Uncertainty Principles and Signal Recovery

By

David L. Donoho
Philip B. Stark

Department of Statistics
University of California
Berkeley, California

Technical Report No. 94
June 1987

Department of Statistics
University of California
Berkeley, California

Uncertainty Principles and Signal Recovery

David L. Donoho
Philip B. Stark

Department of Statistics
University of California
Berkeley, California

Submitted to SIAM Journal on Applied Mathematics

ABSTRACT

The uncertainty principle can easily be generalized to cases where the 'sets of concentration' are not intervals. We present such generalizations for continuous and discrete-time functions, and for several measures of 'concentration' (*e.g.* L_2 and L_1 measures). The generalizations explain interesting phenomena in signal recovery problems where there is an interplay of missing data, sparsity and bandlimiting.

Key words. uncertainty principle, signal recovery, unique recovery, stable recovery, bandlimiting, timelimiting, L_1-methods, sparse spike trains

AMS(MOS) subject classifications. 42A05, 94A11

1. Introduction

The classical uncertainty principle says that if a function $f(t)$ is essentially zero outside an interval of length Δt and its Fourier transform $\hat{f}(\omega)$ is essentially zero outside an interval of length $\Delta\omega$, then

$$\Delta t \cdot \Delta\omega \geq 1 ; \tag{1.1}$$

a function and its Fourier transform cannot both be highly concentrated. The uncertainty principle is widely known for its 'philosophical' applications: in quantum mechanics, of course, it shows that a particle's position and momentum cannot be determined simultaneously (Heisenberg, 1930); in signal processing it establishes limits on the extent to which the 'instantaneous frequency' of a signal can be measured (Gabor, 1946). However, it also has technical applications, for example in the theory of partial differential equations (Fefferman, 1983).

We show below that a more general principle holds: it is not necessary to suppose that f and $\hat{f}$ are concentrated *on intervals*. If f is practically zero outside a measurable set T and $\hat{f}$ is practically zero outside a measurable set W, then

$$|T|\,|W| \geq 1-\delta \tag{1.2}$$

where $|T|$ and $|W|$ denote the measures of the sets T and W, and δ is a small number bound up in the definition of the phrase 'practically zero'–a precise definition will be given later. In short, f and $\hat{f}$ cannot both be highly concentrated, no matter what 'sets of concentration' T and W we pick. The assertion (1.2) appears to be new.

The uncertainty principle also applies to discrete functions. Let $(x_t)_{t=0}^{N-1}$ be a sequence of length N and let $(\hat{x}_w)_{w=0}^{N-1}$ be its discrete Fourier transform. Suppose that (x_t) is not zero at N_t points and that $(\hat{x}_w)$ is not zero at N_w points. Then

$$N_t \cdot N_w \geq N \ . \tag{1.3}$$

The inequality (1.3) makes no reference to the kind of sets where (x_t) and $(\hat{x}_w)$ are nonzero: these may be intervals or any other sets. This assertion (1.3) also appears to be new.

The usual approaches to the uncertainty principle, via either Weyl's inequality or via the prolate spheroidal wave functions, involve rather sophisticated methods: eigenfunctions of the Fourier transform (Weyl, 1928); eigenvalues of compact operators (Landau and Pollak, 1961). In contrast, the more *general* principles (1.2) and (1.3) we introduce here have elementary proofs. The discrete-time uncertainty principle follows from the fact that a certain Vandermonde determinant does not vanish; the proof could be taught in an undergraduate linear algebra course. The continuous-time uncertainty principle requires only the introduction of the Hilbert-Schmidt norm of an operator, and could be taught in an introductory functional analysis course.

Principles (1.2) and (1.3) have applications in signal recovery. The continuous-time principle shows that missing segments of a bandlimited function can be restored stably in the presense of noise if (total measure of the missing segments)·(total bandwidth) < 1. The discrete-time principle suggests that a wide-band signal can be reconstructed from narrow-band data -- provided the wide-band signal to be recovered is sparse or 'impulsive'. The classical uncertainty principle does not apply in these examples.

The discrete-time principle (1.3) is proved in section 2; section 3 proves a continuous-time principle for L_2 theory. These are then applied to signal recovery problems in sections 4 and 5. Section 6 proves another version of the continuous-time principle using L_1 theory; this has the rather remarkable application that a bandlimited function corrupted by noise of unknown properties can be restored *perfectly*, without error, if the noise is 'sparse': zero outside an (unknown) set of measure $< 1/(2\cdot\text{bandwidth})$--a phenomenon first discovered by B.F. Logan. We show here that the phenomenon derives from the L_1-uncertainty principle.

The l_1 version of Logan's phenomenon (*i.e.* the version of Logan's phenomenon for discrete time) can be used in the study of an l_1-algorithm for restoring a sparse signal from narrowband data. It shows that the l_1-algorithm recovers a wideband signal *perfectly* from noiseless narrowband data, provided the signal is sufficiently sparse. This fact about the l_1-algorithm was demonstrated by Santosa and Symes (1986); we show here that it derives from Logan's phenomenon and the l_1 uncertainty principle.

Section 7 discusses the sharpness of the uncertainty principles given here; section 8 supplies a 'counterexample' of sorts, discusses connections with deeper work and mentions generalizations to other settings. Appendix A identifies the extremal functions of the discrete-time principle.

2. The Discrete-Time Uncertainty Principle

Let (x_t) be a sequence of length N and let $(\hat{x}_w)$ be its discrete Fourier transform

$$\hat{x}_w \equiv \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t \, e^{-\frac{2\pi i w t}{N}} \quad , \quad w = 0, \cdots, N-1 \ . \tag{2.1}$$

As above, N_t and N_w count the number of nonzero entries in (x_t) and $(\hat{x}_w)$ respectively.

Theorem 1.

$$N_t \cdot N_w \geq N \ . \tag{2.2}$$

Corollary 1.

$$N_t + N_w \geq 2\sqrt{N} \ .$$

The theorem bounds the time-bandwidth product; the corollary (which follows immediately from the theorem by the geometric mean–arithmetic mean inequality) bounds the total number of nonzero elements. It is easy to construct examples attaining the limits set in (2.2). For any N, the example $\{x_0=1;\ x_t=0,\ t>0\}$ always works. If N is composite and admits the factorization $N=k\cdot l$, the sequence

$$Ш_t^k \equiv \begin{cases} 1, & t=i\cdot l \quad \text{for } i=0,\cdots,N_t-1 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

has k equally-spaced nonzero elements. It is the indicator function of a subgroup of $\{0,\cdots,N-1\}$ and its discrete Fourier transform is, up to a constant factor, the indicator function of the dual subgroup (Dym and McKean, 1972); explicitly,

$$\sqrt{N}\ \hat{Ш}^k = k\,Ш^l\ .$$

The dual subgroup has l nonzero elements, so that $N_t N_w = k\cdot l = N$. We show in Appendix A that apart from simple modifications, these are the only pairs of sequences that attain the bound $N_t N_w = N$; the extremal functions for this uncertainty principle are basically periodic 'spike trains' with an integral number of periods in the length N.

To prove Theorem 1 it is convenient to think of (x_t) and $(\hat{x}_w)$ as defined on the 'discrete circle' $\{0, 1, \cdots, N-1\}$ which 'wraps around' so that $N-1$ and 0 are consecutive sites. The 'wraparound' convention is equivalent to interpreting subscripts modulo N; thus x_{t+N} is identified with x_t. Similarly, $\hat{x}_{w+N}$ is identified with $\hat{x}_w$. The convention is justified by looking at the formula (2.1) and noting that

$$e^{\frac{-2\pi i w(t+N)}{N}} = e^{\frac{-2\pi i (w+N)t}{N}} = e^{\frac{-2\pi i w t}{N}}\ .$$

The proof of Theorem 1 is an application of the following key fact: if (x_t) has N_t nonzero elements, no N_t consecutive elements of $(\hat{x}_w)$ can all vanish. (Here and below, we use the wraparound convention so that $\hat{x}_{N-1}$ and $\hat{x}_0$ are consecutive in $(\hat{x}_w)$.) This will be proved below as a lemma.

To see how this lemma implies the theorem, suppose that N_t divides N. Partition the set $w=0,1,\cdots,N-1$ into $\frac{N}{N_t}$ intervals of length N_t each. By the lemma, in each interval $\hat{x}_w$ can not vanish entirely: each interval contains at least one nonzero element of $\hat{x}_w$. Thus the total number of nonzero elements $N_w \geq \frac{N}{N_t}$ and we are done.

For equality $N_t N_w = N$ to be attained, the N_w nonzero elements of $(\hat{x}_w)$ must be equally spaced; otherwise there would be more than N_t consecutive zeros between some pair of nonzero elements of $(\hat{x}_w)$--but the lemma disallows such gaps of length $> N_t$. This 'gap argument' also shows that $N_t N_w > N$ when N_t does not divide N. Let $L \equiv \left\lceil \frac{N}{N_t} \right\rceil$, where $\lceil r \rceil$ denotes the smallest integer greater than or equal to r. There is no way to spread out fewer than L elements among N places without leaving a gap longer than N_t. Thus $N_w \geq L$, so $N_t N_w > N$. We now prove the lemma.

Lemma 1. *If (x_t) has N_t nonzero elements, then $\hat{x}_w$ can not have N_t consecutive zeros.*

Proof. Let $\tau_1, \cdots, \tau_{N_t}$ be the sites where (x_t) is nonzero, and let $b_j \equiv x_{\tau_j}, j = 1, \cdots, N_t$ be the corresponding nonzero elements of (x_t). Denote by $z_1, \cdots, z_{N_t}$ the Fourier transform elements: $z_j \equiv \exp\{-\frac{2\pi i}{N} \cdot \tau_j\}$. Let $w = m+1, \cdots, m+N_t$ be the frequency interval under consideration. Define

$$g_k \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N_t} b_j \, (z_j)^{m+k}, \quad k = 1, \cdots, N_t \; . \tag{2.4}$$

As $\hat{x}_{m+k} \equiv g_k$, the lemma says that $g_k \neq 0$ for some k in the range $1, \cdots, N_t$. We rewrite the assertion (2.4) in terms of matrices and vectors. Define the matrix Z with elements

$$Z_{kj} \equiv \frac{(z_j)^{m+k}}{\sqrt{N}}$$

and the vectors $\boldsymbol{g} \equiv (g_j)$ and $\boldsymbol{b} \equiv (b_j)$. Equation (2.4) takes the form

$$\boldsymbol{g} = Z \, \boldsymbol{b} \; .$$

The conclusion of the lemma is that $g \neq 0$ and we know that $\boldsymbol{b} \neq 0$ by construction. Thus, the lemma is true if the system

$$0 = Z \, \boldsymbol{b}$$

has no solution $b \neq 0$, *i.e.* if Z is nonsingular.

