right\} dtds \\ &= \mu R(0) \int_{A\cap B} h^2(t)dt \\ &\quad + \mu^2 \int_A \int_B h(t)h(s)R(t-s)\{c(t-s)+1\}dtds \end{aligned}$$

We shall use compound processes of this form as a class of estimators of X_t:

$$\begin{aligned} \check{X}_t &= \frac{1}{\mu} \int_A X_s\, h(t,s)\ N(ds) \\ &= \frac{1}{\mu} \sum_{s_i \in A} h(t,s_i) X_{s_i} \end{aligned}$$

The constant $\frac{1}{\mu}$ is a matter of future convenience. The set A will be treated as fixed Borel set, but otherwise unspecified (it may or may not be compact), as will the variable t. The sample path X is assumed to be continuous. In this paradigm it is easy to see that the sampled process X should be path continuous as should the function h. This makes the estimator and the sampled process path continuous. Conditions will be delineated in order to assure the continuity of h.

A natural direction to proceed is to minimize the estimation error — that is choosing h as to minimize the estimation error covariance.

$$
\begin{aligned}
J[h(.,.)] &= E\left\{\left(X_t - \frac{1}{\mu}\int_A X_s h(t,s)\ N(ds)\right)^2\right\} \\
&= R(0) - 2\int_A R(t-s)h(t,s)\ ds \\
&+ \int_A\int_A R(s'-s)(c(s'-s)+1)h(t,s')h(t,s)\ ds'ds \\
&+ \frac{R(0)}{\mu}\int_A h^2(t,s)\ ds
\end{aligned}
$$

An obvious requirement is to determine for which functions $J[h(.,.)] < \infty$. In order to determine such conditions the following lemma is needed.

Lemma 11 *The function c is positive definite.*

Proof.

$Cov(N(ds)N(dt))$ is a positive definite (p.d.) measure $\forall \mu > 0$ i.e.

$$\int\int f(t)f(s)\ cov(N(ds)N(dt)) \geq 0 \quad \forall f \in C_c$$

Hence

$$\{\mu\delta(t-s) + \mu^2 c(t-s)\}dtds\ is\ p.d.$$

This implies:

$$\mu\int\int f(t)f(s)\ c(t-s)dtds \geq -\int f^2(t)\ dt$$

Assume that c is not p.d. then $\exists f_0$ s.t.

$$\int\int f_0(t)f_0(s)c(t-s)\ dtds = -\epsilon < 0$$

Then $-\mu\epsilon \geq -|f_0|^2_{L_2}$ or $\mu\epsilon \leq |f_0|^2_{L_2}$ $\forall \mu \in I\!R^+$. Contradiction. ■

This does not imply that c is continuous; it does imply that c is the Fourier transform of a tempered distribution [7]. However c will be taken as bounded and continuous in the subsequent discussion. Since R and $R(c+1)$ are bounded, $h(t,.) \in L_1(A) \cap L_2(A) \Rightarrow J[h] < \infty$. Hence functions in C_c are allowable e.g. $h(t,.) \in C_c \Rightarrow J[h] < \infty$. This allows use of the techniques of calculus of variations in the determination of the minimum of the functional.

2.1 Calculation of the Frechet Derivative

The minimum of the cost functional $J[.]$ is determined by the calculus of variations. Specifically, necessary conditions for the minimization of the cost functional can be derived from: $(\frac{d}{dt}J[h +$

$\epsilon g])|_{\epsilon=0} = 0 \;\; \forall$ admissible g. The perturbation functions are taken as $g(t,.) \in C_c$ and the first variable is held fixed for this argument.

$$\frac{d}{d\epsilon}J[h+\epsilon g]|_{\epsilon=0} = \frac{d}{d\epsilon}E\left\{\left(X_t - \frac{1}{\mu}\int_A X_s(h(t,s)+\epsilon g(t,s))\; N(ds)\right)^2\right\}|_{\epsilon=0}$$

$$\begin{aligned}
&= \frac{d}{d\epsilon}\{R(0) - 2\int_A R(t-s)(h(t,s)+\epsilon g(t,s))\; ds \\
&\quad + \frac{1}{\mu}\int_A\int_A R(s'-s)(h(t,s')+\epsilon g(t,s'))(h(t,s)+\epsilon g(t,s))\delta(s-s')\; ds'ds \\
&\quad + \int_A\int_A R(s'-s)(c(s'-s)+1)(h(t,s')+\epsilon g(t,s')) \\
&\qquad\qquad (h(t,s)+\epsilon g(t,s))\; ds'ds\}|_{\epsilon=0}
\end{aligned}$$

