Signal Recovery and the Large Sieve

By

D.L. Donoho
University of California, Berkeley

B.F. Logan
A.T.T. Bell Laboratories, Murray Hill, N.J.

Technical Report No. 271
September 1990

Department of Statistics
University of California
Berkeley, California 94720

Signal Recovery and the Large Sieve

D. L. Donoho
University of California, Berkeley

B.F. Logan
A.T.T. Bell Laboratories, Murray Hill, N.J.

March, 1990
Revised December, 1990

Abstract

We develop inequalities for the fraction of a bandlimited function's L_p norm which can be concentrated on any set of small 'Nyquist density'. We mention two applications. First, that a bandlimited function corrupted by impulsive noise can be reconstructed perfectly, provided the noise is concentrated on a set of Nyquist density $< 1/\pi$. Second, that a wideband signal supported on a set of Nyquist density $< 1/\pi$ can be reconstructed stably from noisy data, even when the low frequency information is completely missing.

Key Words and Phrases. Entire Functions of Exponential Type. Trigonometric Polynomials. Signal Recovery. L_1 recovery methods. Logan's Phenomenon. Nyquist Density.

Acknowledgements. This work was supported in part by National Science Foundation grant DMS 84-51753, and by grants from Schlumberger-Doll Research and from Western Geophysical. The first-named author would like to thank Ytzhak Katznelson, H.L. Montgomery, and Philip Stark for helpful discussions.

1 Introduction

The 'Analytic Principle of the Large Sieve', as Montgomery [13] calls it, is a family of inequalities for trigonometric polynomials which has found a variety of applications in analytic number theory. From the point of view of the present paper, such inequalities are interesting because they control the size of trigonometric polynomials on "sparse" sets. For example, Montgomery [13] page 562 gives the following inequality, whose form he attributes to unpublished work by E. Bombieri. Let μ be a positive measure of period 1, and let $S(\alpha) = \sum_{k=m+1}^{m+n} a_k e^{2\pi i k\alpha}$ be a trigonometric polynomial of degree n and period 1. Then

$$\int_0^1 |S(\alpha)|^2 d\mu \leq (n + 2\delta^{-1}) \left(\sup_{\alpha} \mu[\alpha, \alpha + \delta] \right) \sum |a_k|^2 \tag{1}$$

As $\int_0^1 |S(\alpha)|^2 d\alpha = \sum_k |a_k|^2$, this inequality has the following interpretation. If T is a periodic set of period 1, and obeys the sparsity condition

$$n \sup_{\alpha} |T \cap [\alpha, \alpha + \frac{1}{n}]| \leq \epsilon \tag{2}$$

with ϵ a small positive number, then, by taking $\mu(d\alpha) = 1_T(\alpha)d\alpha$ in (1) we have

$$\int_{T\cap[0,1]} |S(\alpha)|^2 d\alpha / \int_{[0,1]} |S(\alpha)|^2 d\alpha \leq 3\epsilon. \tag{3}$$

In short, only a small fraction of a trigonometric polynomial's "energy" is localized to the set T. Thus if T is "sparse" in precisely the sense that (2) holds, then S cannot be concentrated to T. We will show below that, by adapting an argument of Selberg, the constant in Bombieri's inequality (1) can be replaced by $(n - 1 + \delta^{-1})$, and so the right side of (3) can be improved.

Independently of developments in analytic number theory, inequalities analogous to (1) have been established for entire functions of exponential type, which of course are close cousins of trigonometric polynomials. Let $B_p(\Omega)$ denote the class of entire functions of exponential type Ω which are in L_p on the real axis. Such functions have Fourier transforms $\hat{f}(\omega)$ which vanish for $|\omega| > \Omega$, and hence are generally called *Bandlimited*; compare [9, 11] and references there. Boas [1] showed that if μ is a positive, sigma-finite measure, placing mass less than ϵ in every interval of length 1, then

$$\int |f|^p d\mu \leq C(p, \Omega)\, \epsilon \int_{-\infty}^{\infty} |f|^p dt \tag{4}$$

for every $f \in B_p(\Omega)$, $\Omega < \pi$. Again, if T is a subset of the real line satisfying the sparsity condition

$$\sup_t |T \cap [t, t+1]| \leq \epsilon \tag{5}$$

this implies that

$$\int_T |f|^p dt / \int |f|^p dt \leq C(p, \Omega)\epsilon. \tag{6}$$

In words, only a small fraction of a bandlimited function's L_p-norm can be concentrated on sets which are sparse in the sense that (5) holds. We show below that an adaptation of Selberg's argument for the large sieve gives a substantial improvement on the coefficient C in Boas' inequality in case $p = 2$; and that a different but related argument gives a substantial improvement in the case $p = 1$.

Our interest in such inequalities, and in getting good constants for them, comes from signal recovery problems, which we describe in section 2 below. Briefly, the existence of constants for which such inequalities hold implies the existence of certain surprising phenomena in signal recovery. Most importantly for us, the better the constants in such inequalities, the broader the range of cases in which those phenomena are known to occur.

In section 2, we sketch the signal recovery motivation of our study; in section 3 we discuss L_2 inequalities related to (4); in section 4, L_1 inequalities; in section 5, the limits of our technique; finally in section 6, we return to inequalities of type (1) and applications to discrete-time signal recovery.

2 Sparsity and Signal Recovery

We are interested in three specific phenomena, which are discussed at length in [5], where further references are given. Here we permit ourselves only a brief discussion.

2.1 Perfect Recovery of a Bandlimited signal

Logan [9] discovered an interesting phenomenon. Suppose we measure a noisy version r of a bandlimited signal $b \in B_1(\Omega)$:

$$r = b + n.$$

Here n is the noise, about which we know only that it has finite L_1 norm, and that its support, although unknown, must be a sparse set. Think of highly impulsive noise.

Logan [9] proposed estimating b by the minimum L_1 projection:

$$\beta_1(r) = \arg \min_{\tilde{b} \in B_1(\Omega)} ||r - \tilde{b}||_1. \tag{7}$$

Here $|| \cdot ||_1$ denotes the L_1 norm; and $\beta_1(r)$ in general depends nonlinearly on r. Surprisingly, under certain conditions $\beta_1(r) = b$ *perfectly*, whatever be the size $||n||_1$ of the noise.

