Edinburgh Research Explorer

Single channel nonstationary signal separation using linear time-varying filters

Single Channel Nonstationary Stochastic Signal Separation using Linear Time Varying Filters

James R. Hopgood* and Peter J. W. Rayner

*Corresponding author. e-mail: jrh1008@eng.cam.ac.uk

The authors are with the Signal Processing Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge. CB2 1PZ. UK. Tel: +44 1223 335562. Fax: +44 1223 332662. e-mail: jrh1008@eng.cam.ac.uk, pjwr@eng.cam.ac.uk

EDICS: 2-TIFR

Abstract

Separability of signal mixtures given only one mixture observation is defined as the identification of the accuracy to which the signals can be separated. The paper shows that when signals are separated using the generalised Wiener filter, the degree of separability can be deduced from the signal structure. To identify this structure, the processes are represented on an general spectral domain, and a sufficient solution to the Wiener filter is obtained. The filter is composed of a term independent of the signal values, corresponding to regions in the spectral domain where the desired signal components are not distorted by interfering noise components, and a term dependent on the signal correlations, corresponding to the region where components overlap. An example of determining perfect separability of modulated random signals is given with application in radar and speech processing.

Keywords

I. Introduction

This paper investigates the *separability* of signal mixtures given, at each time instance, only *one* observation of the mixture in the temporal domain. In this context, separability means identifying whether the mixture of two signals can be separated to a given degree of accuracy. The problem of actually separating the signals is referred to as *signal separation*. Formally:

Definition 1 (**Signal Separability**) Suppose a desired signal, $d(t)$,[1] is corrupted by an additive noise signal, $n(t)$, such that the observation, $x(t)$, of the desired signal is given by $x(t) = d(t) + n(t), \forall t \in \mathcal{T} \subset \mathbb{R}$. The problem of separability is to determine conditions on $d(t)$ and $n(t)$ such that an estimate of the desired signal, $\hat{d}(t), \forall t \in \mathcal{T}$, can be obtained, from $x(t)$, to a given degree of accuracy using a time-varying filter.

A. Structure of Paper

This paper presents a detailed exposition of separating nonstationary stochastic signals using linear time-varying (LTV), and is structured as follows. The introductory section and §II continue with an overview of the factors that affect the classes of signals which are separable by LTV filters. In §III, the generalised power spectrum (GPS) is introduced, and its relationships with

---

[1]Scalar stochastic processes are denoted by a roman font, while scalar deterministic processes are represented by normal type; *e.g.* $x(t)$ denotes a random variable. Bold symbols denote vectors of stochastic processes.
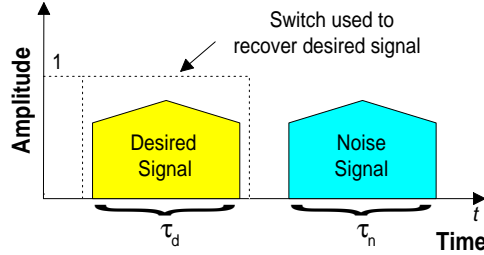
Fig. 1.   Separability of signals in the time domain.

the stochastic spectral representation of a random process is discussed, and in §IV, the *generalised bifrequency transfer function* is presented. Zadeh's deterministic *ideal filter* is reviewed in §V-A, and §V-B presents solutions (first published in [1, 2]) of the nonstationary Wiener-Hopf filter from a viewpoint that leads to a formal criterion for determining when two classes of signals are separable. Finally, in §V-D to §VII the separation of 'modulated' random signals is discussed with application in radar and speech processing.

B. Separation Techniques

A general separability criterion is difficult to derive since any such result would be a function of the separation method. To illustrate this, consider the separability of two signals that do not overlap in the time domain. These signals can be separated using a temporal switch – a special case of a LTV filter. Separability criteria for these signals depend on the times for which they have non-zero components; *i.e.* $\tau_d = \{t \in \mathbb{R} : \mathrm{d}(t) \neq 0\} \subset \mathbb{R}$ and $\tau_n = \{t \in \mathbb{R} : \mathrm{n}(t) \neq 0\} \subset \mathbb{R}$. To recover $\mathrm{d}(t)$, the impulse response of the LTV filter, as defined in §IV, is given by $h(t, \tau) = \delta(t - \tau)\mathbb{I}_{\tau_d}(t)$, where $\mathbb{I}_{\tau_d}(t) = 1$ if $t \in \tau_d$, and 0 otherwise. These signals cannot be separated using a linear time-invariant (LTI) filter and, therefore, in such a case, no separability criteria would exist and separation would appear to be impossible. Separability criteria are derived assuming that separation is achieved using LTV filters. However, even if signals are not separable under the criterion derived here, they are not necessarily inseparable: it merely implies they are not separable using deterministic LTV filters designed under the specific choice of *cost function* discussed below.

Signal separation with one observation sensor can only be achieved by exploiting prior knowledge of the signal structure, since the separation problem is inherently under-constrained. Moreover, it is desirable that any prior knowledge necessary for signal separation can be specified such
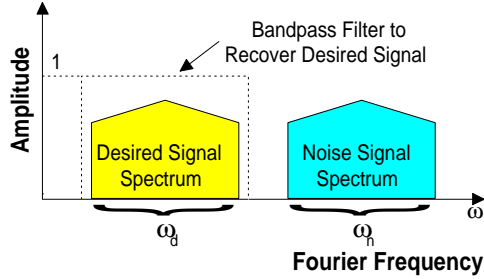
Fig. 2. Separability of signals in the Fourier domain.

that it is common to a 'class' of stochastic processes. This prior information must somehow be determined for a particular problem. So, for example, it is well known that stationary stochastic signals are 'perfectly' separable if their power spectra do not overlap in the Fourier domain. Separation can be achieved using a bandpass filter as shown in Figure 2, when the frequencies in the pass-band, $\omega_d = \{\omega \in \mathbb{R} : |\mathrm{D}(\omega)| > 0\} \subset \mathbb{R}$ and $\omega_n = \{\omega \in \mathbb{R} : |\mathrm{N}(\omega)| > 0\} \subset \mathbb{R}$, are known *a priori*, where $\mathrm{d}(t) \rightleftharpoons \mathrm{D}(\omega)$ and $\mathrm{n}(t) \rightleftharpoons \mathrm{N}(\omega)$. It is not, however, necessary to know the particular set of values of $\mathrm{D}(\omega)$ and $\mathrm{N}(\omega)$.

Since finite bandwidth non-overlapping signals in the Fourier domain can be separated using a LTI bandpass filter, and finite duration non-overlapping signals in the time-domain can be separated using a LTV filter (a switch), this suggests there might exist an arbitrary domain such that the representation of two classes of signals are disjoint and, therefore, separation can be achieved using a generalised bandpass filter on that domain. If such a domain does exist, then the question of how it can be determined from the data must be considered.

II. Signal Separation using LTV filters

Definition 2 (**Perfect Signal Separation**) 'Perfect' separation of deterministic signals is achieved when the estimate of the desired signal identically equals the desired signal for all desired time instances. 'Perfect' separation of stochastic signals is achieved when the mean squared error (MSE), $\sigma^2(t)$, between the estimate of the desired signal, $\hat{\mathrm{d}}(t)$, and the actual desired signal, $\mathrm{d}(t)$,

$$\sigma^2(t) \triangleq \mathbb{E}\left\{|\hat{\mathrm{d}}(t) - \mathrm{d}(t)|^2\right\} \tag{1}$$

is zero at each desired time instance; $\{\sigma^2(t) = 0; \forall t \in \mathcal{T}\}$. $\mathbb{E}\{\cdot\}$ denotes expectation.

As a result of using a second-order cost function, it proves necessary to use the second-order statistics of random signals. The autocorrelation function (ACF) describes a stochastic process

in the time domain and, at times $(t, \tau) \in T^2 = T \times T$, $T \subset \mathbb{R}$, is defined by:

$$R_{xx}(t, \tau) = \mathbb{E}\left\{ \mathrm{x}(t)\,\mathrm{x}^*(\tau) \right\}, \quad (t, \tau) \in T^2 \tag{2}$$

Estimation of the correlation functions of a nonstationary process is hindered by the fact that ensemble averages are extremely burdensome or impossible to obtain [3–5]. Separation is achieved by LTV filtration of $\mathrm{x}(t)$:

$$\hat{\mathrm{d}}(t) = \int_{\mathcal{T}} h(t, \alpha)\, \mathrm{x}(\alpha)\, d\alpha, \quad \forall t \in T \tag{3}$$

where the impulse response of the filter, $h(t, \tau)$, is the response at time t given an impulse occurred at its input at time τ. The cost function dictates that the filter is designed to minimise $\sigma^2(t)$, leading to the nonstationary Wiener-Hopf filter (WHF):

Theorem 1 (**Nonstationary Wiener-Hopf Filter**) The filter, $h(t, \tau)$, minimising the MSE between $\hat{\mathrm{d}}(t)$ and $\mathrm{d}(t)$ of (3) is called the Wiener-Hopf filter and is given, for all $(t, \tau) \in T \times \mathcal{T}$, by the solution of the convolution [6]:

$$R_{dx}(t, \tau) = \int_{\mathcal{T}} h(t, \alpha)\, R_{xx}(\alpha, \tau)\, d\alpha \tag{4a}$$

with the MSE given, for all $t \in T$, by:

$$\sigma^2(t) = R_{dd}(t, t) - \int_{\mathcal{T}} h(t, \alpha)\, R_{dx}(t, \alpha)\, d\alpha \tag{4b}$$

For 'perfect separation' $\{\sigma^2(t) = 0;\ \forall t \in T\}$ and, thus, the ACFs must satisfy:

$$R_{dd}(t, t) = \int_{\mathcal{T}} h(t, \alpha)\, R_{dx}(t, \alpha)\, d\alpha, \quad \forall t \in T \tag{4c}$$

III. Power Spectra for Nonstationary Stochastic Processes

The *power spectrum* of a stationary stochastic process is a powerful tool in the analysis of LTI systems; it would be useful to extend the definition of the power spectrum for stationary processes to the nonstationary case.

A. Notion of Power Spectra for Nonstationary Signals

The notion of *power spectra* for nonstationary signals has been investigated in great depth, beginning with the classic works of Gabor [7]. Cohen [8, 9], gives an excellent overview of time-frequency analysis for deterministic signals. More recently, Hammond and White [10] also give an excellent review discussing further time-frequency distributions including, for example,

Loèves' Harmonizable processes [11–15]. A comprehensive and far reaching contribution was made by Mark [16] in which the 'physical' spectrum was introduced; Loynes [17] and Turner [18], *inter alios*, discuss the concept of the power spectrum for nonstationary processes and, in general, the discussion centres around the concept of 'time-varying spectra'. Priestley [15] introduces the notion of the evolutionary spectrum (ES), and numerous researchers have discussed its theoretical framework [17, 19] and its generalisations [20–22]. Lampard [23] generalises the Wiener-Khintchine theorem to nonstationary processes, and that definition of the power spectrum is essentially equivalent to the Page distribution [8]. Cohen [24] reviews a general approach for obtaining joint signal representations using general linear decompositions which analyse signals in terms of physical quantities other than just time and frequency, for example, time and scale in wavelet transforms; in [25] Sayeed and Jones discuss the connection between current approaches to joint signal representations. Cohen [26] also discusses the generalisation of the stationary autocorrelation function and its power spectral density in terms of a general basis set, giving a corresponding generalisation of the Wiener-Khintchine theorem. Silverman [11] defines a *locally stationary process* through a separable correlation function, and generalises the Wiener-Khintchine theorem for such processes. A number of fractional-time-frequency (FTF) decompositions, such as the fractional-Fourier transform are discussed in Ozaktaz *et al.* [27]. Further investigations of correlation functions and power spectra for nonstationary processes can be found in [13, 18, 28–32] and references therein.