$$\begin{aligned}
&= \{-2\int_A R(t-s)g(t,s))\; ds + 2\frac{1}{\mu}\int_A R(0)(h(t,s)+\epsilon g^2(t,s))\; ds \\
&\quad + \int_A\int_A R(s'-s)(c(s'-s)+1) \\
&\qquad \cdot\{g(t,s')h(t,s') + h(t,s')g(t,s) + \epsilon g(t,s)g(t,s')\}\; ds'ds\}|_{\epsilon=0}
\end{aligned} \tag{2}$$

$$\begin{aligned}
&= -2\int_A R(t-s)g(t,s)\; ds + 2\frac{R(0)}{\mu}\int_A h(t,s)g(t,s)\; ds \\
&\quad + \int_A\int_A R(s'-s)(c(s'-s)+1) && (3)\\
&\quad (g(t,s')h(t,s) + h(t,s')g(t,s))\; ds'ds && (4)
\end{aligned}$$

The necessary condition for the minimization of the functional is that $\forall\; g(t,.) \in C_c$:

$$\begin{aligned}
0 &= -2\int_A R(t-s)g(t,s)\; ds + 2\frac{R(0)}{\mu}\int_A h(t,s)g(t,s)\; ds \\
&\quad + \int_A\int_A R(s'-s)(c(s'-s)+1)(g(t,s')h(t,s) + h(t,s')g(t,s))\; ds'ds \\
&= -2\int_A \\
&\quad \left\{R(t-s) - \frac{R(0)}{\mu}h(t,s) - \int_A R(s'-s)(c(s'-s)+1)h(t,s')\; ds'\right\} \\
&\qquad g(t,s)ds
\end{aligned} \tag{5}$$

since:

$$\begin{aligned}
&\int_A\int_A R(s'-s)(c(s'-s)+1)(g(t,s')h(t,s) + h(t,s')g(t,s))\; ds'ds \\
&\quad = \int_A\int_A R(s'-s)(c(s'-s)+1)h(t,s')g(t,s)\; ds'ds
\end{aligned}$$

$$\begin{aligned} &+ \int_A \int_A R(s-s')(c(s-s')+1)g(t,s)h(t,s')\, ds'ds \\ &= 2\int_A \int_A R(s'-s)(c(s'-s)+1)h(t,s')g(t,s)\, ds'ds \end{aligned}$$

By invoking the fundamental lemma of the calculus of variations [14] in equation 5, the minimization condition becomes:

$$\begin{aligned} 0 &= R(t-s) - \frac{R(0)}{\mu}h(t,s) - && (6) \\ & \int_A R(s'-s)(c(s'-s)+1)h(t,s')\, ds' && (7) \\ & \forall s \in A \end{aligned}$$

This is the Wiener-Hopf equation. The cost associated with using the function h^o which satisfies equation 7 can be calculated (the uniqueness of h^o will be shown):

$$\begin{aligned} J[h^o] &= E\{(X_t - \frac{1}{\mu}\int_A X_s h^o(t,s)\, N(ds))^2\} \\ &= R(0) - 2\int_A R(t-s)h^o(t,s)\, ds && (8) \\ &\quad +\frac{1}{\mu}\int_A\int_A R(s-s')h^o(t,s)h^o(t,s')\, E\{N(ds)N(ds')\} \\ &= R(0) - 2\int_A R(t-s)h^o(t,s)\, ds + \frac{R(0)}{\mu}\int_A (h^o(t,s))^2\, ds \\ &\quad + \int_A\int_A R(s-s')h^o(t,s')h^o(t,s)(c(s-s')+1)\, dsds' \\ &= R(0) + \mu\int_A\{-R(t-s) + \frac{R(0)}{\mu}h^o(t,s) \\ &\quad + \int_A R(s-s')h^o(t,s')(c(s-s')+1)\, ds'\}h(t,s)ds \\ &\quad - \int_A R(t-s)h^o(t,s)\, ds \\ &= R(0) - \int_A R(t-s)h^o(t,s)\, ds && (9) \end{aligned}$$

