Non-parametric Wiener filter for reducing noise on reproducible pure signals

1999 J. Phys. A: Math. Gen. 32 3075

(http://iopscience.iop.org/0305-4470/32/17/301)

J. Phys. A: Math. Gen. **32** (1999) 3075–3094. Printed in the UK PII: S0305-4470(99)95528-2

Non-parametric Wiener filter for reducing noise on reproducible pure signals

Robert S Caprari†

Defence Science and Technology Organisation, PO Box 1500, Salisbury SA 5108, Australia

Received 29 June 1998, in final form 21 January 1999

Abstract. This paper communicates a novel method of Wiener filtering signals from scientific instruments to reduce their noise content. The main prerequisite for the applicability of this technique is that the pure (i.e. noiseless) signal is reproducible, a condition that is satisfied by a wide range of experimental measurements. The benefits of this Wiener filter design approach are its simplicity, generality and practicality. In particular, signal-dependent or multiplicative noise are accommodated without complication. The non-parametric filter does not require models of pure signal or noise statistics, and is exactly optimal for the observed (i.e. noisy) signal ensemble. Implementation of this Wiener filter only requires measurement of observed signal correlation functions. After deriving the classical stationary signal Wiener filter, the analysis is extended to derive the reproducible stationary pure signal Wiener filter. A distinction is made between non-adaptive Wiener filters derived from correlations computed as ensemble averages, and adaptive Wiener filters derived from correlations computed as time averages. The analysis is extended again to encapsulate non-stationary signal ensembles and further extended to synthesize Wiener filters based on statistically good observed signal correlation estimates for arbitrarily many pure signal reproductions.

1. Introduction

Measurements made using scientific instruments are always afflicted by 'noise', by which we mean any signal that is not experimentally reproducible and that we cannot ascribe to a specific physical mechanism. Experimenters and instrument designers often take deliberate action to suppress noise. There are three distinct approaches to achieving this. One is to physically reduce the amount of noise present by judiciously designing the experimental apparatus (e.g. cooling and shielding of critical electronic components). Another is to acquire measurements in a noise insensitive manner (e.g. modulating inputs, and detecting outputs using lock-in amplifiers). The third noise reduction technique is to statistically analyse the direct experimental results to better infer their underlying physical information content. We address the last method here.

Specifically, we consider measurements that take the form of an observed quantity depending on a second controlled quantity, the instrument being such that the controlled quantity is caused to vary over a substantial domain. The observed quantity is assumed to be a single-valued function of the controlled quantity (e.g. no hysteresis). Examples of controlled/observed quantities are wavelength/intensity (light spectrometers), energy/counts (electron spectrometers), angle/intensity (telescopes), time/intensity (time-resolved reflectometers), mass/counts (mass spectrometers) and position/height (scanning

† E-mail address: robert.caprari@dsto.defence.gov.au

0305-4470/99/173075+20$19.50

tunnelling microscopes). In this paper, the controlled quantity is generically denoted 'time' and the variation in time of the observed quantity is the observed 'signal'.

We assume that the controlled quantity is devoid of any experimental error (systematic or random) and that the observed quantity is subject to random error, but not systematic error. The importance of this assumption is that there is no need for deconvolution along with the noise reduction (Wertheim [1] reviews deconvolution of experimental measurements), although a generalization of our approach accomplishes both [2, section 13.3].

We propose a signal processing approach to noise suppression, which entails analysing a measured signal in a way that estimates the noise-free signal value at a particular time instant from the noisy signal values in a time interval, possibly with the assistance of prior knowledge of the measurement process. Within this filtering model a statistically optimal linear operator for reducing noise on signals is the classical *Wiener filter*, which has its roots in statistical communication theory, but also has a history of use in experimental data analysis. Examples of the use of Wiener filters for experimental signal noise reduction include spectroscopy [3–6], electron microscopy [7, 8], diffraction [9], scanning tunnelling microscopy [10–13], physiological imaging [14–17], film photography [18] and astronomical imaging [19].

The following presentation pertains to the classical Wiener filter, but from a quite novel perspective that leads to a simple, general and practical Wiener filter algorithm for experiments whose noise-free output signals (i.e. pure signals) are reproducible. When we refer to 'reproducible' pure signals in this paper we mean that any given observation of the pure signal can be invoked at least twice, but each time degraded by an observation from a different noise ensemble, with all noise ensembles having the same statistics. In essence, reproducible pure signals result from repeated noisy experimental measurements of the same physical phenomenon. In tune with reality, noise is not considered to be reproducible and consequently neither is the observed signal.

Our Wiener filter methodology has several attractive features not usually associated with Wiener filters. First, the filter function is completely and exactly determined only from signals observed during the normal course of experimentation. There is no need for a separate noise measurement experiment, which may not even be possible. Second, the filter function is automatically optimized for the actual pure signal and noise statistics encountered in practice, and not just some hypothesized parametric model for the pure signal or noise statistics. Third, this Wiener filter algorithm is inherently applicable to signal-dependent or multiplicative noise, without extra complication beyond the additive, signal-independent noise case. The only prerequisite for these desirable features is that the pure signal output of the scientific instrument be reproducible at least once beyond its original occurrence.

The simplicity of this Wiener filtering approach is apparent in all circumstances under which it is applicable. Its generality is best appreciated in those circumstances that usually make Wiener filtering problematic, that is, signal-dependent or multiplicative noise. Because these types of noise cannot be observed in isolation from the pure signal, traditional approaches to Wiener filter design tend to hypothesize a parametric model of signal statistics, so that the derived filter can only be approximately optimal. In contrast, our Wiener filter automatically yields the exact filter function (assuming that experimental estimates of signal statistics are exact), without the need for deliberate action to accommodate the particular kind of noise. Practical examples of signal-dependent or multiplicative noise include *speckle* [20–22] (interference of scattered coherent illumination) and *shot*, *Poisson* or *photon* noise [23–25] (random arrival times of particles).

The layout of the paper is as follows. We introduce our notation and method of analysis in section 2 by deriving the general stationary signal Wiener filter. The analysis is extended to the special case of stationary reproducible pure signals in section 3, where the central result is derived. Section 4 presents a discussion of two alternative classes of Wiener filters that may be derived from this analysis depending on whether signal statistics (specifically correlations) are computed by ensemble or time averages—ensemble averages yield Wiener filters that are non-adaptive with respect to individual pure signal and noise realizations, and time averages yield adaptive Wiener filters. In section 5 the analysis is extended again to make it applicable to a pertinent class of non-stationary signals. A further extension of the analysis is presented in section 6, which solves the practically important problem of constructing Wiener filters whose filter function is computed from statistically good observed signal correlation estimates for arbitrarily many pure signal reproductions.

2. General discrete time Wiener filter

To familiarize ourselves with the type of analysis needed to develop the specialized Wiener filter for reproducible pure signals, let us start by developing the general discrete time Wiener filter. The theoretical foundations of our analysis are described in texts such as that of Papoulis [26], and the general discrete time Wiener filter is examined by Haykin [27].

Statistical signal analysis theory in general, and Wiener filter theory in particular, are formulated in the context of ensembles of all possible signals of a given class, which are formally represented by stochastic processes. Even if in practice it transpires that a deterministic signal is to be filtered, the deterministic signal is formally treated as a single realization of a notional stochastic process, as described in section 4. Therefore, whether one or many members of an ensemble of experimental measurements is available, it is always appropriate to represent signals by stochastic processes and our analysis is undertaken on this basis.

Consider the discrete time stochastic processes:

$$\boldsymbol{x}(\xi, i) \equiv \text{ observed signal}$$
$$\boldsymbol{s}(\xi, i) \equiv \text{ pure signal}$$
$$\boldsymbol{n}(\xi, i) \equiv \text{ noise signal}$$
$$\boldsymbol{y}(\xi, i) \equiv \text{ filtered signal}$$

where i is the integer time index and ξ is the generally continuous sample space variable. The sample space is the totality of different possible signal functions in the ensemble. Our convention is to denote random variables and stochastic processes by bold symbols. The sample space argument ξ will usually be omitted, in which case it will be implied. Sometimes the time index i will also be omitted, in which case it will be implied that the stochastic process variable alone (e.g. $\boldsymbol{x}$) refers to all random variables $\boldsymbol{x}(i)$ at all time indices i.