For many of these time-frequency distributions, much emphasis is placed on ensuring there is a meaningful physical interpretation of the power spectra. Each of these power spectral representations have the same motivation behind their development: to find another domain in which the statistical properties of a signal can be represented in a more 'useful' form, or to provide an environment in which a problem becomes easier to solve. The development of the generalised power spectrum (GPS) in the following sections is based on the theory of linear signal decomposition for representing a stochastic process on an arbitrary power spectral domain.

B. Spectra of Deterministic Processes

B.1 Continuous Time and Continuous Spectral Domains

The representation of a *continuous* time deterministic signal, $x(t)$, on an arbitrary *continuous* spectral domain, $\lambda \in \Lambda \subset \mathbb{C}$, is $X(\lambda)$, defined by the integral transform:

$$X(\lambda) = \int_T x(t) \, K(t, \lambda) \, dt \, , \ \forall \lambda \in \Lambda \tag{5a}$$

where Λ is the region in the signal space in which the representation $X(\lambda)$ lies, and the function $K(t, \lambda)$ is called the *direct transform basis kernel*. Conversely, $X(\lambda)$ may be represented on the time-domain $t \in T \subset \mathbb{R}$ as $x(t)$:

$$x(t) = \int_\Lambda X(\lambda) \, k(\lambda, t) \, d\lambda \, , \ \forall t \in T \tag{5b}$$

where $k(\lambda, t)$ is called the *inverse transform* or *reciprocal basis kernel* of $K(t, \lambda)$. The transformation is assumed isomorphic so, for a given $x(t)$, there exists a unique $X(\lambda)$ and, conversely, for a given $X(\lambda)$, there exists a unique $x(t)$. By extending the definition of a function to include the Dirac δ function, it can be shown that $k(\lambda, t)$ and $K(t, \lambda)$ must satisfy:

$$\delta(t - \tau) = \int_\Lambda K(t, \lambda) \, k(\lambda, \tau) \, d\lambda \, , \ \forall (t, \tau) \in T^2 \tag{6a}$$

$$\delta(\lambda - \hat{\lambda}) = \int_T k(\lambda, t) \, K(t, \hat{\lambda}) \, dt \, , \ \forall (\lambda, \hat{\lambda}) \in \Lambda^2 \tag{6b}$$

where $\Lambda^2 = \Lambda \times \Lambda$. Any pair of such functions satisfying these equations for $(t, \lambda) \in T \times \Lambda$ is called a transform pair over $T \times \Lambda$.

C. Stochastic Spectral Transforms

The spectral transform for deterministic signals may be extended to random signals, with a particular realisation of a stochastic process decomposed as the superposition or integral of basis functions with stochastic coefficients. The stochastic spectral representation of a *continuous* stochastic process, x(t), on an arbitrary *continuous* spectral domain, $\lambda \in \Lambda \subset \mathbb{C}$, is X($\lambda$), defined by:

$$\mathrm{X}(\lambda) = \int_T \mathrm{x}(t) \, K(t, \lambda) \, dt \, , \quad \forall \lambda \in \Lambda \tag{7a}$$

where this integral is interpreted in a mean square (MS) limit. Conversely, X(λ) may be represented on the time domain, $t \in T \subset \mathbb{R}$, as $\hat{\mathrm{x}}(t)$:

$$\hat{\mathrm{x}}(t) = \int_\Lambda \mathrm{X}(\lambda) \, k(\lambda, t) \, d\lambda \, , \quad \forall t \in T \tag{7b}$$

Theorem 2 (**MS Equality of Integral Transforms**) The representation $\hat{x}(t)$ equals $x(t)$ in the mean square (MS) sense:

$$\mathbb{E}\left\{|x(t) - \hat{x}(t)|^2\right\} = 0, \quad \forall t \in T \tag{8}$$

Proof: This is a generalisation of the proof given in [33, Chapter 12, pp. 414]. ∎

D. Generalised Power Spectrum

The spectral decomposition of (7a) implies the ACF of the random variable $X(\lambda)$, $\mathcal{P}_{\hat{x}\hat{x}}(\lambda, \hat{\lambda}) \triangleq \mathbb{E}\left\{X(\lambda)X^*(\hat{\lambda})\right\}$, may be written $\forall(\lambda, \hat{\lambda}) \in \Lambda^2$ as:

$$\mathcal{P}_{\hat{x}\hat{x}}(\lambda, \hat{\lambda}) = \iint_{T^2} R_{xx}(t, \tau)\, K(t, \lambda)\, K^*(\tau, \hat{\lambda})\, dt\, d\tau \tag{9}$$

The ACF of a nonstationary stochastic process is a well defined 2-D deterministic function and, therefore, as a natural extension to the concept of the stationary power spectrum, it may be expressed on an arbitrary spectral domain using a 2-D integral transform:

$$\mathcal{P}_{xx}(\lambda, \hat{\lambda}) = \iint_{T^2} R_{xx}(t, \tau)\, \hat{K}(t, \tau, \lambda, \hat{\lambda})\, dt\, d\tau\,, \ \ \forall(\lambda, \hat{\lambda}) \in \Lambda^2$$

Considering the form of the ACF of $X(\lambda)$ in (9), it is natural to assume $\hat{K}(t, \tau, \lambda, \hat{\lambda})$ admits a separable form: *i.e.* $\hat{K}(t, \tau, \lambda, \hat{\lambda}) \triangleq K(t, \lambda)\, K(\tau, \hat{\lambda})$, $(\lambda, \hat{\lambda}, t, \tau) \in \Lambda^2 \times T^2$.

Definition 3 (**Generalised Power Spectrum**) The generalised power spectrum (GPS) of the process $x(t)$ and its inverse relationship are defined as:

$$\mathcal{P}_{xx}(\lambda, \hat{\lambda}) = \iint_{T^2} R_{xx}(t, \tau)\, K(t, \lambda)\, K^*(\tau, \hat{\lambda})\, dt\, d\tau \tag{10a}$$

$$R_{xx}(t, \tau) = \iint_{\Lambda^2} \mathcal{P}_{xx}(\lambda, \hat{\lambda})\, k(\lambda, t)\, k^*(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{10b}$$

where $K(t, \lambda)$ and $k(\lambda, t)$ are related by (6). Some basic relations follow: 1) $\mathcal{P}_{xx}^*(\lambda, \hat{\lambda}) = \mathcal{P}_{xx}(\hat{\lambda}, \lambda)$, 2) $\mathcal{P}_{\hat{x}\hat{x}}(\lambda, \hat{\lambda}) = \mathcal{P}_{xx}(\lambda, \hat{\lambda})$, and 3) $R_{\hat{x}\hat{x}}(t, \tau) = R_{xx}(t, \tau)$.

Finally, suppose that a signal $y(t)$ admits a similar representation to (7), so that $\{y(t), t \in T\} \rightleftharpoons \{Y(\lambda), \lambda \in \Lambda_Y\}$, then $R_{\hat{y}\hat{x}}(t, \tau) = \mathbb{E}\{\hat{y}(t)\hat{x}^*(\tau)\}$ may be written as:

$$R_{\hat{y}\hat{x}}(t, \tau) = \iint_{\hat{\Lambda}_X \hat{\Lambda}_Y} \mathcal{P}_{\hat{y}\hat{x}}(\lambda, \hat{\lambda})\, k(\lambda, t)\, k^*(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{11}$$

$\forall (t, \tau) \in T^2$, where $\hat{\Lambda}_X \equiv \Lambda_X \oplus \Lambda_0$, $\hat{\Lambda}_Y \equiv \Lambda_Y \oplus \Lambda_0$ and $\mathcal{P}_{\hat{y}\hat{x}}(\lambda, \hat{\lambda}) = \mathbb{E}\left\{ Y(\lambda) X^*(\hat{\lambda}) \right\}$, and $\oplus$ is the orthogonal direct sum. Here, $\Lambda_X \subset \Lambda$ and $\Lambda_Y \subset \Lambda$ are the regions over which the spectral components of $\hat{x}(t)$ and $\hat{y}(t)$, respectively, do *not* overlap, and $\Lambda_0 \subset \Lambda$ is the region over which they *do*. The limits in (11) reduce to $\Lambda_0 \times \Lambda_0$ since $\mathcal{P}_{\hat{y}\hat{x}}(\lambda, \hat{\lambda}) = \mathbb{E}\left\{ Y(\lambda) X^*(\hat{\lambda}) \right\} = 0$ if $(\lambda, \hat{\lambda}) \in \Lambda_X \times \Lambda_Y$, *etc.*

Definition 4 (**Generalised Cross Power Spectrum**) The generalised cross power spectrum (GCPS) of the processes $y(t)$ and $x(t)$, $\forall (\lambda, \hat{\lambda}) \in \Lambda_0^2 = \Lambda_0 \times \Lambda_0$, $\Lambda_0 \subset \Lambda$, and its inverse relationship are defined $\forall (t, \tau) \in T^2$ as:

$$\mathcal{P}_{yx}(\lambda, \hat{\lambda}) = \iint\limits_{T^2} R_{yx}(t, \tau) K(t, \lambda) K^*(\tau, \hat{\lambda}) \, dt \, d\tau \tag{12a}$$

$$R_{yx}(t, \tau) = \iint\limits_{\Lambda_0^2} \mathcal{P}_{yx}(\lambda, \hat{\lambda}) k(\lambda, t) k^*(\hat{\lambda}, \tau) \, d\lambda \, d\hat{\lambda} \tag{12b}$$

IV. TRANSFER FUNCTIONS

In signal theory, the concept of a transfer function of a LTI system is very important and, thus, it is natural to extend the definition of a transfer function to LTV systems. The approach used here takes advantage of a generalisation of the *bifrequency transfer function* of a filter [34–37], which is of fundamental importance in the study of LTV systems. The generalised bifrequency transfer function (GBTF) is sometimes called the *kernel function* of the system [38]. An alternative definition of a transfer function, called the *general system function*, of which the well-known Zadeh's transform is a special case [34], is not pursued here, although direct relationships exist [2]. References to other work in the literature discussing generalised frequency response concepts for LTV systems can be found in, for example, [34, 35, 39–43].

The input–output relation for a LTV system may be specified by the well known time-domain *superposition integral* [41]:

$$y(t) = \int_T h(t, \tau) x(\tau) \, d\tau, \quad \forall t \in T \tag{13}$$

where $y(t)$ is the response of the linear system to the input $x(t)$, and $h(t, \tau)$, is the LTV impulse response at time t to an impulse occurring at the system input at time τ. The term *impulse response* is not uniquely defined in the LTV system literature and, in many publications, the impulse response is also given by $\hat{h}(t, \tau) = h(t, t - \tau)$. The form in (13) is used throughout this work since it corresponds to the usual definition for the kernel of a linear integral operator [21].

Let $X(\lambda)$ be the representation of the signal $x(t)$ on the domain λ. If the input signal, $x(t)$, is filtered by a linear operator, $\mathcal{L}$, then the output is given by $y(t) = \mathcal{L}\, x(t)$, and represented on the domain λ using equations (7). It is assumed that the initial state of the filters discussed throughout this paper, is rest (zero initial conditions).

Theorem 3 (**Transfer Function of a LTV System**) The impulse response of the system $\mathcal{L}$ may be written in the separable form:

$$h(t,\tau) = \iint_{\Lambda^2} H(\lambda,\hat{\lambda})\, k(\lambda,t)\, K(\tau,\hat{\lambda})\, d\lambda\, d\hat{\lambda} \tag{14a}$$

$\forall (t,\tau) \in T^2$, where $\{k(\lambda,t),\, K(t,\lambda),\, (t,\lambda) \in T \times \Lambda\}$ is a transform kernel pair given by (6), and $H(\lambda,\hat{\lambda})$ is the generalised bifrequency transfer function of $\mathcal{L}$, given by,

$$H(\lambda,\hat{\lambda}) = \iint_{T^2} h(t,\tau)\, K(t,\lambda)\, k(\hat{\lambda},\tau)\, dt\, d\tau \tag{14b}$$

$\forall (\lambda,\hat{\lambda}) \in \Lambda^2$.