The second variation can be calculated from equation 2 as:

$$\begin{aligned} \delta^2 J[g,g] &= 4\frac{R(0)}{\mu}\int_A g^2(t,s)dt \\ &+ 2\int_A\int_A R(s'-s)(c(s'-s)+1)g(t,s)g(t,s')\, dsds' \\ &\geq 4\frac{R(0)}{\mu}|g|^2_{L_2} \end{aligned}$$

Or

$$\geq 2\frac{R(0)}{\mu}(c(0)+1)|g|^2_{L_1}$$

This equation meets the sufficiency condition for the minimization of the cost functional in the spaces L_2 and L_1 [14]. Thus the critical point is a minimum of the functional in these spaces.

Let $k(t-s) = R(t-s)(c(t-s)+1)$, and

$$K[f](s) = \int_A k(s'-s)f(s')\,ds',$$

then equation 7 is written as:

$$R(t-s) = (\frac{R(0)}{\mu}I + K)h(t,s) \;\; \forall s \in A$$

There is a well developed literature on equations of this type [15], using operator factorization methods. In the examination of Equation 7, four questions are of interest:

(i) What is the behavior of the equation as $\mu \to \infty$?

(ii) If A is compact (with non-empty interior) — what is the behavior as the region grows, $A \to I\!R^n$?

(iii) How does the solution depend on t? Is it continuous or a difference kernel?

(iv) What is the value of $J[h^o]$? Does path-wise aliasing occur?

2.2 Solutions for A Compact

If A is compact, then equation 7 can be taken as being on $L_2(A)$ (i.e. $R|_A \in L_2(A)$). In this case the equation has a unique solution for any function R under consideration.

We need the following fact:

Lemma 12 *K is a nonnegative definite linear functional on $L_2(A)$.*

Proof.

Recall that R is a covariance kernel and therefore it is nonnegative definite.

$$\begin{aligned} < K[f],\, f > &= \int_A\int_A f(t)f(s)k(t-s)\,dtds \\ &\geq \int_A\int_A R(t-s)f(t)f(s)\,dtds \\ &\geq 0 \quad \forall\, f \in L_2(A) \end{aligned}$$

■

Theorem 13 *If A is a compact then the Wiener-Hopf equation has a unique solution, $h^o \in L_2(A)$, for all continuous positive definite R and c. Additionally, h^o is continuous in $s \in A$ and $t \in \mathbb{R}^n$.*

Proof.

Since R, $R(c+1)$ are bounded and A is compact $\frac{1}{\mu}R(t-.)$, $R(c+1) \in L_2(A)$. Then $(\frac{R(0)}{\mu}+K)$ is a bounded linear operator on $L_2(A)$. Since K is non-negative definite, $-K$ is non-positive definite and the spectrum of $-K$ is negative, thus $\frac{R(0)}{\mu}$ is in the resolvent set and $(\frac{R(0)}{\mu}I-(-K))$ is invertible on $L_2(A)$.

Rewriting the Wiener-Hopf equation:

$$\begin{aligned} h(t,s) &= \frac{\mu}{R(0)}R(t-s)+\frac{\mu}{R(0)}\int_A R(s'-s)(c(s'-s)+1)h(t,s')\,ds' \\ &\qquad \forall s \in A \end{aligned}$$

so that $h(t,\cdot)$ is the sum of a continuous function and the convolution of two L_2 functions – hence $h(t,s)$ is continuous in $\forall s \in A$. Thus far it has been established that the operator $(\frac{R(0)}{\mu}I+K)^{-1}$ acts on a continuous L_2 function to yield a continuous function. A similar argument shows that it maps bounded functions into bounded functions. This will be used to show that h is continuous in t as well. Let $|t-t'|$ be chosen s.t. $|R(t-s)-R(t'-s)|<\epsilon$. Then:

$$\begin{aligned} \left|h(t,s)-h(t',s)\right. &= \left|(\frac{R(0)}{\mu}I+K)^{-1}(R(t-.)-R(t'-.))\right|\Big| \\ &\leq \left|(\frac{R(0)}{\mu}I-K)^{-1}\epsilon\right| = \epsilon\left|(\frac{R(0)}{\mu}I+K)^{-1}1\right| \\ &= \epsilon c_0 \end{aligned}$$