Our analysis certainly will be valid for *additive noise*, that is,

$$\boldsymbol{x}(i) = \boldsymbol{s}(i) + \boldsymbol{n}(i). \tag{1}$$

By making the analysis sufficiently general, it will also be valid for *multiplicative noise*. Without loss of generality, let the multiplicative noise have unit mean, in which case

$$\boldsymbol{x}(i) = (1 + \boldsymbol{n}'(i))\boldsymbol{s}(i) \tag{2}$$

where, on introduction of the notation $\langle\cdot\rangle$ for expectation or ensemble average,

$$\langle \boldsymbol{n}'(i) \rangle = 0. \tag{3}$$

Then

$$\boldsymbol{x}(i) = \boldsymbol{s}(i) + \boldsymbol{n}'(i)\boldsymbol{s}(i) \tag{4}$$

and defining the new noise stochastic process

$$n(i) \equiv n'(i)s(i) \tag{5}$$

we substitute (5) into (4) to obtain (1).

Consequently, the basic relationship (1) applies whether the noise is additive or multiplicative, and we shall proceed on the basis of relation (1). The only possible complication arising from multiplicative noise is that (5) implies that the equivalent additive noise n in (1) must be dependent on the pure signal s, which need not be the case for inherently additive noise. Our analysis will be conducted to be inclusive of realistic signal-dependent noise processes, and we shall see that the reproducible pure signal Wiener filter is inherently applicable to such signal-dependent noise without any special precautions. So the Wiener filter designed on the basis of additive noise models is also valid for multiplicative noise models. Most physical noise processes are intrinsically additive in nature, but photographic film grain noise and speckle are two examples of noise influences that are often modelled at least in part as multiplicative noise. The design of Wiener filters for multiplicative noise has received some attention [28].

To proceed, we make two very reasonable assumptions about pure signals and noise:

Assumption A. n is zero mean:

$$\langle n(i) \rangle = 0 \quad \forall\, i.$$

For the multiplicative noise model, if the zero mean n' is uncorrelated with, but possibly dependent on, the pure signal s, then the effective additive noise n (5) is zero mean.

Assumption B. s and n are mutually uncorrelated, but possibly dependent:

$$\langle s(i)n(j) \rangle = \langle s(i) \rangle \langle n(j) \rangle = 0 \quad \forall\, i,\, j.$$

For the multiplicative noise model, if the zero mean n' is independent of s, then the effective additive noise n (5) is dependent on, but uncorrelated with, the pure signal s.

For the multiplicative noise model to satisfy both assumptions A and B, the multiplicative noise must be independent of the pure signal.

The conventional Wiener filter is time invariant and linear, and we will also impose upon it a finite impulse response (we shall often refer to the impulse response as the ‘filter function’). That is, the discrete time Wiener filter function has the finite domain (or ‘region of support’) $m \in [-M, M]$, within which the filter elements are f_m. The linear Wiener filter operates on x to yield y, as

$$y(i) = \sum_{m=-M}^{M} f_m x(i+m). \tag{6}$$

Time invariance of the filter is exhibited in the independence of filter elements f_m on time index i. The Wiener filter operation can be enhanced from the bare convolution of (6) to construct a ‘variationally optimal’ Wiener filter [29].

For the present assume that both the pure signal and noise are stationary stochastic processes, in which case the observed and filtered signals are also stationary. The stationarity assumption will be relaxed in section 5. If y is to estimate s, then the instantaneous square error is the stochastic process

$$\varepsilon^2(i) \equiv (y(i) - s(i))^2 \tag{7}$$

and the mean square error is the deterministic number

$$\langle \varepsilon^2 \rangle \equiv \langle (y(i) - s(i))^2 \rangle. \tag{8}$$

$\langle \varepsilon^2 \rangle$ is independent of time index i by virtue of the stationarity of signals. Substituting (6) into (8) explicitly shows the dependence of the mean-square error on the filter elements:

$$\langle \varepsilon^2 \rangle(f_m) \equiv \left\langle \left(\sum_{m=-M}^{M} f_m \boldsymbol{x}(i+m) - \boldsymbol{s}(i) \right)^2 \right\rangle. \tag{9}$$

The Wiener filter is the linear (time invariant) minimum mean-square error filter, so its filter function f_m is chosen so that the gradient of $\langle \varepsilon^2 \rangle(f_m)$ vanishes, that is,

$$\frac{\partial \langle \varepsilon^2 \rangle}{\partial f_m} = 0 \qquad m \in [-M, M]. \tag{10}$$

Equation (10) expands into an inhomogeneous system of linear equations, the general discrete time *Wiener–Hopf* equation:

$$\sum_{n=-M}^{M} \langle \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \rangle f_n = \langle \boldsymbol{s}(i)\boldsymbol{s}(i+m) \rangle \qquad m \in [-M, M]. \tag{11}$$

Apart from the most pathological and unrealistic of cases, the matrix with mnth element being $\langle \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \rangle$ is positive definite, hence non-singular, so the Wiener–Hopf equation (11) has a unique solution for the filter function. $\langle \varepsilon^2 \rangle(f_m)$ (9) is a quadratic in the filter elements that is bounded from below by zero, but is not bounded from above; consequently its unique stationary point must be a global minimum, as desired.

Notice that the general Wiener–Hopf equation (11) cannot be solved explicitly without some *a priori* information, because it makes reference to $\boldsymbol{s}$, which is not necessarily observable (if it were conveniently observable, there probably would be little need for a Wiener filter to estimate it!). It is this need to somehow estimate the autocorrelation of the pure signal when the pure signal is not directly observable that has received most attention in the diversity of design approaches to Wiener filtering. Note that the problem may be equivalently thought of as one of estimating the unobserved noise autocorrelation, since solution of this latter problem together with measurement of the observed signal autocorrelation yields the pure signal autocorrelation. Our solution to this problem in the case of reproducible pure signals will be revealed in section 3.

Introducing conventional definitions of pure signal and noise correlation functions (whenever we refer to a ‘correlation function’, it is implied that we are referring to an *auto*correlation function, unless otherwise stated)

$$R_{ss}(m) \equiv \langle \boldsymbol{s}(i)\boldsymbol{s}(i+m) \rangle \tag{12}$$

$$R_{nn}(m) \equiv \langle \boldsymbol{n}(i)\boldsymbol{n}(i+m) \rangle \tag{13}$$

we may also define the observed signal correlation function and by appealing to assumptions A and B derive the following relations:

$$\begin{aligned} R_{xx}(m-n) &\equiv \langle \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \rangle \\ &= R_{ss}(m-n) + R_{nn}(m-n). \end{aligned} \tag{14}$$

Stationarity of signals ensures that correlations are independent of time offset i.

Substituting (12) and (14) into (11) yields the simplified Wiener–Hopf equation:

$$\sum_{n=-M}^{M} R_{xx}(m-n) f_n = R_{ss}(m) \qquad m \in [-M, M]. \tag{15}$$

Correlation functions (12)–(14) have even parity,

$$\left.\begin{aligned} R_{ss}(-m) &= R_{ss}(m) \\ R_{nn}(-m) &= R_{nn}(m) \\ R_{xx}(-m) &= R_{xx}(m) \end{aligned}\right\} \quad \forall m. \tag{16}$$

Of necessity, this symmetry is also a property of the filter function,

$$f_{-n} = f_n \qquad n \in [-M, M] \tag{17}$$

thus showing that the Wiener–Hopf equations (15) have considerable redundancy. Removing this redundancy converts the system of $(2M+1)$ linear equations in $(2M+1)$ variables (15) into a formally equivalent, but computationally more efficient, system of $(M+1)$ linear equations in $(M+1)$ variables:

$$R_{xx}(m)f_0 + \sum_{n=1}^{M}(R_{xx}(m-n) + R_{xx}(m+n))f_n = R_{ss}(m) \qquad m \in [0, M]. \tag{18}$$

Substituting the solution of the Wiener–Hopf equation (15) into the definition of the mean-square error (9) yields the minimum mean-square error as

$$\langle \varepsilon^2 \rangle_{\min} = \sum_{m=-M}^{M} R_{nn}(m) f_m. \tag{19}$$

Note that for filter elements that are not the solution of the Wiener–Hopf equation, (19) does not give the mean-square error (9), so (19) does not have universal applicability beyond the context in which it is presented here.