Rescale each column of Z by dividing by its leading entry $z_j^m / \sqrt{N}$. Since $z_j^m \neq 0$, this produces a matrix V which is nonsingular if and only if Z is. The matrix V is

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ z_1 & \cdots & \cdots & \cdots & z_{N_t} \\ z_1^2 & \cdots & \cdots & \cdots & z_{N_t}^2 \\ \cdot & \cdots & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \cdots & \cdot \\ z_1^{N_t-1} & \cdots & \cdots & \cdots & z_{N_t}^{N_t-1} \end{bmatrix}$$

But this is just the usual N_t by N_t Vandermonde matrix, which is known to be nonsingular (Hoffman and Kunze, 1971). Its nonsingularity is equivalent to saying that given N_t data (y_j, z_j), $j=1, \cdots, N_t$, with the z_j distinct, there is a polynomial in z of degree N_t-1 that takes the values (y_j) at (z_j)--a fact which can be demonstrated by the Lagrange interpolation formula. □

3. The Continuous-Time Principle

Let $f(t)$ be a complex-valued function of $t \in \mathbf{R}$, with Fourier transform

$$\hat{f}(\omega) \equiv \int_{-\infty}^{\infty} f(t)\, e^{-2\pi i \omega t} dt \ .$$

We suppose below that the L_2-norm of f, $\|f\| \equiv \left(\int_{-\infty}^{\infty} |f(t)|^2 dt \right)^{½}$ is equal to 1. We may also take the norm of $\hat{f}$; Parseval's identity $\int |f(t)|^2 = \int |\hat{f}(\omega)|^2$ says that $\|\hat{f}\| = 1$ as well.

We say that f is ε–concentrated on a measurable set T if there is a function $g(t)$ vanishing outside T such that $\|f - g\| \le \varepsilon$. Similarly, we say that $\hat{f}$ is ε–concentrated on a measurable set W if there is a function $h(\omega)$ vanishing outside W with $\|\hat{f} - h\| \le \varepsilon$.

Theorem 2. *Let T and W be measurable sets and suppose there is a Fourier transform pair $(f, \hat{f})$, with f and $\hat{f}$ of unit norm, such that f is ε_T*–concentrated *on T and $\hat{f}$ is ε_W*–concentrated *on W. Then*

$$|W| \cdot |T| \ge (1 - (\varepsilon_T + \varepsilon_W))^2. \tag{3.1}$$

Before beginning the proof we introduce two operators; the first is the time-limiting operator

$$(P_T f)(t) \equiv \begin{cases} f(t), & t \in T \\ 0, & \text{otherwise} \end{cases}$$

This operator kills the part of f outside T. Moreover, it gives the closest function to f (in the L_2-norm) that vanishes off T. Thus, f is ε–concentrated on T if and only if $\| f - P_T f \| \le \varepsilon$.

The second operator is the frequency-limiting operator

$$(P_W f)(t) \equiv \int_W e^{2\pi i \omega t} \hat{f}(\omega) \, d\omega .$$

Comparing this with the Fourier inversion formula

$$f(t) = \int_{-\infty}^{\infty} e^{2\pi i \omega t} \hat{f}(\omega) \, d\omega$$

we see that $P_W f$ is a partial reconstruction of f using only frequency information from frequencies in W. If $g = P_W f$ then $\hat{g}$ vanishes outside W. Moreover, g is the closest function to f with this property: $\hat{f}$ is ε–concentrated on W if and only if $\| f - P_W f \| \le \varepsilon$. (This last statement is an application of Parseval's identity.)

We also need to define the norm of an operator Q; it is

$$\| Q \| \equiv \sup_g \frac{\| Qg \|}{\| g \|} .$$

For example, $\| P_W \| = 1$, as we may see by the following argument: if g has energy at frequencies outside W, P_W eliminates it and $\| P_W g \| < \| g \|$; if g has energy only within W then $\| P_W g \| = \| g \|$. In no event is $\| P_W g \| > \| g \|$, and so $\| P_W \| = 1$.

With these definitions in place, the proof of Theorem 2 takes only a few lines. Consider the operator $P_W P_T$ that first time-limits then frequency-limits. By the triangle inequality and the fact that $\| P_W \| = 1$, if f is ε_T–concentrated on T and $\hat{f}$ is ε_W–concentrated on W, then

$$\| f - P_W P_T f \| \le \varepsilon_T + \varepsilon_W . \tag{3.2}$$

We shall see that (3.2) places a rather strict requirement on $|T| \cdot |W|$. Combined with the inequality $\| f - g \| \ge \| f \| - \| g \|$ and the fact that f is of norm 1, (3.2) implies that

$$\| P_W P_T f \| \geq \|f\| - \varepsilon_T - \varepsilon_W$$
$$= 1 - \varepsilon_T - \varepsilon_W ,$$

or equivalently, that

$$\frac{\| P_W P_T f \|}{\|f\|} \geq 1 - \varepsilon_T - \varepsilon_W . \tag{3.3}$$

In terms of the operator norm defined above, we conclude that a pair $(f, \hat{f})$ with f ε_T–concentrated on T and $\hat{f}$ ε_W–concentrated on W can only exist if

$$\| P_W P_T \| \geq 1 - \varepsilon_T - \varepsilon_W . \tag{3.4}$$

We will see below that the norm of $P_W P_T$ obeys the bound

$$\|P_W P_T\| \leq \sqrt{|W||T|} . \tag{3.5}$$

This inequality, together with (3.4), implies the theorem.

Our proof of (3.5) utilizes the so-called Hilbert-Schmidt norm of $P_W P_T$. Define the operator Q

$$(Qf)(t) \equiv \int_{-\infty}^{\infty} q(s, t) f(s)\, ds .$$

The Hilbert-Schmidt norm of Q is just

$$\| Q \|_{HS} \equiv \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |q(s, t)|^2 ds\, dt \right]^{1/2} .$$

It turns out that $\| Q \|_{HS} \geq \| Q \|$ (Halmos and Sunder, 1978), and so (3.5) and Theorem 2 follow from the calculation of $\|P_W P_T\|_{HS}$:

Lemma 2. $\| P_W P_T \|_{HS} = \sqrt{|T|\,|W|}$.

Proof.

$$(P_W P_T f)(s) = \int_W e^{2\pi i \omega s} \int_T e^{-2\pi i \omega t} f(t)\, dt\, d\omega$$
$$= \int_T \left[\int_W e^{2\pi i \omega (s-t)} d\omega \right] f(t)\, dt \tag{3.6}$$

So that

$$(P_W P_T f)(s) = \int_{-\infty}^{\infty} q(s,t) f(t)\, dt$$

where

$$q(s,t) \equiv \begin{cases} \int_W e^{2\pi i \omega(s-t)} d\omega \,, & t \in T \\ 0, & \text{otherwise} \end{cases}$$

Now

$$\| P_W P_T \|_{HS}^2 = \int_T \int_{-\infty}^{\infty} |q(s,t)|^2 ds\, dt \ .$$

Let $g_t(s) \equiv q(s,t)$ and note that $\hat{g}_t = 1_W \cdot e^{-2\pi i t}$, where 1_W is the indicator function for the set W. By Parseval's identity,

$$\begin{aligned} \int_{-\infty}^{\infty} |g_t(s)|^2 ds &= \int_{-\infty}^{\infty} |\hat{g}_t(\omega)|^2 d\omega \\ &= \int_W 1\, d\omega \\ &= |W| \,. \end{aligned}$$

Thus $\int_{-\infty}^{\infty} |q(s,t)|^2 ds = |W|$ and hence

$$\| P_W P_T \|_{HS}^2 = |T|\,|W| \,.$$

In retrospect, this answer is clear since (3.6) shows that $P_W P_T$ involves the integral of a unimodular kernel ($e^{2\pi i \omega(s-t)}$) over a set of measure $|W|\,|T|$.

This proof gives much more information than just the stated conclusion. The norm $\|P_T P_W\|$ (which appeared in an unmotivated fashion) actually satisfies the identity

$$\|P_T P_W\| = \sup_{f \in L_2} \frac{\|P_T P_W f\|}{\|P_W f\|}$$

which relies on the fact that $\|P_W f\| \le \|f\|$. Let $B_2(W)$ denote the set of functions in L_2 that are bandlimited to W (*i.e.* $g \in B_2(W)$ implies $P_W g = g$). Then this identity becomes

$$\|P_T P_W\| = \sup_{g \in B_2(W)} \frac{\|P_T g\|}{\|g\|} \ . \tag{3.7}$$

Thus $\|P_T P_W\|$ in fact measures how nearly concentrated on T a bandlimited finction $g \in B_2(W)$ can be. The inequality $|W||T| > \|P_T P_W\|^2$ thus implies, for example, that if $|W||T| = ½$, no bandlimited function can have more than 50% of its 'energy' concentrated to T. This is a quantitative refinement to (3.1) and is often much more useful. We regard any quantitative bound $c > \|P_T P_W\|$ where $c < 1$ as an expression of the uncertainty principle. In many of the applications we give below, the standard result involves $|W||T|$, but the proof shows the key quantity is $\|P_T P_W\|$.

Theorem 2 has an analog for discrete time. The proof is the same, step-by-step: one merely translates it into the language of finite-dimensional vectors and matrices. The sets T and W become index sets and concentration is defined in terms of the Euclidean norm on $\mathbf{R}^N$. The Frobenius matrix norm (Golub and Van Loan, 1983) provides an analog of the Hilbert-Schmidt norm for matrices. The resulting theorem is slightly more general than Theorem 1 because it does not require (x_t) and $(\hat{x}_w)$ to be perfectly concentrated on T and W; however the proof is not useful in identifying the extremal functions of the inequality, which we show in Appendix A are simple spike trains. We state this theorem without proof.