The term $\left|(\frac{R(0)}{\mu}I+K)^{-1}1\right|$ is bounded $\forall s \in A$ because 1_A is a bounded function. From this it is easy to see that h is continuous $\forall t \in \mathbb{R}^n$. ∎

2.3 Solution for $R \in L_2(\mathbb{R}^n)$

Henceforth we shall consider explicitly the case $R \in L_2$. The additional restriction of $R \in L_1$ has several other implications – it is a sufficient condition for the process, X_t, to be ergodic (the necessary and sufficient condition is that the spectral distribution function be continuous [16]). The Wiener-Hopf equation in $\mathbb{R}^n$ is

$$\begin{aligned} h(t,s) &= \frac{\mu}{R(0)}R(t-s) \\ &\quad +\frac{\mu}{R(0)}\int_{\mathbb{R}^n} R(s'-s)(c(s'-s)+1)h(t,s')\,ds' \\ &\qquad \forall s \in \mathbb{R}^n \end{aligned} \tag{10}$$

The candidate solution is found by taking the Fourier Transform of both sides:

$$\hat{h}(t,\lambda) = \left(\frac{R(0)}{\mu} + \hat{k}(\lambda)\right)^{-1} e^{-i2\pi t}\hat{R}(\lambda). \tag{11}$$

The multiplier on the right-hand-side of equation 11 is bounded above by $\frac{\mu}{R(0)}$ and below by zero. This operator is a bounded linear operator on $L_2(\mathbb{R}^n)$. The multiplier is zero only at points where $\widehat{R}$ is unbounded. Hence the candidate h is in $L_2(\mathbb{R}^n)$ and it does satisfy the equation.

One easy lemma concerning the nature of convolution operators will prove useful in showing that the function h^o is continuous.

Lemma 14 *Let $k,\ h \in L_2(\mathbb{R}^n)$ then $f = k * h$ is a continuous function in $L_1(\mathbb{R}^n)$ If $\hat{k}$ is bounded then $f \in L_2$. If k and h are positive (i.e. greater that or equal to zero a.e.) so is f.*

Proof.

Since $h,\ k \in L_2 \Rightarrow \hat{k}\hat{h} \in L_1 \Rightarrow k * h \in C$, by the Riemann–Lebesgue lemma. If $\hat{k}$ is bounded then clearly $||f||_{L_2} < \max(\hat{k})||h||_{L_2}$. Positivity is obvious. ∎

Hence the function h exists in $L_2(\mathbb{R}^n)$ and is a difference kernel (from the form of the dependence on t). Note that based on the equation 10, h is the difference of two continuous functions, hence it is continuous. We have proved the main theorem for $R \in L_2(\mathbb{R}^n)$, which is:

Theorem 15 *If $R \in L_2(\mathbb{R})$ then the Wiener-Hopf equation has a unique solution, $h^o \in L_2$, which is a continuous difference kernel in s and t.*

2.4 The limiting case: $A \to I\!R^n$

The arguments presented demostrate that the solution exists for A compact and for $I\!R^n$ with $R \in L_2(I\!R^n)$. Nothing is said about the relationship between the solutions (as $A \to I\!R^n$) or the structure of the problem for $R \notin \mathrm{L}_2(I\!R^)$. Both of these questions can be addressed using classical Factorization techniques.

For sufficiently regular sets (half–spaces, etc.), Wiener–Hopf factorization methods can be used to give results. A variant of the basic method for the analysis of positive operators is described in Shinbrot ([17] and [18]); the notation of the latter is used here. This analysis is applicable to unbounded positive operators, K is a bounded operator if and only if $\hat{R} * \hat{c} + \hat{R}$ is essentially bounded (see Balakrishnan [19], p.90).

The following is a brief description (without proofs) of the ideas for bounded strongly positive operators.

Let P be an orthogonal projector in $H = L_2(I\!R^n)$. Let L have positive range[3] and be selfadjoint on H. Consider the Hilbert spaces $H_{\frac{1}{2}} \subset H$ which is the space of functions for which $< Lf,\ f > < \infty$ and $H \subset H_{-\frac{1}{2}} = Dual(H_{\frac{1}{2}})$. Then $\tilde{L}$ has an extension $L :\ H_{\frac{1}{2}} \to H_{-\frac{1}{2}}$ and the extension is isometric and onto. Denote the (closed, orthogonal) extension of P to $H_{\pm\frac{1}{2}}$ by $P^{\pm}$. Let $L_{|_{Ra(P^{\pm})}}$ denote the restriction of L to the range of $P^{\pm} \subset H_{\pm\frac{1}{2}}$.