If the filter region of support is unbounded (i.e. $M \to \infty$), then there are no circular convolution problems in Fourier transforming the Wiener–Hopf equation and the frequency domain filter function (also called the filter 'transfer function') is directly analogous to that for continuous time Wiener filters:

$$\begin{aligned} F_k &\equiv \text{ Fourier transform of } f_n \\ &= S_{ss}(k)/S_{xx}(k) \\ &= S_{ss}(k)/(S_{ss}(k) + S_{nn}(k)). \end{aligned} \tag{20}$$

Functions of the form $S(k)$ in (20) are power spectra, which are Fourier transforms of the corresponding correlation functions $R(n)$. Some universal relations between spectra and correlations have been analysed by Caprari [30]. Equation (20) is the most familiar form of the Wiener filter, but it is valid only for an unbounded filter region of support (this assumption is usually implied, rather than explicitly stated). We do not use (20) in our analysis, because we impose bounds on the filter region of support and notionally solve the Wiener–Hopf equation (18) directly in the time domain by matrix algebra. This policy is not only a matter of choice, but, as we shall see in section 5, is mandatory for designing Wiener filters for the non-stationary signal ensembles frequently encountered in experimentation. If the filter bounded region of support is much longer than the signal correlation times, imposition of bounds has an insignificant effect on the Wiener filter function and the direct solution of (18) gives essentially the time domain version of (20).

3. Reproducible pure signal Wiener filter

The Wiener filter derived in section 2 is quite general for stationary signals, since assumptions A and B are quite mild. However, we have already alluded to a fundamental problem in its application to practical situations: the Wiener–Hopf equation (15) makes reference to the statistics of the unobservable pure signal $\boldsymbol{s}$ through the correlation function $R_{ss}(m)$. It may be possible to specify R_{ss} by hypothesizing an analytic form based on defensible reasoning and estimating the values of any free parameters from the observed signal $\boldsymbol{x}$. It may be possible, with the expenditure of considerable time and effort, to actually observe sufficient samples of the pure signal s to estimate R_{ss} with reasonable confidence. It may be possible to indirectly measure R_{ss} from separate observations of x and n, but n is not separately observable for signal-dependent or multiplicative noise. None of these avenues may exist, and even if they do, the accuracy achieved in practice may not be satisfactory or the required effort may be excessive.

In this section we shall present an exact and simple solution to the problem of Wiener–Hopf equation specification for the special case of reproducible pure signals. Other research with a similar approach, although very different detail, is that of Galatsanos and Chin [31], in which they derive Wiener filters for pure signals that are multiply realizable as outputs of multiple linear systems, where distinct linear systems have different transfer functions; and Viberg *et al* [32] carry out parameter estimation for similar types of imperfectly reproducible pure signals. The essential distinction between imperfect reproducibility and noise is that the former pure signal perturbation is deterministic and the latter stochastic. Unlike noise, imperfect reproducibility of the pure signal destroys the simplicity and elegance of the following analysis, so it will not be admitted.

In most types of scientific instrumentation, the pure signal may be reproduced by repeating the measurement on the same sample, under the same conditions, in which case any discrepancy between the two measurements is attributable to noise, which is not reproducible. The truth of this assertion follows directly from the tenet of experimental reproducibility that is one of the pillars of the scientific method. Exceptions to this rule include instruments in which the sample is destroyed by the measurement process (e.g. some modes of electron microscopy on biological specimens, mass spectrometry on trace samples), measurements on rare and transient natural phenomena (e.g. supernovas, gravitational waves, earthquakes) and instruments that take a long time to acquire their measurement (e.g. coincidence spectrometers [33]). However, for the preponderance of circumstances the following assumption is valid for scientific instrumentation:

Assumption C. The pure signal is reproducible at least once, in addition to its original occurrence. This is formally accommodated by introducing two noise stochastic processes n_0 and n_1, with the same statistics as n, thus yielding in place of x the two stochastic processes

$$x_0 \equiv s + n_0 \qquad x_1 \equiv s + n_1.$$

$x_0(\xi, \cdot)$ and $x_1(\xi, \cdot)$ are two observed signals for the same state of a system ('state', as specified by a specific value of ξ, is defined by the specimen being probed and the properties of the probing instrument), being the state that produces the pure signal output $s(\xi, \cdot)$.

Assumption C prompts a further reasonable assumption about noise signals:

Assumption D. Noise signals n_0 and n_1 accompanying different occurrences of the same pure signal s are mutually uncorrelated, but possibly dependent:

$$\langle n_0(i) n_1(j) \rangle = \langle n_0(i) \rangle \langle n_1(j) \rangle = 0 \quad \forall\, i, j.$$

For the multiplicative noise model (2), if the zero mean n_0' and n_1' are independent of s and mutually uncorrelated (maybe dependent), then the effective additive noise signals n_0 and n_1 (5) are mutually uncorrelated, but certainly dependent.

Assumption D formalizes the usual situation of random errors in different experiments not influencing each other, which will be the case unless the noise has an extremely long correlation time extending from one experiment to the next, which may happen with $1/f$ noise.

We briefly consider why there is no inconsistency in the notion of a pure signal that is both stochastic and reproducible. The stochasticity of the pure signal reflects the fact that there is a multitude of pure signal realizations that emanate from possible experimental measurements, and all of these realizations collectively may be considered as the ensemble that defines the pure signal stochastic process $\boldsymbol{s}$. Different members of the ensemble would typically correspond to different samples being examined, under different experimental conditions and with different instrument settings. The reproducibility of the pure signal reflects the fact that any given experimental measurement that can be done once, may be repeated again using the same sample, same experimental conditions and same instrument settings; or stated formally, if a member of the pure signal ensemble is invoked once, it may be invoked again. Accordingly, the two observed signals $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ in assumption C may refer to the same pure signal $\boldsymbol{s}$, which is actually a stochastic process, without violating the bounds of what is possible in practice. Although the concept of a Wiener filter formally implies the existence of an ensemble of signal realizations, a meaningful Wiener filter may be derived from as few as one realization from the signal ensemble, this being the adaptive Wiener filter described in section 4.

The best Wiener filter operates on both observed signals $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ to yield the one filtered signal

$$\boldsymbol{y}(i) = \sum_{m=-M}^{M} f_m(\boldsymbol{x}_0(i+m) + \boldsymbol{x}_1(i+m)). \tag{21}$$

Since $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are statistically identical, by symmetry their Wiener filter coefficients are identical, as enforced in (21).

Introduce the two derived observed stochastic processes

$$\boldsymbol{x}_+(\xi, i) \equiv \text{ observed sum signal}$$
$$\boldsymbol{x}_-(\xi, i) \equiv \text{ observed difference signal}$$

defined as

$$\boldsymbol{x}_+ \equiv \boldsymbol{x}_1 + \boldsymbol{x}_0 = 2\boldsymbol{s} + \boldsymbol{n}_1 + \boldsymbol{n}_0 \tag{22}$$
$$\boldsymbol{x}_- \equiv \boldsymbol{x}_1 - \boldsymbol{x}_0 = \boldsymbol{n}_1 - \boldsymbol{n}_0. \tag{23}$$

In practice it is simple to compute these sum and difference signals from the directly observed signals $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, so $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ are truly observed signals. The reproducible pure signal Wiener filter operation (21) simplifies to

$$\boldsymbol{y}(i) = \sum_{m=-M}^{M} f_m \boldsymbol{x}_+(i+m). \tag{24}$$

An analogous derivation to that of section 2, supplemented with assumptions C and D, and definitions (22) and (23), gives the *reproducible pure signal Wiener–Hopf* equation:

$$\sum_{n=-M}^{M} R_{++}(m-n) f_n = \frac{1}{2}(R_{++}(m) - R_{--}(m)) \qquad m \in [-M, M]. \tag{25}$$

As usual, the sum and difference correlation functions are defined as

$$R_{++}(m-n) \equiv \langle \boldsymbol{x}_+(i+m)\boldsymbol{x}_+(i+n) \rangle \tag{26}$$
$$R_{--}(m) \equiv \langle \boldsymbol{x}_-(i)\boldsymbol{x}_-(i+m) \rangle. \tag{27}$$

The reproducible pure signal Wiener–Hopf equation (25) has a unique solution for the filter function. The elementary reproducible pure signal Wiener filter derived in this section is extended to the case of non-stationary signals in section 5, and multiply reproduced pure signals in section 6.

