N72- 21120

ESL-P-473

February, 1972

OPTIMAL FILTER DESIGN SUBJECT TO OUTPUT SIDELOBE CONSTRAINTS*

by

CASE FILE COPY

Thomas E. Fortmann
University of Newcastle
New South Wales, Australia

and

Michael Athans
Massachusetts Institute of Technology
Cambridge, Massachusetts

MAR 1972
RECEIVED
NASA STI FACILITY
INPUT BRANCH

* This research was supported by the National Science Foundation under grant GK-2645, by the National Aeronautics and Space Administration under grant NGL-22-009(124), and by the Australian Research Grants Committee.

ABSTRACT

The design of filters for detection and estimation in radar and communications systems is considered, with inequality constraints on the maximum output sidelobe levels. A constrained optimization problem in Hilbert space is formulated, incorporating the sidelobe constraints via a partial ordering of continuous functions. Generalized versions (in Hilbert space) of the Kuhn-Tucker and Duality Theorems allow the reduction of this problem to an unconstrained one in the dual space of regular Borel measures. A convergent algorithm is presented for computational solution of the dual problem.

1. INTRODUCTION

In many radar and communications systems, pulses of relatively long time duration are transmitted because of peak power limitations, and an operation known as "pulse compression" is performed at the receiver. This compression is most commonly achieved with a matched filter [1-7], that is, by correlating the incoming signal with time- and/or frequency-shifted copies of the transmitted waveform, a technique which is well-known to be optimal with respect to various performance criteria. The presence of a signal is detected when the matched filter output exceeds a threshold value, and parameters such as time delay and (Doppler) frequency shift are estimated by locating the peak output in time and frequency. One example of such a system is a pulse-position modulation communication scheme [7-10], where information is coded in the time delays of individual pulses and then decoded by estimating the delay with a matched filter. Another is a linear FM, or "chirp" radar [2-4], where the frequencies of transmitted pulses are swept linearly with time and the receiver has a linear time delay vs. frequency characteristic of the opposite slope.

If a signal s, non-zero on the interval (t_0, t_0+T), appears at the input to a pulse-compression filter, the output* typically consists of a main peak surrounded by sidelobes, as shown in Figure 1, and it is desirable in certain cases to reduce or restrict the height of these sidelobes. This becomes

---

* In the case of a matched filter, this is simply the autocorrelation function of s.

important, for instance, when a radar must distinguish among multiple targets, or when a pulse-position modulation system operates in an environment of severely degraded signal-to-noise ratio [8]. This paper is concerned with the design of optimal "mismatched" filters for detection and estimation, subject to inequality constraints on the sidelobe levels.

Previous work in this area has been principally concerned with analogous problems of designing antenna arrays [2-3] or with discrimination against statistically-distributed clutter [2-5], an approach which is often equivalent to reducing the sidelobe energy [12]. Another method [9-11] involves constraining the sidelobes approximately, with inequalities at a finite set of times, and leads to an optimization problem in terms of ordinary differential equations with delays.

In the next section attention is restricted to finding an optimal filter for detection, and a problem is formulated with the objective of maximizing the probability of detection subject to constraints on the sidelobe levels. Section 3 summarizes some results of constrained optimization theory in function space, which is applied in Section 4 to reduce the detection filter problem to an unconstrained minimization in the dual space. An algorithm for the solution of this dual problem is developed in Section 5. Finally, the problem of determining an optimal filter for estimation is formulated in Section 6, and its solution is seen to be analogous to that of the detection filter problem.

2. Receiver model and formulation of detection filter optimization problem

The receiver model to be considered is shown in Figure 2, its possible interpretations as part of a communication, radar, or other system [1-7] being left unspecified. The signal s is assumed to be a given real continuous function of time with support in $[0,T]$ and unit energy, that is

$$||s||^2 = \int_{-\infty}^{\infty} s^2(t)\, dt = \int_0^T s^2(t)\, dt = 1 \tag{2-1}$$

It appears at the input to the receiver, after a delay t_0, corrupted by zero-mean, stationary, gaussian white noise n with power spectral density N_0. The impulse response of the linear, time-invariant filter is to be determined, and is restricted to be square-integrable with support in $[0,T]$. The two terms y and ξ of the output, due respectively to the signal and noise, are given by

$$y(t) = \int_{-\infty}^{\infty} s(\tau-t_0)h\,(t-\tau)\, d\tau \tag{2-2}$$

$$\xi(t) = \int_{-\infty}^{\infty} n(\tau)h\,(t-\tau)\, d\tau \tag{2-3}$$

and it is an easy matter to show that the output noise power is

$$E\{\xi^2(t)\} = N_0\int_{-\infty}^{\infty} h^2(t-\tau)\, d\tau = N_0\int_{-\infty}^{\infty} h^2(\tau)\, d\tau \qquad \text{for all } t \tag{2-4}$$

where E denotes the expectation operator.

It is convenient at this point to replace the filter impulse response h with an equivalent "receiver function" u defined by*

$$u(t) \triangleq h(T-t), \tag{2-5}$$

shift the time origin to $t_0 + T$, and consider the output of the filter to be the cross-correlation function ψ defined by

$$\begin{aligned}\psi(t) &\triangleq y(t+t_0+T)\\ &= \int_{-\infty}^{\infty} s(\tau-t_0)h(t+t_0+T-\tau)\,d\tau\\ &= \int_{-\infty}^{\infty} s(\tau+t)h(T-\tau)\,d\tau\\ &= \int_{-\infty}^{\infty} s(\tau+t)u(\tau)\,d\tau\end{aligned} \tag{2-6}$$

Because s and u both have support in $[0,T]$, ψ must have support in $[-T,T]$ and

$$\psi(t) = \int_0^T u(\tau)s(\tau+t)\,d\tau = \langle u, s_t\rangle \tag{2-7}$$

---

* By an abuse of terminology, u will be referred to as "the filter."

$$E\{\xi^2(t)\} = N_0 \int_0^T u^2(\tau)\, d\tau = N_0 ||u||^2 \quad \text{for all} \quad t \tag{2-8}$$

where* s_t is the shifted time function defined by

$$s_t(\tau) \triangleq \begin{cases} s(\tau+t) & \tau \in [0,T] \\ 0 & \text{otherwise} \end{cases} \tag{2-9}$$

For this model it is well-known that the detection scheme which maximizes the probability of detection consists of processing the input with a matched filter ($u = s$) and comparing the output $[\psi(0)+\xi(0)]$ to a threshold value. Assuming that the threshold is specified independently (often so as to ensure an acceptable "false alarm" probability), the detection probability depends only upon the output signal-to-noise ratio, in this case $1/N_0$.

More generally, if a threshold-type detection scheme is used with a filter which is not necessarily matched to the signal, the probability of detection (at $t = 0$) for this model is easily shown to depend monotonically upon the output signal-to-noise ratio

$$(S/N) = \psi^2(0) \,/\, E\{\xi^2\} = \langle u,s\rangle^2 \,/\, N_0 ||u||^2 \tag{2-10}$$

The problem here will be to find a filter u which maximizes this value subject to the sidelobe constraint**

---

* All norms and inner products refer to $L^2[0,T]$ unless labelled otherwise by a subscript.