Proof: The proof centres around the proof of (14a): given the definition of the GBTF, $H(\lambda,\hat{\lambda})$, (14b) follows naturally from the definition of the integral transforms (5) as shown below. Since $\mathcal{L}$ is a linear operator on functions of t only, then:

$$y(t) = \mathcal{L}\, x(t) = \int_{\Lambda} X(\lambda)\, \mathcal{L}\left\{k(\lambda,t)\right\}\, d\lambda \tag{15}$$

Substituting (7a) into (15) gives,

$$y(t) = \iint_{T\Lambda} x(\tau)\, \mathcal{L}\left\{k(\lambda,t)\right\}\, K(\tau,\lambda)\, d\lambda\, d\tau \tag{16}$$

Comparing with equation (13) the impulse response may be written as:

$$h(t,\tau) = \int_{\Lambda} \mathcal{L}\left\{k(\lambda,t)\right\}\, K(\tau,\lambda)\, d\lambda \tag{17}$$

Considering $\mathcal{L}\{k(\lambda,t)\}$ as a function of t, and treating λ as a parameter, it may be represented by the general transform (5b) with spectrum $H(\lambda,\hat{\lambda})$ defined by

$$\mathcal{L}\left\{k(\hat{\lambda},t)\right\} \triangleq \int_{\Lambda} H(\lambda,\hat{\lambda})\, k(\lambda,t)\, d\lambda \tag{18a}$$

Substituting (18a) into (17) gives (14a). Moreover, $\mathcal{L}\{k(\hat{\lambda},t)\}$ may also be expressed as:

$$\mathcal{L}\left\{k(\hat{\lambda},t)\right\} = \int_{T} h(t,\tau)\, k(\hat{\lambda},\tau)\, d\tau \tag{18b}$$

This expression is obtained by either inverting transform (17), with τ considered as a parameter, or by using the superposition integral, (13), with $\hat{k}(\lambda, t)$ as the input. By inverting (18a), with $\hat{\lambda}$ as a parameter, gives,

$$H(\lambda, \hat{\lambda}) = \int_T \mathcal{L}\left\{k(\hat{\lambda}, t)\right\} K(t, \lambda) \, dt \tag{18c}$$

which, along with (18b), gives equation (14b). ∎

Theorem 4 (**Spectral Convolution**) The relationship between the spectral functions $Y(\lambda)$ and $X(\lambda)$ is:

$$Y(\lambda) = \int_\Lambda H(\lambda, \hat{\lambda}) \, X(\hat{\lambda}) \, d\hat{\lambda} \tag{19}$$

Gersho [40] provides equivalent forms of spectral convolution for the case when alternative definitions of the impulse response are used; a discussion of analogous results may also be found in recent work by Margrave [43].

Proof: This follows by substituting (15) into the expression for $Y(\lambda)$ as determined from equations (7), rearranging the order of integration, and using (18c). ∎

V. Separation Techniques

A. The Ideal Filter

The concept of the bandpass filter is commonly understood to be a system which passes, without distortion, all Fourier frequency components falling in a certain frequency range, the filter's pass-band, and rejects all frequency components falling outside this range, the stop-band. The ideal bandpass filter is often referred to as an *ideal filter*. Hence, a generalised ideal filter is defined on any arbitrary domain as an operator which passes, without distortion, all generalised spectral components within a given range, and rejects all others [44–48]:

Definition 5 (**Spectral Response**) A linear operator, $\mathcal{L}$, is said to be an ideal filter over the λ-domain for an input space $\mathcal{X}$ if, for any $x(t) \in \mathcal{X}$, the representation of $y(t) = \mathcal{L}x(t)$, $Y(\lambda)$, on the domain λ is given by:

$$Y(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin \Lambda_H \\ X(\lambda) & \text{if } \lambda \in \Lambda_H \end{cases} \tag{20}$$

where $X(\lambda)$ is the representation of $x(t)$ on the λ-domain, and the acceptance region of $\mathcal{L}$, $\Lambda_H \subset \Lambda$, is a subset of the generalised frequency space Λ.

The definition of an ideal filter applies to random signals by considering the interpretation of an ensemble space of a stochastic process [33].

Theorem 5 (**Existence of an Ideal Filter**) A linear operator, $\mathcal{L}$, is said to be an ideal filter if, and only if, there exists a basis, $K(t, \lambda)$, and reciprocal basis, $k(\lambda, t)$, defining the transform pair (5), and satisfying (6), such that:

$$\mathcal{L}\{k(\lambda, t)\} = \begin{cases} 0 & \text{if } \lambda \notin \Lambda_H \\ k(\lambda, t) & \text{if } \lambda \in \Lambda_H \end{cases} \tag{21}$$

Proof: **Sufficiency**

Let $\mathcal{L}$ be a linear operator satisfying (21) with respect to the basis set $k(\lambda, t)$. Then, substituting (21) into (18c) and using (6b), gives:

$$H(\lambda, \hat{\lambda}) = \begin{cases} 0 & \text{if } \lambda \notin \Lambda_H \\ \delta(\lambda - \hat{\lambda}) & \text{if } \lambda \in \Lambda_H \end{cases} \tag{22}$$

which, after substitution into (19), gives Definition 5. Hence, $\mathcal{L}$ is an ideal filter.

Necessity

Let $\mathcal{L}$ be an ideal filter over the domain of $\lambda \in \Lambda$ for an input space $\mathcal{X}$. Then, from (19), and by Definition 5, $H(\lambda, \hat{\lambda})$ must satisfy (22) and, consequently, from (6b) and (18c), $\mathcal{L}\{k(\lambda, t)\}$ is given by (21). ∎

Theorem 6 (**Impulse Response**) An ideal filter over the domain $\lambda \in \Lambda$ with impulse response $h(t, \tau)$ may be represented in the 'separable' form

$$h(t, \tau) = \int_{\Lambda_H} k(\lambda, t) K(\tau, \lambda) \, d\lambda, \quad (t, \tau) \in T^2 \tag{23}$$

where Λ_H is the filter's passband.

Proof: Follows from (17) and Theorem 5. ∎

B. Spectral Solution of the Nonstationary Wiener Filter

General solutions to the nonstationary Wiener-Hopf filter defined in (4a) are not easily found, although it is easy to show [49] that the solution of the Wiener-Hopf equations can be reduced to the factorisation, or decomposition, of the correlation functions $R_{xx}(t, \tau)$ and $R_{dx}(t, \tau)$ [49–54]. Further, techniques which attempt to estimate the correlation functions of nonstationary processes from a single realisation rely on a similar decomposition [3–5].

B.1 Spectral Solution

A sufficient solution to the nonstationary Wiener-Hopf filter equations can be readily obtained by the factorisation of the ACFs into generalised power spectra. A necessary condition for a solution is not presented, inasmuch as there may be solutions for signal structures other than that defined in equation (7); this result is presented in [1, 2].

Theorem 7 (**WHF Solution**) Suppose $R_{dd}(t, \tau)$, $R_{nn}(t, \tau)$ and $R_{dn}(t, \tau)$ can be written as the generalised spectral decompositions of (10b) and (12b) respectively, such that:

$$R_{dd}(t, \tau) = \iint_{\hat{\Lambda}_d^2} \mathcal{P}_{dd}(\lambda, \hat{\lambda}) \, k(\lambda, t) \, k^*(\hat{\lambda}, \tau) \, d\lambda \, d\hat{\lambda} \tag{24a}$$

$$R_{nn}(t, \tau) = \iint_{\hat{\Lambda}_n^2} \mathcal{P}_{nn}(\lambda, \hat{\lambda}) \, k(\lambda, t) \, k^*(\hat{\lambda}, \tau) \, d\lambda \, d\hat{\lambda} \tag{24b}$$

$$R_{dn}(t, \tau) = \iint_{\Lambda_0^2} \mathcal{P}_{dn}(\lambda, \hat{\lambda}) \, k(\lambda, t) \, k^*(\hat{\lambda}, \tau) \, d\lambda \, d\hat{\lambda} \tag{24c}$$

where $(t, \tau) \in T^2$, $\hat{\Lambda}_d \equiv \Lambda_d \oplus \Lambda_0$ and $\hat{\Lambda}_n \equiv \Lambda_n \oplus \Lambda_0$. Here, $\Lambda_d \subset \Lambda$ and $\Lambda_n \subset \Lambda$ are the regions of the $\Lambda \subset \mathbb{C}$ space over which the spectral components of $\{d(t), \, t \in T\}$ and $\{n(t), \, t \in T\}$, respectively, do *not* overlap, and $\Lambda_0 \subset \Lambda$ is the region over which spectral components of $d(t)$ and $n(t)$ *do* overlap. A sufficient solution, $h(t, \tau)$, to the WHF equation (4a) for the additive case, when $\mathcal{T} \equiv T \subset \mathbb{R}$, is given by (14a) where the GBTF $H(\lambda, \hat{\lambda})$, is given by:

$$H(\lambda, \hat{\lambda}) = \begin{cases} \delta(\lambda - \hat{\lambda}) & \text{for } \{\lambda, \hat{\lambda}\} \in \Lambda_d^2, \\ H_0(\lambda, \hat{\lambda}) & \text{for } \{\lambda, \hat{\lambda}\} \in \Lambda_0^2, \\ 0 & \text{elsewhere.} \end{cases} \tag{25a}$$

where $\Lambda_d^2 = \Lambda_d \times \Lambda_d$, $\Lambda_0^2 = \Lambda_0 \times \Lambda_0$, and the function $H_0(\lambda, \hat{\lambda})$ is the solution of

$$\mathcal{P}_{dx}(\lambda, \hat{\lambda}) = \int_{\Lambda_0} H_0(\lambda, \bar{\lambda}) \, \mathcal{P}_{xx}(\bar{\lambda}, \hat{\lambda}) \, d\bar{\lambda} \tag{25b}$$

$\forall (\lambda, \hat{\lambda}) \in \Lambda_0^2$. Further, $H_0(\lambda, \hat{\lambda}) = 0$, $\forall (\lambda, \hat{\lambda}) \notin \Lambda_0^2$. The filter, $h(t, \tau)$, may, therefore, be written as

$$\begin{aligned} h(t, \tau) = &\int_{\Lambda_d} k(\lambda, t) \, K(\tau, \lambda) \, d\lambda \\ &+ \iint_{\Lambda_0^2} H_0(\lambda, \hat{\lambda}) \, k(\lambda, t) \, K(\tau, \hat{\lambda}) \, d\lambda \, d\hat{\lambda} \end{aligned} \tag{25c}$$

where $\mathcal{P}_{xx}(\cdot) = \mathcal{P}_{dx}(\cdot) + \mathcal{P}_{nx}(\cdot)$, $\mathcal{P}_{dx}(\cdot) = \mathcal{P}_{dd}(\cdot) + \mathcal{P}_{dn}(\cdot)$ and $\mathcal{P}_{nx}(\cdot) = \mathcal{P}_{nd}(\cdot) + \mathcal{P}_{nn}(\cdot)$. The resulting MSE is:

$$\sigma^2(t) = \iint\limits_{\Lambda_0^2} \mathcal{P}_{\sigma\sigma}(\lambda, \hat{\lambda})\, k(\lambda, t)\, k^*(\hat{\lambda}, t)\, d\lambda\, d\hat{\lambda} \qquad (25d)$$

where the spectrum of $\sigma^2(t)$ is given by:

$$\mathcal{P}_{\sigma\sigma}(\lambda, \hat{\lambda}) = \mathcal{P}_{dd}(\lambda, \hat{\lambda}) - \int_{\Lambda_0} H_0(\lambda, \bar{\lambda})\, \mathcal{P}_{dx}(\bar{\lambda}, \hat{\lambda})\, d\bar{\lambda} \qquad (25e)$$

$$= \int_{\Lambda_0} H_0(\lambda, \bar{\lambda}) \left\{ \mathcal{P}_{nd}(\bar{\lambda}, \hat{\lambda}) + \mathcal{P}_{nn}(\bar{\lambda}, \hat{\lambda}) \right\} d\bar{\lambda} \qquad (25f)$$

It is assumed the initial state of the filter $h(t, \tau)$ is at *rest*.