Let $W = \Omega/\pi$, so that W^{-1} is the usual Nyquist interval for entire functions of type Ω. Given a set $T \subset \mathbf{R}$, let P_T be the operator that restricts support to T, so that $(P_T f)(t) = f(t)1_T(t)$. Define the operator norm

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

Logan [9] showed that if $T \equiv \mathit{support}(n)$ and $||n||_1 < \infty$, the condition $\mu_0(T, W) < 1/2$ ensures that $\beta_1(r) = b$ exactly, even though n may be of arbitrarily high energy and of arbitrary form (subject to the constraint on its support). He proved that $\mu_0(T, W) \leq W|T|$, which shows that if the set T is sparse in the sense of small total measure this perfect recovery phenomenon occurs. He asked the question for what other sets T the inequality $\mu_0(T, W) < 1/2$ would hold. See [5] for further discussion, and explanation of the connection of $\mu_0(T, W) \leq W|T|$ with the uncertainty principle.

In section 4 below we give Theorem 7, which may be used to considerably extend the range of cases where Logan's phenomenon is known to occur. Define the *Maximum Nyquist Density* by

$$\rho(T, W) \equiv W \sup_t \left| T \cap [t, t + \frac{1}{W}] \right|.$$

Combining inequality (21) of Theorem 7 below with the notation of this section gives

$$\begin{aligned} \mu_0(T, W) &= \sup_{f \in B_1(\Omega)} \frac{\int_T |f| dt}{\int |f| dt} \\ &\leq \frac{\Omega/2}{\sin(\Omega/(2W))} \sup_t |T \cap [t, t + 1/W]| \\ &= \frac{\pi}{2} \rho(T, W). \end{aligned}$$

Since $\rho(T,W) \leq W|T|$, and since one can easily construct examples where $\rho(T,W)$ is arbitrarily small yet $W|T| = \infty$, this inequality can be a considerable improvement over $\mu_0(T,W) \leq W|T|$. It implies

Corollary 1 *Suppose that $r = b + n$, $b \in B_1(\Omega)$, that $||n||_1 < \infty$, and that $supp(n) = T$ has Nyquist density $\rho(T,W) < \frac{1}{\pi}$. Then $\mu_0(T,W) < \frac{1}{2}$; the solution $\beta_1(r)$ of the minimum L_1-problem (7) is unique, and recovers the signal perfectly: $\beta_1(r) = b$.*

In other words, the noise can be supported on a set of infinite measure, yet if the support occupies a fraction less than $1/\pi$ of each Nyquist interval, the original bandlimited signal will be recovered perfectly by the L_1 technique.

Earlier work on inequalities for entire functions, such as Nikol'skiĭ's [14] could be adapted to show that the density threshold for Logan's phenomenon is at least as large as $1/(2+2\pi)$; this is considerably weaker than the result here. Improving the constant in the inequality (21) would raise the known threshold for this phenomenon from $1/\pi$ to something larger. On the other hand, the condition $\rho(T,W) \leq 1/2$ is easily seen to be not sufficient to ensure perfect recovery, so there is limited room for further improvement in this direction.

For later use, it is convenient to have the following stability result.

Lemma 2 *Let $r = b + n$ with $b \in B_1(\Omega)$ and suppose that for some set T satisfying $\mu_0(T,W) < 1/2$ we have $||n - P_T n||_1 \leq \epsilon$. Then for any solution $\beta_1(r)$ of the minimum L_1-problem (7),*

$$||\beta_1(r) - b||_1 \leq 2(1 - 2\mu_0(T,W))^{-1}\epsilon$$

Thus, if the noise is *almost* concentrated to a set of Nyquist density $\rho < 1/\pi$, the L_1 method *almost* recovers b.

Although the argument for this Lemma is a simple extension of the argument for uniqueness, it does not seem to have been recorded in either [9] or [5]. We prove it in the appendix, section 7.

2.2 Recovery of a Sparse Signal

In exploration seismology, there arises the problem of recovering a wideband signal from noisy observations when it is essentially impossible to obtain reliable low frequency information [5, 16, 8, 19].

This situation may be modelled as follows. Suppose that we are interested in recovering a signal s, when our measurements are

$$r = s + z + b,$$

where $z(t)$ is noise, known to be small in L_1-norm, and $b(t)$ is an unknown bandlimited function $b \in B_1(\Omega)$, of arbitrary but finite L_1 norm. Because the nuisance function b is unknown, and may be much larger in norm than s, there is no reliable information in r about the low frequency behavior of s.

The problem as stated is obviously ill-posed. Luckily, with an additional piece of prior information - *sparsity* of s - stable recovery is possible. Consider the following reconstruction rule (essentially the continuous time version of one discussed by Santosa and Symes, [16]). Let $\sigma_1(r)$ be the minimal L_1 norm reconstruction:

$$\sigma_1(r) = \arg\min ||\tilde{s}||_1 : r = \tilde{s} + b, b \in B_1(\Omega);$$

in the case of more than one minimizer, select any minimizer.

Corollary 3 *Suppose that in the above model $||z||_1 \leq \eta_1$ and that s is approximately concentrated on a set of low Nyquist density: i.e.*

$$||s - P_T s||_1 \leq \eta_2$$

for some T satisfying

$$\rho(T, W) < 1/\pi.$$

Then any solution $\sigma_1(r)$ as above satisfies

$$||s - \sigma_1(r)||_1 \leq 2(1 - \pi\rho(T, W))^{-1}(\eta_1 + \eta_2) + \eta_1.$$

In short, approximate sparsity of s allows approximate recovery from noisy data even though the low frequencies are missing.

This result follows from the stability result, Lemma 2, and a few simple observations. First, we have the identity $\sigma_1(r) = r - \beta_1(r)$. Second, under the given conditions, putting $n \equiv s + z$, we have $r = b + n$, and $||n - P_T n||_1 \leq \eta_1 + \eta_2$. Hence the stability Lemma applies and $||\beta_1(r) - b||_1 \leq 2(1 - 2\mu_0(T, W))^{-1}(\eta_1 + \eta_2)$. The result then follows from the triangle inequality.