** It is a straightforward generalization to replace ε by $\varepsilon(t)$

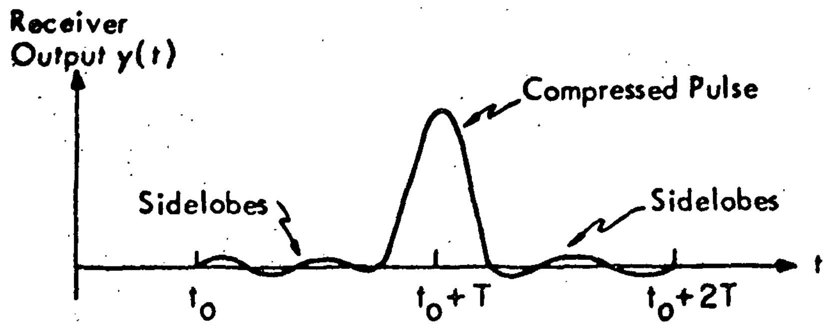

Fig. 1 Typical Compressed Pulse

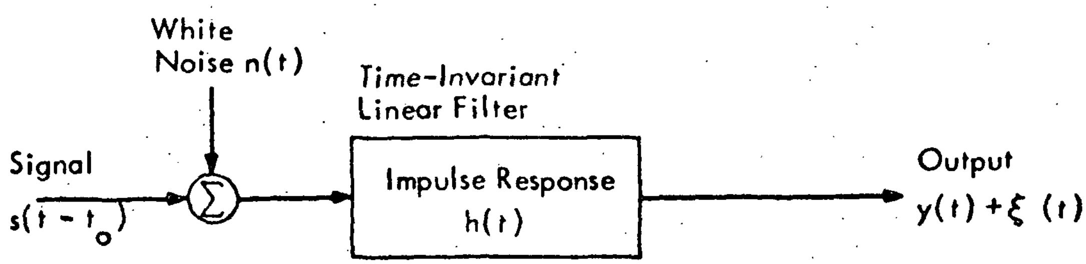

Fig. 2 Receiver Model

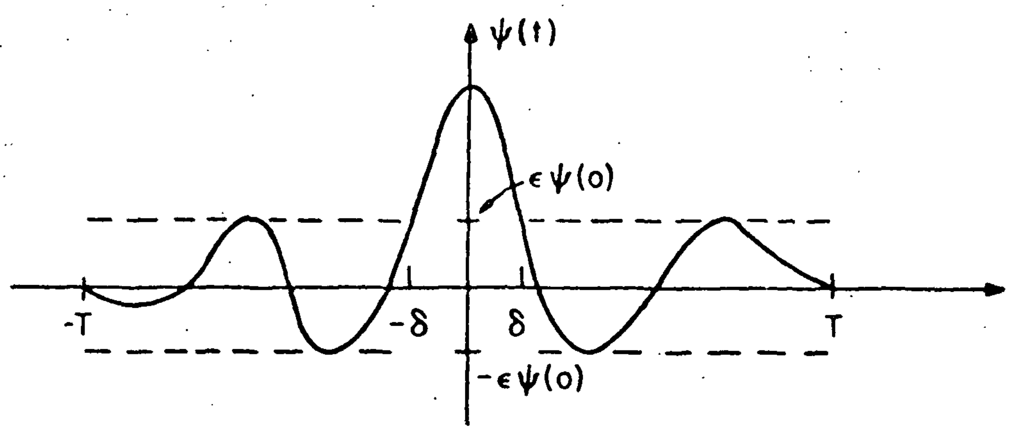

Fig. 3 Sidelobe Constraints

$$\max_{\delta \leq |t| \leq T} \; |\,\psi(t)\,| \leq \varepsilon\,|\,\psi(0)| \tag{2-11}$$

which is illustrated in Figure 3. This allows for a central lobe of width 2δ about $t = 0$ and restricts the magnitude of $\psi(t)$ elsewhere to be less than a fraction ε of $\psi(0)$, where δ and ε are parameters to be chosen by the designer.

It is implicit in the constraint (2-11) that the peak value of $|\psi(t)|$ in fact occurs at $t = 0$, an assumption which is justified intuitively by the fact that $|\psi(0)|$ will be the performance index to be maximized. This in turn is justified by the observation that maximizing at any $t \neq 0$ corresponds either to making a decision before the entire signal has arrived or to delaying the decision until some time after it has.

It can be assumed with no loss of generality that $\psi(0) = \langle u,s\rangle \geq 0$, since (2-10) and (2-11) are the same for both u and $-u$. But then the maximization of (2-10) subject to (2-11) is equivalent to maximizing

$$\langle u/||u||,s\rangle \tag{2-12}$$

subject to the constraints

$$\left.\begin{array}{r} \langle u,s_t\rangle \leq \varepsilon\,\langle u,s\rangle \\ -\quad \langle u,s_t\rangle \leq \varepsilon\,\langle u,s\rangle \end{array}\right\} \;\delta < |t| \leq T \tag{2-13}$$

Next, note that scaling u has no effect on (2-13), so that maximizing (2-12) is equivalent to maximizing $\langle u,s\rangle$ over all u of unit energy $(||u||^2 = 1)$. In fact, it is also equivalent (and more convenient) to maximize $\langle u,s\rangle$ over the unit ball $\{u : ||u||^2 \le 1\}$, since a nonzero solution of this latter problem will clearly have $||u||^2 = 1$. Thus, the detection filter optimization problem has assumed the following form:

$$\left.\begin{array}{ll} \text{maximize} & \langle u,s\rangle \ , \quad u \in L^2[0, T] \\ \text{subject to} & \langle u,s_t\rangle \le \varepsilon \langle u,s\rangle \ , \ \delta \le |t| \le T \\ & -\langle u,s_t\rangle \le \varepsilon \langle u,s\rangle \ , \ \delta \le |t| \le T \\ & ||u||^2 \le 1 \end{array}\right\} \tag{2-14}$$

3. Constrained optimization theory

In this section some results from the theory of constrained minimization in normed linear spaces are stated, and in the next section they will be applied to the filter optimization problem just formulated. This presentation follows Luenberger [13], but the theory may also be found in Hurwicz [14] and elsewhere.

Let U and Z_i, $i = 1,2, \cdots n$ be real normed linear spaces, with convex cones $P_i^+ \subseteq Z_i$, $i = 1,2, \cdots n$. Each cone P_i^+ generates a partial ordering† on Z_i, denoted $\preccurlyeq$ and defined by

$$x \preccurlyeq y \overset{\Delta}{\Longleftrightarrow} y-x \in P_i^+ , \quad x,y \in Z_i \tag{3-1}$$

P_i^+ is called the positive cone of Z_i, since $P_i^+ = \{z \in Z_i: \ z \succcurlyeq 0\}$, and the notation $x \prec y$ indicates that $y-x$ is an interior point of P_i^+. The dual positive cone $P_i^{\oplus}$ in the dual space Z_i^* is defined by

$$P_i^{\oplus} \triangleq \{z^* \in Z_i^* : \ z^* \cdot z \geq 0 \quad \forall z \in P_i^+ \} \tag{3-2}$$

The function $G_i : U \rightarrow Z_i$ is said to be convex (with respect to the partial ordering $\preccurlyeq$) if

$$G_i[\alpha u+(1-\alpha)v] \preccurlyeq \alpha G_i(u) + (1-\alpha)G_i(v), \quad u,v \in U, \ \alpha \in [0, 1] \tag{3-3}$$

---

† A partial ordering is defined as in Dunford and Schwartz [15] to be a transitive and reflexive but not necessarily antisymmetric relation. Thus $x \preccurlyeq y$ and $y \preccurlyeq x$ may not imply that $x = y$ (i.e. Z_i may not be well-ordered).

If the ordering sign is reversed in (3-3), then the function is said to be <u>concave</u>. Note that if $G_i : U \to Z_i$ is convex in the sense of (3-3) and $p_i \in P_i^{\oplus} \subseteq Z_i^*$, (i.e. $p_i \succcurlyeq 0$), then the functional $g_i : U \to R^1$ defined by

$$g_i(u) \triangleq p_i \cdot G_i(u), \qquad u \in U \tag{3-4}$$

is convex in the ordinary sense.

Consider the optimization problem

$$\left.\begin{array}{ll} \text{minimize} & f(u)\ ,\ u \in U \\ \text{subject to} & G_i(u) \preccurlyeq 0\ ,\ i=1,\ 2,\ \cdots\ n \end{array}\right\} \tag{3-5}$$

where $f : U \to R^1$ and $G_i : U \to Z_i$, $i=1, 2, \cdots n$ are all convex. The <u>Lagrangian functional</u> $L : U \times Z_1^* \times Z_2^* \times \cdots \times Z_n^* \to R^1$ for problem (3-5) is defined by

$$L(u\ ;\ p_1,\ p_2,\ \cdots\ p_n) \triangleq f(u) + \sum_{i=1}^{n} p_i \cdot G_i(u) \tag{3-6}$$

If $p_i \in P_i^+ \subseteq Z_i^*$, $i = 1, 2, \cdots n$, are all given then $L(\cdot;\ p_1,\ p_2,\ \cdots\ p_n)$ is a <u>convex</u> functional on U.
The <u>dual functional</u> $\phi : Z_1^* \times Z_2^* \times \cdots \times Z_n^* \to R^1$ for problem (3-5) is defined by

$$\phi\ (p_1,\ p_2,\ \cdots\ p_n) \triangleq \inf_{u \in U} L(u;\ p_1,\ p_2,\ \cdots\ p_n)\ , \tag{3-7}$$

and it is easily verified [13,12] that ϕ is a <u>concave</u> functional.