The case of interest is for P given by $Pf(t) = 1_A(t)f(t)$ and $\tilde{L} = \frac{R(0)}{\mu} + K$, where $A = I\!R^n$. This operator has strictly positive numerical range, $W(\tilde{L}) \subset (\frac{R(0)}{\mu}, \infty]$.

A Wiener-Hopf equation is an equation of the form:

$$P^- L_{|_{Ra(P^+)}} h = P^- g \tag{12}$$

which is exactly the form of equation 7, with $g(\cdot) = R(t - \cdot)$. For positive operators the proof is constructive, and the following theorem is from [18]:

Theorem 16 *Let L be a strongly positive operator, and P be an orthogonal projection. Then the Wiener–Hopf operator $P^- L_{Ra(P+)}$ is one–to–one and onto $Ra(P^-)$. Let $\{u_n\}$ be any orthonormal sequence in $H_{\frac{1}{2}}$, total in $Ra(P^+)$. Then the solution to 12 is given by:*

$$h = \sum (P^- g, u_n) u_n \tag{13}$$

Since $0 \geq \frac{\hat{R}}{1+\hat{R}} \leq 1$, then $\int \frac{\hat{R}^2}{1+\hat{R}} d\lambda < \infty$, if $R(0) < 0$, e.g. $R \in H$ and $R(0) < \infty$ imply $R \in H_{\frac{1}{2}}$. This is about as weak a condition as one might expect to find. Notice that nothing precludes $R \notin H$.

[3]The numerical range is defined as : $W(L) = \left\{ \frac{<Lf,\ f>}{||f||^2} : f \in Do(L) \right\}$.

If $Do(L_{\frac{1}{2}}P)$ and $Do(L_{\frac{-1}{2}}(I-P))$ are dense in H, then the solution is given by:

$$h = L_{\frac{1}{2}}^{-1} P L_{\frac{-1}{2}}^{-1} P^{-} R \tag{14}$$

In the case that A is regular and $R \in H$, $L_2(\mathbb{R}^n) \cap L_1(\mathbb{R}^n)$ is in both domains and $L_2(\mathbb{R}^n) \cap L_1(\mathbb{R}^n) \subset H_{\frac{1}{2}} \subset H \subset H_{-\frac{1}{2}}$, so the domains are dense and and all of the operators on the right hand side are bounded. Equation 14 is the main tool for our discussion of the behavior of dependence of equation 7 on the region A.

$$\begin{aligned}
&||L_{\frac{1}{2}}^{-1} P L_{-\frac{1}{2}}^{-1} P^{-} R - L_{\frac{1}{2}} - 1 L_{-\frac{1}{2}}^{-1} R||^2 \leq ||L_{\frac{1}{2}}^{-1}||^2 ||P L_{-\frac{1}{2}}^{-1} P^{-} R - L_{-\frac{1}{2}}^{-1} R||^2 \\
&= \; ||L_{\frac{1}{2}}^{-1}||^2 ||P L_{-\frac{1}{2}}^{-1} P^{-} R - P L_{-\frac{1}{2}}^{-1} R + P L_{-\frac{1}{2}}^{-1} R - L_{-\frac{1}{2}}^{-1} R||^2 \\
&\leq \; ||L_{\frac{1}{2}}^{-1}||^2 \{ ||P L_{-\frac{1}{2}}^{-1} P^{-} R - P L_{-\frac{1}{2}}^{-1} R||^2 + ||P L_{-\frac{1}{2}}^{-1} R - L_{-\frac{1}{2}}^{-1} R||^2 \} \\
&\leq \; ||L_{\frac{1}{2}}^{-1}||^2 \{ ||P L_{-\frac{1}{2}}^{-1}||^2 ||P^{-} R - R||^2 + ||P L_{-\frac{1}{2}}^{-1} R - L_{-\frac{1}{2}}^{-1} R||^2 \} \\
&\leq \; ||L_{\frac{1}{2}}^{-1}||^2 \{ ||P L_{-\frac{1}{2}}^{-1}||^2 ||(I - P^{-}) R||^2 + ||(I - P) L_{-\frac{1}{2}}^{-1} R||^2 \}
\end{aligned}$$

If $P = P_A \xrightarrow{\mathrm{W}} I$, and the extension converges ($P^{-} \xrightarrow{\mathrm{W}} I$), then both of the terms above vanish. Therefore the solution for A approaches the solution for $\mathbb{R}^n$ as "$A \to \mathbb{R}^n$."