2.3 Recovery of Missing Data

We now describe an application of L_2-type inequalities. Suppose we have noisy observations on a bandlimited signal $b \in B_2(\Omega)$. Unfortunately, there is a set T of time indices where observations are missing, so we only observe

$$r(t) = b(t) + n(t), \quad t \notin T;$$

where $||n||_2 \leq \epsilon$, say. The problem is to recover b for all $t \in \mathbf{R}$ from these partial observations, and to do so stably, i.e. with error at worst proportional to ϵ. See [5], section 4.

Define the ideal low-pass operator

$$(P_W f)(t) = \frac{1}{2\pi} \int_{-\pi W}^{\pi W} \hat{f}(\omega) e^{i\omega t} d\omega$$

with $\hat{f}(\omega) = \int f(t) e^{-i\omega t} dt$ the Fourier transform. Define the operator norm

$$\lambda_0(T, W) \equiv ||P_W P_T||_2,$$

where P_T is the support restriction operator introduced earlier. [5], section 4, describes how a convenient iterative technique (the method of alternating projections), may be used to recover b for all t, stably, provided $\lambda_0(T, W) < 1$.

The uncertainty principle is developed in [5], to get the inequality

$$\lambda_0(T, W) \leq \sqrt{W|T|}. \tag{8}$$

Theorem 4 below affords a considerable improvement on (8). Let us see how. First, if $f = P_W P_T s$ where $s \in L_2$, then $f \in B_2(\Omega)$; and as P_W and P_T are both projections, $||f||_2 \leq ||s||_2$. Second, as P_W and P_T are self-adjoint operators on L_2,

$$\begin{aligned} ||P_W P_T||_2^2 = ||P_T P_W||_2^2 &= \sup_{s \in L_2} \frac{||P_T P_W s||_2^2}{||s||_2^2} \\ &\leq \sup_{f \in B_2(\Omega)} \frac{\int_T |f|^2}{\int |f|^2}. \end{aligned}$$

The inequality (11) of Theorem 4 below, taking $d\mu(t) = 1_T(t) dt$, gives that for every f in $B_2(\Omega)$ we have

$$\int_T |f|^2 \leq (W + \delta^{-1}) \sup_t |T \cap [t, t+\delta]| \int |f|^2. \tag{9}$$

Substituting in the definition of Nyquist density, we get

$$\lambda_0^2(T, W) \leq 2\,\rho(T, W). \tag{10}$$

The improvement from inequality (8), which uses the total measure of T, to (10), which uses density, can be significant. In the missing data problem, (10) allows data to be missing on a set of infinite total measure, yet still guarantees that the alternating method will recover the bandlimited signal stably, provided the Nyquist density of the missing observations is less than 1/2 .

It may be instructive to compare this result with what we could conclude by applying earlier inequalities such as (4). The coefficient Boas gives is $C(p, \Omega)^{1/p} = 7 + \frac{2}{\pi(\pi-\Omega)}$ for $\Omega < \pi$. From this we could conclude only that a density at the half-Nyquist smaller than .02 is sufficient for stable recovery. One may adapt results of Plancherel and Pólya [15] to get an even weaker conclusion. Results of Duffin and Schaeffer [6] and of Nikolskii [14] can be adapted to get the improved constant $C(2, \pi) = (1 + \pi)$, but even this gives only the conclusion that Nyquist density smaller than about 1/4 is sufficient for recovery.

We now turn away from the signal recovery setting, and focus on the inequalities for bandlimited functions which drive these results.

3 Concentration in L_2 Norm

Theorem 4 *Let μ be a a positive sigma-finite measure, and $f \in B_2(\Omega)$.*

$$\int |f|^2 d\mu \leq (\Omega/\pi + \delta^{-1}) \left(\sup_t \mu[t, t+\delta]\right) \int |f|^2. \tag{11}$$

The theorem follows immediately from the next two lemmas.

Lemma 5 *Suppose that g is supported on $[-\delta/2, \delta/2]$, and that convolution with g is a continuous, invertible linear operator on $B_2(\Omega)$. Then*

$$\int |f|^2 d\mu \leq C_2(g, \Omega)(\sup_t \mu[t, t+\delta]) \int |f|^2 \tag{12}$$

where

$$C_2(g, \Omega) = ||g||_2^2 \left(\sup\{\frac{||\varphi||_2}{||g \star \varphi||_2} : \varphi \in B_2(\Omega)\}\right)^2 .$$

Proof. Under the hypotheses, given $f \in B_2(\Omega)$, there exists a unique $f^* \in B_2(\Omega)$ with $g \star f^* = f$. By Cauchy-Schwartz

$$\begin{aligned}
\int |f|^2 d\mu &= \int |g \star f^*|^2 d\mu = \int |\int_{-\delta/2}^{\delta/2} g(h) f^*(t-h) dh|^2 d\mu(t) \\
&\leq \int \int_{-\delta/2}^{\delta/2} |g(h)|^2 dh \int_{t-\delta/2}^{t+\delta/2} |f^*(u)|^2 du d\mu(t) \\
&= ||g||_2^2 \int_{-\infty}^{\infty} |f^*|^2(u) \int_{u-\delta/2}^{u+\delta/2} d\mu(t)\, du \\
&\leq ||g||_2^2 ||f^*||_2^2 \left(\sup_t \mu[t, t+\delta] \right)
\end{aligned}$$

By definition $||g||_2^2 ||f^*||_2^2 / ||f||_2^2 \leq C_2(g, \Omega)$ and the lemma follows.