Using these definitions, the following theorems apply to problem (3-5)

Theorem 3.1: (Generalized Kuhn-Tucker conditions) For $i = 1, 2, \cdots n$, let U, Z_i be real normed linear spaces, as above, with each of the positive cones $P_i^+ \subseteq Z_i$ having a non-empty interior. Let $f: U \to R^1$ and $G_i: U \to Z_i$ be convex, as above, and assume the existence of $\bar{u} \in U$ such that $G_i(\bar{u}) \prec 0$ for all i.

Then if $u^0 \in U$ solves the problem

$$\left.\begin{aligned} &\text{minimize } f(u), \quad u \in U \\ &\text{subject to } G_i(u) \preccurlyeq 0, \quad i=1,2,\ldots,n \end{aligned}\right\} \tag{3.8}$$

there must exist Lagrange multipliers $0 \preccurlyeq p_i^0 \in Z_i^*$, $i=1, 2, \cdots n$, such that

(a) u^0 solves the problem

$$\text{minimize } L(u\,;\,p_1^0, p_2^0, \cdots p_n^0)\,,\; u \in U \tag{3-9}$$

where L is the convex Lagrangian functional defined in (3-6)

(b) $$p_i^0 \cdot G_i(u^0) = 0\,,\; i=1, 2, \cdots n \tag{3-10}$$

Proof: See [13, p.217], where it is proved for a single inequality constraint $G(u) \preccurlyeq 0$ in the space Z with positive cone P^+. This theorem becomes equivalent if one identifies Z as the Cartesian product space

$$Z = Z_1 \times Z_2 \times \cdots \times Z_n\,, \tag{3-11}$$

$G: U \to Z$ as the mapping

$$G(u) = [G_1(u), G_2(u), \cdots G_n(u)], \tag{3-12}$$

and the primal and dual positive cones as

$$\begin{aligned} P^+ &= P_1{}^+ \times P_2{}^+ \times \cdots \times P_n{}^+ \\ P^\oplus &= P_1{}^\oplus \times P_2{}^\oplus \times \cdots \times P_n{}^\oplus \end{aligned} \tag{3-13}$$

The details are straightforward and will be omitted.

The above theorem generalizes the well-known Kuhn-Tucker conditions for convex nonlinear programming problems in finite-dimensional spaces. Condition (3-10) expresses the fact that a Lagrange multiplier is non-zero only where the corresponding constraint is "active". The following duality theorem provides a method for finding the Lagrange multipliers.

<u>Theorem 3.2</u>: (Duality) In Theorem 3.1, the Lagrange multipliers $p_i^0 \in Z_i{}^*$, i=1, 2, $\cdots$ n, solve the problem

$$\left.\begin{array}{l} \text{maximize} \quad \phi(p_1, p_2, \cdots p_n) \\ \text{subject to} \quad p_i \succcurlyeq 0, \ i=1, 2, \cdots n \end{array}\right\} \tag{3-14}$$

and

$$\min_{G_i(u) \preccurlyeq 0} f(u) = \max_{p_i \succcurlyeq 0} \phi(p_1, p_2, \cdots p_n) \tag{3-15}$$

where ϕ is the concave dual functional defined in (3-7). Moreover, conditions (a) and (b) of Theorem 3.1 hold for <u>any</u> $p_1, p_2, \cdots p_n$ which solve (3-14).

Proof: See [13, p.224]. The last statement follows by examining the proof [13, p.218] of Theorem 3.1, which consists of showing the existence of a hyperplane, defined by $p_1^0, p_2^0, \cdots p_n^0$, which separates two sets in $R^1 \times Z_1 \times Z_2 \times \cdots \times Z_n$. It is easily verified that any $p_1, p_2, \cdots p_n$ which solve (3-14) also define a separating hyperplane, and hence must yield u^0 through (3-9).

4. Application to detection filter optimization problem

The optimization theory of the previous section will now be applied to the detection filter problem formulated in Section 2. In order to consider (2-14) in the form (3-5) let

$$\left.\begin{aligned} &U = L^2[0, T] \\ &Z_1 = Z_2 = C[-T, T] \\ &Z_3 = R^1 \end{aligned}\right\} \qquad (4\text{-}1)$$

and define f, G_1, G_2, G_3 as follows:

$$\left.\begin{aligned} &f(u) = -\langle u, s\rangle \\ &[G_1(u)](t) = \langle u, s_t\rangle - \varepsilon \langle u, s\rangle \;,\; t\epsilon[-T, t] \\ &[G_2(u)](t) = -\langle u, s_t\rangle - \varepsilon \langle u, s\rangle \;,\; t\epsilon[-T, t] \\ &G_3(u) = ||u||^2 - 1 \end{aligned}\right\} \qquad (4\text{-}2)$$

The fact that G_1 and G_2 map L^2 functions into continuous functions is easily established by noting that the translation defined in (2-9) is a uniformly continuous mapping from R^1 to $L^2[-\infty, \infty]$ [16, p.183]. The four mappings in (4-2) are all convex, the first three by virtue of their linearity and the last because the norm is convex.

The detection filter optimization problem (2-14) can now be written as

$$\left.\begin{aligned} &\text{minimize} \quad f(u) \;,\; u \in L^2[0, T] \\ &\text{subject to} \quad G_i(u) \preccurlyeq 0 \;,\; i=1, 2, 3 \end{aligned}\right\} \qquad (4\text{-}3)$$

where the ordering in $Z_3 = R^1$ is the natural one (i.e. $P_3^{+} = R^{+}$, the nonnegative reals, and the partial ordering $\preccurlyeq$ in $Z_1 = Z_2 = C[-T, T]$ is generated by the positive cone

$$P_1^+ = P_2^+ = P_C^+ \triangleq \{x \in C[-T, T] : x(t) \geq 0, \ \delta \leq |t| \leq T\}, \tag{4-4}$$

In other words,

$$x \preccurlyeq y \iff x(t) \leq y(t) \ , \ \delta \leq |t| \leq T \tag{4-5}$$

It is easy to verify that P_C^+ is a closed, convex cone with nonempty interior † [12].

The dual space $C^*[-T, T]$ is well-known [15,16] to be represented by $M[-T, T]$, the space of regular Borel measures on $[-T, T]$, where the linear functional $\phi \in C^*[-T, T]$ corresponding to $m \in M[-T, T]$ is obtained by Lebesgue integration,

$$\phi \cdot x = \int_{-T}^{T} x \ dm \ , \ x \in C[-T, T] \tag{4-6}$$

The norm of a measure in $M[-T, T]$ is given by its total variation,

$$||m||_M = \sup \sum_{i=1}^{n} | \ m(E_i) \ | \ , \ m \in M[-T, T] \tag{4-7a}$$

where the supremum is taken over all Borel partitions $\{E_i\}_{i=1}^n$ of $[-T, T]$. In the case of a <u>positive measure</u> p (i.e. $p(E) \geq 0$ for $E \subseteq [-T, T]$, the norm is

$$||p||_M = p([-T, T]) = \int_{-T}^{T} 1 \ dp \tag{4-7b}$$

† A non-empty interior is essential, since the proof of Theorem 1 relies on the Separation Theorem. The corresponding positive cones in $L^p[-T, T]$, $1 \leq p < \infty$, have no interior points.

The dual positive cone consists of positive measures which are zero on $(-\delta, \delta)$, as follows:

<u>Lemma 4.1</u>: If the positive cone $P_C^+ \subseteq C[-T, T]$ is given by (4-4), then the dual positive cone $P_M^{\oplus} \subseteq M[-T, T]$ is

$$P_M^{\oplus} = \{m \in M[-T, T] : m(E) \geq 0 \ , \ E \subseteq [-T, T] \text{ and } m(F) = 0 \ , \ F \subseteq (-\delta, \delta)\} \tag{4-8}$$

<u>Proof</u>: Suppose $x \in P_C^+$, so that $x(t) \geq 0$ for $\delta \leq |t| \leq T$, and $m \in P_M^{\oplus}$ as defined by (4-8). Then it is clear that

$$\int_{-T}^{T} x \, dm = \int_{-T}^{-\delta} x \, dm + \int_{\delta}^{T} x \, dm \geq 0 \ , \tag{4-9}$$

so that m must belong to the dual positive cone as defined by (3-2).