Theorem 3. *Let $((x_t), (\hat{x}_w))$ be a Fourier transform pair of unit norm, with (x_t)* ε_T–concentrated *on the index set T and $(\hat{x}_w)$* ε_W–concentrated *on the index set W. Let N_t and N_w denote the number of elements of T and W, respectively. Then*

$$N_t N_w \geq N(1-(\varepsilon_T+\varepsilon_W))^2 . \tag{3.8}$$

4. Recovering Missing Segments of a Bandlimited Signal.

Often the uncertainty principle is used to show that certain things are impossible (*e.g.* determining the momentum and position of a particle simultaneously; measuring the 'instantaneous frequency' of a signal). In this section and the next we present two examples where the generalized uncertainty principle shows something unexpected is *possible*; specifically, the recovery of a signal or image despite significant amounts of missing information.

The following example is prototypical. A signal $s(t)$ is transmitted to a receiver who knows that $s(t)$ is bandlimited, meaning that s was synthesized using only frequencies in a set W (which for our purposes may be an interval or some other measurable set). Equivalently, $P_W s = s$, where P_W is the bandlimiting operator of the previous section. Now suppose the receiver is unable to observe all of s; a certain subset T (*e.g.* a collection of intervals) of t-values is unobserved. Moreover, the observed signal is contaminated by observational noise $n(t)$. Thus the received signal $r(t)$ satisfies

$$r(t) = \begin{cases} s(t) + n(t)\,, & t \in T^c \\ 0, & t \in T \end{cases} \tag{4.1}$$

where T^c is the complement of the set T. Equivalently,

$$r = (I - P_T)(s) + n$$

where I denotes the identity operator $(If)(t) = f(t)$.

The receiver's aim is to reconstruct the transmitted signal s from the received signal r. Although it may seem that information about $s(t)$ for $t \in T$ is completely unavailable, the uncertainty principle says recovery is possible provided $|T|\,|W| < 1$!

To see that this is true intuitively, consider what could go wrong. If there were a bandlimited function h completely concentrated on T, the measured data would show no trace of h. The data would be the same, regardless of whether the true signal was $s(t)$ or $s(t) + \alpha h(t)$, $\alpha \in \mathbf{R}$. Thus on the basis of the data and the knowledge that s is bandlimited to W, we would have no way of discriminating between the competing reconstructions $s_0 = s$ and $s_1 = s + \alpha h$. At the very least, our uncertainty about the reconstruction would be $\|s_0 - s_1\| = \|s - (s + \alpha h)\| = |\alpha|\,\|h\|$, where $|\alpha|$ could be arbitrarily large: our uncertainty would be completely unbounded.

However, Theorem 2 says that there is no such function h if $|W|\,|T| < 1$. It seems surprising, but the lack of such a function implies that reconstruction is possible. We shall say that s can be stably reconstructed from r if there exists a linear operator Q and a constant C such that

$$\| s - Qr \| \le C \|n\| \tag{4.2}$$

for all s, r, and n obeying (4.1). The operator Q is the *reconstruction method*; its action produces an estimate of s based on r that is in error by at most some multiple C of the noise level $\|n\|$. Of course, we would like C to be small.

Theorem 4. *If W and T are arbitrary measurable sets with $|T||W| < 1$, s can be stably reconstructed from r. The coefficient C in (4.2) is not larger than* $(1-\sqrt{|T||W|})^{-1}$.

Proof. We propose the operator $Q \equiv (I - P_T P_W)^{-1}$ as a reconstruction method. First we should establish that Q exists. Note that $\|P_T P_W\| = \|P_W P_T\| \le \sqrt{|T||W|} < 1$. By a well-known argument, the linear operator $I - L$ is invertible if $\|L\| < 1$. Invertibility can also be seen by a direct formal argument. We actually have a formula for Q:

$$Q \equiv (I - P_T P_W)^{-1} \overset{formally}{=} \sum_{k=0}^{\infty} (P_T P_W)^k. \tag{4.3}$$

But since $\|P_T P_W\| < 1$, the sum converges. Now $(I - P_T)s = (I - P_T P_W)s$ for every bandlimited s, because $P_W s = s$. Thus,

$$\begin{aligned} s - Qr &= s - Q(I - P_T)s - Qn \\ &= s - (I - P_T P_W)^{-1}(I - P_T P_W)s - Qn \\ &= 0 - Qn \end{aligned}$$

so

$$\begin{aligned} \|s - Qr\| &\le \|Qn\| \\ &\le \|Q\| \, \|n\|. \end{aligned}$$

Now as $Q = I + P_T P_W + (P_T P_W)^2 + \cdots$, and the operator norm is subadditive,

$$\begin{aligned} \|Q\| &\le \sum_{k \ge 0} \|P_T P_W\|^k \\ &= (1 - \|P_T P_W\|)^{-1}. \end{aligned}$$

It follows from (3.5) that we may take $C = (1-\sqrt{|T||W|})^{-1}$ in (4.2). □

The identity (4.3) suggests an algorithm for computing Qr. Put $s^{(n)} = \sum_{k=0}^{n} (P_T P_W)^k r$; then $s^{(n)} \to Qr$ as $n \to \infty$. Now

$$\begin{aligned} s^{(0)} &= r \\ s^{(1)} &= r + P_T P_W s^{(0)} \\ s^{(2)} &= r + P_T P_W s^{(1)} \end{aligned}$$

$$\cdots \tag{4.4}$$

and so on. So $s^{(n)}$ is the result of bandlimiting then timelimiting $s^{(n-1)}$, then adding the result back to the original data r. This very simple iteration recovers s. The iterations converge at a geometric rate to the fixed point

$$s^* = r + P_T P_W s^* .$$

On T^c (the complement of the set T) where the data were observed, $s^{(n)} = r$ at each iteration n, while on the unobserved set T the missing values are magically filled in by a gradual adjustment, iteration after iteration.

The algorithm (4.4) is an instance of the *alternating projection* method: it alternately applies the bandlimiting projector P_W and the timelimiting projector P_T. Algorithms of this type have been applied to a host of problems in signal recovery: for beautiful and illuminating applications see the papers of Landau and Miranker (1961), Gerchberg (1974), and Papoulis (1975); for a more abstract treatment, Youla, 1978; Youla and Webb, 1982; Schafer *et al.* (1981) give a nice review.

Note that the original uncertainty principle, which requires both W and T to be intervals, would help here only if the set T where data are missing consisted of a single interval.

5. Recovery of a 'sparse' wide-band signal from narrow-band measurements

In several branches of applied science, instrumental limitations make the available observations bandlimited, even though the phenomena of interest are definitely wide-band. In astronomy, for example, diffraction causes bandlimiting of the underlying wideband image, despite the fact that the image is a superposition of what are nearly Dirac delta functions which, by the uncertainty principle, have extremely broad Fourier transforms. The same is true in spectroscopy, where the 'image' is itself a spectrum.

Although it may seem that accurate reconstruction of a wideband signal from narrowband data is impossible--'the out-of-band data were never measured, so they are lost forever'--workers in a number of fields are trying to do exactly this. They claim to be able to recover the missing frequency information using side information about the signal to be reconstructed, such as its 'sparse' character in the cases mentioned. We first became aware of these efforts in seismic exploration (the interested reader is referred to the papers of Levy and Fullagar (1981), Oldenburg *et al.* (1983), Walker and Ulrych (1983), and Santosa and Symes (1986)), but later found examples in other fields such as medical ultrasound (Papoulis and Chamzas, 1979).

The discrete-time uncertainty principle suggests how sparsity helps in the recovery of missing frequencies. Suppose that the discrete-time measurement (r_t) is a noisy, bandlimited version of the ideal signal (s_t):

$$r = P_B s + n \tag{5.1}$$

where $n = (n_t)$ denotes the noise and P_B is the operator that limits the measurements to the passband B of the system. Here we let P_B be the 'ideal' bandpass operator:

$$P_B s = \frac{1}{\sqrt{N}} \sum_{w \in B} \hat{s}_w e^{\frac{2\pi i w t}{N}} . \tag{5.2}$$

If we take discrete Fourier transforms, (5.1) becomes

$$\hat{r}_w = \begin{cases} \hat{s}_w + \hat{n}_w , & w \in B \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

where we have assumed (without loss of generality) that the noise n is also bandlimited.

Let W denote the set of unobserved frequencies $W \equiv B^c$, and let N_w denote its cardinality. The equation (5.3) represents a frequency-domain missing data problem analogous to the time-domain missing data problem of section 4.

It may seem that the the problem is hopeless. After all, as (5.3) shows, the data $(\hat{r}_w : w \in W)$ are not observed. Even if there were no noise, one might be skeptical that anything could be done. Enter the uncertainty principle.

Theorem 5. *Suppose there is no noise in (5.1), so that $r = P_B s$. If it is known that s has only N_t nonzero elements, and if*

$$2N_t N_w < N \ , \tag{5.4}$$

then s can be uniquely reconstructed from r.

To prove this, we first show that s is the unique sequence satisfying (5.4) that can generate the given data r. Suppose that s_1 also generates r, so $P_B s_1 = r = P_B s$. Put $h = s_1 - s$, so that $P_B h = 0$. Now s_1 also has fewer than N_t nonzero elements, so h has fewer than $N'_t = 2N_t$ nonzero elements; because $P_B h = 0$, its Fourier transform has at most N_w nonzero elements. Then h must be zero, for otherwise it would violate Theorem 1 (since $N'_t N_w < N$). Thus $s_1 = s$; we have established uniqueness.

To actually reconstruct s from r, one could use a 'closest point' algorithm: let $\tilde{s}$ be the sequence minimizing $||r - P_B s'||$ among all sequences s' with N_t or fewer nonzero elements. From the last paragraph, we know that $\tilde{s} = s$.