   Proof: See appendix.   ■

B.2 Signal Separability

   The first term in (25c) corresponds to the expression for an ideal filter which passes the generalised frequency range Λ_d and thus corresponds to perfect separation of the non-overlapping signal components. Moreover, this term, unlike the second, does not depend on *full* knowledge of the autocorrelation functions, inasmuch as it does not depend on the actual values of the signals n(t) or d(t), but only on the 'generalised' frequency bands over which the signal components do not overlap.

   Theorem 8 (**Ideal Filter Component**) The WHF, $h(t, \tau)$, which gives perfect separation, is an ideal filter (23), assuming that the signals have disjoint generalised spectral components.

   Proof: This is a direct corollary of Theorem 7. The MSE (25d) is zero, if $\Lambda_0 \equiv \{\emptyset\}$. This implies d($t$) and n($t$) are uncorrelated.   ■

Since the ideal filter is independent of the signal values, it separates the class of all nonstationary stochastic signals that are disjoint in the filter's domain, provided the desired signal components lie in its passband.

C. Selecting Transform Kernels

   In §V-B, it is demonstrated that if two signals lie in some domain, λ, and are disjoint, then *perfect signal separation* is achievable given appropriate prior knowledge of the signal structure. However, no mention has yet been made of how the transform domains are chosen. One valid solution to the signal separability problem is to choose all possible kernels, $\{k(\lambda, t), (\lambda, t) \in \Lambda \times$

T}, and classify as separable all pairs of stochastic processes admitted by this representation that are disjoint in this domain. Such an approach is exhaustive and does not provide a satisfactory solution to the more general question of whether two processes are separable given the form of their autocorrelation functions. Another approach, presented here, relies on finding a common domain in which the representation of two signals are disjoint. This is achieved by *concatenating* the basis functions for each individual signal representation to find a common domain.

The characteristics of the Wiener-Hopf filter in Theorem 7 were derived on the assumption that the spectral representations of d(t) and n(t) are defined on a common domain. This, however, appears restrictive and seems to indicate, for example, that a stationary process can be separated from a nonstationary process if, and only if, they are disjoint in the Fourier domain; this is clearly not the case. This apparent restriction is lifted by concatenating the different spectral representations of each signal, d(t) and n(t), provided each representation is bandlimited in their respective domains. So, for example, suppose d(t) is a stationary process expressed in the Fourier domain, $\omega \in \Omega \subset \mathbb{R}$, with basis set $\{k_d(\omega, t) = e^{-j\omega t}, (\omega, t) \in \Omega \times T\}$, and n($t$) is a nonstationary process, where d(t) and n(t) overlap in the Fourier domain. If n(t) is expressed in a domain, $\lambda \in \Lambda$, with basis set $\{k_n(\lambda, t), (\lambda, t) \in \Lambda \times T\}$, where its representation N(λ) is bandlimited, these domains can be combined as follows: the kernel in this *concatenated* domain, $\hat{\lambda} \in \hat{\Lambda} \subset \mathbb{C}$, denoted by $\{k(\hat{\lambda}, t), (\hat{\lambda}, t) \in \hat{\Lambda} \times T\}$, is constructed as:

$$k(\hat{\lambda}, t) = \begin{cases} k_n(\lambda(\hat{\lambda}), t), & \{\hat{\lambda} : |\mathrm{N}(\lambda(\hat{\lambda}))| \geq 0\} \\ k_d(\omega(\hat{\lambda}), t), & \{\hat{\lambda} : |\mathrm{D}(\omega(\hat{\lambda}))| \geq 0\} \end{cases}$$

where $\lambda(\hat{\lambda})$ and $\omega(\hat{\lambda})$ are isomorphic mappings of $\lambda : \hat{\Lambda} \to \Lambda$ and $\omega : \hat{\Lambda} \to \Omega$, with $\hat{\Lambda} \subset \mathbb{C}$. In other words, the kernel in the region of $\hat{\Lambda}$ on which the stationary signal is represented corresponds to the Fourier kernel, while the kernel in the region of $\hat{\Lambda}$ on which the nonstationary signal is represented corresponds to $k_n(\lambda(\hat{\lambda}), t)$. Assuming these regions are disjoint, this domain exists if the transform is unique and, in such a case, the two signal classes are separable. If the signals overlap in the $\hat{\lambda}$-domain, then a transform pair cannot be derived since there is no unique transform. The idea behind this 'concatenation' is summarised in Figure 3.
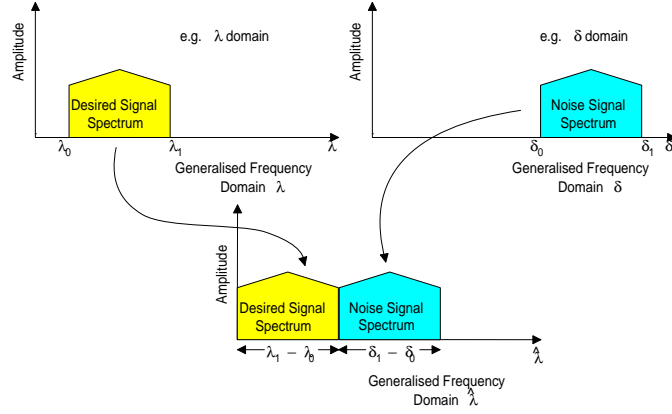
Fig. 3.   Concatenating power spectra.

Formalising this idea, suppose the ACFs of $d(t)$ and $n(t)$ admit the spectral representations:

$$R_{dd}(t,\tau) = \iint\limits_{\Lambda_D^2} \mathcal{P}_{dd}(\lambda, \hat{\lambda})\, k_d(\lambda, t)\, k_d^*(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{26a}$$

$$R_{nn}(t,\tau) = \iint\limits_{\Lambda_N^2} \mathcal{P}_{nn}(\lambda, \hat{\lambda})\, k_n(\lambda, t)\, k_n^*(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{26b}$$

for any $\Lambda_d \subset \Lambda$ and $\Lambda_n \subset \Lambda$. Now, assume it is possible to apply a change of variables $\lambda_d' : \Lambda_d \to \Lambda_d'$ and $\lambda_n' : \Lambda_n \to \Lambda_n'$ such that equations (26) become:

$$R_{dd}(t,\tau) = \iint\limits_{\Lambda_D'^2} \mathcal{P}_{dd}'(\lambda, \hat{\lambda})\, k_d'(\lambda, t)\, k_d'^{*}(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{27a}$$

$$R_{nn}(t,\tau) = \iint\limits_{\Lambda_N'^2} \mathcal{P}_{nn}'(\lambda, \hat{\lambda})\, k_n'(\lambda, t)\, k_n'^{*}(\hat{\lambda}, \tau)\, d\lambda\, d\hat{\lambda} \tag{27b}$$

where $\Lambda_D' \cap \Lambda_N' = \{\emptyset\}$. In such a case, the spectral representations can be concatenated by defining the *concatenated kernel*:

$$k(\lambda, t) = \sum_i \mathbb{I}_{\Lambda_i}(\lambda)\, k_i'(\lambda, t) \tag{28}$$

$\forall (\lambda, t) \in \Lambda \times T$, $i \in \boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the set of indices for the various spectral subsets, taking on the elements 'd', 'n', and 'v' so $\boldsymbol{\theta} = \{d, n, v\}$; correspondingly, $\boldsymbol{\Theta} \triangleq \{D, N, V\}$ contains the elements 'D', 'N' and 'V'. The indicator function $\mathbb{I}_\Lambda(\lambda)$ is defined as $\mathbb{I}_\Lambda(\lambda) = 1$ if $\lambda \in \Lambda$, and 0 otherwise. Note further that Λ_V is 'unused' space, such that $\bigoplus_{i \in \boldsymbol{\Theta}} \Lambda_i \triangleq \Lambda_D \oplus \Lambda_N \oplus \Lambda_V = \Lambda$. Assuming this transform domain exists, it follows that the autocorrelation functions of $d(t)$ and $n(t)$ possess

spectral representations on a domain with kernel $k(\lambda, t)$ as defined in equation (28), rather than the kernels $k'_d(\lambda, t)$ and $k'_n(\lambda, t)$ in equation (27). Since each of the autocorrelation functions have been represented on the same transform domain, assuming one exists, it follows that these processes can be separated by the ideal filter in equation (23). Ergo, from §III-B, such a domain exists when $k(\lambda, t)$ and $K(t, \lambda)$ satisfy equations (6).

D. Separating Modulated Signals

Using the method in the previous section, separability constraints can be determined for the class of *filtered modulated* signals which exhibit the general form:

$$\left.\begin{aligned}
\mathrm{d}(t) &= \int_T h_d(t, \tau)\, \mathrm{a}(\tau)\, d\tau \\
\mathrm{n}(t) &= \int_T h_n(t, \tau)\, \mathrm{b}(\tau)\, d\tau
\end{aligned}\right\} \quad t \in T \tag{29}$$

where, assuming some prior knowledge regarding the structure of the signals $\mathrm{d}(t)$ and $\mathrm{n}(t)$, it is known that $\mathrm{d}(t)$ and $\mathrm{n}(t)$ overlap in the time and Fourier spectral domains, $\mathrm{a}(t)$ and $\mathrm{b}(t)$ are bandlimited to $\pm\omega_c$ but otherwise unknown, and $h_d(t, \tau)$ and $h_n(t, \tau)$ are known deterministic signals.

Since $\mathrm{a}(t)$ and $\mathrm{b}(t)$ are bandlimited in the Fourier domain, they admit the representation

$$\mathrm{a}(t) = \int_{-\omega_c}^{\omega_c} \mathrm{A}(\omega)\, e^{j\omega t}\, d\omega, \; \mathrm{b}(t) = \int_{-\omega_c}^{\omega_c} \mathrm{B}(\omega)\, e^{j\omega t}\, d\omega \tag{30a}$$

and, therefore, after substitution into (29), a little rearrangement, and use of the mappings $\lambda_d : \Omega \to \Omega + \omega_c$ for $\mathrm{d}(t)$ and $\lambda_n : \Omega \to \Omega - \omega_c$ for $\mathrm{n}(t)$, it follows that $\mathrm{d}(t)$ and $\mathrm{n}(t)$ admit the representations:

$$\begin{aligned}
\mathrm{d}(t) &= \int_0^{2\omega_c} \mathrm{D}(\lambda)\, k(\lambda, t)\, d\lambda \\
\mathrm{n}(t) &= \int_{-2\omega_c}^0 \mathrm{N}(\lambda)\, k(\lambda, t)\, d\lambda
\end{aligned} \tag{30b}$$

where $\lambda \in \Lambda \subset \mathbb{R}$, and the kernel is defined as:

$$k(\lambda, t) = \begin{cases}
\int_T h_d(t, \tau)\, e^{j(\lambda - \omega_c)\tau}\, d\tau & \text{for } \lambda \in [0, 2\omega_c) \\
\int_T h_n(t, \tau)\, e^{j(\lambda + \omega_c)\tau}\, d\tau & \text{for } \lambda \in [-2\omega_c, 0) \\
k_v(\lambda, t) & \text{for } \lambda \notin (-2\omega_c, 2\omega_c)
\end{cases} \tag{31}$$

with the generalised spectral components given by $D(\lambda) = A(\lambda - \omega_c)$ and $N(\lambda) = B(\lambda + \omega_c)$. If $d(t)$ and $n(t)$ are separable, $k(\lambda, t)$ and $K(t, \lambda)$ must satisfy (6): substitution of (31) into (6) creates constraints on the filters $h_d(t, \tau)$ and $h_n(t, \tau)$.