Conversely, suppose m belongs to the dual positive cone. If $m(E) < 0$ for any $E \subseteq [-T, T]$, then by Urysohn's Lemma [15,16] there must exist $x \in P_C^+$ such that

$$\int_{-T}^{T} x \, dm < 0 \ , \tag{4-10}$$

which contradicts (3-2). Similarly, if $m(F) \neq 0$ for any $F \subseteq (-\delta, \delta)$, then there exists $y \in C[-T, T]$, with support in $(-\delta, \delta)$, such that

$$\int_{-T}^{T} y \, dm \neq 0 \tag{4-11}$$

Since y and $-y$ both belong to P_C^{+}, this also contradicts (3-2). Thus m must belong to $P_M^{\oplus}$.

Q.E.D.

In order to apply Theorem 3.1, it is necessary to assume the existence of a feasible solution which satisfies the inequalities strictly:

Assumption 4.2: There exists $\bar{u} \in L^2\,[0, T]$ such that

$$G_i\,(\bar{u}) \prec 0\ ,\ i=1,\ 2,\ 3 \tag{4-12}$$

Because of the nature of the sidelobe constraints (2-11), one expects that if ε or δ is chosen too small, there will be no non-zero feasible solutions at all. On the other hand, if there is a feasible solution u, i.e.

$$G_i\,(u) \preccurlyeq 0\ ,\ i=1,\ 2,\ 3, \tag{4-13}$$

then it is clear from the definition (4-2) that an infinitesimal increase in ε will allow $\frac{1}{2}u$ to satisfy (4-12). Thus Assumption 4.2 is not a particularly strong additional restriction. With this assumption, the existence and uniqueness of a nontrivial optimal solution may be established:

Theorem 4.3: Under Assumption 4.2, a unique nonzero optimal solution exists for the detection filter optimization problem (4-3).

Proof: The problem is to minimize f on the constraint set

$$S \triangleq \{u \in L^2[0, T] : G_i(u) \preccurlyeq 0, \; i=1, 2, 3\}$$

$$= G_1^{-1}(-P_C^{\;+}) \cap G_2^{-1}(-P_C^{\;+}) \cap \{u : ||u|| \leq 1\} \qquad (4\text{-}14)$$

Note that $S \neq \{0\}$ by Assumption 4.2. Since $-P_C^{\;+}$ is closed and convex, the continuity and linearity of G_1 and G_2 imply that $G_1^{-1}(-P_C^{\;+})$ and $G_2^{-1}(-P_C^{\;+})$ are also closed and convex, and the intersection S of these two sets with the unit ball $\{u : ||u|| \leq 1\}$ is closed, convex, and bounded. But this implies that S is weakly closed†, and, $L^2[0, T]$ being reflexive, Alaoglu's Theorem† says that it is weakly compact. The continuous linear functional f is also weakly continuous†, so $f(S)$ is compact in R^1 and must contain its infimum, which establishes the existence of an optimal $u^0 \in S$.

Now suppose that $u^1 \in S$ is also optimal. Assumption 4.2 implies that u^0 and u^1 are non-zero, so that $||u^0|| = ||u^1|| = 1$, according to the discussion preceding (2-14). The element $u = \frac{1}{2}u^0 + \frac{1}{2}u^1$ satisfies the constraints because they are convex, and the linearity of f implies that u is also optimal:

$$f(u) = f(\tfrac{1}{2}u^0 + \tfrac{1}{2}u^1)$$

$$= \tfrac{1}{2}f(u^0) + \tfrac{1}{2}f(u^1) = f(u^0) = f(u^1) \qquad (4\text{-}15)$$

Thus u must also have unit norm,

---

† See [15], pp. 422-424

$$||u||^2 = ||\tfrac{1}{2}u^0 + \tfrac{1}{2}u^1||^2 = \tfrac{1}{4}||u^0||^2 + \tfrac{1}{2} < u^0, u^1 > + \tfrac{1}{4}||u^1||^2$$

$$= \tfrac{1}{2} + \tfrac{1}{2} < u^0, u^1 > = 1 \tag{4-16}$$

so that the Schwartz inequality becomes an equality:

$$< u^0, u^1 > = 1$$

This implies that $u^1 = u^0$ so that the optimal u^0 must be unique.

Q.E.D.

Necessary conditions and dual problem

Theorem 3.1 may now be applied to the detection filter optimization problem (4-3). The spaces (4-1) and positive cones (4-4) are as specified by the theorem, and Assumption 4.2 asserts the existence of a feasible solution which satisfies the constraints with strict inequality. Theorem 4.3 guarantees the existence of a unique optimal solution u^0, so by Theorem 3.1 there must exist nonnegative Lagrange multipliers

$$\begin{aligned} &p_1^0, p_2^0 \in P_M^{\oplus} \subseteq M[-T, T] \\ &p_3^0 \in R^+ \end{aligned} \tag{4-17}$$

such that

(a) u^0 minimizes

$$L(u;\, p_1{}^0,\, p_2{}^0,\, p_3{}^0) = f(u) + \sum_{i=1}^{3} p_i{}^0 \cdot G_i(u)$$

$$= - \langle u,s\rangle + \int_{-T}^{T} [\langle u,s_t\rangle - \varepsilon \langle u,s\rangle] dp_1{}^0(t)$$

$$+ \int_{-T}^{T} [-\langle u,s_t\rangle - \varepsilon \langle u,s\rangle] dp_2{}^0(t) + p_3{}^0[||u||^2 - 1] \tag{4-18}$$

over all $u \in L^2[0, T]$, and

(b)
$$\left.\begin{aligned} p_1{}^0 \cdot G_1(u^0) &= \int_{-T}^{T} [\langle u^0, s_t\rangle - \varepsilon \langle u^0, s\rangle] dp_1{}^0(t) = 0 \\ p_2{}^0 \cdot G_2(u^0) &= \int_{-T}^{T} [-\langle u^0, s_t\rangle - \varepsilon \langle u^0, s\rangle] dp_2{}^0(t) = 0 \end{aligned}\right\} \tag{4-19}$$

$$p_3{}^0 \cdot G_3(u^0) = p_3{}^0[||u^0||^2 - 1] = 0 \tag{4-20}$$

Changing the order of integration† and rearranging terms in (4-18) yields

$$L(u;\, p_1,\, p_2,\, p_3) = - \langle u,\, \hat{u}(p_1,\, p_2)\rangle + p_3||u||^2 - p_3 \tag{4-21}$$

where $\hat{u}\,(p_1,\, p_2) \in L^2[0, T]$ is defined by

$$\hat{u}(p_1,\, p_2) = s + \int_{-T}^{T} (-s_t + \varepsilon\, s) dp_1(t) + \int_{-T}^{T} (s_t + \varepsilon\, s) dp_2(t)$$

$$= s - \int_{-T}^{T} s_t\, dp_1(t) + \int_{-T}^{T} s_t\, dp_2(t) + \varepsilon\, s\Big[\int_{-T}^{T} dp_1(t) + \int_{-T}^{T} dp_2(t)\Big]$$

---

† Since s and s_t have been assumed continuous on $[-T, T]$, this is a trivial application of Fubini's Theorem [15, 16].

$$= s[1 + \varepsilon||p_1||_M + \varepsilon\ ||p_2||_M] - \int s_t\, dp_1(t) + \int s_t\, dp_2(t) \tag{4-22}$$

If $p_3{}^0 = 0$, then either $L(u;\ p_1{}^0,\ p_2{}^0,\ p_3{}^0)$ has a minimum value of $-\infty$ or $u(p_1{}^0,\ p_2{}^0) = 0$, but in the latter case the duality theorem implies that $f(u^0) = 0$, contradicting the existence of a unique non-zero optimal solution. Thus it suffices to consider only $p_3 > 0$, and completing the square in (4-21) yields

$$L(u;\ p_1,\ p_2,\ p_3) = p_3||u - \hat{u}(p_1,\ p_2)/2p_3||^2 - p_3 - ||\hat{u}(p_1,\ p_2)||^2/4p_3 \tag{4-23}$$

This is minimized over u by

$$u = \hat{u}(p_1,\ p_2)/2p_3 \tag{4-24}$$

so that the dual functional is

$$\phi(p_1,\ p_2,\ p_3) = -p_3 - ||\hat{u}(p_1,\ p_2)||^2/4p_3 \tag{4-25}$$

for $p_1,\ p_2 \succcurlyeq 0$ and $p_3>0$.