An algorithm for obtaining $\tilde{s}$ is combinatorial in character. Let N_t be given and let Π denote the $\binom{N}{N_t}$ subsets τ of $\{0, \cdots, N-1\}$ having N_t elements. For a given subset $\tau \in \Pi$ let $\tilde{s}_\tau$ be the sequence supported on τ that comes closest to generating the data r, *i.e.* the solution to the least squares problem

$$\min \{ ||r - P_B s'|| : P_\tau s' = s' \} \ .$$

Then $\tilde{s} = \tilde{s}_{\tau_0}$ for some $\tau_0 \in \Pi$; one merely has to find which:

$$\tilde{s} = \arg \min_{\tilde{s}_\tau, \tau \in \Pi} ||r - P_B \tilde{s}_\tau|| \ .$$

This algorithm requires solving $\binom{N}{N_t}$ sets of linear least-squares problems, each one requiring $O(N_t^3)$ operations, so that it is totally impractical for large N. A much better approach will emerge in section (6.3); for its justification, we will need yet another uncertainty principle.

Theorem 5 establishes uniqueness; can one establish stability in the presense of noise? Any sparsity constrained reconstruction method will be nonlinear, so such a result will be delicate. We state one without proof.

Theorem 6. *Suppose that the 'signal-to-noise' ratio*

$$\frac{\min \{ |s_t| : s_t \neq 0 \}}{||n||} > \left[1 - \sqrt{\frac{2N_t N_w}{N}} \right]^{-1} .$$

Then s can be stably reconstructed from r with stability coefficient $C \le \left[1 - \sqrt{\frac{2N_t N_w}{N}}\right]^{-1}$.

The result is not really practical; we give it only to illustrate two features we expect to be present in a practical result. First, that stability depends on having a large enough signal-to-noise ratio; second, that the uncertainty principle helps to determine not only the stability coefficient but also the signal-to-noise threshold. A practical result would probably use a different notion of signal-to-noise ratio.

The bound $2N_w N_t < N$ is rather disturbing. It demands an unusual degree of sparsity: even if only 10% of the frequencies are missing, it requires that s contain no more than 5 nonzero entries in the time domain. Can one have uniqueness and stability if $2N_w N_t \ge N$? Closer study shows that the right necessary condition is of the form $\|P_W P_T\| < c < 1$, where T is a set of $2N_t$ sites, N_t of which are the support of the signal s to be recovered, and the other N_t of which are arbitrary. Theorem 3 (or more precisely, its proof) shows that $\|P_W P_T\|$ can be bounded in this way if $2N_w N_t < N$, but this bound is not sharp.

In section 7.2 we will consider the sharpness of the discrete uncertainty principle and come to the conclusion that the condition $2N_t N_w < N$ may be relaxed if the locations of the nonzero entries in s are known to be widely scattered. On the other hand, if the entries might be close together, (5.4) is essentially the best one can do.

6. An L_1 Uncertainty Principle and Applications

6.1 The L_1 Principle

For signal reconstruction problems like those of sections 4 and 5, it is also useful to have uncertainty principles for the L_1-norm. The results are not as neat as the L_2 results, but they have quite remarkable applications.

The L_1-norm of the function f is, of course, $\|f\|_1 \equiv \int |f(t)|\, dt$; we also will need the L_∞-norm $\|f\|_\infty \equiv \sup_t |f(t)|$. As before, we say that f is ε-concentrated to T if $\|f - P_T f\|_1 < \varepsilon$. Let $B_1(W)$ denote the set of functions $f \in L_1$ that are bandlimited to W. We say that f is ε-bandlimited to

W if there is a $g \in B_1(W)$ with $\|f - g\|_1 < \varepsilon$. With this equipment, the statement of the theorem is as expected . . . except for a factor $(1-\varepsilon)^{-1}$.

Theorem 7. *Let f be of unit L_1-norm. If f is ε_T-concentrated to T and ε_W-bandlimited to W, then*

$$|W|\,|T| \geq \frac{1 - \varepsilon_T - \varepsilon_W}{1 - \varepsilon_W} . \tag{6.1}$$

We will prove the theorem in two steps, first assuming that $\varepsilon_W = 0$. If $\varepsilon_W = 0$ then $f \in B_1(W)$. By hypothesis,

$$\frac{\|P_T f\|_1}{\|f\|_1} \geq 1 - \varepsilon_T .$$

Define the operator norm

$$\mu_0(W, T) \equiv \sup_{f \in B_1(W)} \frac{\|P_T f\|_1}{\|f\|_1} \tag{6.2}$$

(note that the analog for L_2 is just the operator norm $\|P_T P_W\|$ of equation (3.7)); then

$$\mu_0(W, T) \geq 1 - \varepsilon_T .$$

The desired result then follows from

Lemma 3.

$$\mu_0(W, T) \leq |W|\,|T| .$$

Proof. For $f \in B_1(W)$ we have

$$\begin{aligned} f(t) &= \int_W e^{2\pi i \omega t} \hat{f}(\omega)\, d\omega \\ &= \int_W \int_{-\infty}^{\infty} e^{2\pi i \omega (t-s)} f(s)\, ds\, d\omega \\ &= \int f(s) \int_W e^{2\pi i \omega (t-s)} d\omega\, ds . \end{aligned}$$

So that

$$|f(t)| \leq \int |f(s)| \int_W 1\, d\omega\, ds$$

or

$$\|f\|_\infty \leq |W|\, \|f\|_1 . \tag{6.3}$$

On the other hand, for $f \in L_1$

$$||P_T f||_1 = \int_T |f| \le ||f||_\infty |T| . \tag{6.4}$$

Combining (6.3) and (6.4) we have for $f \in B_1(W)$

$$\frac{||P_T f||_1}{||f||_1} \le \frac{||f||_\infty |T|}{||f||_\infty / |W|} = |W||T| . \square$$

Now suppose that $\varepsilon_W \neq 0$. If f is ε_W-bandlimited, by definition there is a g in $B_1(W)$ with $||g - f||_1 \le \varepsilon_W$. For this g, we have

$$\begin{aligned} ||P_T g||_1 &> ||P_T f||_1 - ||P_T(g - f)||_1 \\ &> ||P_T f||_1 - \varepsilon_W \end{aligned}$$

and also

$$||g||_1 < ||f||_1 + \varepsilon_W ,$$

so that

$$\frac{||P_T g||_1}{||g||_1} \ge \frac{||P_T f||_1 - \varepsilon_W}{||f||_1 - \varepsilon_W} = \frac{1 - \varepsilon_T - \varepsilon_W}{1 - \varepsilon_W} .$$

Thus $\mu_0(W, T) > (1 - \varepsilon_T - \varepsilon_W)/(1 - \varepsilon_W)$; this combined with Lemma 3 proves (6.1). $\square$

6.2 Logan's Phenomenon

Consider the following continuous-time signal reconstruction problem. The bandlimited signal s is transmitted to the receiver who measures s *perfectly* except on a set T, where the signal has been distorted by a noise n. The set T is *unknown* to the receiver, and the noise is *arbitrary* except that $||n||_1 < \infty$. In short, the received signal r satisfies

$$r = s + P_T n .$$

The aim is to reconstruct s.

The method to be used is L_1-reconstruction, letting $\tilde{s}$ be the closest bandlimited function to r in the L_1-norm

$$\tilde{s} = \arg \min_{s' \in B_1(W)} ||r - s'||_1 .$$

One might suppose that reconstruction is made difficult because n may be very large: we have not constrained $\|n\|$. Enter the uncertainty principle, again.

Theorem 8. *If $|W||T| < ½$, the L_1 method recovers* perfectly*: $\tilde{s} = s$, whatever be n.*

To see why this is true, consider the special case where $s = 0$. Thus $r = n$. Theorem 8 requires that the best bandlimited approximation to r be zero.

Here is where the uncertainty principle acts. As $|W||T| < ½$, every bandlimited function $g \in B_1(W)$ is less than 50% concentrated on T: $\|P_T g\|_1 < ½\|g\|_1$, and so $\|P_T g\|_1 < \|P_U g\|_1$ where $U = T^c$ and $P_U = I - P_T$, projection on the complement of T.

This will imply that the best bandlimited approximation to n is zero. Indeed, if $g \in B_1(W)$ and $g \neq 0$

$$\begin{aligned} \|n - g\|_1 &= \|P_T(n - g)\|_1 + \|P_U g\|_1 \\ &\geq \|P_T n\|_1 - \|P_T g\|_1 + \|P_U g\|_1 \\ &> \|P_T n\|_1 = \|n\|_1 . \end{aligned}$$

We have just proved

Lemma 4. *Let $|W||T| < ½$. If n vanishes outside of T, then its best bandlimited approximation from $B_1(W)$ is 0.*

The role of the uncertainty principle in this lemma should be emphasized: it says that because no bandlimited function is as much as 50% concentrated to T, any effort to approximate n well on T incurs such a penalty on T^c that one does better not to try.

To prove Theorem 8, suppose that $s \neq 0$, and let $g \in B_1(W)$.

$$\begin{aligned} \|r - g\|_1 &= \|s + n - g\|_1 \\ &= \|n + (s - g)\|_1 . \end{aligned}$$

We want to minimize this expression over all $g \in B_1(W)$. Now $(s - g)$ is bandlimited; the lemma says that this expression is minimized if $(s - g) = 0$. It follows that $\tilde{s} = s$.

This rather striking phenomenon first appeared in B.F. Logan's Thesis (Logan, 1965). He proved the theorem for the case where $W = [-\Omega/2, \Omega/2]$, without explicitly noting the connection to the uncertainty principle. We think our proof makes the result rather intuitive.

One might call this result "Logan's Certainty Principle", as it says that s can be reconstructed with *certainty* when the set T where s is corrupted is smaller than some critical threshhold value. Logan envisions applying this result to the problem of smoothing away high-energy impulsive noise. By using the L_1-technique, the effects of such noise can be entirely removed, provided the total duration of the noise bursts is short. In our view, this is a powerful, novel property of L_1-methods in signal processing.

The crucial and surprising thing here is that T is unknown and may be totally arbitrary (provided $|T|$ is small), and n may be arbitrary as well. In contrast the application in section 4 required that T be known and that $\|n\|$ be small.