We remark that the constant C_2 has the following alternate definition:

$$C_2(g, \Omega) = ||g||_2^2 \sup_{[-\Omega,\Omega]} 1/|\hat{g}(\omega)|^2. \tag{13}$$

Lemma 6 *There exists g of support $[-\delta/2, \delta/2]$ satisfying the hypotheses of Lemma 5 for which*

$$C_2(g, \Omega) = (\Omega/\pi + \delta^{-1}). \tag{14}$$

Proof. We use a function g constructed by Selberg in connection with the Large Sieve; once again, see the article by Montgomery [13]. Our account of the construction of g follows closely that of Vaaler ([18], pages 185-186). The construction uses Beurling's function [3]:

$$B(x) = \left(\frac{\sin \pi x}{\pi} \right)^2 \{ \sum_{n=0}^{\infty} (x-n)^{-2} - \sum_{m=-\infty}^{-1} (x-m)^{-2} + 2x^{-1} \}.$$

Among other properties, B is entire of type 2π and majorizes the signum function. Selberg uses this function as follows. Let $E = [\alpha, \beta]$ and put

$$C_E(x) = \frac{1}{2} \{ B(\beta - x) + B(x - \alpha) \}.$$

It turns out that

$$C_E(x) \geq 1_E(x) \tag{15}$$

$$\int_{-\infty}^{\infty} C_E(x) - 1_E(x) dx = 1 \tag{16}$$

and, for a certain function c_E supported on $[-2\pi, 2\pi]$

$$C_E(x) = \int_{-2\pi}^{2\pi} c_E(t)e^{-ixt}dt \tag{17}$$

(i.e. C_E is entire of type 2π.)

To apply this, fix δ and set $\lambda = \delta/2\pi$. Let $\Gamma(x) = C_E(\lambda x)$ where $E = [-\lambda\Omega, \lambda\Omega]$. Then, by (17)

$$\gamma(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Gamma(x)e^{itx}dx \tag{18}$$

is supported in $[-2\pi\lambda, 2\pi\lambda] = [-\delta, \delta]$. An explicit formula for $\gamma(t)$ in terms of elementary functions is given in Vaaler. We need only the value $\gamma(0)$,

$$\begin{aligned}\gamma(0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \Gamma(x)dx = \frac{1}{2\pi\lambda}\int_{-\infty}^{\infty} C_E(y)dy \\ &= \Omega/\pi + \delta^{-1},\end{aligned}$$

the last step using (16). Now Γ is nonnegative and entire of type δ. By a theorem of Fejer there exists a function G such that $\Gamma = |G|^2$ and G is entire of type $\delta/2$. Put $g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(x)e^{itx}dx$; then g is supported in $[-\delta/2, \delta/2]$, and by Parseval and definition of G,

$$\int_{-\delta/2}^{\delta/2} |g|^2 = \gamma(0) = \Omega/\pi + \delta^{-1}. \tag{19}$$

Given $f \in B_2(\Omega)$, define

$$f^*(t) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} \hat{f}(\omega)/G(\omega)e^{it\omega}d\omega.$$

This defines a bounded linear operator on $B_2(\Omega)$. Indeed, by (15), $1/G(\omega) \in [0, 1]$ for $\omega \in [-\Omega, \Omega]$ and so, by Parseval

$$||f^*||^2 \leq ||f||^2. \tag{20}$$

Combining (20) and (19) gives the desired result (14).

4 Concentration in L_1 norm

Theorem 7 *Let μ be a positive, sigma-finite measure, and let $f \in B_1(\Omega)$. If $\delta\Omega < 2\pi$ then*

$$\int |f| d\mu \leq \frac{\Omega/2}{\sin(\Omega\delta/2)} \left(\sup_t \mu[t, t+\delta] \right) \int |f| \tag{21}$$

The proof is structurally analogous to that of Theorem 4; one simply combines Lemmas 8 and 9 below.

Lemma 8 *Suppose that g is supported on $[-\delta/2, \delta/2]$, and that convolution with g is a continuous, invertible linear operator on $B_1(\Omega)$. Then for $f \in B_1(\Omega)$*

$$\int |f| d\mu \leq C_1(g, \Omega)(\sup_t \mu[t, t+\delta]) \int |f| \tag{22}$$

where

$$C_1(g, \Omega) = ||g||_\infty \left(\sup\{ \frac{||\varphi||_1}{||g \star \varphi||_1} : \varphi \in B_1(\Omega)\} \right).$$

Proof. By hypothesis, given $f \in B_1(\Omega)$, there exists a unique element f^* of $B_1(\Omega)$ satisfying $f = g \star f^*$. We have

$$\begin{aligned} \int |f| d\mu &= \int | \int_{-\delta/2}^{\delta/2} g(h) f^*(t-h) dh | d\mu(t) \\ &\leq ||g||_\infty \int |f^*(u)| \int_{u-\delta/2}^{u+\delta/2} d\mu(t)\, du \\ &\leq ||g||_\infty ||f^*||_1 \left(\sup_t \mu[t, t+\delta] \right). \end{aligned}$$

As $||g||_\infty ||f^*||_1 / ||f||_1 \leq C_1(g, \Omega)$, the lemma is established.

There is an equivalent definition of the constant $C_1(g, \Omega)$ in terms of Beurling's theory of *minimal extrapolation* [2, 4]. We recall some definitions. Let $h(\omega)$ be a continuous function given for $\omega \in [-\Omega, \Omega]$. Let $M(h, \Omega)$ denote the collection of all finite signed measures on the line with $\hat{\nu}(\omega) = h(\omega)$ for $\omega \in [-\Omega, \Omega]$. The transforms of these measures all agree with h on $[-\Omega, \Omega]$ and so are all extrapolations of h. Minimal extrapolation is defined as follows.

By definition, each measure in $M(h,\Omega)$ has a finite total variation $Var(\nu)$. Let T_Ω denote the smallest such variation:

$$T_\Omega(h) = \inf\{Var(\nu) : \nu \in M(h,\Omega)\},$$

with $T_\Omega(h) = \infty$ if $M(h,\Omega)$ is empty. Using these definitions, we may write

$$C_1(g,\Omega) = T_\Omega(1/\hat{g})||g||_\infty.$$

Up to normalization, $C_1(g,\Omega)$ is the T_Ω norm of the function $1/\hat{g}(\omega)$.

For general h, it is difficult or impossible to compute minimal extrapolations. In this sense, $C_1(g,\Omega)$ is much harder to work with than $C_2(g,\Omega)$. However, in certain special cases minimal extrapolations are known. Beurling, in the Mittag-Leffler lectures [4], gives some general ideas. Logan [11] gives several examples and specific computational tools. By combining some of the ideas presented there with a few new ones, it is possible to compute the minimal extrapolation of $1/\hat{g}$ in the case where g is a "boxcar".