Since the sum and difference correlation functions have even parity,

$$\begin{aligned} R_{++}(-m) &= R_{++}(m) \\ R_{--}(-m) &= R_{--}(m) \end{aligned} \quad \forall\, m \tag{28}$$

it follows that the filter function also has even parity, just as for the general Wiener filter (17). These symmetry properties allow the $(2M+1)\times(2M+1)$ Wiener–Hopf matrix equation (25) to be reduced to a computationally preferable $(M+1)\times(M+1)$ matrix equation:

$$R_{++}(m)\,f_0 + \sum_{n=1}^{M}(R_{++}(m-n) + R_{++}(m+n))\,f_n = \frac{1}{2}(R_{++}(m) - R_{--}(m)) \qquad m \in [0, M]. \tag{29}$$

The minimum mean-square error is

$$\langle\varepsilon^2\rangle_{\min} = \frac{1}{2}\sum_{m=-M}^{M} R_{--}(m)\,f_m. \tag{30}$$

In the case of unbounded filter region of support ($M \to \infty$), the Wiener filter transfer function is

$$F_k = \frac{1}{2}\frac{(S_{++}(k) - S_{--}(k))}{S_{++}(k)}. \tag{31}$$

The Wiener–Hopf equation (15) is applicable for general stationary signals or, more precisely, those signals satisfying assumptions A and B. For stationary reproducible pure signals, being those for which assumptions A to D are satisfied, the general Wiener–Hopf equation (15) is exactly equivalent to the reproducible pure signal Wiener–Hopf equation (25). The major practical advantage of (25) over (15) is that (25) makes reference to only observed signals $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ through their correlation functions R_{++} and R_{--}. Equation (25) contains no reference to the unobservable pure signal, nor to noise, which is often unobserved, and sometimes unobservable, as for signal-dependent or multiplicative noise. This is the central result of the analysis: that in the reproducible pure signal case, the Wiener–Hopf equation is exactly expressible in terms of only observable signal statistics. Hence, to the extent that the correlation functions may be accurately estimated from practical signals, the reproducible pure signal Wiener–Hopf equation (25) is inherently realizable without further assumption. This is the great appeal of the reproducible pure signal Wiener filter.

The fact that the correlation functions in (25) can be experimentally measured obviates the need to hypothesise analytic models for the correlation functions, whose free parameters must then be estimated from the observed signals. This feature makes the reproducible pure signal Wiener filter *non-parametric* and is very desirable for reasons of simplicity of design. If experimental measurements of correlation functions yield superior accuracy to that available from analytic models of signal statistics, the non-parametric Wiener filter is also more robust than its parametric counterpart.

One further point that needs clarification is whether the reproducible pure signal Wiener filter derived in this section requires twice as much time to implement as the general Wiener filter of section 2, the former requiring the observation of two signals $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, while the latter only requiring observation of one signal $\boldsymbol{x}$. A doubling of measurement time may be mandated by the instrument design, but the way we conduct the filtering (21) implies that the integration time for acquiring $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ individually need only be half that of $\boldsymbol{x}$ for the unfiltered signals to be essentially equivalent. Apart from the overheads associated with acquiring two signals instead of one, the experiment exposure time is no longer for the reproducible pure signal case than for the single observation case. Yet only in the double observation case are the undoubted benefits of the reproducible pure signal Wiener filter available.

4. Adaptive and non-adaptive Wiener filters

As stated in section 2, the preceding analysis strictly only pertains to stationary stochastic processes, but is equally applicable to both ergodic and non-ergodic stochastic processes. In this section we elaborate on the implications for Wiener filters of ergodicity or non-ergodicity of stochastic processes. Although ergodic stochastic processes are conceptually simpler than non-ergodic ones, the latter type often arise from the multitude of pure signals that are yielded by experimental measurements (e.g. a stationary but non-ergodic ensemble can be as prosaic as an ensemble of images of outside scenes, some being day-time scenes and others night-time scenes). The implications for Wiener filters of non-stationarity of stochastic processes are examined in section 5.

By modelling the pure signal as a stochastic process, which is represented by an ensemble of realizations that may occur in experiments, the Wiener filter that we derive will be the best (in terms of minimum square error) linear estimator on average for all realizations in the ensemble, and not necessarily the best linear estimator for any individual member of the ensemble. Specifically, for a non-ergodic ensemble, each individual member of the ensemble in general will have its own individual optimum linear estimator that differs from the Wiener filter corresponding to the whole ensemble, the latter being the single linear estimator that is best on average over all members of the ensemble. Conversely, for an ergodic ensemble in which ensemble averages such as correlation functions are equivalent to the corresponding time averages along any individual member of the ensemble, the best linear estimator for any individual member of the ensemble is identical to the single best linear estimator on average for the whole pure signal ensemble, which is the Wiener filter resulting from the foregoing analysis.

In some practical circumstances, especially when the rate of acquiring signals is high or when real-time filtering is required, it is important to minimize the computational effort devoted to evaluating the Wiener filter function. Situations such as these mandate that a single Wiener filter be applicable to the complete pure signal ensemble, in which case sum and difference correlation functions should be estimated by appropriate ensemble averages over many different realizations of the observed sum and difference signals $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$. The separate realizations are obtained from separate experimental measurements with a diversity of samples being investigated, under a diversity of experimental conditions, and with a diversity of instrument settings. Following this prescription yields the single Wiener filter that is optimal on average for the totality of members of the ensemble. However, for a non-ergodic pure signal ensemble the derived Wiener filter will not necessarily be the optimal linear estimator for individual members of the ensemble. Only for an ergodic pure signal ensemble is the Wiener filter also the optimal linear estimator for individual members of the ensemble. This single Wiener filter for the overall pure signal ensemble is in a sense ‘non-adaptive’, because it does not adapt to the peculiarities of individual members of the ensemble.

It is always possible to derive the optimal linear estimator for a specific realization in the pure signal ensemble by notionally ‘constructing’ an ergodic pure signal ensemble whose ensemble averages (specifically correlations) are precisely the corresponding time averages of the pure signal realization whose estimate is being sought. Such a notional pure signal ensemble may be as simple as being composed of individual members that are distinct arbitrary time displacements of the true pure signal realization whose estimate is being sought. The Wiener filter derived from the preceding analysis conducted on this notional pure signal ensemble is not only the optimal linear estimator on average for all members of the ensemble, but by virtue of the ergodicity of the ensemble, the Wiener filter is also the optimal linear estimator of every individual member of the ensemble, and by construction this includes the specific realization of the actual pure signal ensemble whose estimate is being sought. Given that the notional pure signal ensemble is ergodic, its ensemble averages (specifically correlations) are equal to the corresponding time averages for any individual member of the notional ensemble. However, one member of this notional ensemble is the actual pure signal realization that arose in an experiment and was manifested in the corresponding members of the observed sum and difference signal ensembles that were measured. Consequently, time averages of a single realization of both the observed sum and observed difference signals are sufficient to derive the Wiener filter for the notional pure signal ensemble, which is also the optimal linear estimator of the specific realization in the actual pure signal ensemble whose estimate is being sought.

Thus we have established a formal procedure for deriving the optimal linear estimator for an individual realization of the actual pure signal ensemble. Of course, every distinct realization of the actual pure signal ensemble requires repetition of this procedure to derive a new optimal linear estimator. So there is a computational complexity penalty for seeking separate Wiener filters customized to each individual realization of the pure signal ensemble, as opposed to the single Wiener filter that is optimal on average over all members of the pure signal ensemble. These separate Wiener filters for individual members of the pure signal ensemble are in a sense 'adaptive', because each filter adapts to the peculiarities of its corresponding member of the ensemble.