VI. Discrete Case

A. Discrete Time and Discrete Spectral Domains

The spectral decomposition in the finite *discrete* time, and finite *discrete* spectral case is conceptually simpler than the continuous case, since a transform from one domain to another can be viewed as a change in basis vectors. The representation of a discrete time deterministic signal $\{x(n),\ n \in \mathcal{N} \subset \mathbb{Z}\}$ on an arbitrary finite support discrete spectral domain is $\{X(p),\ p \in \mathcal{P} \subset \mathbb{Z}\}$, where

$$X(p) = \sum_{n \in \mathcal{N}} x(n)\, K(n, p), \quad x(n) = \sum_{p \in \mathcal{P}} X(p)\, k(p, n) \tag{32}$$

corresponding to (5) for the continuous case. Assuming a *non-degenerate* transform, the kernels satisfy:

$$\sum_{p \in \mathcal{P}} K(n, p)\, k(p, \hat{n}) = \delta(n - \hat{n})\ ,\ \ (n, \hat{n}) \in \mathcal{N} \times \mathcal{N} \tag{33a}$$

$$\sum_{n \in \mathcal{N}} k(p, n)\, K(n, \hat{p}) = \delta(p - \hat{p})\ ,\ \ (p, \hat{p}) \in \mathcal{P} \times \mathcal{P} \tag{33b}$$

where $\delta(p) = 1$ if $p = 0$ and 0 otherwise. Defining the vectors $[\mathbf{x}]_n = x(n)$, $[\mathbf{X}]_n = X(n)$, $n \in \mathcal{N}$, and the kernel matrices $[\mathrm{k}]_{pn} = k(p, n)$, $[\mathrm{K}]_{np} = K(n, p)$, $(n, p) \in \mathcal{N} \times \mathcal{P}$, then (32) can be written compactly as:

$$\mathbf{X} = \mathrm{K}^T \mathbf{x} \qquad \text{and} \qquad \mathbf{x} = \mathrm{k}^T \mathbf{X} \tag{34}$$

where T denotes the matrix transpose. Furthermore,

$$\mathrm{k}\,\mathrm{K} = \mathrm{K}\,\mathrm{k} = \mathrm{I}_N \tag{35}$$

where $\mathrm{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix, and (35) is equivalent to (33) written in matrix form. Observe that (35) implies the basis kernels must have full rank.

B. Concatenating Discrete Spectra

Following the same line of argument as in the continuous case in §V-D, suppose that the autocorrelation functions of the two processes $d(n)$ and $n(n)$ can be written as two-dimensional

transform pairs with kernels $k_d(p,n)$ and $k_n(p,n)$, respectively, and domains of summation given by $\mathcal{P}_D$ and $\mathcal{P}_N$, respectively, where $\bigoplus_{i \in \boldsymbol{\Theta}} \mathcal{P}_i = \mathcal{P}$, $\mathcal{P}_V$ is unused space, and $\mathcal{P}_D \cap \mathcal{P}_N = \{\emptyset\}$ *etc.* For such a domain to exist, the concatenated kernel $k(p,n)$ must have an inverse. In the discrete case, $\{k(p,n), (p,n) \in \mathcal{P} \times \mathcal{N}\}$ can be written as a matrix, and thus the existence of the transform can be found by using linear algebra techniques. Assume, for clarity, that $\{\mathcal{P}_i, i \in \boldsymbol{\Theta}\}$ are contiguous regions, given by $\mathcal{P}_D = \{0, \ldots, P_D - 1\}$, $\mathcal{P}_N = \{P_D, \ldots, P_N - 1\}$, and $\mathcal{P}_V = \{P_N, \cdots, P - 1\}$, where P denotes the number of element in $\mathcal{P}$; *i.e.* $P = \dim \mathcal{P}$. If the matrices k_d, k_n and k_v, whose elements are the values of their respectively named kernels, are appropriately defined then:

$$
k\,K = \left[\begin{array}{c} k_d \\ \hline k_n \\ \hline k_v \end{array} \right] \left[\begin{array}{c|c|c} K_d & K_n & K_v \end{array} \right] = I_P \tag{36}
$$

C. Discrete Modulation

The results from §V-D carry across, *mutatis mutandis*, to the discrete-time discrete-spectrum case, and it can be shown that by defining $W_N^{np} = \frac{1}{N} e^{jnp\frac{2\pi}{N}}$, where $N = \dim \mathcal{N}$, and the matrices

$$
\left[\hat{k}_d\right]_{pn} = W_N^{n(p-q_c)},\ p \in \mathcal{P}_D,\ \left[\hat{k}_v\right]_{pn} = k_v(p,n),\ p \in \mathcal{P}_V,
$$
$$
\left[\hat{k}_n\right]_{pn} = W_N^{n(p+q_c)},\ p \in \mathcal{P}_N,\ \left[H_{[\boldsymbol{\theta}]_i}\right]_{n\hat{n}} = h_{[\boldsymbol{\theta}]_i}(n,\hat{n})
$$

where $(n, \hat{n}) \in \mathcal{N} \times \mathcal{N}$, and H_d, $H_n \in \mathbb{R}^{N \times N}$, then k, defined in §VI-B, can be partitioned as:

$$
k^T = \left[H_d\,\hat{k}_d^T\ \middle|\ H_n\,\hat{k}_n^T\ \middle|\ \hat{k}_v^T \right] \tag{37}
$$

where $\hat{k}_v$ are the unused basis vectors. The problem now is to find constraints on the matrices H_d and H_n such that k has full rank. One physical way of interpreting the rôle of H_d and H_n in (37) is to consider them as linear transformations of $a(n)$ and $b(n)$, mapping them to *different subspaces*. Thus, $d(n)$ and $n(n)$ are not separable when these linear transformations map $a(n)$ and $b(n)$ to *overlapping subspaces*.

Calculating Ideal Filter Matrix

Using the results in §VI-C, the ideal filter matrix to recover the signal $d(t)$ can be calculated as detailed in Algorithm 1. Provided the kernel matrix in (37) has full rank, the signals are separable. An estimate of the desired signal, $d(t)$, can be calculated by evaluating:

Algorithm 1 Calculating Ideal Filter Matrix

1:   Evaluate: $H_d = \text{diag}\,[h_d(t)]$, $H_n = \text{diag}\,[h_n(t)]$ and $[k_v]_{pt} = W_N^{tp}$, $p \in \mathcal{P}_V$, $t \in \mathcal{T}$, $N = \dim \mathcal{T}$.

2:   Construct the inverse-kernel:

$$k^T = \left[\, H_d\,\hat{k}_d^T \,\middle|\, H_n\,\hat{k}_n^T \,\middle|\, \hat{k}_v^T \,\right] \tag{37}$$

3:   **if** rank [k] = N **then**

4:       Construct the generalised bandpass filter:

$$H = \text{diag}\,[\,\mathbf{1}_{2p_c} \mid \mathbf{0}_{N-2p_c}\,] \tag{38}$$

   where $\mathbf{1}_{2p_c} \in \mathbb{R}^{2p_c}$ and $\mathbf{0}_{N-2p_c} \in \mathbb{R}^{N-2p_c}$ are vectors of ones and zeros respectively.

5:       Evaluate the ideal filter matrix:

$$A = k^T\,H\,K^T \tag{39}$$

6:   **else**

7:       Signals belonging to the class defined by equations (43) and (45) are inseparable.

8:   **end if**

$$\hat{\mathbf{d}} = A\,\mathbf{x} \tag{40}$$

where A is given by equation (39).

D. Noise Gain for Filter Matrix

   If an additional noise signal, $\mathbf{w}$, appears at the input of the filter then, in general, it is passed to the filter's output. The *noise gain* of an ideal filter is defined as the gain in noise energy when white noise passes through it.

   Theorem 9 (Noise Gain) In the case of white noise, $\mathbf{w}$, with variance σ_w^2 appearing at the input of the filter defined by equation (39), the noise gain is given by:

$$\eta_{WGN} \triangleq \frac{\mathbb{E}\left\{\mathbf{y}\,\mathbf{y}^T\right\}}{\mathbb{E}\left\{\mathbf{w}\,\mathbf{w}^T\right\}} = \frac{1}{N}\,\text{trace}\left[A\,A^T\right] \tag{41}$$

where $\|B\|_F$ denotes the Frobenius norm of a matrix B, and $\mathbf{y}$ is the output.

   Proof: The output of the filter is given by $\mathbf{y} = A\,\mathbf{w}$, and hence the total noise power at the output is given by $\text{trace}\left[\mathbf{y}\,\mathbf{y}^T\right] = \text{trace}\left[A\,\mathbf{w}\,\mathbf{w}^T\,A\right]$. Taking expectations of both sides, noting the linearity of the trace function, and noting $\mathbb{E}\left\{\mathbf{w}\,\mathbf{w}^T\right\} = \sigma_w^2 I$, gives the desired result. ∎

VII. Uniform Modulation

As an example of single channel separation, consider the special class of nonstationary signals known as *uniformly modulated processes*, or amplitude modulated time series, characterised by the form

$$y(t) = c(t) \, x(t), \qquad t \in \mathcal{T} \tag{42}$$

where $x(t)$ is a stochastic process and $c(t)$ is a known deterministic signal [15]. Such nonstationary processes have been shown to describe the behaviour of seismic reflectivity data [15, 55] where, for example, earthquake and explosion data have been modelled by (42), with $x(t)$ as a zero mean first-order autoregressive (AR) process, and the deterministic function given by $c(t) = t^{\alpha} \nu^{-\beta t}$, with α, β and ν as known constants. The model in equation (42) is also appropriate for modelling deterministic signals that have been corrupted by multiplicative noise, for example, in radar applications where a target is illuminated and the signal experiences random time-varying amplitude distortion caused either by target fluctuation, or scattering of the medium (*e.g.* fading) [56, 57].

In this special case of uniformly modulated signals, the general form of (29) reduces to:

$$\left. \begin{aligned} d(t) &= h_d(t) \, a(t) \\ n(t) &= h_n(t) \, b(t) \end{aligned} \right\} \quad t \in T \tag{43}$$

where $a(t)$ and $b(t)$ are stochastic processes bandlimited to $\pm \omega_c$. The first two expressions of the transform kernel in (31) reduce to:

$$k(\lambda, t) = \begin{cases} h_d(t) \, e^{j(\lambda - \omega_c)t} & \text{for } \lambda \in [0, 2\omega_c) \\ h_n(t) \, e^{j(\lambda + \omega_c)t} & \text{for } \lambda \in [-2\omega_c, 0) \end{cases} \tag{44}$$

It can be shown that in the range $\lambda \in [-2\omega_c, 2\omega_c]$, $k(\lambda, t)$ is a linear combination of the basis set $\{e^{j\omega t}, \forall \omega \in \hat{\Omega} \subset \Omega\}$. To complete the basis set, $k_v(\lambda, t)$, which is itself a linear combination of the set $\{e^{j\omega t}, \forall \omega \in \Omega\}$, must not be a linear combination of $k(\lambda, t)$ for $\lambda \in \{-2\omega_c, 2\omega_c\}$. It can thence be deduced that if $k_v(\lambda, t) = e^{jg(\lambda) t}$, for some function $g(\lambda)$, the basis set can be completed provided that $k(\lambda, t)$ is not degenerate in the range $\lambda \in \{-2\omega_c, 2\omega_c\}$. In many cases, $\{k_v(\lambda, t) = e^{j\lambda t}, \forall \lambda \notin (-2\omega_c, 2\omega_c)\}$ suitably extends the basis. Assuming that the set of exponential basis functions for the unused transform space completes the set of basis functions then, since the rank of the kernel matrix can readily be determined, the inverse kernel matrix in

(37) can be readily constructed and its inverse calculated, facilitating the separation of separable discrete–time signals. This is demonstrated by the following specific examples.