2.5 The Error

Recall that the estimation error can be written as (equation 9):

$$\begin{aligned} J[h^o](t) &= R(0) - \int_{\mathbb{R}^n} R(t-s)h^o(t-s)\, ds \\ &= R(0) - (R * h^o)(0) \\ &= (R(t) - (R * h^o)(t))\,|_{t=0} \;\equiv\; j(t)|_{t=0}. \end{aligned}$$

The Fourier transform of $j(t)$ is:

$$\begin{aligned} \widehat{j}(\lambda) &= \widehat{R}(\lambda) - \widehat{R}(\lambda)h^o(\lambda) \\ &= \widehat{R}(\lambda)(1 - \widehat{h}^o)(\lambda) \\ &= \widehat{R}(\lambda)(1 - \frac{\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{k}(\lambda)}) \\ j_\mu(0) &= \int \widehat{R}(\lambda)(1 - \frac{\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{k}(\lambda)})\, d\lambda \end{aligned}$$

The expression

$$\frac{\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{k}(\lambda)},$$

is always positive and strictly less than 1 on the support of $\widehat{R}$. So for $R \in L_1 \cap \mathrm{L}_2(\mathbb{R}^n$, the error is finite and nonzero. Note that $\widehat{J}_\mu$ goes to zero as $\mu \to \infty$. Further since $\hat{R}(\lambda) \leq \hat{R}(\lambda) + (\hat{R} * \hat{c})(\lambda)\ \forall\ \lambda$, "pure" Poisson sampling ($c \equiv 0$) is the best sampling possible. Thus we have:

Theorem 17 *Poisson sampling is asymptotically alias free. Aliasing occurs for all finite values of* μ.

This theorem and theorem 3 stand in marked contrast to theorem 1 and theorem 2, in the latter case spectral–alias free implies pathwise–alias free in this case it does not.

2.6 Examples

This section presents several solutions to the Wiener-Hopf equation using equation 2.6. Two of the examples are taken from optical systems, and two are taken from turbulence models. Several other examples, some of which have easy analytic solutions are given.

The general computation is based on:

$$\widehat{h}(t,\lambda) = \frac{e^{-2\pi i t\cdot\lambda}\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{k}(\lambda)}$$

where $\widehat{k}(\lambda) = \widehat{R}(\lambda) * (\widehat{c}(\lambda) + \delta(\lambda))$. Since the nature of the t dependence is obvious, the equations will be examined at $t = 0$ for simplicity. We take $c \equiv 0$ in all of the examples.

Let $\widehat{R} = \frac{P}{Q}$ is a rational function. Then the equation for h is:

$$\widehat{h}(\lambda) = \frac{\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{R}(\lambda)} = \frac{\widehat{P}(\lambda)}{\frac{R(0)}{\mu} Q(\lambda) + P(\lambda)}$$

Another example is $R(t) = R(0)e^{-a|t|}$ for $a > 0$, $t \in I\!R^1$. This is the covariance of the Ornstein–Uhlenbeck process [16]. With the appropriate choice of a it represents the Dryden spectral density for turbulence [20]. Then $\widehat{R}(\lambda) = \frac{2a}{a^2 + 2\pi\lambda^2}$. Hence h is given by:

$$\widehat{h}(\lambda) = \frac{\mu a}{\sqrt{a^2 + 2a\mu}} \frac{\sqrt{a^2 + 2a\mu}}{(a^2 + \mu) + 2\pi\lambda^2}$$

so that:

$$h(t) = \frac{\mu a}{\sqrt{a^2 + 2a\mu}} \cdot e^{-\sqrt{a^2+2a\mu}|t|}$$

As $\mu \to \infty$, $h(t) \to \frac{\mu a}{\sqrt{a^2+2a\mu}} \cdot e^{-\sqrt{a^2+2a\mu}|t|}$ which is a "delta convergent" sequence [7], i.e. $h \to \delta(t)$ in the sense of distributions. The cost (error covariance) of the last example can be computed (and it is nonzero at the origin).