Lemma 9 *Let $g(t) = \delta^{-1}1_{\{|t|\le\delta/2\}}$ with $\delta < 2$, and $1 \le p \le \infty$. There exists a signed measure ν representing the convolution inverse of g; i.e. if $f^* = f \star \nu (\equiv \int f(\cdot - u)d\nu(u))$ with $f \in B_p(\pi)$ then $g \star f^* = f$. Convolution with ν is a bounded linear operator of $B_p(\pi)$, with operator norm*

$$\sup\{\frac{||f \star \nu||_p}{||f||_p} : f \in B_p(\pi)\} = \frac{\pi\delta/2}{\sin(\pi\delta/2)}. \tag{23}$$

Proof. The Fourier transform of g is $\hat{g}(\omega) = \frac{\sin(\omega\delta/2)}{\omega\delta/2}$. Below we will construct a finite signed measure ν so that

$$\hat{g}(\omega)\hat{\nu}(\omega) = 1 \quad \omega \in [-\pi,\pi]. \tag{24}$$

As $\int |d\nu|(u)$ is finite, and

$$||f \star \nu||_p \le \int ||f(\cdot - u)||_p |d\nu|(u) = ||f||_p \int |d\nu|(u), \tag{25}$$

convolution with ν is a bounded linear operator on $B_p(\pi)$. By the convolution theorem and (24), ν is a convolution inverse to g on $B_p(\pi)$. The measure ν has the special property that

$$\max_{\omega\in[-\pi,\pi]} |\hat{\nu}(\omega)| = \int |d\nu|(u); \tag{26}$$

from this, (24), and the formula for $\hat{g}$ it follows that the operator norm of convolution with ν is not larger than $\frac{\pi\delta/2}{\sin(\pi\delta/2)}$. Theorem A of Logan [11] shows that equality holds in (23).

Construction of ν requires some preparation. Let $\tilde{g}(\omega)$ denote the 2π-periodic extension of $\hat{g}(\omega)$ away from the fundamental interval $[-\pi,\pi]$, and define $\hat{h}(\omega) = 1/\tilde{g}(\pi+\omega)$. Then, as $\delta < 2$, $\hat{h}(\omega)$ is defined on the whole real line, continuous, and 2π periodic. Now

$$\left[\log \frac{\lambda}{\sin\lambda}\right]'' = \frac{1}{\sin^2\lambda} - \frac{1}{\lambda^2} \geq 0, \quad \lambda \in (-\pi,\pi).$$

and so $\hat{h}(\omega)$ is convex on $(0,2\pi)$. Moreover, $\hat{h}$ is even, so by e.g. Theorem 25, page 25 of Hardy and Rogosinski [7], the Fourier coefficients of $\hat{h}$,

$$h_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{h}(\omega)e^{ik\omega}d\omega$$

are nonnegative, sum to $\hat{h}(0)$, and the inversion formula

$$\hat{h}(\omega) = \sum_{k=-\infty}^{\infty} h_k e^{-ik\omega}$$

holds, showing that $\hat{h}$ is indeed a Fourier transform.

Define now

$$\nu = \sum_{k=-\infty}^{\infty} (-1)^k h_k \delta_k$$

where δ_k is the Dirac mass at k. The measure ν is finite since $\int |d\nu| = \sum |h_k| = \sum h_k = \hat{h}(0)$. Note that

$$\begin{aligned} \hat{\nu}(\omega) &= \int e^{-i\omega u}d\nu(u) = \sum_{k=-\infty}^{\infty} (-1)^k h_k e^{-ik\omega} \\ &= \sum_{k=-\infty}^{\infty} h_k e^{-ik(\omega-\pi)} \\ &= \hat{h}(\omega-\pi). \end{aligned}$$

(24) follows, since by definition of $\hat{h}$ we get $\hat{\nu}(\omega) = 1/\tilde{g}(\omega)$. Also (26) follows, since

$$\hat{\nu}(\pi) = \hat{h}(0), \tag{27}$$

and, as we have seen, $\hat{h}(0) = \int |d\nu|$. The properties claimed for ν have now been verified, and the proof is complete.

Remark. The minimal extrapolation property of the measure ν constructed in this proof follows from (26).

5 Optimality

The particular kernels we have introduced above, via Lemmas 6 and 9, are in certain senses best possible. That is, for certain combinations of δ, Ω, they are best possible for use with Lemmas 5 and 8.

Let us first describe the optimality of the Selberg function. In our terminology, Selberg [17] showed that if $\delta\Omega$ is an integral multiple of π then

$$\inf\{C_2(g,\Omega) : supp(g) \subset [-\delta/2, \delta/2]\} = (\Omega/\pi + \delta^{-1});$$

compare especially [18]. For $\delta\Omega/\pi$ nonintegral, though, the Selberg function is suboptimal. The best kernel in such cases has been characterized by Logan (see the announcement [10]). If $\delta\Omega < \pi$ the analysis is particularly simple:

Lemma 10 *For* $\Omega\delta < \pi$

$$\inf\{C_2(g,\Omega) : supp(g) \subset [-\delta/2, \delta/2]\} = (\delta/2 + \frac{\sin(\Omega\delta)}{\Omega})^{-1}. \qquad (28)$$

and an optimal kernel is

$$k(t) = 1_{\{|t| \leq \delta/2\}} \cos(\Omega t) \qquad (29)$$

Proof. It is enough to restrict attention to kernels in L_1. For any L_1 kernel g,

$$C_2(g,\Omega) \geq ||g||^2/|\hat{g}(\Omega)|^2.$$

Define

$$I_2 = \sup\{|\hat{g}(\Omega)|^2 : supp(g) \subset [-\delta/2, \delta/2] \text{ and } ||g||^2 = 1\}.$$

Then for all g supported in $[-\delta/2, \delta/2]$

$$C_2(g,\Omega) \geq I_2^{-1}. \qquad (30)$$

Recalling that

$$\hat{g}(\Omega) = \int_{-\delta/2}^{\delta/2} g(t)\cos(\Omega t)dt,$$

it follows from Cauchy-Schwartz that the optimization problem defining I_2 has for its unique solution a function proportional to (29).