For definiteness, the preceding discussion is in terms of only the pure signal ensemble, but exactly the same implications hold if the noise ensemble is substituted for the pure signal ensemble in the argument. Unlike the pure signal, signal-independent noise is often ergodic, and it is only the non-ergodicity of the pure signal that imposes the distinction between adaptive and non-adaptive Wiener filters that has been identified here. However, signal-dependent noise and multiplicative noise (5) necessarily will also be non-ergodic if the pure signal is non-ergodic, which is often the case.

An alternative but equivalent interpretation of adaptive Wiener filters is that they are a means of applying the concept of statistically optimal linear estimation to a deterministic signal, which is generalized to one realization of a notional ergodic stochastic process by the type of reasoning described in this discussion. The time invariant property imposed on Wiener filters implies that the derived filter is the linear estimator of the pure signal that is optimal on average over all times, rather than a time-varying instantaneously optimal linear estimator, but is otherwise customized to the specific characteristics of the deterministic signal being filtered.

5. Generalized correlations for non-stationary signals

As is conventional for Wiener filters, the preceding analysis has assumed stationary signal ensembles. However, ensembles of signals emanating from measuring instruments are not necessarily stationary. The most obvious non-stationarity—that otherwise stationary signals have a finite domain—usually poses no difficulty other than close to the boundaries of the domain, where the local inappropriateness of the Wiener filter is tolerated. A more subtle form of non-stationarity in instrument signal ensembles arises because the human operator often adjusts the instrument to put the most 'interesting' part of the signal towards the middle of the signal domain; a practice that is not conducive to attaining stationarity. This behaviour may produce an intrinsic non-stationarity in the instrument signal ensemble that should be taken into account in the Wiener filter design. However, the stationarity assumption is fundamental to the Wiener filter concept. We resolve this dilemma by invoking a generalization to the notion of stationarity that is compatible with both non-stationary signal ensembles and Wiener filter analysis. In effect, the strategy is to time integrate 'out' the non-stationarity and then formulate an average Wiener filter around the ensuing average signal statistics.

We demonstrate the approach only for the general Wiener filter of section 2, but the reasoning is entirely analogous for the reproducible pure signal Wiener filter that we introduced in section 3. The following analysis explicitly assumes a signal domain $i \in [-I, I]$, but it is implied that the signal is actually measured throughout the interval $i \in [-I-M, I+M]$. In practice, the signal measurement interval (i.e. $I+M$) and filter region of support (i.e. M) are chosen, thereby specifying the signal domain (i.e. I).

Integrate the instantaneous square error stochastic process $\varepsilon^2(i)$ (7) over the signal domain to obtain the time-averaged square error random variable

$$\overline{\varepsilon^2} \equiv \frac{1}{(2I+1)} \sum_{i=-I}^{I} (\boldsymbol{y}(i) - \boldsymbol{s}(i))^2. \tag{32}$$

The mean time-averaged square error is the deterministic number

$$\left\langle \overline{\varepsilon^2} \right\rangle \equiv \left\langle \frac{1}{(2I+1)} \sum_{i=-I}^{I} (\boldsymbol{y}(i) - \boldsymbol{s}(i))^2 \right\rangle \tag{33}$$

which, on substitution of (6), is expressed as an explicit function of the filter elements:

$$\left\langle \overline{\varepsilon^2} \right\rangle (f_m) \equiv \left\langle \frac{1}{(2I+1)} \sum_{i=-I}^{I} \left(\sum_{m=-M}^{M} f_m \boldsymbol{x}(i+m) - \boldsymbol{s}(i) \right)^2 \right\rangle. \tag{34}$$

The Wiener filter is the linear (time invariant) minimum mean (time-averaged) square error filter, obtained as the solution of

$$\frac{\partial \left\langle \overline{\varepsilon^2} \right\rangle}{\partial f_m} = 0 \qquad m \in [-M, M] \tag{35}$$

which expands into the non-stationary signal Wiener–Hopf equation:

$$\sum_{n=-M}^{M} \left(\frac{1}{(2I+1)} \sum_{i=-I}^{I} \langle \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \rangle \right) f_n = \frac{1}{(2I+1)} \sum_{i=-I}^{I} \langle \boldsymbol{s}(i)\boldsymbol{s}(i+m) \rangle \qquad m \in [-M, M]. \tag{36}$$

Equation (36) has a unique solution for the filter function.

To make the Wiener filter more computationally tractable, we make two further very reasonable assumptions about pure signals and noise:

Assumption E. The pure signal ensemble $\boldsymbol{s}$ is ‘stationary’ sufficiently close to the lower and upper bounds of the signal domain $i \in [-I, I]$, which is precisely stated as

$$\langle \boldsymbol{s}(-I+m)\boldsymbol{s}(-I+n) \rangle = \langle \boldsymbol{s}(I+m)\boldsymbol{s}(I+n) \rangle = \text{function of } (n-m) \qquad m, n \in [-M, M].$$

Assumption F. The noise ensemble $\boldsymbol{n}$ satisfies the same boundary stationarity condition as $\boldsymbol{s}$, that is,

$$\langle \boldsymbol{n}(-I+m)\boldsymbol{n}(-I+n) \rangle = \langle \boldsymbol{n}(I+m)\boldsymbol{n}(I+n) \rangle = \text{function of } (n-m) \qquad m, n \in [-M, M].$$

For the multiplicative noise model, if the zero mean $\boldsymbol{n}'$ is independent of $\boldsymbol{s}$, and satisfies this assumption, then the effective additive noise $\boldsymbol{n}$ (5) also satisfies this assumption.

Assumptions E and F are formal definitions of ‘generalized’ stationarity.

Note that these boundary stationarity assumptions are conditions on the average properties of complete ensembles; there is no requirement for individual realizations of the ensembles to somehow reflect these conditions. Stationary stochastic processes definitely satisfy these conditions. So do stochastic processes that are very non-stationary within interval $i \in (-I + M, I - M)$, but whose sample signals all have the same constant value (e.g. zero) everywhere outside this interval. Thus the type of non-stationarity that we have identified as being typical of instrument signal ensembles is compatible with assumptions E and F. Many other non-stationary stochastic processes also satisfy these conditions, so these conditions may justifiably be seen as generalizations of the stationarity condition that is conventionally imposed in Wiener filter derivations. In fact, any finite domain non-stationary stochastic process can be made compatible with assumptions E and F by zero padding sufficiently far on either side of the original domain and appropriately redefining the signal domain.

Quite generally we may define the pure signal, noise and observed signal generalized correlation functions:

$$\begin{aligned} R_{ss}(m) &\equiv \left\langle \frac{1}{(2I+1)} \sum_{i=-I}^{I} \boldsymbol{s}(i)\boldsymbol{s}(i+m) \right\rangle \\ &= \frac{1}{(2I+1)} \sum_{i=-I}^{I} \langle \boldsymbol{s}(i)\boldsymbol{s}(i+m) \rangle \end{aligned} \tag{37}$$

$$\begin{aligned} R_{nn}(m) &\equiv \left\langle \frac{1}{(2I+1)} \sum_{i=-I}^{I} \boldsymbol{n}(i)\boldsymbol{n}(i+m) \right\rangle \\ &= \frac{1}{(2I+1)} \sum_{i=-I}^{I} \langle \boldsymbol{n}(i)\boldsymbol{n}(i+m) \rangle \end{aligned} \tag{38}$$

$$\begin{aligned} R_{xx}(m,n) &\equiv \left\langle \frac{1}{(2I+1)} \sum_{i=-I}^{I} \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \right\rangle \\ &= \frac{1}{(2I+1)} \sum_{i=-I}^{I} \langle \boldsymbol{x}(i+m)\boldsymbol{x}(i+n) \rangle. \end{aligned} \tag{39}$$

By appealing to assumptions A, B, E and F we derive the relations

$$R_{xx}(m,n) = R_{xx}(m-n) = R_{ss}(m-n) + R_{nn}(m-n) \qquad m,n \in [-M, M] \tag{40}$$

just as for stationary stochastic processes (14).