According to Theorem 3.2, the Lagrange multipliers $p_1{}^0,\ p_2{}^0,\ p_3{}^0$ for the primal problem may be found by solving the dual problem of maximizing ϕ over $p_1,\ p_2,\ p_3 \succcurlyeq 0$. It has already been argued that $p_3{}^0 > 0$ and $\hat{u}(p_1{}^0,\ p_2{}^0) \neq 0$, so the maximization over p_3 for any p_1, p_2 is easily accomplished by differentiation:

$$\frac{d\phi}{dp_3} = -1 + ||\hat{u}(p_1, p_2)||^2/4p_3^2 = 0$$

$$p_3^0(p_1, p_2) = \tfrac{1}{2}||\hat{u}(p_1, p_2)|| > 0 \tag{4-26}$$

This must be a maximum, since

$$\frac{d^2\phi}{dp_3^2} = -||\hat{u}(p_1, p_2)||^2/2p_3^3$$

$$= -4/||\hat{u}(p_1, p_2)|| < 0 \tag{4-27}$$

Substituting (4-26) into (4-25), the dual problem reduces to one of maximizing

$$\phi[p_1, p_2, p_3^0(p_1, p_2)] = -||\hat{u}(p_1, p_2)|| \tag{4-28}$$

over $p_1, p_2 \succcurlyeq 0$ or equivalently,* of minimizing

$$||u(p_1, p_2)||^2 = ||s[1 + \varepsilon||p_1||_M + \varepsilon||p_2||_M] - \int_{-T}^{T} s_t dp_1(t) + \int_{-T}^{T} s_t dp_2(t)||^2 \tag{4-29}$$

over $p_1, p_2 \in P_M^{\oplus} \subseteq M[-T, T]$.

---

* The quantity to be minimized is squared for convenience in taking differentials.

Unconstrained minimization

This constrained minimization problem will now be reduced to an unconstrained one by combining p_1 and p_2 into a single real measure.

Recalling Lemma 4.1, the constraint $p_1, p_2 \in P_M^{\oplus}$ means that p_1 and p_2 are positive measures on $[-T, T]$ which are zero on $(-\delta, \delta)$. But the sidelobe constraint (2-13) implies that the integrands in the necessary conditions (4-19) are nonpositive, so that in each case the measure must be zero wherever the integrand is non-zero. Put another way, the Lagrange multipliers p_1 and p_2 are nonzero only where their respective constraints are active. Since the sidelobe constraints (2-13) cannot be simultaneously active, the measures p_1 and p_2 must be <u>mutually singular</u> [15, 16], that is, each must be zero wherever the other is not.

Consider the real measure

$$p \triangleq p_1 - p_2 , \tag{4-30}$$

which belongs to the subspace of $M[-T, T]$ defined by

$$M_0[-T, T] = \{m \in M[-T, T] : m(E) = 0, E \subseteq (-\delta, \delta)\}, \tag{4-31}$$

since both p_1 and p_2 are zero on $(-\delta, \delta)$.

Because of the mutual singularity of p_1 and p_2, the norm of p is given by

$$||p||_M = ||p_1||_M + ||p_2||_M \tag{4-32}$$

and $\hat{u}(p_1, p_2)$ in (4-22) can be written as

$$\hat{u}(p_1, p_2) = \hat{u}(p) = s + \varepsilon||p||_M s - \int_{-T}^{T} s_t \, dp(t) \tag{4-33}$$

Moreover, it follows from the Hahn Decomposition Theorem [15, 16] that every element in $M_0[-T, T]$ is uniquely decomposable into two such mutually singular measures. Therefore, minimizing $||\hat{u}(p_1, p_2)||^2$ over $p_1, p_2 \in P_M^{\oplus}$ is completely equivalent to minimizing $||\hat{u}(p)||^2$ over $p \in M_0[-T, T]$.

The original problem (2-14) has thus been reduced to an unconstrained* dual problem:

$$\left.\begin{array}{l} \text{minimize} \quad \eta(p) = ||\hat{u}(p)||^2, \; p \in M_0[-T, T] \\ \text{where} \quad \hat{u}(p) = s + \varepsilon||p||_M s - \int_{-T}^{T} s_t \, dp(t) \end{array}\right\} \tag{4-34}$$

Once p^0 is found to minimize η, Theorems 3.1 and 3.2 establish that the optimal solution of (2-14) is

$$u^0 = \hat{u}(p^0)/||\hat{u}(p^0)|| \tag{4-35}$$

---

* The stipulation that $p \in M_0[-T, T]$ is not really a constraint, since this subspace is in fact equivalent to the measure space $M\{[-T,-\delta] \cup [\delta,T]\}$.

or

$$u^0(\tau) = \frac{s(\tau) + \varepsilon||p^0||s(\tau) - \int_{-T}^{T} s(\tau+t)dp^0(t)}{||s + \varepsilon||p^0||s - \int_{-T}^{T} s_t \; dp^0(t)||}, \; \tau\epsilon[0,T] \tag{4-36}$$

Note that the denominator is nonzero in (4-35) because

$$\langle u^0,s\rangle = -f(u^0) = -\phi(p_1{}^0, p_2{}^0, p_3{}^0) = \eta^{\frac{1}{2}}(p^0) = ||\hat{u}(p^0)|| \tag{4-37}$$

by (3-15), and Assumption 4.2 implies that this quantity will be > 0.

The duality exhibited in (4-37) also provides a measure of optimality for approximate solutions, since

$$\langle u,s\rangle \leq \langle u^0,s\rangle = ||\hat{u}(p^0)|| \leq ||\hat{u}(p)|| \tag{4-38}$$

where p and u are any suboptimal solutions of the dual and primal problems, respectively. Thus, a filter

$$u^1 = \hat{u}(p^1)/||\hat{u}(p^1)|| \tag{4-39}$$

may be considered "approximately optimal" if it satisfies the constraints (2-13) to within some acceptable tolerance e_1 and

$$||\hat{u}(p^1)|| - \langle u^1,s\rangle \leq e_2 \tag{4-40}$$

for some tolerance e_2.

5. Computational solution of dual problem

In the previous section the detection filter optimization problem was reduced to an unconstrained dual problem (4-34). An iterative scheme will now be developed for its solution.

In order to utilize a digital computer for the solution, it is necessary to "discretize" the measure space $M_0[-T,T] = M\{[-T,-\delta]\cup[\delta,T]\}$ defined in (4-31). This will be done by partitioning the time set into intervals $(t_j, t_{j+1}]$, where

$$t_j \triangleq \begin{cases} -T + j\Delta n, & j = 0, 1, 2, \dots n \\ \delta + (j-n-1)\Delta_n, & j = n+1, n+2, \dots 2n+1 \end{cases} \tag{5-1}$$

and

$$\Delta_n \triangleq (T-\delta)/n \tag{5-2}$$

Next, define the finite-dimensional subspace

$$M_0^n[-T,T] \triangleq \{p \in M_0 : \int_{-T}^{T} x(t)dp(t) = \int_{-T}^{T} x(t)\dot{p}(t)dt,\ x \in C[-T,T]\} \tag{5-3}$$

where $\dot{p}(t)$ (the Radon-Nikodym derivative of the measure p) has the form

$$\dot{p}(t) = p_j\ , \quad t \in (t_j, t_{j+1}],\ j = 0, 1, 2, \dots 2n \tag{5-4}$$

and $p_n = 0$ so that

$$\dot{p}(t) = 0, \; t \in (t_n, t_{n+1}] = (-\delta, \delta] \tag{5-5}$$

This is illustrated in Figure 4.

In other words, a measure in $M_0^n[-T,T]$ is represented by a "staircase" function (its derivative) which is constant on each of the $2n+1$ intervals of length Δ_n. Such a measure is completely specified by the weights $\{p_j\}_{j=0}^{2n}$, and

$$\int_{-T}^{T} x(t)dp(t) = \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} x(t)dt \tag{5-6}$$

for any $x \in C[-T,T]$. Its norm is given by

$$||p||_M = \int_{-T}^{T} |\dot{p}(t)|dt = \Delta_n \sum_{j=0}^{2n} |p_j| \tag{5-7}$$

It will now be shown that for large enough n, the approximating subspace always contains near-optimum elements. An equivalent statement is that the subspace $\bigcup_{n=1}^{\infty} M_0^n[-T,T]$ is dense in the weak-star topology [15] of $M_0[-T,T]$.