For the kernel (29), we have

$$\hat{k}(\omega) = \int_{-\delta/2}^{\delta/2} \cos(\Omega t)\cos(\omega t)dt,$$

so that

$$\frac{\partial^2}{\partial\omega^2}\hat{k}(\omega) = -\int_{-\delta/2}^{\delta/2} \cos(\Omega t)t^2\cos(\omega t)dt.$$

Since $\delta\Omega < \pi$, we have $\cos(\omega t) > \cos(\Omega t) > 0$ for $t \in [-\delta/2, \delta/2]$ and $|\omega| < \Omega$, and so

$$\frac{\partial^2}{\partial\omega^2}\hat{k}(\omega) < 0 \text{ for } \omega \in [-\Omega, \Omega].$$

Hence $\hat{k}$ is concave on $[-\Omega, \Omega]$. But as $\hat{k}$ is even we get

$$\inf_{\omega\in[-\Omega,\Omega]} |\hat{k}(\omega)| = \hat{k}(\Omega)$$

so

$$C_2(k, \Omega) = I_2^{-1};$$

that is, (29) attains the lower bound (30).

We calculate

$$\begin{aligned} I_2 &= \int_{-\delta/2}^{\delta/2} \cos^2(\Omega t)dt \\ &= \delta\left(1/2 + \frac{\sin(\Omega\delta)}{\Omega\delta}\right), \end{aligned} \tag{31}$$

and (28) follows.

We now turn to optimality of the boxcar function for use with Lemma 8.

Lemma 11 *If $\Omega\delta < \pi$ then*

$$\inf\{C_1(g, \Omega) : supp(g) \subset [-\delta/2, \delta/2]\} = \frac{\Omega/2}{\sin(\Omega\delta/2)}$$

and any optimal kernel is proportional to $1_{[-\delta/2,\delta/2]}$.

Proof. For any kernel g

$$C_1(g,\Omega) \geq ||g||_\infty / |\hat{g}(\Omega)|.$$

Letting I_∞ denote the value of the optimization problem .

$$\sup\{|\hat{g}(\Omega)| : supp(g) \subset [-\delta/2, \delta/2] \text{ and } ||g||_\infty = 1\},$$

we have that for all g supported in $[-\delta/2, \delta/2]$

$$C_1(g,\Omega) \geq I_\infty^{-1}.$$

For $\Omega\delta < \pi$ the function $\cos(\Omega t)$ is nonnegative on $[-\delta/2, \delta/2]$. It follows that

$$\hat{g}(\Omega) = \int_{-\delta/2}^{\delta/2} g(t)\cos(\Omega t)dt$$

is maximized, subject to $|g(t)| \leq 1$, by taking $g(t) = 1_{[-\delta/2,\delta/2]}$. Hence $I_\infty = \frac{\sin(\Omega\delta/2)}{\Omega/2}$. Lemma 9 shows, in effect, that with this choice of g we actually have equality $C_1(g,\Omega) = I_\infty^{-1}$ – hence the optimality of this kernel, for small δ.

6 Discrete Time

For readers interested in the geophysical prospecting problem which motivates section 2.2, it might be useful to have analogs of the above results for discrete time. Such results would provide a significant strengthening of results of Santosa and Symes [16] and help explain the conditions sufficient for success of methods of Levy and Fullagar [8] and Walker and Ulrych [19].

6.1 The l_2 setting on the discrete circle

We first state an improvement of the Bombieri-Montgomery inequality (1) based on Selberg's function.

Let $S(\alpha) = \sum_{k=m+1}^{m+n} a_k e^{2\pi i k\alpha}$ be a trigonometric polynomial of degree n and period 1, and let μ be a nonnegative measure of period 1.

$$\int_0^1 |S(\alpha)|^2 d\mu \leq (n - 1 + \delta^{-1}) \left(\sup_\alpha \mu[\alpha, \alpha+\delta]\right) \sum |a_k|^2 \qquad (32)$$

The proof of (32) is analogous to the proof of (11). First, in an analog of Lemma 5, one shows that

$$\int_0^1 |S(\alpha)|^2 d\mu \leq C_2(g, \pi(n-1)) \left(\sup_\alpha \mu[\alpha, \alpha+\delta]\right) \sum |a_k|^2$$

where C_2 has the same meaning as in section 3; one then invokes Lemma 6, with argument $\Omega = \pi(n-1)$. By the above discussion, for $(n-1)\delta$ integral, the stated result uses the best possible value for C_2. Compare also [13, 18].

Our application is as in [5], section 4. Suppose that $r = (r_t, t = 0, \ldots, N-1)$ is a measured discrete signal. It contains noise and is missing low frequency information; thus $r = (I - P_K)s + n$ where $n = (n_t, t = 0, \ldots, N-1)$ is a noise sequence, and P_K is a circular bandlimiting operator, the matrix that operates as least squares projector onto the span of the sinusoids with frequencies in $\{\frac{2\pi j}{N} : \frac{-K}{2} \leq j \leq \frac{K}{2}\}$.

In this setting, which is analogous to section 2.1, recovery of s is possible if s is sufficiently sparse. A mathematical statement is as follows. Suppose we know *a priori* that the support $T = \mathrm{supp}(s)$ belongs to a certain class $\mathcal{T}$ of sets; then the support of the difference $s_1 - s_0$ of any two proposed reconstructions belongs to the class

$$\mathcal{T}_2 \equiv \{T_1 \cup T_2 : T_i \in \mathcal{T}\}.$$

Define $\lambda_0(T, K) = ||P_K P_T||_2$, where the usual spectral norm is implied. Donoho and Stark [5] show that if

$$\Lambda(\mathcal{T}_2, K) \equiv \sup_{T \in \mathcal{T}_2} \lambda_0(T, K) < 1,$$

then *a priori* knowledge that $supp(s) \in \mathcal{T}$ enables stable recovery of s from r, with stability coefficient $2(1-\Lambda^2)^{-1/2}$.