$$\begin{aligned}
J(t) &= R(t) - \int R(t-s)h^o(t-s)\, ds \\
J(0) &= R(0) - \frac{R(0)\mu a}{\sqrt{a^2 + 2a\mu}} \int e^{-a|t|} \cdot e^{-\sqrt{a^2+2a\mu}|t|}\, dt \\
&= R(0) - \frac{R(0)\mu a}{\sqrt{a^2 + 2a\mu} \cdot (a + \sqrt{a^2 + 2a\mu})} \\
&= R(0)\left(1 - \frac{2a\mu}{2a\mu + a^2 + a\sqrt{a^2 + 2a\mu}}\right)
\end{aligned}$$

Here it can be seen that the second term $\to 1$ as $\mu \to \infty$, hence Poisson sampling is pathwise alias free only asymptotically, as proved earlier.

An example which is directly computable in $I\!R^N$ is:

$$R(t) = \prod_{i=1}^{N} \frac{sin(2W^i\pi t^i)}{2W^i\pi t^i}$$

$$\widehat{R}(\lambda) = \frac{1}{|2W|^N} \prod_{i=1}^{N} 1_{(-W^i, W^i)}(\lambda^i)$$

So that $\widehat{h}$ is computed as:

$$\begin{aligned}
\widehat{h} &= \frac{\widehat{R}(\lambda)}{\frac{R(0)}{\mu} + \widehat{R}(\lambda)} \\
&= \frac{\mu}{|2W|^N(\mu+1)} \prod_{i=1}^{N} 1_{(-W^i, W^i)}(\lambda^i),
\end{aligned}$$

hence:

$$h(t) = \frac{\mu}{\mu+1} \prod_{i=1}^{N} \frac{sin(2W^i \pi t^i)}{2W^i \pi t^i}$$

The error variance is: $J(0) = \frac{1}{|2W|^N \mu}$.

3 Remarks

3.1 Gaussian Random Fields

One special case that comes to mind is the case in which the process to be sampled is Gaussian (i.e. the $\{X_{t_1}, X_{t_2}\}$ are jointly Gaussian $\forall t_1,\ t_2$). In this case the conditional mean is an important estimator — it minimizes the mean square error over *all* estimators (not just linear estimators). The conditional mean is a (bounded) linear function in the data. If one assumes that any (bounded) linear functional of the data (or at least a dense subset) can be written in the form: $\int X_\tau h(\tau) N(d\tau)$ the the usual condition that the error be independent of the data becomes:

$$\underset{\forall g}{0 = E}\left\{\left(X_t - \int X_\tau h(t,\tau) N(d\tau)\right) \int X_\sigma g(t,\sigma) N(d\tau)\right\}$$

This equation leads precisely to equation 7, the Wiener—Hopf equation. The estimator under discussion is the conditional mean. The only question is: is the class of functions for which $E\{(\int X_\tau g(t,\tau) N(d\tau))^2\} < \infty$ sufficiently inclusive? Modifications of the arguments contained in Section 2.3 of M. H. A. Davis' book ([21]) show that this class is sufficiently rich.

Also the formalism can be used to derive Theorem 2. Consider the Wiener–Hopf equation:

$$\begin{aligned}
0 &= E\left\{\left(X_t - \sum_j X_{\frac{j}{2W}} h(t, \frac{j}{2W})\right) \sum_k X_{\frac{k}{2W}} g(t, \frac{k}{2W})\right\} \\
&= \sum_k \left\{R(t - \frac{k}{2W}) - \sum_j R(\frac{j}{2W} - \frac{k}{2W}) h(t, \frac{j}{2W})\right\} g(t, \frac{k}{2W}) \qquad (15)
\end{aligned}$$

An argument similar to the fundamental lemma in calculus of variations: shows that equation 15 is equivalent to requiring that $\forall k$:

$$\begin{aligned} R(t-\frac{k}{2W}) &= \sum_j R(\frac{j}{2W}-\frac{k}{2W})h(t,\frac{j}{2W}) \\ &= \sum_l R(\frac{l}{2W})h(t,\frac{l+k}{2W}) \end{aligned}$$

which, by the Whittaker–Shannon Theorem (Theorem 1), is true if $h(t,\frac{l+k}{2W}) = \frac{\sin 2W\pi(t-\frac{l+k}{2W})}{2W\pi(t-\frac{l+k}{2W})}$.