A. Separation of Chirp Modulated Signals

Suppose the desired and noise signals can be expressed in the form of (43), where $\mathcal{T} = \{0, \ldots, T-1\}$, and the deterministic functions, $h_d(t)$ and $h_n(t)$, are given by:

$$
\left.
\begin{aligned}
h_d(t) &= \cos\left[2\pi f_i \frac{t}{f_s}\left(1 + \frac{1}{\tau_d}\frac{f_f - f_i}{f_i}\frac{t}{f_s}\right)\right] \\
h_n(t) &= \cos\left[2\pi f_i' \frac{t}{f_s}\left(1 + \frac{1}{\tau_n}\frac{f_f' - f_i'}{f_i'}\frac{t}{f_s}\right)\right]
\end{aligned}
\right\} \quad t \in \mathcal{T}
\tag{45}
$$

where a(t) and b(t) are bandlimited to f_c, the sampling frequency $f_s \geq 4f_c$, the *start* and *stop* sweep frequencies are given by $(f_i, f_i') \in \mathbb{R}^2$ and $(f_f, f_f') \in \mathbb{R}^2$ respectively, and the frequency sweep times are $\boldsymbol{\tau} = (\tau_d, \tau_n) \in \mathbb{R}^2$. For convienience, define $\mathbf{f} = \{f_i, f_i', f_f, f_f'\} \in \mathbb{R}^4$.

If $\{h_d(t) = h_n(t), \forall t \in \mathcal{T}\}$, separation is not possible since no additional signal structure distinguishing d(t) and n(t) is available. However, by a continuity argument, it is reasonable to assume that a small perturbation in one of the functions is not likely to make separation possible either. Furthermore, if $f_c \gg f, \forall f \in \mathbf{f}$, such that a$(t)$ and b(t) are only slightly modulated, it is also reasonable to assume that there are not enough distinguishing features to achieve separation. Continuity arguments suggest that d(t) and n(t) can be separated provided that the spectra of the modulating signals, $h_d(t)$ and $h_n(t)$, differ *sufficiently*, but are not, in general, disjoint. Further, the boundary frequencies, (f_i, f_i') and (f_f, f_f'), must be in the *region* of f_c: *i.e.* to ensure a(t) and b(t) are modulated such that d(t) and n(t) have significantly different characteristics.

As a particular example, Figure 4 shows the Fourier spectra of $h_d(t)$ and $h_n(t)$ of equation (45) with $f_i = f_f' = 4250$ Hz, $f_f = f_i' = 5750$ Hz, $f_s = 22.050$ kHz, and $\tau_d = \tau_n = 70$ msec; *i.e.* these sweeps are identical, except they are in opposite directions. Consider taking $T = 1000$ samples or 45 msec worth of data, with $f_c = 5$ kHz ($p_c = 227$ frequency bins), such that there will be $T - 4p_c = 92$ unused generalised frequency bins. The ideal filter matrix to recover the signal d(t) can be calculated as detailed in Algorithm 1. Provided the parameters $\mathbf{f}$ and $\boldsymbol{\tau}$ are such that the kernel matrix in (37) has full rank, the signals are separable. This is true in the case outlined above, and the impulse response of the resulting ideal filter is shown in Figure 5.

Since the Fourier spectra of the modulating signals in Figure 4 overlap, so do the Fourier spectra of d(t) and n(t). To emphasis this, consider the unknown bandlimited signals a(t)
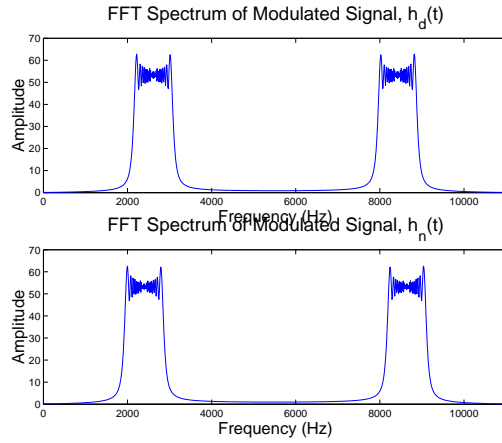
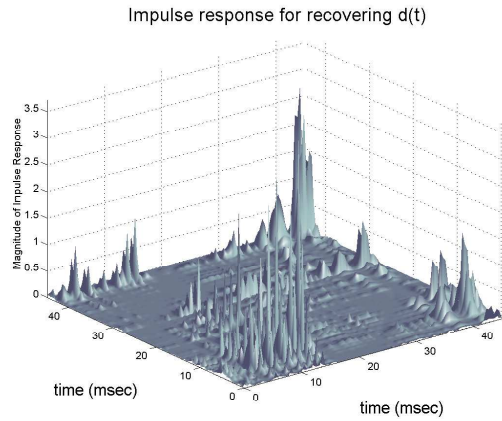Fig. 4. Fourier spectra of $h_d(t)$, $h_n(t)$; $f_s = 22.050\,\mathrm{kHz}$.

Fig. 5. The response of the ideal filter which recovers d(t) for the chirp modulated case.

and b(t) shown in Figure 6. The spectra of the resulting desired and noise signals, d(t) and n(t), are shown in Figure 7, and clearly overlap, with the Fourier spectrum of x(t) shown in Figure 8. Therefore, conventionally, d(t) and n(t) would be considered inseparable; however, taking a generalised spectral transform using equation (34) with the kernel matrices derived in equation (37), gives the generalised spectrum of x(t) shown in Figure 8. This generalised spectrum can then be bandpass filtered to recover the generalised spectrum of either d(t) or n(t), and the inverse transform can be taken to obtain the time–domain representations of the signals; this is shown in Figure 9. Since perfect separation is possible in this case, the estimate identically equals the actual desired signal.

In simulations, where the desired signal d(t) is available, the error between the actual signal,
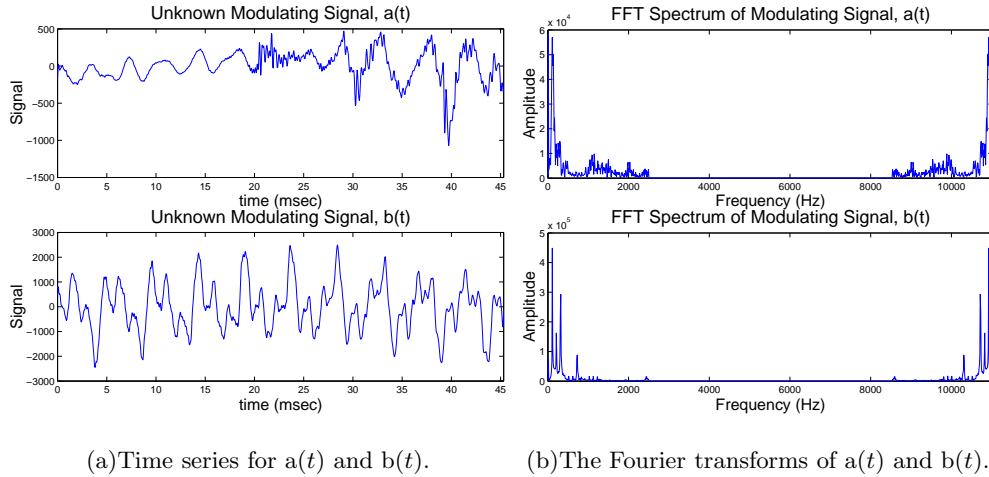
(a)Time series for a(t) and b(t).   (b)The Fourier transforms of a(t) and b(t).

Fig. 6.   The unknown stochastic modulating signals a(t), b(t). The sampling frequency, $f_s = 22.050 \, \text{kHz}$,
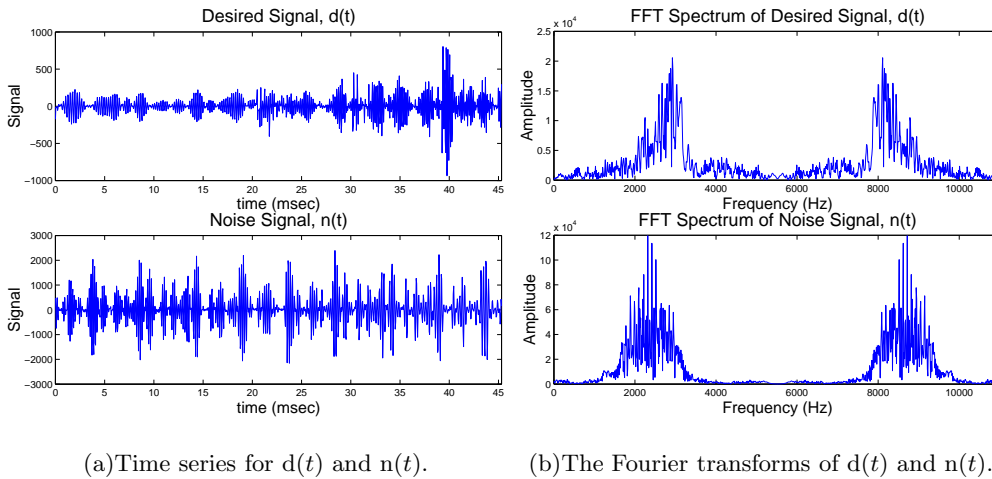and they are bandlimited to $5 \, \text{kHz}$, as emphasised in Figure 6(b).

(a)Time series for d(t) and n(t).   (b)The Fourier transforms of d(t) and n(t).

Fig. 7.   The desired, d(t), and noise, n(t), signals resulting from multiplying the signals in Figure 6(a)
by the chirp signals of Figure 6. The Fourier transforms of d(t) and n(t) are overlapping, and so
conventionally, would be considered inseparable.

and the estimate obtained in equation (40), can be evaluated as $\mathbb{E}\left\{\hat{\mathbf{d}} - \mathbf{d}\right\}$. A number of
different realisations were used for the modulating signals d(t) and n(t), and the expected error
was extremely small $\approx 10^{-8}$; this is a result of numerical accuracy. Furthermore, the noise gain
of the filter matrix A is around 35, and thus any parameter errors due to noise increase linearly
as a function of the variance of the input noise, as discussed in §VI-D. This has been confirmed
by experimentation, although the results are trivial and ommitted here.
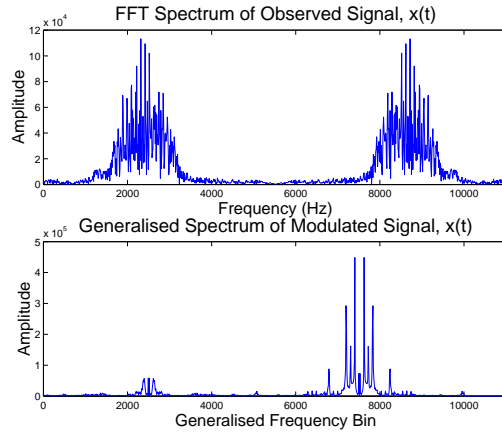
Fig. 8.   *Top:* The Fourier transform of the observed signal $x(t) = d(t) + n(t)$, $\forall t \in \mathcal{T}$ for the resulting signals in Figure 7(a); *Bottom:* The Generalised transform of $x(t)$.
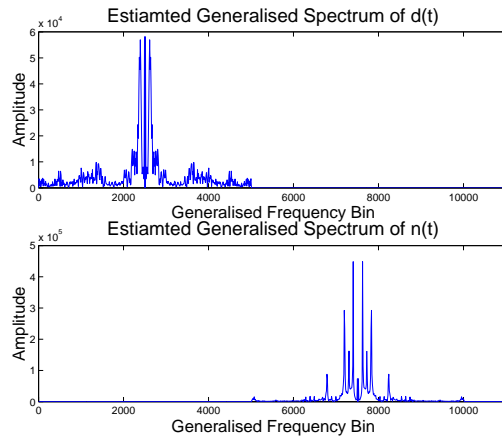

Fig. 9. The Generalised Spectra of $d(t)$ and $n(t)$, obtained by bandpass filtering the Generalised Spectrum of $x(t)$ as shown in Figure 8 with pass-bands $\{0, 2f_c - 1\}$ and $\{2f_c, 4f_c - 1\}$ respectively.