Lemma 5.1 To every $e > 0$ there corresponds N such that for $n \geq N$, $M_0^n[-T,T]$ contains an element p^n for which

$$|u^n(\tau) - u^0(\tau)| < e, \qquad \tau \in [0, T] \tag{5-8}$$

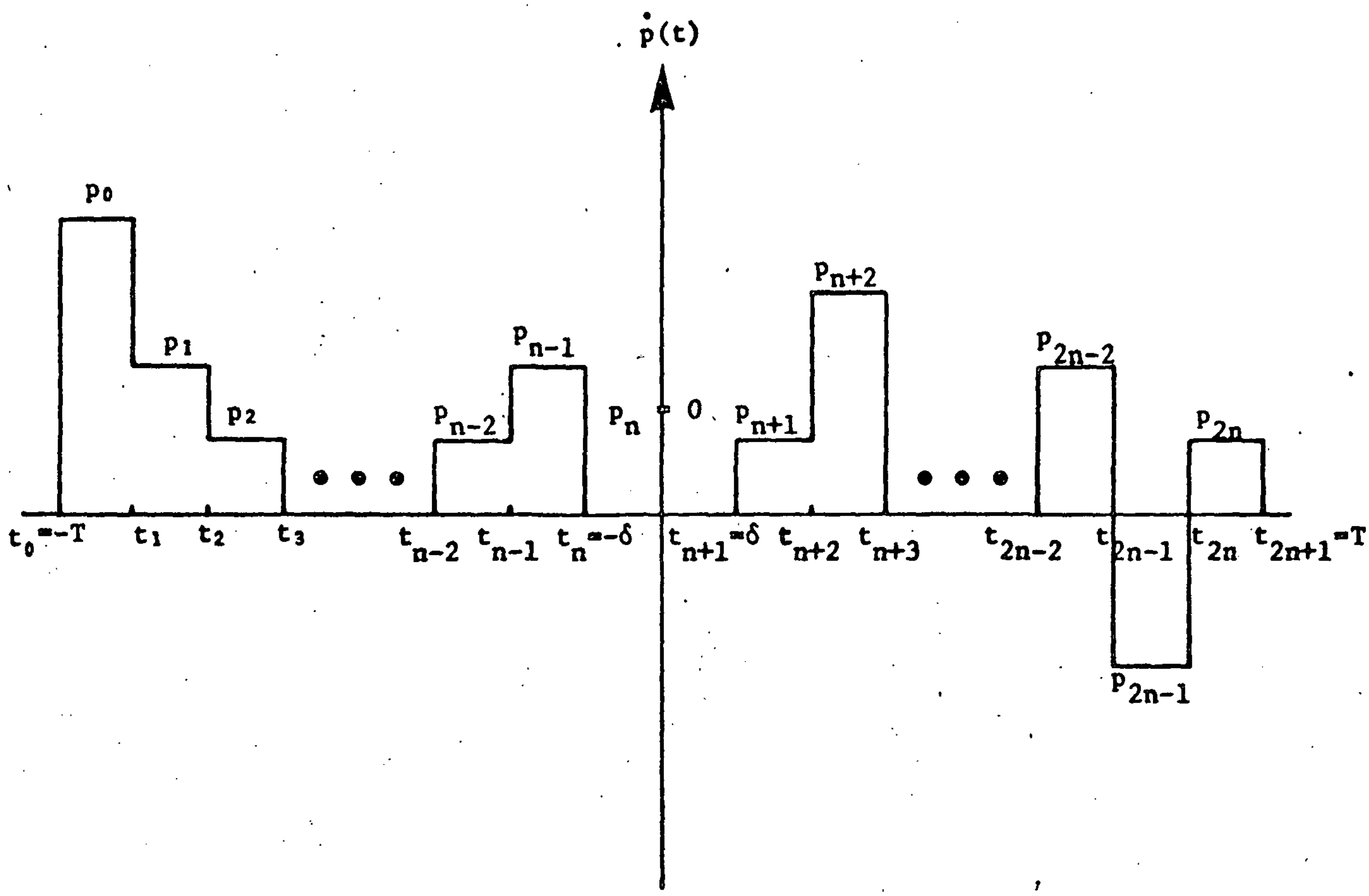

FIGURE 4 Approximating measure $p \in M_0^n[-T, T]$

and

$$\eta(p^n) - \eta(p^0) < e \tag{5-9}$$

where u^0 is given by (4-35) and

$$u^n = \hat{u}(p^n)/||\hat{u}(p^n)|| \tag{5-10}$$

Proof: A measure $p^n \in M_0^n[-T,T]$ will be constructed such that

$$\left|[\hat{u}(p^0)](\tau) - [\hat{u}(p^n)](\tau)\right| \leq \varepsilon\left|s(\tau)\right| \cdot \left|\,||p^0||_M - ||p^n||_M\right| + \left|\int_{-T}^{T} s_t(\tau)dp^0(t) - \int_{-T}^{T} s_t(\tau)dp^n(t)\right| \leq e, \quad \tau \in [0,T] \tag{5-11}$$

for large enough n. The last integral may be expressed for any fixed τ as a limit of integrals of a sequence of simple functions [15, 16] which approach $s_t(\tau)$. Because s is continuous, such a sequence of simple functions $\{s_t^n(\tau)\}_{n=1}^{\infty}$ is defined by

$$s_t^n(\tau) = \int_{t_j}^{t_{j+1}} s_\xi(\tau)d\xi, \; t \in (t_j, t_{j+1}], \qquad j=0, 1,2, \ldots 2n \tag{5-12}$$

where the intervals $(t_j, t_{j+1}]$ were defined in (5-1). Thus there must exist an n (independent of τ, because of the continuity of s) such that

$$\left|\int_{-T}^{T} s_t(\tau)dp^0(t) - \int_{-T}^{T} s_t^n(\tau)dp^0(t)\right|$$

$$= \left|\int_{-T}^{T} s_t(\tau)dp^0(t) - \sum_{j=0}^{2n} p^0((t_j, t_{j+1}]) \int_{t_j}^{t_{j+1}} s_t(\tau)dt\right| < e/2, \tau \in [0, T] \tag{5-13}$$

Referring to (5-4) and (5-6), a measure $p^n \in M_0^n[-T,T]$ may now be specified by choosing

$$p_j^n = p^0((t_j, t_{j+1}]), \qquad j = 0, 1, 2, \ldots 2n \tag{5-14}$$

so that

$$\left|\int_{-T}^{T} s_t(\tau)dp^0(t) - \int_{-T}^{T} s_t(\tau)dp^n(t)\right| < e/2, \qquad \tau \in [0,T] \tag{5-15}$$

From the definition (4-7a) of the norm in $M[-T,T]$, it is clear that

$$\Delta_n \sum_{j=0}^{2n} |p_j^n| = ||p^n||_M \leq ||p^0||_M \tag{5-16}$$

If equality in (5-16) is not approached with increasing n, then $||p^n||_M$ can be increased, for instance by adding to p^n a measure whose weight is γ in $(t_k, t_{k+1}]$, $-\gamma$ in $(t_{k+1}, t_{k+2}]$, and zero elsewhere for some integer $k < 2n$ and scalar γ. Such a perturbation can be used to make

$$\varepsilon s(\tau)\Big|\, ||p_0||_M - ||p^n||_M \Big| < e/2, \qquad \tau \in [0,T] \tag{5-17}$$

and the continuity of s ensures that for large enough n, (5-15) remains valid.

This establishes (5-11), from which (5-8) and (5-9) follow immediately.

Q.E.D.

<u>Steepest Descent Algorithm</u>

A steepest descent algorithm will now be proposed for minimizing $\eta(p)$ in the subspace $M_0^n[-T,T]$. Lemma 5.1 establishes that the true optimum in $M_0[-T,T]$ can be approached as closely as desired by increasing n. The functional to be minimized is

$$\begin{aligned}\eta(p) &= ||s + \varepsilon||p||_M s - \int_{-T}^{T} s_t dp(t)||^2 \\ &= ||s + \varepsilon\Delta_n \sum_{j=0}^{2n} |p_j| s + \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} s_t dt||^2\end{aligned} \tag{5-18}$$

(recall that $(t_n, t_{n+1}] = (-\delta,\delta]$ and $p_n = 0$).