The conventional definitions of correlations (12)–(14) omit the summation over i and the associated $1/(2I + 1)$ factor, which would give correlation functions that are independent of time index i only for stationary stochastic processes. Since we admit non-stationary stochastic processes, we must explicitly include the i summation (i.e. time integration) to remove the i dependence (i.e. time dependence) of the correlation functions. Even with this i summation, the correlation functions are in general functions of the ordered pair of time displacements (m, n). Assumptions E and F allow an i summation over a finite domain to give correlation functions that are functions of only the difference of time displacements $(m - n)$, just as for stationary stochastic processes. Correlations that are functions of only the difference of time displacements are estimated with better accuracy and less computational complexity than those that are functions of both time displacements separately.

Signals are observed for only finite time intervals in practical situations and it is only through assumptions E and F that we can make the same simplifications that are made under the more restrictive assumption of stationarity. The i summations in our definitions of generalized correlations, combined with assumptions E and F, effectively allow our analysis to formally proceed as if the stochastic processes were stationary, even though they are not necessarily so. In essence, we derive the one stationary signal Wiener filter that is on average best suited to the actual non-stationary signals. A much more complicated method of Wiener filtering non-stationary signals is to derive at each time sample the Wiener filter that is best suited to the instantaneous signal statistics, effectively yielding a time variant filter [19].

Substituting (37), (39) and (40) into (36) yields exactly the stationary signal Wiener–Hopf equation (15) of section 2, with correlations interpreted as the generalized correlations that we have introduced in this section. The remainder of the analysis proceeds exactly as in section 2, since assumptions E and F ensure the same symmetry properties (16) and (17) as in section 2.

Assumptions E and F maintain their validity for an unbounded filter region of support ($M \to \infty$) only for stationary stochastic processes, so the frequency domain form of the Wiener filter (20) only exists for stationary signals. We now see that the Wiener filter is generalizable to some non-stationary signals only when it is expressed in the time domain, but not the more common frequency domain.

The reproducible pure signal Wiener filter of section 3 is generalizable to non-stationary signals satisfying assumptions E and F in an analogous way to that revealed in this section. Once again, the same reproducible pure signal Wiener–Hopf equation (25) is obtained, where now the correlations are interpreted as generalized correlations, analogous to (39).

6. Statistically good observed signal correlation estimates for multiply reproduced pure signals

The reproducible pure signal Wiener filter synthesis of section 3 was predicated upon the existence of a single pure signal reproduction in addition to the original pure signal occurrence. We denote the ensuing filter as corresponding to pure signal 'multiplicity' or 'degeneracy' $D = 2$ (i.e. original plus one reproduction). $D = 2$ is the simplest possible reproducible pure signal Wiener filter and we now generalize to arbitrarily higher pure signal multiplicities ($D > 2$; original plus more than one reproduction). In principle, this problem is trivial, since the Wiener–Hopf equation (25) is still exactly correct for higher multiplicities, where $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are any two of the observed signals.

To make the problem theoretically interesting and more practically relevant, we assume that correlations are estimated with a degree of statistical uncertainty from a given finite amount of observed data (i.e. finite number of distinct pure signals multiplied by finite multiplicity multiplied by finite signal domain). This is how correlations are obtained in practice, and we assume that there is a practical need to reduce the amount of statistical uncertainty in correlation estimates. Accordingly we seek a formally correct multiply reproduced pure signal Wiener filter that is evaluated from statistically good observed signal correlation estimates.

Assumption C of section 3 generalizes by associating D statistically identical noise signals $\boldsymbol{n}_k$ and D observed signals $\boldsymbol{x}_k$ with a single reproducible pure signal $\boldsymbol{s}$ (it is now convenient for k to start counting up from 1):

$$\boldsymbol{x}_k \equiv \boldsymbol{s} + \boldsymbol{n}_k \qquad k \in [1, D]. \tag{41}$$

Assumption D generalizes to expressing the mutual orthogonality of all noise signal pairs:

$$\langle \boldsymbol{n}_k(i)\boldsymbol{n}_l(j)\rangle = \langle \boldsymbol{n}_k(i)\rangle\langle \boldsymbol{n}_l(j)\rangle = 0 \qquad \forall\, i, j, k, l \quad \text{with } k \neq l. \tag{42}$$

The best multiply reproduced pure signal Wiener filter operates on the 'observed sum signal'

$$\boldsymbol{x}_+ \equiv \sum_{j=1}^{D} \boldsymbol{x}_j = D\boldsymbol{s} + \sum_{j=1}^{D} \boldsymbol{n}_j \tag{43}$$

by the standard convolution operation (24). $\boldsymbol{x}_+$ in (43) is the arbitrary D generalization of the $D = 2$ observed sum signal in (22). $\boldsymbol{x}_+$ is characterized by the sum correlation function R_{++} (26) for stationary signals or an analogue of the observed signal generalized correlation (39) for non-stationary ensembles that conform to the generalized stationarity criteria of assumptions E and F. Using assumption B, (42), and the statistical identicality of all noise signals, the sum correlation becomes

$$R_{++}(m) = D^2 R_{ss}(m) + DR_{nn}(m). \tag{44}$$

Statistical estimators of $R_{++}(m)$ at every argument value m are constructed from observations of $\boldsymbol{x}_+(i)$ at all samples i. Details of the several possible estimators are an intricate matter that we will not address here. Assume momentarily, for the sake of argument, that $\boldsymbol{s}$ is a deterministic signal rather than a stochastic process; effectively we are only interested in a single realization s from the ensemble $\boldsymbol{s}$. Then the standard deviation of any estimator of $R_{++}(m)$ increases as $O(D^{3/2})$ as $D \to \infty$, while $R_{++}(m)$ (44) itself increases as $O(D^2)$ as $D \to \infty$, yielding the result

$$R_{++}(m) \text{ estimator: relative uncertainty } = O\left(\frac{1}{D^{1/2}}\right) \quad \text{as } D \to \infty. \tag{45}$$

In practice, only small values of D would be encountered, so the achievable decline of estimator relative uncertainty with increasing D does not strictly conform to (45). Nevertheless, (45) gives a rough measure of the benefit of increasing pure signal multiplicity for the precision of sum correlation estimates.

Although the observed sum signal extends trivially from the $D = 2$ case (22) to arbitrary D (43), extension of the observed difference signal from the $D = 2$ case (23) to arbitrary D is a more complicated matter. We begin with a prescription for the construction of a coefficient matrix that will be needed later in (52). Starting from the basis of R^D defined by the rows of the $D \times D$ matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{46}$$

use the Gram–Schmidt process of linear algebra [34] to derive the $D \times D$ orthogonal matrix

$$\Psi_D \equiv \left(\frac{D-1}{D}\right)^{1/2} \begin{bmatrix} 1/(D-1)^{1/2} & 1/(D-1)^{1/2} & \cdots & 1/(D-1)^{1/2} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2D} \\ \vdots & \vdots & & \vdots \\ \psi_{D1} & \psi_{D2} & \cdots & \psi_{DD} \end{bmatrix}. \tag{47}$$

Ψ_D matrices are effortlessly and exactly computed by symbolic mathematics packages [35], but for reference purposes we quote Ψ_D for multiplicities $D = 3$–8 in the appendix. The important features of matrix (46) are that the top row elements are all equal and that its rows in total constitute any basis of R^D. Any change in the starting basis that complies with these requirements produces another equally suitable starting basis, which will orthogonalize to different but equally suitable Ψ_D matrix elements ψ_{ij}.

Denoting the transpose matrix by a T superscript, Ψ_D (47) has the property

$$\Psi_D^{-1} = \Psi_D^{\mathrm{T}} \tag{48}$$

being an orthogonal matrix. More important for our purposes is the orthonormality of the row vectors of Ψ_D:

$$\sum_{j=1}^{D} \psi_{kj}^2 = D/(D-1) \qquad k \in [2, D] \tag{49}$$

$$\sum_{j=1}^{D} \psi_{kj} = 0 \qquad k \in [2, D] \tag{50}$$

$$\sum_{j=1}^{D} \psi_{kj}\psi_{lj} = 0 \qquad k \neq l \in [2, D]. \tag{51}$$

Equation (50) expresses the orthogonality of the first row of Ψ_D with respect to all other rows.