B. Prior Knowledge and Parameter Sensitivity

The *prior* knowledge needed to separate the signals includes:

1. requiring $d(t)$ and $n(t)$ are of the form equation (43) with $h_d(t)$ and $h_n(t)$ given by equation (45);

2. knowing $a(t)$ and $b(t)$ must be known to be bandlimited to f_c;

3. ensuring $f_s \geq 4f_c$, $\mathbf{f}$ and $\boldsymbol{\tau}$ in equation (45) must all be known.

These parameters must somehow be estimated from the observed signal and, although this is a non-trivial problem, for the purpose of signal separation it is reasonable to assume that these
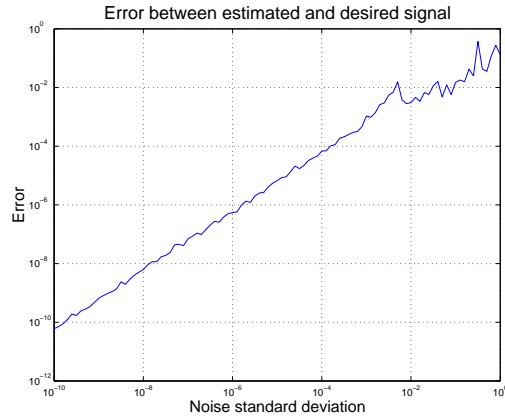
Fig. 10.   The mean squared error between the actual desired signal, $\mathbf{d}$, and the estimated desired signal, $\hat{\mathbf{d}}$, for the example in §VII-A.

few parameters are known *a priori*. These can be obtained, for example, by modelling the unknown modulating signals $\mathrm{a}(t)$ and $\mathrm{b}(t)$ parametrically, and estimating all the parameters for the unknown signals $\mathrm{d}(t) \triangleq h_d(t)\,\mathrm{a}(t)$ and $\mathrm{n}(t) = h_n(t)\,\mathrm{b}(t)$ using, for example, Bayesian methods. The parameter estimates for $h_d(t)$ and $h_n(t)$ can then be used to create the ideal-filer in Algorithm 1, and all signals in this class can be separated without any need to place a model on the unknown modulating signals.

For the example discussed in §VII-A, consider the case where an error is made in the estimate of $h_d(t)$ and $h_n(t)$, such that

$$
\begin{aligned}
\mathrm{h}_d^{\mathrm{est}}(t) &= h_d(t) + \sigma_w^t \mathrm{w}(t) \\
\mathrm{h}_n^{\mathrm{est}}(t) &= h_n(t) + \sigma_w^t \hat{\mathrm{w}}(t)
\end{aligned}
\tag{46}
$$

In this case, suppose the kernel-matrix and ideal-filer matrix of equations (37) and (39), respectively, are evaluated. This, naturally, will generate errors in the estimate of the desired signal, $\mathrm{d}(t)$, since the noise signal, $\mathrm{n}(t)$, no longer lies in the null-space of the matrix A. Furthermore, $\mathrm{d}(t)$ is no longer an eigenvector of the matrix A. Figure 10 shows the mean squared error between the actual desired signal, $\mathbf{d}$, and the estimated desired signal, $\hat{\mathbf{d}}$, for the example in §VII-A. This log-log plot indicates that the performance of the ideal filter falls off dramatically as the signal-to-noise ratio (SNR) decreased. This is problematic practically, and future research is needed to investigate methods for compensating errors in the estimate of the modulating signals, $h_d(t)$ and $h_n(t)$.
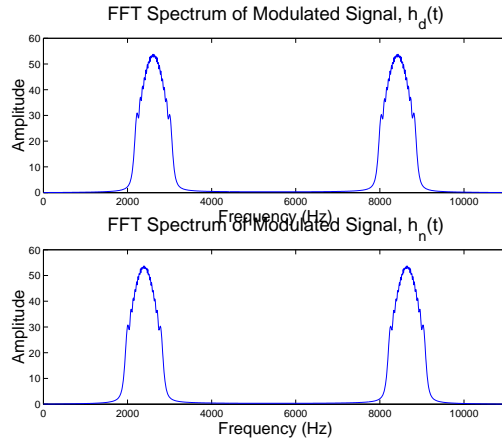
Fig. 11. Fourier spectra of $h_d(t)$, $h_n(t)$; $f_s = 22.050$ kHz.

C. Separation of Gaussian-Chirp Modulated Signals

As a second example, the chirp in §VII-A is modified by multiplying equation (45) by a Gaussian. Hence, $h_d(t)$ and $h_n(t)$, are given by:

$$\left.\begin{aligned}
h_d(t) &= \exp\left\{-\frac{(t-\hat{\tau}_d)^2}{\sigma_d^2}\right\} \cos\left[2\pi f_i \frac{t}{f_s}\left(1 + \frac{1}{\tau_d}\frac{f_f - f_i}{f_i}\frac{t}{f_s}\right)\right] \\
h_n(t) &= \exp\left\{-\frac{(t-\hat{\tau}_n)^2}{\sigma_n^2}\right\} \cos\left[2\pi f_i' \frac{t}{f_s}\left(1 + \frac{1}{\tau_n}\frac{f_f' - f_i'}{f_i'}\frac{t}{f_s}\right)\right]
\end{aligned}\right\} \quad t \in \mathcal{T} \qquad (47)$$

Using the same parameter values for $\{f_i, f_i', f_f, f_f', \tau_d, \tau_n\}$ as in the previous section, and $\hat{\tau}_d = \hat{\tau}_n = \frac{T}{2}$, and $\sigma_d = \sigma_n = \frac{1}{20\sqrt{5}}$, the modulated signals $h_d(t)$, $h_n(t)$ are shown in Figure 11. The ideal filter matrix is again calculated using Algorithm 1, and as the kernel matrix in (37) has full rank, the signals are separable. The impulse response of the resulting ideal filter is shown in Figure 12.

VIII. Conclusions

Nonstationarity is useful in applications where the mixture of two or more signals is observed and, at each time instance, there is only a *single* observation of the mixture in the time domain. Separation of the signals has important applications where background noises must be removed in, for example, mobile telephones, hearing aids, conference telephony, forensic science, restoration of black box recordings, speech recoginition, and biomedical data analysis. Signal separation with one observation sensor can only be achieved by exploiting prior knowledge of the signal structure. Using the assumption that signal separation is possible using the Wiener-Hopf
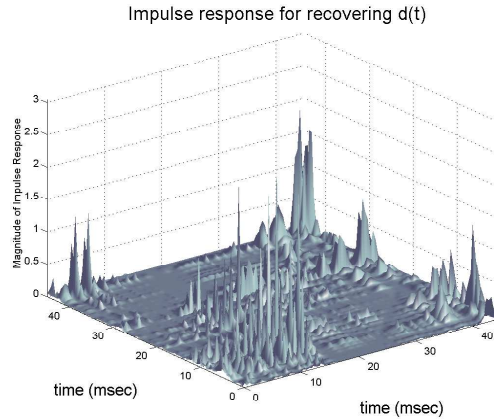
Impulse response for recovering d(t)

Fig. 12.   The response of the ideal filter which recovers d(t) for the Gaussian-chirp modulated case.

filter, an intuitive solution to the Wiener-Hopf filter equations is obtained, and relies on the factorisation of the ACFs into *generalised power spectra*. The solution to the Wiener-Hopf filter equations leads naturally to a criterion for signal separation with perfect precision, and is a natural extension to the condition for separating stationary signals. This paper has also introduced a signal separation technique by *concatenating* the domains on which two signal classes can be represented on with a finite number of basis functions. The technique has been specifically applied to uniformly modulated signals, and an example has been discussed of separating chirp signals embedded in multiplicative noise. The known modulated signals can be obtained, for example, by modelling the unknown modulating signals parametrically, and estimating all the parameters for the desired and noise signals using, for example, Bayesian methods. The parameter estimates for the "known" modulated signals can then be used to construct the ideal-filer, and all signals in this class can be separated without any need to place a model on the unknown modulating signals.

Appendix

This section contains a proof of sufficiency of Theorem 7.

Proof: Consider taking the general spectral transform of the Wiener-Hopf filter equation (4a) using equation (10a) in Definition 3:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint_{\mathcal{T}\mathcal{T}\mathcal{T}} h(t, \alpha) \, R_{xx}(\alpha, \tau) \, K(t, \lambda_t) \, K^*(\tau, \lambda_\tau) \, d\alpha \, dt \, d\tau$$

substituting equation (14a) gives ($\forall \lambda_t, \lambda_\tau \in \Lambda$),

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiint\limits_{\mathcal{T}\mathcal{T}\mathcal{T}} \iint\limits_{\Lambda\Lambda} H(\lambda_1, \lambda_2)\, R_{xx}(\alpha, \tau)\, k(\lambda_1, t)$$

$$\times\, K(\alpha, \lambda_2)\, K(t, \lambda_t)\, K^*(\tau, \lambda_\tau)\, d\lambda_1\, d\lambda_2\, d\alpha\, dt\, d\tau$$

and rearranging the integral w. r. t. t gives:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \iiiint\limits_{\mathcal{T}\mathcal{T}\Lambda\Lambda} H(\lambda_1, \lambda_2)\, R_{xx}(\alpha, \tau)\, K(\alpha, \lambda_2)$$

$$\times\, K^*(\tau, \lambda_\tau) \int\limits_{\mathcal{T}} k(\lambda_1, t)\, K(t, \lambda_t)\, dt\, d\lambda_1\, d\lambda_2\, d\alpha\, d\tau$$

Thus, simplifying using equation (6b), the sifting property, and the definition for the spectral decomposition, yields:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \int_{\Lambda} H(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda, \ \forall (\lambda_t, \lambda_\tau) \in \Lambda^2 \tag{48a}$$

Finally, assuming equation (25a) is the solution of the Wiener-Hopf equations, then equation (48a) may be written in the form:

$$\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = \int_{\Lambda_d} \delta(\lambda_t - \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda$$

$$+ \int_{\Lambda_0} H_0(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda \tag{48b}$$

Noting that $\Lambda_d \oplus \Lambda_n = \{\emptyset\}$, $\Lambda_d \oplus \Lambda_0 = \{\emptyset\}$ then, using the sifting property, the first term in equation (48b) is equivalent to $\mathcal{P}_{xx}(\lambda_t, \lambda_\tau)$ if $\lambda_t \in \Lambda_d$, and zero otherwise. Moreoever, if $\lambda_t \in \Lambda_d$ then, since $H_0(\lambda_1, \lambda_2) = 0$ if $(\lambda_1, \lambda_2) \notin \Lambda_0^2$, the second term is zero. However, if $\lambda_t \in \Lambda_0$, the first term is zero, while the second term is nonzero. Hence, using the relationships in Definition 3 and noting $P_{dx}(\lambda_t, \lambda_\tau) = \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) + \mathcal{P}_{dn}(\lambda_t, \lambda_\tau)$, it follows, after some slight rearrangement, that:

$$\mathcal{P}_{dd}(\lambda_t, \lambda_\tau) =$$

$$\begin{cases} \mathcal{P}_{xx}(\lambda_t, \lambda_\tau) \equiv \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) & (\lambda_t, \lambda_\tau) \in \Lambda_d^2 \\[2mm] \displaystyle\int_{\Lambda_0} H_0(\lambda_t, \lambda)\, \mathcal{P}_{xx}(\lambda, \lambda_\tau)\, d\lambda - \mathcal{P}_{dn}(\lambda_t, \lambda_\tau) & (\lambda_t, \lambda_\tau) \in \Lambda_0^2 \\[2mm] 0 & \text{elsewhere} \end{cases} \tag{48c}$$

where it has been noted that if $\lambda_t \notin \{\Lambda_0 \cap \Lambda_d\}$, then $\mathcal{P}_{dx}(\lambda_t, \lambda_\tau) = 0 \Rightarrow \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) = -\mathcal{P}_{dn}(\lambda_t, \lambda_\tau) = 0$ since d(t) and n(t) only have spectral components which overlap in the re-

gion Λ_0. This proves the first part of the theorem. Since $\mathrm{d}(t)$ is a real process, the mean squared error from (4b) may be written as:

$$\sigma^2(t) \equiv R_{dd}(t, t) - \int_{\mathcal{T}} h(t, \alpha) R_{dx}(\alpha, t) \, d\alpha \qquad (49)$$

Writing equation (24a) in the expanded form

$$R_{dd}(t, \tau) = \iint_{\Lambda_d \Lambda_d} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) \, k(\lambda_t, t) \, k^*(\lambda_\tau, \tau) \, d\lambda_t \, d\lambda_\tau$$

$$+ \iint_{\Lambda_0 \Lambda_0} \mathcal{P}_{dd}(\lambda_t, \lambda_\tau) \, k(\lambda_t, t) \, k^*(\lambda_\tau, \tau) \, d\lambda_t \, d\lambda_\tau$$

and substituting this and the expanded expression in (25c) into (4b), gives, after the use of (6b) and (24c), the fact that $\Lambda_d \cap \Lambda_0 = \{\emptyset\}$, and considerable rearrangement, the desired expression in (25d). Further details are given in [2]. ∎

References

[1] J. R. Hopgood and P. J. W. Rayner, "Single channel signal separation using linear time-varying filters: Separability of non-stationary stochastic signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, Mar. 1999, vol. 3, pp. 1449–1452.