The functional (5-18) is not Fréchet (strong sense) differentiable because it contains a term

$$||p||_M = \Delta_n \sum_{j=0}^{2n} |p_j| \tag{5-19}$$

This norm does, however, possess a <u>directional Gateaux (weak sense) differential</u> [13, 17, 18] which is convex (but not always linear) in its increment. This differential at p in the direction h is defined as

$$\delta^+||p;\ h||_M = \lim_{\alpha\to 0^+} \frac{1}{\alpha}[||p + \alpha h|| - ||p||],\ p,\ h \in M_0^n[-T,T] \tag{5-20}$$

For each j, it is easily verified that

$$\delta^+|p_j;\ h_j| = \begin{cases} h_j, & x_j > 0 \\ |h_j|, & x_j = 0 \\ -h_j, & x_j < 0 \end{cases} \tag{5-21}$$

from which it follows that

$$\delta^+||p;\ h||_M = \Delta_n \sum_{j\in S^+} h_j - \Delta_n \sum_{j\in S^-} h_j + \Delta_n \sum_{j\in S^0} |h_j| \tag{5-22}$$

where

$$S^+ \triangleq \{j:\ p_j > 0\} \tag{5-23}$$

$$S^- \triangleq \{j:\ p_j < 0\} \tag{5-24}$$

$$S^0 \triangleq \{j:\ p_j = 0\} \tag{5-25}$$

and $$S^+ \cup S^- \cup S^0 = \{0,\ 1,\ 2,\ \ldots\ 2n\} \tag{5-26}$$

It is a straightforward matter to show that the differential of (5-18) is

$$\delta^+\eta(p;\ h) = 2\left\langle [s + \varepsilon||p||_M s - \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} s_t dt], \right.$$

$$\left. [\varepsilon\delta^+||p;\ h||_M s - \sum_{i=0}^{2n} h_i \int_{t_i}^{t_{i+1}} s_\tau d\tau] \right\rangle \tag{5-27}$$

Interchanging integrals, this becomes

$$\delta^+\eta(p;\ h) = 2\varepsilon[1 + \varepsilon||p||_M - \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} <s,s_t>dt]\delta^+||p;\ h||_M$$

$$-2(1 + \varepsilon||p||_M) \sum_{i=0}^{2n} h_i \int_{t_i}^{t_{i+1}} <s,s_t>dt$$

$$+ 2 \sum_{j=0}^{2n} \sum_{i=0}^{2n} p_j h_i \int_{t_j}^{t_{j+1}} \int_{t_i}^{t_{i+1}} <s_t,s_\tau>dtd\tau$$

$$= K\delta^+||p;\ h||_M + \sum_{i=0}^{2n} h_i L_i \tag{5-28}$$

where

$$K \triangleq 2\varepsilon[1 + \varepsilon||p||_M - \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} <s,s_t>dt] \tag{5-29}$$

$$L_i \triangleq 2[\sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} \int_{t_i}^{t_{i+1}} <s_t,s_\tau>dtd\tau$$

$$- (1 + \varepsilon||p||_M) \int_{t_i}^{t_{i+1}} <s,s_t>dt],\quad i=0,\ 1,\ \dots\ 2n \tag{5-30}$$

and $||p||_M$ is given by (5-19).

Finally, substitution of (5-22) yields

$$\delta^+\eta(p;\, h) = \sum_{j\in S^+} (L_j + K\Delta_n)h_j$$

$$+ \sum_{j\in S^-} (L_j - K\Delta_n)h_j$$

$$+ \sum_{j\in S^0} (L_j h_j + K\Delta_n |h_j|) \tag{5-31}$$

The object is now to iteratively decrease the cost $\eta(p)$ by choosing a sequence of increments h for which the differential $\delta^+\eta(p;h)$ is negative. Suppose for a moment that S^0 is empty, so that (5-31) is a linear functional of h,

$$\delta^+\eta(p;\, h) = \sum_{j\in S^+\cup S^-} N_j h_j \tag{5-32}$$

where $N_j \triangleq L_j \pm K\Delta_n$ for $j \in S^+$ and $j \in S^-$, respectively.

Then $-\delta^+\eta(p;\, h)$ will be maximized over all h of unit norm whenever h is <u>aligned</u>* with it. This means that h_j should be nonzero only where $|N_j|$ is a maximum, and opposite in sign to N_j.

The generalization to non-empty S^0 is straightforward if one notes that the last term in (5-31) can be written for any $S \subseteq S^0$ as

$$\sum_{j\in S} (L_j H_j + K\Delta_n |h_j|) = \begin{cases} \sum_{j\in S} (L_j + K\Delta_n)h_j & \text{if } h_j \geq 0 \text{ on } S \\ \\ \sum_{j\in S} (L_j - K\Delta_n)h_j & \text{if } h_j \leq 0 \text{ on } S \end{cases} \tag{5-33}$$

---

* An element $x \in X$ and a functional $f \in X^*$ must satisfy the inequality $f\cdot x \leq ||f||\ ||x||$, by the definition of f. The functional f is said to be *aligned* with x whenever equality is achieved.

On S^0, then, an "aligned" h will have h_j nonzero only where $(-L_j - K\Delta_n)$ or $(L_j - K\Delta_n)$ is positive and a maximum, its sign being positive or negative, respectively.

To summarize, an "aligned" $h \in M_0^n[-T,T]$, resulting in a maximally negative cost differential $\delta^+\eta(p; h)$, will have $h_j = 0$ for $j \notin R$, where* $R \subseteq \{0, 1, 2, \ldots n-1, n+1, \ldots 2n\}$ is the set of integers for which the maximum

$$M = \max \{0, \max_{j \in S^+} |L_j + K\Delta_n|, \max_{j \in S^-} |L_j - K\Delta_n|,$$

$$\max_{j \in S^0} [-L_j - K\Delta_n], \max_{j \in S^0} [L_j - K\Delta_n]\} \qquad (5\text{-}34)$$

is achieved. For $j \in R$, the sign of h_j will be opposite to $(L_j + K\Delta_n)$ or $(L_j - K\Delta_n)$, whichever applies. The magnitudes $|h_j|$, $j \in R$ will be chosen equal for simplicity, although they could be optimized if desired. This leaves only the "step size" $\sum_{j \in R} |h_j|$ to be adjusted at each iteration.

Whenever $M > 0$, there will exist an increment h for which $\delta^+\eta(p; h) < 0$ and hence η can be decreased. On the other hand, if $M = 0$ for some p^0, then

$$\delta^+\eta(p^0; h) \geq 0 \quad \text{for all } h \qquad (5\text{-}35)$$

---

* Recall that by the definition of $M_0^n[-T,T]$, p and h are both zero on $(-\delta,\delta)$, which implies that $p_n = h_n = 0$.

and η must have a minimum at p^0.

Moreover, the second differential can be calculated from (5-28),

$$\delta^{+^2}\eta(p;h,g)$$

$$= 2\varepsilon[1+\varepsilon||p||_M - \sum_{j=0}^{2n} p_j \int_{t_j}^{t_{j+1}} <s,s_t>dt]\delta^{+^2}||p;h,g||_M$$

$$+ 2\varepsilon^2\delta^+||p;g||_M\delta^+||p;h||_M + 2\sum_{j=0}^{2n}\sum_{i=0}^{2n} g_i h_i \int_{t_j}^{t_{j+1}}\int_{t_i}^{t_{i+1}} <s_t,s_\tau>dt d\tau$$

$$- 2\varepsilon\sum_{j=0}^{2n}\int_{t_j}^{t_{j+1}}<s,s_t>dt[g_j\delta^+||p;h||_M + h_j\delta^+||p;g||_M] \quad (5\text{-}36)$$

For any j, it is easy to verify that

$$\delta^{+^2}|p_j;h_j,g_j| = \begin{cases} -\infty & \text{if } p_j = 0,\ h_j g_j < 0 \\ 0 & \text{otherwise} \end{cases} \quad (5\text{-}37)$$

Thus $\delta^{+^2}||p;\ h,g||_M$ is not defined for all h and g, but

$$\delta^{+^2}||p;\ h,h||_M = 0 \qquad \text{for all } h \in M_0^n[-T,T] \quad (5\text{-}38)$$

Substituting this into (5-36) with $h = g$ and again rearranging integrals yields

$$\delta^{+^2}\eta(p;h,h) = 2||\varepsilon\delta^+||p;h||s - \sum_{j=0}^{2n} h_j\int_{t_j}^{t_{j+1}} s_t dt||_2^2$$

$$\geq 0 \quad \text{for all } p,\ h \in M_0^n[-T,T] \quad (5\text{-}39)$$

This establishes that $\delta^{+^2}\eta(p;\, h,h)$ is nonnegative definite for all p, so η must be convex and have no relative minima on $M^n[-T,T]$. The algorithm will decrease η at every iteration and hence approach the minimum.