There are $D-1$ statistically beneficial 'observed compound signals' $\boldsymbol{x}'_k$, given by linear combinations of directly observed signals $\boldsymbol{x}_j$ with coefficients specified by rows 2 to D of Ψ_D (47):

$$\boldsymbol{x}'_k \equiv \sum_{j=1}^{D} \psi_{kj}\boldsymbol{x}_j = \sum_{j=1}^{D} \psi_{kj}\boldsymbol{n}_j \qquad k \in [2, D]. \tag{52}$$

The equality in (52) follows from assumption (41) and orthogonality relation (50). Observed compound signals are arbitrary multiplicity D generalizations of the observed difference signal (23) that arises for $D = 2$.

For stationary signals, observed compound signals (52) are characterized by conventional compound autocorrelation functions

$$R'_{kk}(m) \equiv \langle \boldsymbol{x}'_k(i)\boldsymbol{x}'_k(i+m)\rangle \qquad k \in [2, D] \tag{53}$$

and, by virtue of assumption (42) and orthogonality relation (51), vanishing compound cross-correlations

$$\langle \boldsymbol{x}'_k(i)\boldsymbol{x}'_l(i+m)\rangle = 0 \qquad \forall m \quad \text{and} \quad k \neq l \in [2, D]. \tag{54}$$

Non-stationary signals that conform to the generalized stationarity criteria of assumptions E and F are characterized by generalized compound autocorrelations that are analogous to (39) and vanishing generalized compound cross-correlations. Substituting (52) into (53), expanding, simplifying by (42) and (49) and expressing the stochastic equivalence of all noise signals $\boldsymbol{n}_k$ by endowing them with a common noise correlation $R_{nn}(m)$ (13), gives

$$R'_{kk}(m) = \frac{D}{(D-1)} R_{nn}(m) \qquad k \in [2, D]. \tag{55}$$

Note that all observed compound signals have identical correlations.

Statistical estimators of each compound correlation function (53) at every argument value are constructed from the corresponding observed compound signal. From (55), the superposition of all compound signal correlations has the property

$$\sum_{k=2}^{D} R'_{kk}(m) = D R_{nn}(m). \tag{56}$$

As the statistical estimator of superposition $\sum_k R'_{kk}(m)$, use the sum of estimators of the individual $R'_{kk}(m)$, so that the superposition estimator has a mean value that is $(D-1)$ times greater than the individual summand estimators. To obtain a similarly simple relation between standard deviations of individual summand and superposition estimators, we restrict assumption D of section 3, which generalizes to (42) for arbitrarily many pure signal reproductions, as follows.

Assumption D′. Noise signals $\boldsymbol{n}_k$ accompanying different occurrences of the same pure signal $\boldsymbol{s}$ are pure signal additive, pure signal independent, marginally normally (i.e. Gaussian) distributed and mutually independent.

Assumption D′ ensures that the $\boldsymbol{x}'_k$ are mutually independent and therefore statistical estimators of $R'_{kk}(m)$ are also mutually independent for different k, so that the estimator of $\sum_k R'_{kk}(m)$ has a standard deviation that is $(D-1)^{1/2}$ times greater than the $R'_{kk}(m)$ estimators. Consequently, the relative uncertainty of the estimator of superposition (56) declines with increasing multiplicity D as

$$\sum_{k=2}^{D} R'_{kk}(m) \text{ estimator: relative uncertainty } = O\left(\frac{1}{(D-1)^{1/2}}\right) \quad \forall\, D. \tag{57}$$

We shall see in (58) that estimates of $R_{++}(m)$ and $\sum_k R'_{kk}(m)$ suffice to specify the Wiener filter, and from (45) and (57) respectively, estimators of both quantities approximately decline as $O(1/D^{1/2})$ with increasing pure signal multiplicity D. This behaviour quantifies the benefits of larger D for the attainable precision of Wiener filter specification. Additional to this benefit is the fact that the unfiltered signal, being the observed sum signal (43), has the noise power diminished by a factor of D relative to the pure signal power, compared with the individual observed signals, as expressed by the comparison between (44) and (14). Accordingly, there are demonstrable benefits for the filtered signal quality accruing from increasing pure signal multiplicity. However, the ultimate decision on pure signal multiplicity will often be dictated more by practical considerations than theoretical principles.

A derivation analogous to that in section 2, supplemented by (41), (42), (44) and (56), yields the *multiply reproduced pure signal Wiener–Hopf* equation (cf (15), (25)):

$$\sum_{n=-M}^{M} R_{++}(m-n) f_n = \frac{1}{D}\left(R_{++}(m) - \sum_{k=2}^{D} R'_{kk}(m)\right) \qquad m \in [-M, M]. \tag{58}$$

Filtered signal minimum mean-square error is (cf (19), (30))

$$\langle \varepsilon^2 \rangle_{\min} = \frac{1}{D} \sum_{m=-M}^{M} \sum_{k=2}^{D} R'_{kk}(m) f_m. \tag{59}$$

For stationary signals and an unbounded filter region of support ($M \to \infty$), the filter transfer function is (cf (20), (31))

$$F_k = \frac{1}{D} \frac{\left(S_{++}(k) - \sum_{l=2}^{D} S'_{ll}(k)\right)}{S_{++}(k)}. \tag{60}$$

Acknowledgments

I am grateful for the support given to this research project by Dr Garry Newsam of Defence Science and Technology Organisation (DSTO), especially for his constructive criticism of preliminary drafts of this paper. I thank Mr Alvin Goh of The University of Western Australia for undertaking Monte Carlo simulations for this project during his tenureship of a vacation research scholarship at DSTO.

Appendix. Ψ_D, $D = 3$–8

$$\Psi_3 = \sqrt{\tfrac{2}{3}}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} & 0 & \frac{\sqrt{3}}{2} \end{bmatrix}$$

$$\Psi_4 = \frac{\sqrt{3}}{2}\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{\sqrt{2}}{3} & 0 & \frac{2\sqrt{2}}{3} & -\frac{\sqrt{2}}{3} \\ -\sqrt{\frac{2}{3}} & 0 & 0 & \sqrt{\frac{2}{3}} \end{bmatrix}$$

$$\Psi_5 = \frac{2}{\sqrt{5}}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{4} & 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4}\sqrt{\frac{5}{3}} & 0 & \frac{\sqrt{15}}{4} & -\frac{1}{4}\sqrt{\frac{5}{3}} & -\frac{1}{4}\sqrt{\frac{5}{3}} \\ -\frac{1}{2}\sqrt{\frac{5}{6}} & 0 & 0 & \sqrt{\frac{5}{6}} & -\frac{1}{2}\sqrt{\frac{5}{6}} \\ -\frac{1}{2}\sqrt{\frac{5}{2}} & 0 & 0 & 0 & \frac{1}{2}\sqrt{\frac{5}{2}} \end{bmatrix}$$

$$\Psi_6 = \sqrt{\tfrac{5}{6}}\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{5} & 1 & -\frac{1}{5} & -\frac{1}{5} & -\frac{1}{5} & -\frac{1}{5} \\ -\frac{1}{5}\sqrt{\frac{3}{2}} & 0 & \frac{2\sqrt{6}}{5} & -\frac{1}{5}\sqrt{\frac{3}{2}} & -\frac{1}{5}\sqrt{\frac{3}{2}} & -\frac{1}{5}\sqrt{\frac{3}{2}} \\ -\frac{1}{\sqrt{10}} & 0 & 0 & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{5}} & 0 & 0 & 0 & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\sqrt{\frac{3}{5}} & 0 & 0 & 0 & 0 & \sqrt{\frac{3}{5}} \end{bmatrix}$$