[2] J. R. Hopgood, *Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments*, Ph. D. Thesis, University of Cambridge, UK, Nov. 2000.

[3] V. Z. Marmarelis, "A single-record estimator for correlation functions of nonstationary random processes," *Proc. IEEE*, vol. 69, no. 7, pp. 841–842, July 1981.

[4] V. Z. Marmarelis, "Practical estimation of correlation functions of nonstationary Gaussian processes," *IEEE Transactions on Information Theory*, vol. IT-29, no. 6, pp. 937–938, Nov. 1983.

[5] W. A. Gardner, "Correlation estimation and time-series modeling for nonstationary processes," *Signal Processing*, vol. 15, no. 1, pp. 31–41, July 1988.

[6] J. Bendat, "A general theory of linear prediction and filtering," *Journal of the Society for Industrial and Applied Mathematics*, vol. 4, no. 3, pp. 131–151, Sept. 1956.

[7] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers*, vol. 93, pp. 429–457, 1946.

[8] L. Cohen, "Time-frequency distributions – a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941 – 981, July 1989.

[9] L. Cohen, "Time-frequency analysis," *IEEE Signal Processing Magazine*, vol. 16, no. 1, pp. 22–28, Jan. 1999.

[10] J. K. Hammond and P. R. White, "The analysis of non-stationary signals using time-frequency methods," *Journal of Sound and Vibration*, vol. 190, no. 3, pp. 419–447, Feb. 1996.

[11] R. A. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. IT-13, pp. 579–587, Sept. 1957.

[12] H. Cramér, "On some classes of nonstationary stochastic processes," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 1961, vol. 2, pp. 57–78, University of California Press.

[13] F. Donati, "Finite-time averaged power spectra," *IEEE Transactions on Information Theory*, vol. 17, no. 1, pp. 7–16, Jan. 1971.

[14] O. D. Grace, "Instantaneous power spectra," *Journal of the Acoustical Society of America*, vol. 69, no. 1, pp. 191–198, Jan. 1981.

[15] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis*, Academic Press, Inc., 1988.

[16] W. D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *Journal of Sound and Vibration*, vol. 11, no. 1, pp. 19–64, Jan. 1970.

[17] R. M. Loynes, "On the concept of the spectrum for non-stationary processes (with discussion)," *Journal of the Royal Statistical Society – Series B*, vol. 30, pp. 1 – 30, 1968.

[18] C. H. M. Turner, "On the concept of an instantaneous power spectrum, and its relationship to the autocorrelation function," *Journal Applied Physics*, vol. 23, no. 11, pp. 1347–1351, Nov. 1954.

[19] Y. H. Tsao, "Uncertainty principle in frequency-time methods," *Journal of the Acoustical Society of America*, vol. 75, no. 5, pp. 1532–1540, May 1984.

[20] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of non-stationary random processes," *IEEE Transactions on Signal Processing*, vol. 45, no. 6, pp. 1520–1534, June 1997.

[21] W. Kozek, *Matched Weyl-Heisenberg expansions of non-stationary environments*, Ph. D. Thesis, Vienna Univ. Technol., Vienna, Austria, Mar. 1997, Also appeared as Tech. Rep 96/I, NUHAG, Dept. Math., Univ. Vienna, Sept. 1996.

[22] C. S. Detka and A. El-Jaroudi, "The generalized evolutionary spectrum," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2877–2881, Nov. 1996.

[23] D. G. Lampard, "Generalization of the Wiener-Khintchine theorem to non-stationary processes," *Journal Applied Physics*, vol. 25, no. 6, pp. 802–803, June 1954.

[24] L. Cohen, "A general approach for obtaining joint representations in signal analysis. I. Characteristic function operator method," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1080–1090, May 1996.

[25] A. M. Sayeed and D. L. Jones, "Equivalence of generalized joint signal representations of arbitrary variables," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2959–2970, Dec. 1996.

[26] L. Cohen, "Generalization of the Wiener-Khinchin theorem," *IEEE Signal Processing Letters*, vol. 5, no. 11, pp. 292–294, Nov. 1998.

[27] H. M. Ozaktaz, Z. Zalevsky, and M. Apler-Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*, John Wiley & Sons, Inc., Chichester, UK, 2001.

[28] M. J. Levin, "Instantaneous spectra and ambiguity functions," *IEEE Transactions on Information Theory*, vol. IT-10, pp. 95–97, Jan. 1964, Correspondence.

[29] J. Kempé de Fériet and F. N. Frenkiel, "Correlations and spectra for non-stationary random functions," *Mathematics of Computation*, vol. 16, no. 77, pp. 1–21, 1962.

[30] W. D. Mark, "Power spectral representation of nonstationary random processes defined over semi-infinite intervals," *Journal of the Acoustical Society of America*, vol. 59, no. 5, pp. 1184–1194, May 1976.

[31] M. G. Amin, "Time-varying spectrum estimation for a general class of nonstationary processes," *Proc. IEEE*, vol. 74, no. 12, pp. 1800–1802, Dec. 1986.

[32] L. B. White and B. Boashash, "Cross spectral analysis of nonstationary processes," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 830–835, July 1990.

[33] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Co., Inc., third edition, 1991.

[34] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, pp. 291–299, Mar. 1950.

[35] L. A. Zadeh, "Correlation functions and power spectra in variable networks," *Proc. IRE*, vol. 38, pp. 1342–1345, Nov. 1950.

[36] L. A. Zadeh, "The determination of the impulsive response of variable networks," *Journal Applied Physics*, vol. 21, pp. 642–645, July 1950.

[37] P. A. Bello, "Time-frequency duality," *IRE Transactions on Information Theory*, vol. IT-10, pp. 18–33, Jan. 1964.

[38] C. Tamvaclis, *Separability of Finite Discrete Signals*, Ph. D. Thesis, University of Cambridge, UK, 1981.

[39] L. A. Zadeh, "Time-varying networks, I," *Proc. IRE*, vol. 49, pp. 1488–1503, Oct. 1961.

[40] A. Gersho, "Characterization of time-varying linear systems," *Proc. IEEE*, vol. 51, pp. 238, Jan. 1963.

[41] H. D'Angelo, *Linear Time-Varying Systems: Analysis and Synthesis*, Allyn and Bacon, Boston, 1970.

[42] J. A. Ball, I. Gohberg, and M. A. Kaashoek, "A frequency response function for linear, time-varying systems," *Mathematics of Control Signals and Systems*, vol. 8, no. 4, pp. 334–351, 1995.

[43] G. F. Margrave, "Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering," *Geophysics*, vol. 63, no. 1, pp. 244–259, Jan.-Feb. 1998.

[44] L. A. Zadeh, "A general theory of linear signal transmission systems," *Journal of the Franklin Institute*, vol. 253, pp. 293–312, 1952.

[45] L. A. Zadeh, "General input-output relations for linear networks," *Proc. IRE (Correspondence)*, vol. 40, pp. 103, Jan. 1952.

[46] L. A. Zadeh and K. S. Miller, "Generalised ideal filters," *Journal Applied Physics*, vol. 23, pp. 223–228, 1952.

[47] L. A. Zadeh, "On the theory of filtration of signals," *Zeitschrift fur Angenwante Mathematix und Physick*, vol. 111, pp. 149–156, 1952.

[48] C. Tamvaclis, "Filtration of separable finite discrete signals," *IEE Proceedings, Part F: Communications, Radar and Signal Processing*, vol. 128, no. 1, pp. 1–8, Feb. 1981.

[49] K. S. Miller and L. A. Zadeh, "Solution of an integral equation occurring in the theories of prediction and detection," *IRE Transactions on Information Theory*, vol. IT-2, pp. 72–72, June 1956.

[50] J. H. Laning and R. H. Battin, *Random Processes in Automatic Control*, McGraw-Hill Book Co., Inc., New York, N.Y., 1956.

[51] D. C. Youla, "The solution of a homogeneous Wiener-Hopf integral equation occuring in the expansion

of second-order stationary random processes.," *IRE Transactions on Information Theory*, vol. IT-13, pp. 187–193, Sept. 1957.

[52] M. Shinbrot, "On the integral equation occurring in optimization theory with nonstationary inputs," *Journal of Mathematics and Physics*, vol. 36, pp. 121–129, July 1957.

[53] M. Shinbrot, "A generalization of a method for the solution of the integral equation arising in optimization of time-varying linear systems with nonstationary inputs," *IRE Transactions on Information Theory*, vol. IT-3, pp. 220–224, Dec. 1957.

[54] S. Darlington, "Nonstationary smoothing and prediction using network theory concepts," *IRE Transactions on Circuit Theory (special suppl.)*, vol. CT-6, pp. 1–11, May 1959.

[55] M. Ghogho and B. Garel, "Maximum likelihood estimation of amplitude-modulated time series," *Signal Processing*, vol. 75, no. 2, pp. 99–116, June 1999.

[56] O. Besson and F. Castanie, "On estimating the frequency of a sinusoid in autoregressive multiplicative noise," *Signal Processing*, vol. 30, no. 1, pp. 65–83, Jan. 1993.

[57] O. Besson, M. Ghogho, and A. Swami, "On estimating random amplitude chirp signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, Mar. 1999, vol. 3, pp. 1561–1564.
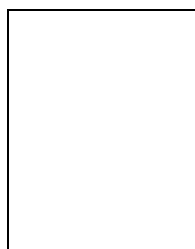
James R. Hopgood received the M. A., M.Eng. degree in Electrical and Information Sciences from the Univeristy of Cambridge in 1997, and a Ph.D. from Cambridge in July 2001. He is currently a Research Fellow at Queens' College, Cambridge, and is with the Signal Processing and Research Group at the Department of Engineering at Cambridge University working on nonstationary signal processing with application to reverberation cancellation in acoustic environments.
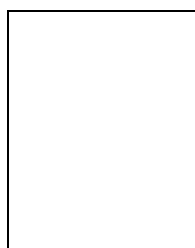
Peter J. W. Rayner received the M.A. degree from the University of Cambridge, UK, in 1968 and the Ph.D. degree from Aston University in 1969. Since 1968 he has been with the Department of Engineering at Cambridge University and is head of the Signal Processing Research Laboratory. In 1999 he was appointed an ad-hominem professorship in Signal Processing. He teaches courses in random signal theory, digital signal processing, image processing and communication systems. His current research interests include: optimal nonlinear filters for signal detection and estimation, nonlinear system identification and equalisation, blind identification of acoustic systems, signal separation, document and image sequence restoration, speech enhancement in low signal-to-noise ratio conditions, restoration of signals in nonstationary non-Gaussian noise, and Bayesian techniques in signal reconstruction and classification.