6.3 The Sparse Spike Train Phenomenon

Theorem 7 has an analog for discrete time in much the same way Theorem 2 has Theorem 3. There is also a discrete-time version of the "Certainty" phenomenon.

To apply these, return to the sparse signal reconstruction problem of section 5. We observe r, a bandlimited version of s; assuming no noise is present, $r = P_B s$, where P_B is the ideal bandlimiter of (5.2). Here B is the system passband and $W = B^c$ is the set of unobserved frequencies.

We saw in section 5 that provided s is sparse with $2N_t N_w < N$, we can recover s from r. However, the combinatorial algorithm we proposed is unnatural and impractical.

Consider instead an l_1-reconstruction algorithm. The l_1-norm of (x_t) is defined to be $\|x\|_1 = \sum_{t=0}^{N-1} |x_t|$. Let $\tilde{s}$ be the signal with smallest l_1-norm generating the observed data r:

$$\tilde{s} = \arg\min_{s'} \|s'\|_1 \text{ subject to } P_B s' = r \ . \tag{6.5}$$

This estimate may be conveniently obtained by linear programming (*e.g.* see Levy and Fullagar (1981), Oldenburg *et al.* (1983), or Santosa and Symes (1986)). In practice it requires $O(N N_t^2 \log N)$ time as compared with $O(\binom{N}{N_t} \cdot N_t^3)$ time for the combinatorial approach. As the cited papers show, the method also has an elegant extension to the case where noise is present. The uncertainty principle shows that the method recovers s in the noiseless case.

Theorem 9. *If* $2N_t N_w < N$, *then*

$$\tilde{s} = s \quad exactly \ .$$

The proof is an application of Logan's phenomenon. First, note that by (6.5), $P_B \tilde{s} = P_B s$, so $\tilde{s} = s + h$ where $P_B h = 0$. Thus h is bandlimited to $W = B^c$. On the other hand, s is sparse--it differs from the sequence $(0, \cdots, 0)$ in only N_t places. As $2N_t N_w < N$, the discrete Logan's phenomenon implies that the best l_1-approximation to s by sequences bandlimited to W is just the zero sequence. In other words,

$$\arg \min_{h \in B_1(W)} \| s + h \|_1 = 0 \ !$$

or

$$\arg\min_{s'} \{ \| s' \| : P_B s' = P_B s \} = s \ .$$

Instances of this phenomenon were demonstrated empirically by Geophysicists in the early 80's. A formal theorem and proof were discovered by Santosa and Symes (1986), who actually mention the uncertainty principle in passing during their proof. (They do not, however, mention exactly which uncertainty principle they are using nor indicate why it is intrinsic to the result.) We think the connection with Logan's phenomenon and with the l_1-uncertainty principle are enlightening here.

7. Sharpness

7.1 Sharpness of the Continuous-Time Uncertainty Principle

Is there a converse to Theorem 2? If T and W are sets with $|T||W| > 1$ will there exist a pair $(f, \hat{f})$ with f practically concentrated on T and $\hat{f}$ practically concentrated on W? This is a difficult question. If T is an interval and W is an interval, the answer is yes (with the appropriate values of ε_T and ε_W). Otherwise, little is known.

As we saw in (3.4), the existence of such a pair is equivalent to

$$\|P_W P_T\| > 1 - \varepsilon_T - \varepsilon_W .$$

Define $\lambda_0(W, T) \equiv \|P_W P_T\|^2$. The work of Slepian, Landau and Pollak provides a great deal of insight into the behavior of λ_0 when W and T are single *intervals* (Slepian and Pollak, 1961; Landau and Pollak, 1961). They show that in this case λ_0 is the largest eigenvalue of the operator $P_W P_T P_W$; and they give a complete eigenanalysis of this operator, identifying its eigenfunctions as the prolate spheroidal wavefunctions. They also show that $\lambda_0(W, T)$ is a function of $c \equiv 2\pi |W| |T|$ alone; call this function $\lambda_0^*(c)$. Slepian and Sonnenblick (1965) give the following table:

c	$\lambda_0^*(c)$
6	0.99990188
7	0.99998546
8	0.99999787
9	0.99999969
10	0.99999996
11	0.99999999

Thus although perfect concentration is impossible when W and T are intervals, if the product of their lengths is at least 1 (*i.e.* if $c > 2\pi$), there is a transform pair $(f, \hat{f})$ such that ε_T and ε_W are quite small (< 0.0001).

For general sets T and W, however, the uncertainty principle can be far from sharp. In particular, if T is the union of very many very 'thin' intervals, then it can be extremely hard to concentrate a bandlimited function on T, even if $|T|$ is quite large.

Theorem 10. *Let W be an interval. Let T be a union of n equal width intervals. Let the minimum separation between subintervals of T tend to ∞; then*

$$\|P_W P_T\| \to \lambda_0^* \left(2\pi \frac{|W| |T|}{n} \right). \tag{7.1}$$

As $\lambda_0^*(2\pi \frac{|W||T|}{n}) \le \frac{|W||T|}{n}$, the right hand side of (7.1) can be small if n is large. The theorem says that for W an interval, there are sets T where $|W| |T| = 1$ but $\lambda_0(W, T)$ is arbitrarily small. It also shows that there are sets T where $\lambda_0(W, T) \le ½$ but $|W| |T|$ is arbitrarily large. In these cases T is a union of many 'thin' intervals.

This result is the best of its kind, in a certain sense. If T is the union of n intervals T_i, it is easy to see that

$$\begin{aligned}\lambda_0(W,T) &\ge \max_i \lambda_0(W,T_i)\\ &= \lambda_o^*(2\pi |W| \max_i |T_i|)\\ &\ge \lambda_0^*\left[2\pi\frac{|W||T|}{n}\right].\end{aligned}$$

Thus the right side of (7.1) is always a lower bound for the left side; the theorem says that it may almost be attained.

Comparing the theorem with Table 1, one gets the impression that when W is an interval, if one varies T keeping $|T|$ fixed, $\lambda_0(W,T)$ would be largest when T is an interval and smallest when T is 'fractured' into many thin sets. Thus intervals would be the easiest sets to concentrate on, and 'thin sets' would be the hardest. This leads to

Conjecture 1. max $\|P_W P_T\|$, *where* W *is an interval and* T *ranges over measurable sets with* $|T||W| = C$, *is attained when both W and T are intervals.*

Ingrid Daubechies has shown us a perturbation theory argument that implies the conjecture is true 'infinitesmally'--that $\|P_W P_T\|$ decreases as T is perturbed away from an interval to a union of intervals having the same total measure but with small gaps between the intervals.

The fact that $\|P_W P_T\| \ll |W||T|$ when T is fractured has positive applications to the problem of section 4. What we really proved in Theorem 4 was that s could be stably reconstructed from r provided $\|P_W P_T\| < 1$, with stability coefficient $(1-\|P_W P_T\|)^{-1}$. In view of Theorem 10 and its corollaries, we can see that, if W is an interval, then we can have $\|P_W P_T\|$ arbitrarily close to zero with $|W||T|$ arbitrarily large. Consequently, when the set T is 'thin enough', s can be stably reconstructed from r even though $|W||T| \gg 1$, and in fact with a stability coefficient close to 1.

7.2 Sharpness of the Discrete-Time Principle

When N is a highly composite number, the periodic spike train examples of section 2 show that many pairs $((x_t), (\hat{x}_w))$ attain equality $N_t N_w = N$. In this sense, the discrete-time principle is sharp.

On the other hand, Appendix A shows that the index sets T and W where the bound is attained are all highly regular (equispaced). For arbitrary index sets T and W with $N_t N_w \geq N$ it could be that no sequences exist that are perfectly concentrated to T in the time domain and to W in the frequency domain.

Defining P_T and P_W for discrete-time signals in the obvious way, it turns out that just as in the continuous time case, there exist transform pairs $((x_t), (\hat{x}_w))$ with (x_t) ε_T–concentrated to T and $(\hat{x}_w)$ ε_W–concentrated to W if and only if $\| P_T P_W \| \geq 1 - \varepsilon_T - \varepsilon_W$.

Thus, defining $\lambda_0 \equiv \| P_W P_T \|^2$, the uncertainty principle (3.8) is nearly sharp if for the sets T and W of interest one has

$$\frac{N_t N_w}{N} > 1 \quad \text{and}$$
$$\lambda_0(W, T) \approx 1. \tag{7.2}$$

As in the continuous-time case, the principle seems nearly sharp if T and W are intervals

$$T = \{t_0, \cdots, t_0 + N_t - 1\}$$
$$W = \{w_0, \cdots, w_0 + N_w - 1\}.$$

For W and T intervals, we have found that $\lambda_0(W, T)$ is very close to 1 when $\frac{N_t N_w}{N} \geq 1$. In the discrete case, $\lambda_0(W, T)$ is just the largest eigenvalue of the matrix $P_W P_T P_W$, and can be computed numerically. Table 2 presents results for the cases N = 64, 96, 128, 192, 256, with (N_t, N_w) chosen so that $N_t = N_w$ and $\frac{N_t N_w}{N} \approx 1$, 2, or 3.

Table 2: λ_0 as a function of $N_t N_w / N$; W and T intervals

$N_w N_t / N \approx$	λ_0 for $N=$ 64	96	128	192
1	0.913	0.918	0.904	0.907
2	0.992	0.988	0.987	0.987
3	0.999	0.999	0.999	0.999

Thus it is rather easy to concentrate on the pairs of sets (T, W) when T and W are both intervals.