The large sieve, described above, affords a useful bound for λ_0 in terms of the discrete Nyquist density. Define the discrete, circular Nyquist density

$$\rho(t, K, N) = K/N \sup_t \#(T \cap [t, t + N/K]) \tag{33}$$

with the interval $[t, t+N/K]$ interpreted circularly. Theorem 12 below implies that $\lambda_0^2 \leq 2\rho$. Define the class $\mathcal{T}$ as the collection of all sets of Nyquist density less than l; then $\mathcal{T}_2$ consists of sets of density less than $2\,l$. Hence we can conclude that $\Lambda(\mathcal{T}_2, K)^2 < 4\,l$; and that for $l < 1/4$ stable recovery is possible.

Theorem 12
$$\lambda_0(T,K)^2 \le 2\ \rho(T,K,N). \tag{34}$$

The argument is as follows. If $(x_t, t = 0, \ldots, N-1)$ is a discrete sequence bandlimited to $\{-K/2, \ldots, K/2\}$, then, for appropriate coefficients (a_k),

$$x_t = \sum_{k=-K/2}^{K/2} a_k e^{2\pi i k t/N}$$

Put $n = K+1$, $m = -K/2-1$, etc. Then, with $S(\alpha) = \sum a_k e^{2\pi i k \alpha}$, $S(\frac{t}{N}) = x_t$. Let ν_α denote the unit Dirac mass at α, and put $\mu = \sum_{t\in T} \nu_{t/N}$, with periodic extension outside $[0,1)$. Then

$$\int_{[0,1)} |S(\alpha)|^2 d\mu = \sum_{t\in T} |x_t|^2$$

and
$$\sup_\alpha \mu[\alpha, \alpha+\delta] = \sup_t \#(T \cap [t, t+N\delta]).$$

By Parseval's relation for the finite discrete Fourier Transform, $\sum_0^{N-1} |x_t|^2 = N \sum_k |a_k|^2$. Using these with (32) gives

$$\frac{\sum_{t\in T} |x_t|^2}{\sum_0^{N-1} |x_t|^2} \le (K+\delta^{-1}) \frac{1}{N} \sup_t \#(T \cap [t, t+N\delta]).$$

Choosing $\delta^{-1} = K$ gives (34).

We summarize our discussion regarding stable recovery.

Corollary 13 *Let $\mathcal{T}$ be the class of sets T with $\rho(T,K) \le l < 1/4$. Then $\Lambda^2(\mathcal{T}_2, K) \le 4l < 1$. Given a priori information that supp$(s) \in \mathcal{T}$ and that $||n||_2 \le \epsilon$, stable recovery of s from $r = (I-P_K)s + n$ is possible. Specifically, there exists a nonlinear mapping $\tilde{s}(r; \epsilon, \mathcal{T})$ so that*

$$||s - \tilde{s}(r)||_2 \le 2(1-4l)^{-1/2}\epsilon. \tag{35}$$

The method which yields stable recovery in this result is a "subset search" algorithm, along the lines of section 5 in [5]. For each subset T, one can find, by a linear least squares projection, a sequence $\tilde{s}_T$ supported on T which best approximates the data, in the sense that $||(I-P_K)(\tilde{s}-r)||_2$

is minimized among all sequences $\tilde{s}$ supported by T. One then searches, among all subsets T satisfying the sparsity constraint $\rho \leq l$, for a subset T with corresponding sequence $\tilde{s}_T$ satisfying $||(I - P_K)(\tilde{s}_T - r)||_2 \leq \epsilon$. Such a set exists, by hypothesis; call it T^*. Then one simply sets $\tilde{s}(r, \epsilon, \mathcal{T}) \equiv \tilde{s}_{T^*}$. Of course, as stated, this procedure is of combinatorial computational complexity. A number of computationally effective methods for recovery have been proposed [16, 8, 19]; it would be interesting to show that these obey stability estimates of the type which we have just established for the subset search method.

6.2 The l_1 setting on the discrete line

Now consider a different discrete time setting, with the time index being all integers rather than the integers mod N. Let $b_1(\Omega)$ be the set of discrete bandlimited sequences: sequences in l_1 whose Fourier transform $\hat{X}(\omega) = \sum_{t=-\infty}^{\infty} x_t e^{-i\omega t}$ vanishes for ω outside of $[-\Omega, \Omega]$. Here we must have the bandlimit $\Omega < \pi$, in fact $\Omega << \pi$ in order for something interesting to happen. Let us in fact assume that $\Omega = \pi/m$ for some odd integer $m > 1$. m is then a sort of discrete Nyquist rate.

Theorem 14 *Let $\Omega = \pi/m$. For $(x_t) \in b_1(\Omega)$ and μ a nonnegative measure supported on the integers*

$$\sum |x_t|\mu(\{t\}) \leq \sin(\frac{\pi}{2m}) \left(\sup_t \mu[t+1, t+m]\right) \sum_t |x_t|. \tag{36}$$

Proof. This is the discrete time analog of (21). The argument is entirely parallel to that for Theorem 7, so we simply mention the steps. Let $g = (g_u)$ be a discrete filter sequence, with $supp(g) = \{1, ..., m\}$. The discrete analog of Lemma 8 holds, and says that if convolution with g is an invertible operator on $b_1(\Omega)$, and if μ is a measure supported on the integers, then

$$\sum_t |x_t|\mu(\{t\}) \leq c_1(g, m) \left(\sup_t \mu[t+1, t+m]\right) \sum_t |x_t|$$

where

$$c_1(g, m) \equiv ||g||_\infty \sup\{\frac{||x||_1}{||g \star x||_1} : \ x \in b_1(\omega)\},$$

l_p norms and discrete convolution being intended in these definitions.

We specialize attention to the discrete boxcar g which is 1 on $\{1, 2, \ldots, m\}$ and zero elsewhere. Defining $\hat{g}(\omega) = \sum_{k=1}^{m} e^{-i\omega k}$, we can show that for this g

$$c_1(g, \Omega) = ||g||_\infty / |\hat{g}(\Omega)|. \tag{37}$$

As $||g||_\infty = 1$ and

$$|\hat{g}(\omega)| = \frac{\sin(\frac{m}{2}\omega)}{\sin(\frac{\omega}{2})},$$

(36) follows.