The steepest descent algorithm described above will produce a sequence of measures $p^i \in M_0^n[-T,T]$, $i = 1, 2, \ldots$ for which the cost $\eta(p^i)$ approaches a minimum on this subspace. Lemma 5.1 verifies that by increasing n (i.e. using finer partitions of $[-T,T]$), this cost may be made to approach the optimal cost $\eta(p^0)$, $p^0 \in M_0[-T,T]$. Moreover,

$$u^i = \hat{u}(p^i)/||\hat{u}(p^i)|| \tag{5-40}$$

approaches the optimal filter u^0.

Recall from the end of Section 4 that duality provides a criterion for stopping the algorithm: the constraints (2-13) should satisfy

$$\left.\begin{array}{l} \langle u^i, s_t\rangle - \varepsilon\langle u^i, s\rangle \leq e_1 \\ \langle u^i, s_t\rangle - \varepsilon\langle u^i, s\rangle \leq e_1 \end{array}\right\} \delta \leq |t| \leq T \tag{5-41}$$

for some tolerance e_1, and the primal and dual costs should satisfy (4-40),

$$||\hat{u}(p^i)|| - \langle u^i, s\rangle \leq e_2 \tag{5-42}$$

for some other tolerance e_2.

6. Estimation filter optimization problem

Sections 2-5 are concerned with determining an optimal filter for detection of a signal, by maximizing the signal-to-noise ratio (2-10) subject to the sidelobe constraint (2-11).

Another task which radar and communications systems must usually perform is the estimation of unknown signal parameters, particularly the arrival time t_0 of the signal. In a radar system the signal delay is proportional to the range of the target. In certain communications systems, such as Pulse Position Modulation (PPM), the signal delay carries some portion of the transmitted information. It is well-known [1] that for a given continuously differentiable signal, the maximum-likelihood estimate of unknown time delay is obtained by processing the input with a matched filter and subtracting T from the time at which the output achieves a maximum.

More generally, if this same estimation scheme is used with a filter which is not necessarily matched to the signal, the accuracy of the estimate has been determined by McAulay and Johnson [11]:

Lemma 6.1 Assuming a large signal-to-noise ratio *a priori*, the above estimate of signal arrival time has a bias proportional to

$$\dot{\psi}(0) = \langle u,\dot{s}\rangle = -\langle \dot{u}, s\rangle \qquad (6\text{-}1)$$

If this bias is zero, the variance of the estimate is the reciprocal of

$$\ddot{\psi}^2(0)/E\{\dot{\xi}^2\} = \langle \dot{u}, \dot{s}\rangle^2/N_0||\dot{u}||^2 \tag{6-2}$$

which may be interpreted as an output signal bandwidth-to-noise bandwidth ratio.

Proof: Using the same assumptions as in Section 2, with the additional provisos that the signal s is twice differentiable and the signal-to-noise ratio is large, an optimal filter for estimating signal arrival time may be sought. The problem will be to choose a filter $\dot{u}$ so that the estimate is unbiased* and its variance (6-2) is minimized, subject to the sidelobe constraint (2-11).

The argument which led to (2-14) may be used [12] to put this estimation filter optimization problem in the following form:

$$\left.\begin{array}{lll} \text{maximize} & \langle \dot{u}, \dot{s}\rangle, & \dot{u} \in L^2[0,T] \\ \text{subject to} & \langle \dot{u}, s\rangle = 0 & \\ & \langle u, s_t\rangle \leq \varepsilon\langle u, s\rangle, & \delta \leq |t| \leq T \\ & -\langle u, s_t\rangle \leq \varepsilon\langle u, s\rangle, & \delta \leq |t| \leq T \\ & ||\dot{u}||^2 \leq 1 & \end{array}\right\} \tag{6-3}$$

* The requirement of zero bias can be dropped, but then the expression for the variance becomes very complex.

The results of Sections 3 and 4 may be used, *mutatis mutandis**, to reduce (6-3) to an unconstrained dual problem analogous to (4-34),

$$\text{minimize} \quad \nu(p) = ||\hat{v}(p)||^2, \quad p \in M_0[-T,T]$$
$$\hat{v}(p) = \dot{s} + \varepsilon||p||_M r_0 - \int_{-T}^{T} r_t dp(t) \tag{6-4}$$

The function r_t is defined by

$$\left.\begin{aligned} r_t(\tau) &\triangleq S_t(\tau) + \langle S_t, s\rangle\, s(\tau) \\ S_t(\tau) &= \int_\tau^T s(\sigma + t)d\sigma \end{aligned}\right\} \tag{6-5}$$

Once p^0 is found to minimize ν, the optimal solution of (6-3) is given by

$$\dot{u}^0 = \hat{v}(p^0)/||\hat{v}(p^0)||, \tag{6-6}$$

and the duality theorem implies that

$$\langle\dot{u}^0, \dot{s}\rangle = \nu^{\frac{1}{2}}(p^0) = ||\hat{v}(p^0)|| \tag{6-7}$$

The algorithm of Section 5 for minimizing η is equally suited to minimizing ν; the function $\hat{u}(p)$ in (4-34) need only be replaced by $\hat{v}(p)$ in (6-4).

---

* Theorems 3.1 and 3.2 must be extended slightly to account for the equality constraint $\langle\dot{u}, s\rangle = 0$. See [12] for details.

7. Conclusion

The problem of determining an optimal detection filter, subject to hard constraints on the magnitude of the output sidelobes, has been formulated and investigated in some detail. Assuming that at least one feasible solution of the constraints exists, it has been shown that this problem has a unique solution which may be found by solving an unconstrained dual problem in a space of measures.

An algorithm for solution of the dual problem, in a "discrete" subspace of measures suitable for computer implementation, has been presented and shown to converge to the optimum.

The optimal estimation filter problem may be solved analogously, as has been indicated in Section 6.

The algorithm for both problems is currently being programmed, and numerical results will be presented in a subsequent publication.

8. Acknowledgements

The authors wish to express their gratitude to Dr. Robert McAulay and Professor Ian Rhodes for various comments and suggestions.

REFERENCES

[1] E. J. Kelly, I. S. Reed and W. L. Root, "The Detection of Radar Echoes in Noise," Parts I and II, Journal of SIAM, v. 8, No. 2-3, June-September 1960, pp. 309-341, 481-507.

[2] M. I. Skolnik, Introduction to Radar Systems, McGraw Hill, 1962.

[3] C. E. Cook and M. Bernfeld, Radar Signals, Academic Press, 1967.

[4] A. W. Rihaczek, Principles of High-Resolution Radar, McGraw-Hill, 1969.

[5] C. W. Helstrom, Statistical Theory of Signal Detection, Pergamon Press, 1960.

[6] H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley and Sons, 1968.

[7] J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, Wiley 1965.

[8] R. J. McAulay and L. P. Seidman, "A Useful Form of the Barankin Lower Bound and Its Application to PPM Threshold Analysis," IEEE Trans. on Information Theory, IT-15, No. 2, March 1969, pp. 273-279.

[9] R. J. McAulay, "Numerical Optimization Techniques Applied to PPM Signal Design," IEEE Trans. on Information Theory, IT-14, No. 5, September 1968, pp. 708-716.

[10] R. J. McAulay, "Optimal Control Techniques Applied to PPM Signal Design," Information and Control, v. 12, No. 3, March 1968, pp. 221-235.

[11] R. J. McAulay and J. R. Johnson, "Optimal Mismatched Filter Design for Radar Ranging, Detection, and Resolution," IEEE Trans. on Information Theory, IT-17, No. 6, November 1971.

[12] T. E. Fortmann, "Optimal Design of Filters and Signals Subject to Sidelobe Constraints," Report ESL-R-400, M.I.T., Cambridge, Massachusetts, September 1969.

[13] D. G. Luenberger, Optimization by Vector Space Methods, Wiley and Sons, 1969.

[14] L. Hurcwicz, "Programming in Linear Spaces", Studies in Linear and Nonlinear Programming, ed. by K. J. Arrow, L. Hurwicz, and H. Uzawa, Stanford University Press, 1958.

[15] N. Dunford and J. Schwartz, Linear Operators, Part I, Interscience, 1958.

[16] W. Rudin, Real and Complex Analysis, McGraw Hill, 1966.

[17] L. V. Kantorovich and G. P. Akilov, Functional Analysis in Normed Spaces, MacMillan, 1964.

[18] D. G. Luenberger, "Control Problems with Kinks," IEEE Trans. Auto. Control, AC-15, #5, October 1970, pp. 570-75.