On the other hand, if W is an interval and T is allowed to range over sets that are not intervals, it can be quite difficult to concentrate on (T, W). Our examples concern the case where W comprises the low frequencies $\{0, 1, \cdots, N_w-1\}$ and T consists of equally-spaced sites. The basic tool here is

Theorem 11. *Let* $m \geq \sqrt{N}$ *divide* N. *Let* $W = \{0, \cdots, m-1\}$ *and* $T = \{0, \frac{N}{m}, \frac{2N}{m}, \cdots\}$. *Then* $N_w = N_t = m$ *and*

$$\lambda_0(W, T) = \frac{m}{N} \leq 1,$$

yet

$$\frac{N_t N_w}{N} = \frac{m^2}{N} \geq 1.$$

This result can be used to construct counter-examples to (7.2) where $\lambda_0 < c < 1$ but $\frac{N_t N_w}{N}$ is arbitrarily large, and examples where $\frac{N_t N_w}{N} = 1$ but λ_0 is arbitrarily small:

Corollary 2. *Let* N *be even. Let* W *consist of the* $\frac{N}{2}$ *lowest frequencies and let* T *be the even numbers* $\{0, 2, \cdots, \frac{N-1}{2}\}$. *Then* $\lambda_0(W, T) = ½$, *but*

$$\frac{N_w N_t}{N} = \frac{N}{4}.$$

Corollary 3. *Let* N *be a perfect square. Let* W *consist of the* $\sqrt{N}$ *low frequencies and let* T *be a set of equispaced points with spacing* $\sqrt{N}$. *Then* $\frac{N_t N_w}{N} = 1$ *but*

$$\lambda_0(W, T) = N^{-1/2}.$$

In short, the uncertainty principle can be arbitrarily 'non-sharp' in this case.

These examples are the best of their kind. From the inequality that the largest eigenvalue of a nonnegative-definite matrix is at least the trace divided by the number of nonzero eigenvalues,

$$\begin{aligned}
\lambda_0(W, T) &\geq \frac{\text{trace } (P_W P_T P_W)}{\text{rank } (P_W P_T P_W)} \\
&= \frac{N_w N_t}{N} \cdot \frac{1}{\min(N_w, N_t)} \\
&= \frac{\max(N_w, N_t)}{N}.
\end{aligned}$$

The fact that trace $P_W P_T P_W = \frac{N_w N_t}{N}$ is the discrete-time analog of Lemma 2. In Theorem 11 and its two corollaries, this lower bound is attained.

The discrete-time case seems to have some interesting structure. Suppose $N_t N_w = N$. If W is an equispaced set then if T is also equispaced, concentration is maximal: $\lambda_0(W, T) = 1$ (by the appendix). If W is still equispaced but now T is an interval, the concentration is minimal (by the previous paragraph). On the other hand if W is an interval the situation is the reverse: T equispaced minimizes the concentration (by the last paragraph), while T an interval seems to maximize it. Numerical experiments support the following

Conjecture 2. *If W is an interval and $N_w N_t = N$, then $\|P_W P_T\|$ is maximized among all sets T of fixed cardinality N_t when T is also an interval.*

If the conjecture is true, intervals and equispaced sets play completely dual roles.

7.3 Lack of Sharpness when T is 'Random'

The lack of sharpness when W is an interval and T is scattered has positive applications to the signal recovery problem of section 5. As we saw there, the uncertainty principle suggests that recovery of a sparse sequence from data missing low frequencies would place severe restrictions on the number of spikes in the sequence. However, we have just seen that the uncertainty principle may be far from sharp, so for some sets T, $N_w N_t$ may be much larger than N without admitting highly concentrated sequences. By an argument we will not repeat here, this suggests recovery is possible. Unfortunately, the examples so far have T equispaced; such perfect spacing is not plausible in practical signal recovery problems.

A more realistic situation is when W consists of the low frequencies $\{0, \cdots, N_w-1\}$ and T is a set of sites chosen at random (*i.e.* by drawing N_t integers from a "hat" containing $\{0, \cdots, N-1\}$). We have investigated this setup on the computer; the results suggest that randomly-selected and equispaced sets T behave similarly.

In our investigation we used several different sequence lengths N: 64, 128, and 256. For each sequence length N, we let N_w and N_t range systematically between 8 and 70, and 2 and 50, respectively. For each choice of N_w and N_t, we let $W = \{0, \cdots, N_w-1\}$ and we randomly generated 20 sets

T with N_t elements. We computed $\lambda_0(W, T)$ for each of the 20 cases; averaging across these we arrived at

$$\overline{\lambda}_0(N_w, N_t, N) = \mathrm{Ave}\{\lambda_0(W, T)\}.$$

In all, we examined more than 30,000 combinations of sets T and W.

Figure 1 shows $\overline{\lambda}_0(N_w, N_t, N)$ versus $\frac{N_w N_t}{N}$. The different symbols are for different length sequences: the dots are results for signals of length 64, the circles are for length 128, and the plusses are for sequences with 256 elements. The plot shows that even with $N_t N_w = 8N$, quite often $\lambda_0 < 0.8$: there is no pair $((x_t), (\hat{x}_w))$ that is simultaneously concentrated on W and T even though the product of their cardinalities is large compared to N. This situation grows more pronounced with N: on the average, the uncertainty principle is less and less sharp for larger and larger N, when W is an interval and T is a random set. This leads us to

Conjecture 3. *Let W be the interval containing the lowest N_w frequencies and let T be a randomly-selected set of N_t sites. Suppose $c = \frac{N_t N_w}{N}$ and $a = \frac{N_t}{N_w}$ are held fixed as $N \to \infty$. Then*

$$E(\lambda_0) \to 0$$

where $E(\lambda_0)$ denotes the expectation of λ_0 under random selection of the sites in T.

Results like this, if true, would suggest that in the problem of section 5, it may be possible to recover many, many times more spikes than (5.4) indicates -- if the spike positions are scattered 'at random'. If they are all together in one interval, however, our uncertainty principle is nearly sharp. Thus those who claim that 'sparsity' allows reconstruction of wideband signals from narrowband data should say that 'scatteredness' is needed also (unless truly extreme sparsity is present).

8. Discussion

a. While the uncertainty principle is true with a great deal of generality, it becomes false if we try to extend its scope to locally finite measures. For example, let δ_t denote the Dirac delta measure $\delta_t(S) = \{1, \text{ if } t \in S;\ 0, \text{ otherwise }\}$. Then fix $h > 0$ and put

$$\nu \equiv \sum_{k=-\infty}^{\infty} \delta_{h \cdot k} .$$

This measure can be viewed as an infinite train of equally-spaced spikes. Formally, the Fourier transform of ν is

$$\hat{\nu} = \sum_{k=-\infty}^{\infty} \delta_{k/h} .$$

This is the Fourier transform of ν in a distributional sense: the Parseval relation

$$\int \psi \, d\nu = \int \hat{\psi} \, d\hat{\nu}$$

holds for every infinitely differentiable test function ψ of compact support (see Katznelson, 1976, for the Fourier theory of locally finite measures). The pair $(\nu, \hat{\nu})$ furnish a 'counterexample' to our uncertainty principle since ν is supported on a set of zero measure, as is $\hat{\nu}$. Therefore

$$|\,\text{support}\,\nu\,|\;|\,\text{support}\,\hat{\nu}\,| = 0 \;< 1 .$$

On the other hand, if one takes a sequence of smooth test functions (f_n) converging in the distributional sense to ν (*i.e.*

$$\int \psi f_n \to \int \psi \, d\nu \quad \text{as } n \to \infty \tag{8.1}$$

for every test function ψ) then each of these functions f_n will satisfy the uncertainty principle (by Theorem 2).

b. Our idea for proving Theorem 1 arose while reading Paul Turán's 'On a new method of analysis and its applications.' The key to our proof is of course the following: let

$$g(\nu) = \sum_{i=1}^{n} b_i \, z_i^{\nu} .$$

If the z_i are distinct then $g(\nu)$ can not vanish at n consecutive integers unless $b_1 = b_2 = \cdots = b_n = 0$. Equivalently,

$$\max_{m \le \nu \le m+n} |g(\nu)| = 0 \quad \text{if and only if } b_1 = \cdots = b_n = 0 .$$

Turán's book is devoted to various quantitative refinements of this and related principles and applications to analysis and number theory. Turán finds bounds for $\max_{m<\nu\le m+n} |g(\nu)|$ under various constraints on (b_i) and (z_i). Our first proof of Theorem 1 obtained Lemma 1 as a special case of Turán's Theorem 11.1. In present terms that inequality can be stated as

$$\max_{m<\nu\le m+n} |g(\nu)| \ge \frac{1}{n}\left[\frac{\delta}{2}\right]^{n-1} \sum_{j=1}^{n} |b_j|$$

where $\delta = \min_{i,j} |z_i - z_j|$. Note that this inequality has application to the problem of section 5; it shows in a qualitative way that it is better for the spikes to be well-separated if they are to generate an appreciable amount of energy in the passband B.

c. Our idea for proving Theorem 2 arose from immersion in papers by Fuchs, Slepian, Landau, Pollak, and Widom. Fuchs (1954) was apparently the first to consider the norm $||P_W P_T||$ where W and T are arbitrary sets of finite measure. Define

$$\lambda_0(W,T) \equiv ||P_W P_T||^2 .$$

Fuchs indicated some uses for $\lambda_0^{1/2}$, including its relation to the uncertainty principle. However, Fuchs (1954) does not contain the key inequality

$$\lambda_0(W,T) \le |W|\,|T| \tag{8.2}$$

which we establish here via our Lemma 2; this, together with his Theorem 1, would have established our Theorem 2.

In much of the work of Slepian, Landau, and Pollak, the sets T and W are restricted to be intervals, so the uncertainty principle they consider is the classical one. They show that λ_0 is the largest eigenvalue of a certain integral operator and explicitly determine the corresponding eigenfunction, one of the prolate spheroidal wavefunctions. However, Landau and Widom (1980) consider the case where W and T are unions of disjoint intervals; they compute explicitly

$$\text{trace } P_T P_W P_T = |W|\,|T|. \tag{8.3}$$

If we use the identity

$$||P_W P_T||_{HS} = \text{trace } P_T P_W P_T$$

(since $P_W^2 = P_W$ and both P_T and P_W are self-adjoint), this gives our Lemma 2. However, Landau and Widom make no connection between (8.3) and the uncertainty principle. We believe Theorem 2 to be new.

d. One can generalize theorems 1 and 2 to other integral transforms. Yitzhak Katznelson and Persi Diaconis, after reading an earlier draft, suggested that an uncertainty principle like Theorem 2 ought to hold for arbitrary transformations $f \to \hat{f}$ satisfying merely

(A) $\| f \|_2 = \| \hat{f} \|_2$ (a Parseval-type identity)

and

(B) $\| \hat{f} \|_\infty \le \| f \|_1$.