The optimal estimator is not of a form that lends itself to recursive estimation. A recursive estimator can be generated by repetitive samplings of the same sample path of the base X_t processes, at least for the case that the covariance is bandlimited. Notice that h^o enjoys the same support in the frequency domain as R, and if R is related to a Reproducing Kernel Hilbert Space[4], the limiting h^o is a reproducing kernel. Let h be any function such that $R(t) = \int R(\tau)h(t-\tau)d\tau$. Then one can easily compute the error using this function as the estimator kernel:

$$\begin{aligned} E\{(X_t-\check{X}_t)^2\} &= R(0) - \frac{1}{\mu}2E\{\int X_tX_\tau h(t-\tau)N(d\tau)\} \\ &\quad +\frac{1}{\mu^2}E\{\int\int X_\sigma X_\tau h(t-\tau)h(t-\sigma)\; N(d\tau)N(d\sigma)\} \\ &= R(0) - 2\int R(t-\tau)h(t-\tau)d\tau \\ &\quad +\frac{1}{\mu^2}\int\int R(\sigma-\tau)h(t-\tau)h(t-\sigma)\{\mu^2+\mu\delta(\tau-\sigma)\}d\tau d\sigma \\ &= -R(0) + \int\int R(\sigma-\tau)h(t-\tau)h(t-\sigma)d\tau d\sigma \\ &\quad +\frac{R(0)}{\mu}\int h(t-\tau)h(t-\tau)d\tau \\ &= \frac{R(0)}{\mu}||h||^2_{L_2} \end{aligned}$$

The advantage of this estimator is that it will work for a whole family of covariances – one only need know the maximum bandwidth of the covariance. It has the same rate of convergence as the optimal estimator!

[4]This occurs if R is bandlimited to a compact set.

References

[1] A. V. Balakrishnan, *Communication Theory.* New York: McGraw-Hill, 1968.

[2] H. S. Shapiro and R. A. Silverman, "Alias-free sampling of random noise," *J. SIAM*, vol. 18, 1960.

[3] J. T. Gillis, *Multidimensional Point Processes and Random Sampling.* PhD thesis, UCLA, 1988.

[4] Y. A. Rozanov, *Markov Random Fields.* Berlin: Springer-Verlag, 1982.

[5] M. I. Yadrenko, *Spectral Theory of Random Fields.* New York: Optimization Software, Inc., 1983.

[6] E. Vanmarcke, *Random Fields: Analysis and Synthesis.* Cambridge, Ma.: The MIT Press, 1983.

[7] I. M. Gel'fand and N. J. Vilenkin, *Generalized Functions, Vol. 4.* New York: Academic Press, 1964.

[8] P. A. W. Lewis, *Stochastic Point Processes: Statistical Analysis, Theory, and Applications.* New York: Wiley, 1972.

[9] A. F. Karr, *Point Processes and Their Statistical Inference.* New York: Dekker, 1986.

[10] E. J. Hannan, *Group Representations and Applied Probability.* London: Methuen, 1965.

[11] R. J. Elliot, *Stochastic Calculus and Applications.* New York: Springer-Verlag, 1982.

[12] K. L. Chung, *Probability Theory.* Reading, Ma.: Addison Wesley, 1974.

[13] A. Papoulis, *Probability, Random Variables and Stochastic Process, 2ed.* New York: McGraw–Hill, 1984.

[14] I. M. Gelfand and S. V. Fomin, *Calculus of Variations.* Englewood Cliffs, N.J.: Prentice–Hall, Inc., 1963.

[15] F.-O. Speck, *General Wiener-Hopf Factorization Methods.* Boston: Pitman, 1985.

[16] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems.* Berlin: Springer-Verlag, 1985.

[17] M. Shinbrot, "On singular integral operators," *Journal of Math. and Mechanics*, vol. 13, pp. 395–406, 1964.

[18] M. Shinbrot, "On the range of general Winer–Hopf operators," *Journal of Math. and Mechanics*, vol. 18, pp. 587–601, 1969.

[19] A. V. Balakrishnan, *Applied Functional Analysis, 2ed.* New York: Springer–Verlag, 1981.

[20] F. C. Tung, *Identification of Aircraft Parameters In Turbulence with Non-rational Spectral Density.* PhD thesis, UCLA, 1978.

[21] M. H. A. Davis, *Linear Estimation And Stochastic Estimation.* London: Halsted Press, 1979.