The identity (37) is established by an argument similar to that for Lemma 9. Recall that $m > 1$ is an *odd* integer. Let $\tilde{g}$ be a 2Ω-periodic extension of g away from the fundamental interval $[-\Omega, \Omega]$. Put $\hat{h}(\omega) = e^{i\omega(m-1)/2} / \tilde{g}(\omega - \Omega)$. One then verifies that

$$[\log \hat{h}(\omega)]'' = \frac{m^2}{4\sin^2(\frac{m}{2}(\omega - \Omega))} - \frac{1}{4\sin^2(\frac{1}{2}(\omega - \Omega))},$$

which implies convexity of h on $[0, 2\Omega]$. The transform $h_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{h}(\omega) e^{i\omega t} d\omega$ is therefore nonnegative, the sequence $v_t = e^{i\Omega t} h_{t+(m-1)/2}$ is a convolution inverse to g on $b_1(\Omega)$, with variation norm $\sum |v_t| = \sum h_t = h(0) = 1/|\hat{g}(\Omega)|$. Both (37) and the Theorem follow.

Here is our application. Suppose we observe $r = b + n$, where b is known to be in $b_1(\pi/m)$ and n is known to lie in l_1. We apply the l_1 projection

$$\beta_1(r) = \arg \min_{\tilde{b} \in b_1(\pi/m)} ||r - \tilde{b}||_1.$$

Let T be the support of n and define the *discrete Nyquist density* of T by

$$\rho(T, m) = m^{-1} \sup_t \#(T \cap [t+1, t+m]).$$

By (36) and an argument similar to Lemma 2, the condition

$$\rho(T, m) < 2m / \sin(\pi/2m)$$

implies $\beta_1(r) = b$ exactly. For large m this condition for Logan's phenomenon is approximately $\rho < 1/\pi$, the same as in the continuous case.

7 Appendix

It is enough to prove Lemma 2 in the case $b = 0$, as the general case follows by translation.

Suppose now that $b = 0$. By definition, $\beta_1(r)$ satisfies $||r - \beta_1(r)||_1 \leq ||r - b||_1$. Now

$$||r - b||_1 = ||n||_1$$

while

$$\begin{aligned} ||r - \beta_1||_1 &\geq ||P_T n||_1 - ||P_T \beta_1||_1 + ||P_{T^c} \beta_1||_1 - ||P_{T^c} n||_1 \\ &= ||n||_1 - 2\epsilon + ||P_{T^c} \beta_1||_1 - ||P_T \beta_1||_1. \end{aligned}$$

By definition of μ_0, and $\beta_1 \in B_1(\Omega)$,

$$||P_{T^c} \beta_1||_1 - ||P_T \beta_1||_1 \geq (1 - 2\mu_0(T, W))||\beta_1||_1.$$

Combining these, we must have

$$||\beta_1(r)||_1 \leq 2(1 - 2\mu_0(T, W))^{-1}\epsilon$$

and the proof is complete.

References

[1] R.P. Boas, Jr. Entire functions bounded on a line. *Duke Math. J.*, 6:148–169, 1940.

[2] A. Beurling. Sur les integrales de Fourier absolument convergentes et leur application á une transformation fonctionelle. *Ninth Scandinavian Math. Congress, Helsingfors*, 1938, pp. 345-366. Reprinted in *Collected Works of Arne Beurling: Volume II Harmonic Analysis*, pp. 39-60. Birkhauser: Boston. 1989.

[3] A. Beurling. On functions with a spectral gap. Seminar on Harmonic Analysis, Univ. of Uppsala. March 24, 1942. Published in *Collected Works of Arne Beurling: Volume II Harmonic Analysis*, pp. 370-372. Birkhauser: Boston. 1989.

[4] A. Beurling. Interpolation on the line. 3. Minimal Extrapolation. Mittag-Leffler Lectures on Harmonic Analysis 1977-78. Published in *Collected Works of Arne Beurling: Volume II Harmonic Analysis*, pp. 359-364. Birkhàuser: Boston. 1989.

[5] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49:906–931, June, 1989.

[6] R.J. Duffin and A.C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Am. Math. Soc.*, 72:341–366, 1952.

[7] G.H. Hardy and W.W. Rogosinksi. *Fourier Series.* Cambridge University Press, 1944.

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

[9] B.F. Logan. *Properties of High-Pass Signals.* PhD thesis, Columbia University, 1965.

[10] B.F. Logan. Bandlimited functions bounded below over an interval. Notices Amer. Math. Soc. **24**, A-331, 1977.

[11] B.F. Logan. The norm of certain convolution transforms on L_p spaces of entire functions of exponential type. *SIAM J. Math. Anal.* **16**, 1, 167-179, 1985.

[12] B.F. Logan. *Analysis in Fourier Spaces.* Book in preparation.

[13] H.L. Montgomery. The Analytic Principle of the Large Sieve. *Bull. Am. Math. Soc.* 84:547-567, 1978.

[14] S.M. Nikolśkiĭ. Inequalities for entire functions of finite degree and their application to the theory of differentiable functions of several variables. In *American Mathematical Society Translations Series 2, Vol. 80, Thirteen Papers on Functions of Real and Complex Variables*, pages 1–38, American Mathematical Society, Providence, R.I., 1969.

[15] M. Plancherel and G. Pólya. Fonctions entières et intégrales de fourier multiples. *Commentarii math. Helv.*, 10:110–163, 1938.

[16] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comp.*, 7:1307–1330, 1986.

[17] A. Selberg. Remarks on Sieves. *Proc. 1972 Number Theory Conference* University of Colorado, Boulder, pp. 205-216, 1972

[18] J. D. Vaaler. Some extremal functions in Fourier Analysis *Bull. Amer. Math. Soc.*, 12:183-216, 1985.

[19] C. Walker and T.J. Ulrych. Autoregressive recovery of the acoustic impedance. *Geophysics*, 48:1338-1350, 1983.