We have been able to give such a result; it involves the use of L_1-concentration in one domain and L_2-concentration in the other.

Theorem 12. *Suppose $f \to \hat{f}$ is a transformation of $L_1 \cap L_2$ into $L_2 \cap L_\infty$ with properties* (**A**) and (**B**). Suppose there is a transform pair $(f, \hat{f})$ *of unit L_2-norm with f ε_T-concentrated to T* **in L_1-norm** *and $\hat{f}$ ε_W-concentrated to W* **in L_2-norm**. *Then*

$$|W| |T| \ge (1 - \varepsilon_W)^2 (1 - \varepsilon_T)^2 .$$

The proof takes only a few lines:

$$\| f \|_2^2 = \| \hat{f} \|_2^2 \le (1 - \varepsilon_W)^{-2} \int_W | \hat{f} |^2$$

$$\le (1 - \varepsilon_W)^{-2} |W| (\| \hat{f} \|_\infty)^2 .$$

The first inequality follows from ε_W-concentration of $\hat{f}$. Now,

$$\| \hat{f} \|_\infty \le \| f \|_1 \le (1 - \varepsilon_T)^{-1} \int_T | f | ,$$

where the last inequality follows from ε_T-concentration of f. By the Cauchy-Schwarz inequality,

$$\int_T | f | \le \sqrt{|T|} \, \| f \|_2 .$$

Combining these inequalities,

$$\| f \|_2^2 \le (1 - \varepsilon_W)^{-2} (1 - \varepsilon_T)^{-2} |W| |T| \; \| f \|_2^2 ,$$

from which the theorem follows.

It seems interesting that the result is "mixed", involving both L_2 and L_1 norms. We wonder if a result using L_2-measures for both frequency and time concentration can be constructed this easily.

The proof is quite general and may be used in other situations. For example, the same reasoning applies if the index set in the transform domain is discrete; thus the proof says something about orthogonal series. Let $\{\phi_k\}$ be an orthonormal basis for $L_2[0, 1]$, let $\hat{f}_k = \int f(t)\overline{\phi_k(t)}\,dt$ be the k-th Fourier-Bessel coefficient of f, and let $(\hat{f})$ denote the sequence of Fourier-Bessel coefficients. Let now $\|\hat{f}\|_2$, $\|\hat{f}\|_1$, and $\|\hat{f}\|_\infty$ denote the l_2, l_1, and l_∞ norms of the sequence $(\hat{f}_k)$, let W denote a subset of $\{1, 2, \cdots\}$, and let $|W|$ denote counting measure ($\#W$). Because the ϕ_k are orthonormal, condition (A) holds for f in the L_2-span of $\{\phi_k\}$; if $\|\phi_k\|_\infty \le 1$ the condition (B) holds as well. With this change in symbolism the same proof establishes

Corollary 4. *Let $\{\phi_k\}$ be an orthonormal set with $\|\phi_k\|_\infty \le 1$ for all k. Let f be of unit L_2-norm and belong to the L_2-span of $\{\phi_k\}$. If f is ε_T-concentrated to $T \subset [0, 1]$ (in L_1-norm) and if the sequence $(\hat{f}_k)$ is ε_W-concentrated to $W \subset \{1, 2, \cdots\}$ (in l_2-norm) then*

$$|T|\,|W| \ge (1-\varepsilon_T)^2(1-\varepsilon_W)^2 . \tag{8.4}$$

This result is now doubly "mixed" in that one domain has a continuous index set, the other is discrete, whereas the measure of concentration in one domain is an L_1-norm while that in the other domain is an l_2-norm.

The special case $\varepsilon_T = \varepsilon_W = 0$ of this result is worth mentioning. Let *support* $f \equiv \{t : f(t) \ne 0\}$, and let *support* $\hat{f} \equiv \{k : \hat{f}_k \ne 0\}$. Then (8.4) implies

$$|\,support\, f\,| \cdot |\,support\, \hat{f}\,| \ge 1 . \tag{8.5}$$

($|\,support\,\hat{f}\,| = \#\{k : \hat{f}_k \ne 0\}$). This can also be proved directly as follows.

$$\|f\|_2^2 \le (\|f\|_\infty)^2\, |\,support\, f\,|$$

$$\begin{aligned}\|f\|_\infty &= \sup_t |\textstyle\sum \hat{f}_k \phi_k(t)| \\ &\le \textstyle\sum |\hat{f}_k|\end{aligned} \tag{8.6}$$

$$\leq \sqrt{\sum |\hat{f}_k|^2} \cdot \sqrt{\# \{ k : \hat{f}_k \neq 0 \}} \tag{8.7}$$

where (8.6) follows from $||\phi_k||_\infty \leq 1$, and (8.7) from Cauchy-Schwarz. Because ϕ_k are orthonormal, Bessel's inequality $\sum |\hat{f}_k|^2 \leq \int |f|^2$ gives

$$||f||_2^2 \leq |\, support\, f \,|\, |\, support\, \hat{f} \,|\, ||f||_2^2$$

which establishes (8.5).

Let ϕ_k be the k-th Rademacher function ($\phi_k(t) = 1$ if the k-th bit of t in binary representation equals 1; $\phi_k(t) = -1$ otherwise); the $\{ \phi_k \}$ are then bounded in absolute value by 1 and are orthonormal on the interval [0, 1]. The functions

$$f_1 = -\phi_0 \quad (\equiv 1)$$

and

$$f_2 = \phi_0 + \phi_1 \quad (\equiv 2 \text{ on } (½, 1);\ \ 0, \text{otherwise})$$

supply two examples where equality obtains in (8.5).

As it turns out, a more useful result would assume that $||\phi_k||_\infty \leq M$ for all k. The last proof of (8.5) then shows immediately that the correct result in this case is

$$|\, support\, f \,|\, |\, support\, \hat{f} \,| \geq \frac{1}{M^2} .$$

We also remark that an argument similar to the proof of (8.5) provides an alternative proof of Theorem 1 ($N_t N_w \geq N$). However, such a proof does not seem to provide direct insight into the nature of the extremal functions.

There are other directions of generalization as well. Persi Diaconis and Mehrdad Shashahani have pointed out to us that an uncertainty principle holds in noncommutative harmonic analysis. Let G be a compact group. Let f be a function on G, and let $\hat{f}_\rho$ be the (matrix-valued) coefficient of f with respect to the unitary representation ρ, via

$$\hat{f}_\rho \equiv \frac{1}{|G|} \int f(g) \overline{\rho}(g)\, dg .$$

Here $|G|$ is the measure of G (*i.e.* 1 for a continuous group or the cardinality of G for a discrete group). Then Diaconis and Shashahani prove that

$$|\, support\, f \,|\, (\textstyle\sum^* dim^2\rho) \geq |G| \tag{8.8}$$

where the sum ranges over irreducible representations of G with $\hat{f}_\rho \neq 0$. When G is continuous, so that $|G| = 1$, this inequality is similar to (8.5), in that $\sum^* dim^2\rho$ is counting the number of nonzero coefficients in the expansion of f (recall that $\rho(g)$ is a $dim\ \rho$ by $dim\ \rho$ matrix). When G is the group of integers modulo N, this inequality, sensibly interpreted, implies $N_t N_w \geq N$. The proof of (8.8), which uses a number of facts about Haar measure and irreducible representations, is not much longer than our proof of (8.5) and (except for terminology) has a flavor similar to the proof of (8.5).

9. Conclusion

We have proven an uncertainty principle in which the sets of concentration need not be intervals. This general principle is easy to prove and has applications in signal recovery. The applications include: analysis of linear recovery problems (section 4) and nonlinear ones (section 6.2); Establishing uniqueness of recovery when no noise is present (section 5) and stability when noise is present (section 4); Establishing that a computationally effective approach to a recovery problem is available (section 6.3). In all these applications, the basic uncertainty principles ($N_t N_w \geq N$; $|W|\,|T| \geq 1 - \delta$) establish that something is possible, but generally much more is possible than these simple inequalities indicate. Better practical results will require seeing how operator norms such as $\|P_T P_W\|$ depend in detail on the sets T and W.

The basic principles also have generalizations to orthogonal series and to harmonic analysis on groups. Perhaps interesting applications of these principles will also be found.

Acknowledgements: This work was supported by National Science Foundation grant NSF DMS-84-51753 and by grants from Schlumberger Computer Aided Systems, Inc., and Sun Microsystems, Inc. The authors would like to thank Maryse Loranger for bibliographic work on the uncertainty principle, and Adele Cutler and Cha Yong Koo for computational assistance. B.F. Logan very kindly gave a bound copy of his doctoral thesis. Persi Diaconis, Yitzhak Katznelson and Henry Landau provided helpful comments on an earlier draft.

References

H. Dym and H.P. McKean (1972). *Fourier Series and Integrals*, Academic Press, N.Y.

C.L. Fefferman (1983). The uncertainty principle, *Bull. Am. Math. Soc., 9,* 129-206.

W.H.J. Fuchs (1954). On the magnitude of Fourier transforms, *Proc. Int. Math. Cong., Amsterdam,* September 1954, 106-107.

D. Gabor (1946). Theory of communication, *J. Inst. Elec. Eng., 93,* 429-457.

R.W. Gerchberg (1974). Super-resolution through error energy reduction, *Opt. Acta, 21,* 709-720.

E.N. Gilbert and D. Slepian (1977). Doubly orthogonal concentrated polynomials, *SIAM J. Appl. Math., 8,* 290-319.

G.H. Golub and C.F. Van Loan (1983). *Matrix Computations*, The Johns Hopkins University Press, MD.

P.R. Halmos and V.C. Sunder (1978). *Bounded Integral Operators on L^2 Spaces*, Springer-Verlag, N.Y.

W. Heisenberg (1930). *The Physical Principles of the Quantum Theory*, C. Eckart and F.C. Hoyt, trans., University of Chicago Press, Chicago.