$$\Psi_7 = \sqrt{\tfrac{6}{7}}\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{6} & 1 & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6}\sqrt{\frac{7}{5}} & 0 & \frac{\sqrt{35}}{6} & -\frac{1}{6}\sqrt{\frac{7}{5}} & -\frac{1}{6}\sqrt{\frac{7}{5}} & -\frac{1}{6}\sqrt{\frac{7}{5}} & -\frac{1}{6}\sqrt{\frac{7}{5}} \\ -\frac{1}{2}\sqrt{\frac{7}{30}} & 0 & 0 & \sqrt{\frac{14}{15}} & -\frac{1}{2}\sqrt{\frac{7}{30}} & -\frac{1}{2}\sqrt{\frac{7}{30}} & -\frac{1}{2}\sqrt{\frac{7}{30}} \\ -\frac{1}{6}\sqrt{\frac{7}{2}} & 0 & 0 & 0 & \frac{1}{2}\sqrt{\frac{7}{2}} & -\frac{1}{6}\sqrt{\frac{7}{2}} & -\frac{1}{6}\sqrt{\frac{7}{2}} \\ -\frac{\sqrt{7}}{6} & 0 & 0 & 0 & 0 & \frac{\sqrt{7}}{3} & -\frac{\sqrt{7}}{6} \\ -\frac{1}{2}\sqrt{\frac{7}{3}} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\sqrt{\frac{7}{3}} \end{bmatrix}$$

$$\Psi_8 = \frac{1}{2}\sqrt{\frac{7}{2}}\begin{bmatrix} \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} & \frac{1}{\sqrt{7}} \\ -\frac{1}{7} & 1 & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{2}{7\sqrt{3}} & 0 & \frac{4\sqrt{3}}{7} & -\frac{2}{7\sqrt{3}} & -\frac{2}{7\sqrt{3}} & -\frac{2}{7\sqrt{3}} & -\frac{2}{7\sqrt{3}} & -\frac{2}{7\sqrt{3}} \\ -\frac{2}{\sqrt{105}} & 0 & 0 & 2\sqrt{\frac{5}{21}} & -\frac{2}{\sqrt{105}} & -\frac{2}{\sqrt{105}} & -\frac{2}{\sqrt{105}} & -\frac{2}{\sqrt{105}} \\ -\sqrt{\frac{2}{35}} & 0 & 0 & 0 & 4\sqrt{\frac{2}{35}} & -\sqrt{\frac{2}{35}} & -\sqrt{\frac{2}{35}} & -\sqrt{\frac{2}{35}} \\ -\sqrt{\frac{2}{21}} & 0 & 0 & 0 & 0 & \sqrt{\frac{6}{7}} & -\sqrt{\frac{2}{21}} & -\sqrt{\frac{2}{21}} \\ -\frac{2}{\sqrt{21}} & 0 & 0 & 0 & 0 & 0 & \frac{4}{\sqrt{21}} & -\frac{2}{\sqrt{21}} \\ -\frac{2}{\sqrt{7}} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\sqrt{7}} \end{bmatrix}$$

References

[1] Wertheim G 1975 Deconvolution and smoothing: applications in ESCA *J. Electron Spectrosc. Relat. Phenom.* **6** 239–51

[2] Press W, Teukolsky S, Vetterling W and Flannery B 1992 *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)

[3] Kosarev E and Pantos E 1983 Optimal smoothing of ‘noisy’ data by fast Fourier transform *J. Phys. E: Sci. Instrum.* **16** 537–43

[4] Seah M, Dench W, Gale B and Groves T 1988 Towards a single recommended optimal convolutional smoothing algorithm for electron and other spectroscopies *J. Phys. E: Sci. Instrum.* **21** 351–63

[5] Seah M and Dench W 1989 Smoothing and the signal-to-noise ratio of peaks in electron spectroscopy *J. Electron Spectrosc. Relat. Phenom.* **48** 43–54

[6] Riris H, Carlisle C, Warren R and Cooper D 1994 Signal-to-noise ratio enhancement in frequency-modulation spectrometers by digital signal processing *Opt. Lett.* **19** 144–6

[7] Lewis B and Sakrison D 1975 Computer enhancement of scanning electron micrographs *IEEE Trans. Circuits Syst.* **22** 267–78

[8] Marks L 1996 Wiener-filter enhancement of noisy HREM images *Ultramicrosc.* **62** 43–52

[9] Mierzwa B and Pielaszek J 1997 Smoothing of low-intensity noisy X-ray diffraction data by Fourier filtering: application to supported metal catalyst studies *J. Appl. Crystallogr.* **30** 544–6

[10] Park S and Quate C 1987 Digital filtering of scanning tunneling microscope images *J. Appl. Phys.* **62** 312–14

[11] Stoll E and Marti O 1987 Restoration of scanning-tunneling-microscope data blurred by limited resolution, and hampered by $1/f$-like noise *Surface Sci.* **181** 222–9

[12] Stoll E and Baratoff A 1988 Restoration and pictorial representation of scanning-tunneling-microscope data *Ultramicrosc.* **25** 149–53

[13] Stoll E 1991 Picture processing and three-dimensional visualization of data from scanning tunneling and atomic force microscopy *IBM J. Res. Develop.* **35** 67–77

[14] Madsen M 1990 A method for obtaining an approximate Wiener filter *Med. Phys.* **17** 126–30

[15] Lee T S and Min B G 1991 Image restoration in digital radiography using dual sensor Wiener filter *Med. Phys.* **18** 1132–40

[16] Jan J 1992 Experimental restoration of ultrasonic tomograms by modified Wiener filters *Arch. Acoust.* **17** 51–65

[17] Buonocore M and Maddock R 1997 Noise suppression digital filter for functional magnetic resonance imaging based on image reference data *Magn. Reson. Med.* **38** 456–69

[18] Naderi F and Sawchuk A 1978 Estimation of images degraded by film-grain noise *Appl. Opt.* **17** 1228–37

[19] Ford S, Welsh B, Roggemann M and Lee D 1997 Reconstruction of low-light images by use of the vector Wiener filter *J. Opt. Soc. Am.* A **14** 2678–91

[20] Goodman J 1976 Some fundamental properties of speckle *J. Opt. Soc. Am.* **66** 1145–50

[21] Sadjadi F 1990 Perspective on techniques for enhancing speckled imagery *Opt. Eng.* **29** 25–30

[22] Harvey E and April G 1990 Speckle reduction in synthetic-aperture-radar imagery *Opt. Lett.* **15** 740–2

[23] Qian H 1990 Inverse Poisson transformation and shot noise filtering *Rev. Sci. Instrum.* **61** 2088–91

[24] Frenkel A, Sartor M and Wlodawski M 1997 Photon-noise-limited operation of intensified CCD cameras *Appl. Opt.* **36** 5288–97

[25] Winzer P 1997 Shot-noise formula for time-varying photon rates: a general derivation *J. Opt. Soc. Am.* B **14** 2424–9

[26] Papoulis A 1991 *Probability, Random Variables, and Stochastic Processes* 3rd edn (New York: McGraw-Hill)

[27] Haykin S 1991 *Adaptive Filter Theory* 2nd edn (Englewood Cliffs, NJ: Prentice-Hall)

[28] Kondo K, Ichioka Y and Suzuki T 1977 Image restoration by Wiener filtering in the presence of signal-dependent noise *Appl. Opt.* **16** 2554–8

[29] Caprari R 1999 Variationally optimal Wiener filters, submitted for publication

[30] Caprari R 1998 Conjugate asymptotic properties of spectra and correlations *J. Phys. A: Math. Gen.* **31** 3205–17

[31] Galatsanos N and Chin R 1989 Digital restoration of multichannel images *IEEE Trans. Acoust. Speech Signal Processing* **37** 415–21

[32] Viberg M, Stoica P and Ottersten B 1995 Array processing in correlated noise fields based on instrumental variables and subspace fitting *IEEE Trans. Signal Processing* **43** 1187–99

[33] Storer P, Caprari R, Clark S, Vos M and Weigold E 1994 Condensed matter electron momentum spectrometer with parallel detection in energy and momentum *Rev. Sci. Instrum.* **65** 2214–26

[34] Anton H 1993 *Elementary Linear Algebra* 7th edn (New York: Wiley)

[35] Wolfram S 1996 *The Mathematica Book* 3rd edn (Cambridge: Wolfram Media/Cambridge University Press)