K. Hoffman and R. Kunze (1971). *Linear Algebra*, Prentice-Hall, Inc., N.J.

Y. Katznelson (1976). *An Introduction to Harmonic Analysis*, Dover Publications, Inc., N.Y.

H.J. Landau (1986). Extrapolating a band-limited function from its samples taken in a finite interval, *IEEE Trans. Info. Th., IT-32,* 464-470.

H.J. Landau and W.L. Miranker (1961). The recovery of distorted band-limited signals, *J. Math. Anal. App., 2,* 97-104.

H.J. Landau and H.O. Pollak (1961). Prolate spheroidal wave functions, Fourier analysis, and uncertainty–II, *B.S.T.J., 40,* 64-84.

H.J. Landau and H. Widom (1980). Eigenvalue distribution of time and frequency limiting, *J. Math. Anal. App., 77,* 469-481.

S. Levy and P.K. Fullagar (1981). Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution, *Geophysics, 46,* 1235-1243.

B.F. Logan (1965). *Properties of High-Pass Signals*, Ph.D. Thesis, Department of Electrical Engineering, Columbia University.

D.W. Oldenburg, T. Scheuer and S. Levy (1983). Recovery of the acoustic impedance from reflection seismograms, *Geophysics, 48,* 1318-1337.

A. Papoulis (1975). A new algorithm in spectral analysis and band-limited extrapolation, *IEEE Trans. Circ. Sys., 22,* 735-742.

A. Papoulis and C. Chamzas (1979). Improvement of range resolution by spectral extrapolation, *Ultrasonic Imaging, 1,* 121-135.

F. Santosa and W.W. Symes (1986). Linear inversion of band-limited reflection seismograms, *SIAM J. Sci. Stat. Comput., 7,* 1307-1330.

R.W. Schafer, R.M. Mersereau and M.A. Richards (1981). Constrained iterative restoration algorithms, *Proc. IEEE, 69,* 432-450.

D. Slepian and H.O. Pollak (1961). Prolate spheroidal wave functions, Fourier analysis and uncertainty–I, *B.S.T.J., 40,* 43-63.

D. Slepian and E. Sonnenblick (1965). Eigenvalues associated with prolate spheroidal wave functions of zero order, *B.S.T.J., 44,* 1745-1759.

P. Turán (1984). *On a New Method of Analysis and its Applications*, John Wiley & Sons, Inc., N.Y.

C. Walker and T.J. Ulrych (1983). Autoregressive recovery of the acoustic impedance, *Geophysics, 48,* 1338-1350.

D.C. Youla (1978). Generalized image restoration by the method of alternating orthogonal projections, *IEEE Trans. Circuits Systems, 25*, 694-702.

D.C. Youla and H. Webb (1982). Image restoration by the method of convex projections. I–theory, *IEEE Trans. Med. Imag., 1*, 81-94.

H. Weyl (1928). *Gruppentheorie und Quantenmechanik*, S. Hirzel, Leipzig.

Appendix A: Extremal Functions of the Discrete-Time Principle.

Theorem 13. *Equality $N_t N_w = N$ is only attained by $Ш^{N_t}$ and sequences (x_t) reducible to it by*

(a) scalar multiplication

(b) cyclic permutation in the time domain

(c) cyclic permutation in the frequency domain, so that

$$\hat{x}_{(t-\tau) \bmod N} = \alpha Ш^{N_w}_{(w-\omega) \bmod N} \tag{A.1}$$

for some $\alpha \neq 0$, and some integers τ and ω. Equivalently,

$$x_{(t-\tau) \bmod N} = \alpha e^{\frac{2\pi i \omega t}{N}} \cdot Ш^{N_t}.$$

Proof. We know that $Ш^{N_t}$ satisfies the equality; by inspection sequences (x_t) that can be written as in (A.1) do also. In the proof of Theorem 1 we showed that equality is only possible if

(1) N is composite with the factorization $N = N_t N_w$ (obviously), and

(2) the N_w nonzero elements of $\hat{x}_w$ are equally spaced.

To these we may add

(3) the N_t nonzero elements of x_t are equally spaced.

The argument for (3) is similar to the argument given for (2). In Lemma 5 (below) we give a result reciprocal to Lemma 1, showing that no N_w consecutive entries of x_t can all vanish. But since x_t has only $N_t = \frac{N}{N_w}$ nonzero elements, they must be equally spaced to avoid a gap more than N_w long.

Let us now see how (1)–(3) imply (A.1). Let (y_t) be a cyclic permutation of (x_t), *i.e.* $y_t = x_{(t-\tau) \bmod N}$, with $y_0 \neq 0$. Henceforth let $k = N_t$, $l = N_w$. By (3), (y_t) has the same support as $Ш^k$. It can therefore be written as the pointwise product

$$y_t = Ш_t^k \cdot e_t$$

where (e_t) is an 'envelope' sequence. The transform $(\hat{y}_w)$ is the circular convolution of the transforms $\hat{Ш}^k = kN^{-1/2} Ш^l$ and $\hat{e}$:

$$\hat{y}_w = \sum_{j=0}^{N-1} \frac{k}{\sqrt{N}} Ш_j^l \cdot \hat{e}_{(w-j) \bmod N} .$$

Now the convolution of the periodic sequence $Ш^l$ with any other sequence yields a periodic sequence with the same period. Thus $(\hat{y}_w)$ has $N_w = \frac{N}{N_t}$ periods of length N_t; to attain equality $N_t N_w = N$ it must have only one nonzero value in each period. By periodicity all the N_t nonzero entries in $\hat{y}_w$ are identical and equally spaced. Thus, for appropriate α, ω,

$$\hat{y}_w = \alpha\, Ш^l_{(w-\omega) \bmod N} .$$

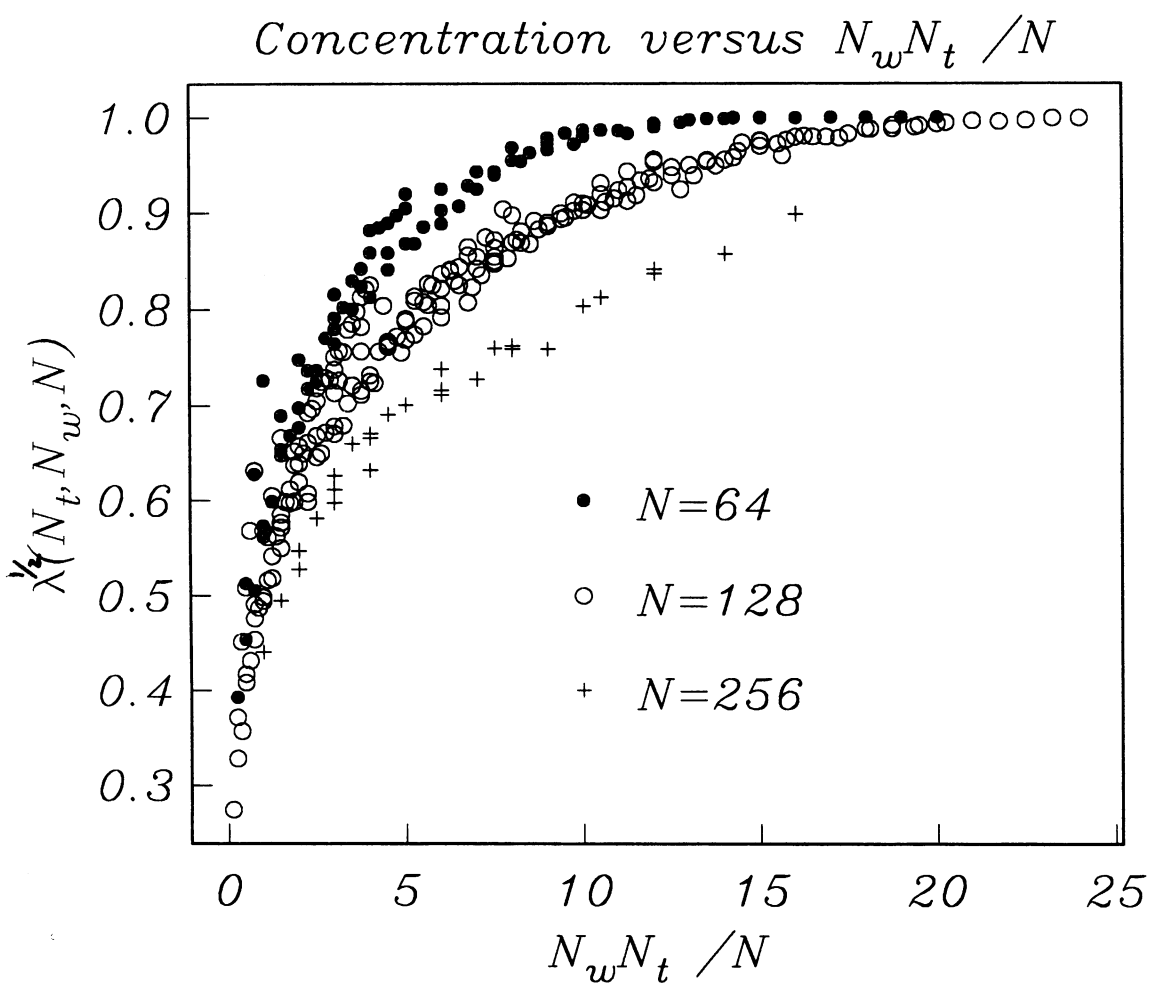
In terms of the original sequence x_t,

$$\widehat{x_{(t-\tau) \bmod N}} = \alpha Ш^l_{(w-\omega) \bmod N} . \square$$

To show that the extremal functions in both the time and frequency domains are equally-spaced spike trains, we used the following lemma in addition to Lemma 1:

Lemma 5. *If* $(\hat{x}_w)$ *has* N_w *nonzero elements, then* x_t *cannot vanish on any interval of* N_w *consecutive times* t.

The proof of Lemma 5 is identical to that of Lemma 1 after interchanging the roles of x and $\hat{x}$, replacing N_t by N_w, and replacing z by z^{-1}.


Concentration versus $N_w N_t / N$
$\chi^{1/2}(N_t, N_w, N)$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
5
10
15
20
25
$N_w N_t / N$
N=64
N=128
N=256