Super-Resolution via Sparsity Constraints

By

D.L. Donoho
University of California, Berkeley

Technical Report No. 285
January 1991

Department of Statistics
University of California
Berkeley, California 94720

Super-Resolution via Sparsity Constraints

David L. Donoho*
Department of Statistics
University of California, Berkeley.

December, 1990

Abstract

Consider the problem of recovering a measure μ supported on a lattice of span Δ, when measurements are only available concerning the Fourier Transform $\hat{\mu}(\omega)$ at frequencies $|\omega| \leq \Omega$. If Ω is much smaller than the Nyquist frequency π/Δ and the measurements are noisy, then, in general, stable recovery of μ is impossible. In this paper we show that if, in addition, we know that the measure μ satisfies certain sparsity constraints, then stable recovery is possible. Say that a set has *Rayleigh index* less than or equal to R if in any interval of length $4\pi/\Omega \cdot R$ there are at most R elements. Indeed, if the (unknown) support of μ is known, *a priori*, to have *Rayleigh index* at most R, then stable recovery is possible with a stability coefficient that grows at most like Δ^{-4R-1} as $\Delta \to 0$. This result validates certain practical efforts, in spectroscopy, seismic prospecting, and astronomy, to provide super-resolution by imposing support limitations in reconstruction. Our results amount to inequalities for interpolation of entire functions of exponential type from values at special point sets which are irregular, yet internally balanced, uniformly discrete, and of uniform density 1.

Key Words and Phrases. Inverse Problems. Spectroscopy. Diffraction-limited imaging. Rayleigh criterion. Nyquist Rate. Super-Resolution. Nonlinear recovery. Entire Functions of Exponential Type. Interpolation. Balayage.

*Supported by NSF grant DMS 88-10192, by NASA contract NCA2-0488, and by grants from Schlumberger-Doll Research, Western Geophysical, and from A.T.T. Foundation.

1 Introduction

Let $\mu = \sum_{k=-\infty}^{\infty} \alpha_k \delta_{k\Delta}$ be a signed measure supported on the lattice $\{k\Delta\}_{k=-\infty}^{\infty}$, with signed mass α_k attached to the point $k\Delta$. We think of the lattice span Δ as a small number $\Delta << 1$. The measure μ may be interpreted as a caricature of certain scientifically interesting objects: for example, a polarized spectrum in a spectroscopy problem; or, in exploration seismology and in medical ultrasound, as the sequence of reflectivities of a layered medium with layers of constant width Δ.

Suppose we obtain noisy measurements on μ in the frequency domain, with frequency cutoff Ω:

$$y(\omega) = \hat{\mu}(\omega) + z(\omega), \ |\omega| \leq \Omega. \tag{1}$$

Here $\hat{\mu}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} d\mu(t) = \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k\Delta}$ is the Fourier transform of μ, and $z(\omega)$ represents noise. Our objective is to recover μ (or, equivalently, the coefficients α_k) from the data (1).

To make the phrase "recovery of μ" precise, we adapt some notions from the theory of *optimal recovery* (compare Micchelli and Rivlin (1977)). We suppose first of all that the noise z can be any function in $L_2[-\Omega,\Omega]$ satisfying $||z||_{L_2[-\Omega,\Omega]} \leq \epsilon$. Let $\tilde{\mu} = \tilde{\mu}(y)$ be a method of recovery. We measure the recovery error with respect to the quadratic "Wiener" norm $||\tilde{\mu}(y) - \mu||_2$, where, for a discrete signed measure ν, we define $||\nu||_2 \equiv (\sum_{t \in supp(\nu)} |\nu(\{t\})|^2)^{1/2}$. We record our *a priori* information about μ by saying that $\mu \in \mathcal{M}$, where $\mathcal{M}$ is a class of measures. For example, if our prior information is that μ is a lattice measure, i.e. a member of the set $\mathcal{L}(\Delta) = \{\mu : \mu = \sum_{k=-\infty}^{\infty} \alpha_k \delta_{k\Delta}\}$, then we set $\mathcal{M} = \mathcal{L}(\Delta)$. In general, we measure the difficulty of recovery with a priori information $\mathcal{M}$ by the minimax error over $\mathcal{M}$:

$$E^*(\epsilon; \mathcal{M}, \Omega) = \inf_{\tilde{\mu}} \sup_{\mu \in \mathcal{M}} \sup_{||z||_{L_2[-\Omega,\Omega]} \leq \epsilon} ||\tilde{\mu}(y) - \mu||_2.$$

In particular, we say that stable recovery of μ is possible, with *a priori* information $\mathcal{M}$, if

$$E^*(\epsilon; \mathcal{M}, \Omega) \leq Const \cdot \epsilon,$$

which indicates the stability property

$$\text{Error} \leq \text{Constant} \cdot \text{Noise Level}.$$

1.1 The stability threshold $\Omega \geq \pi/\Delta$

π/Δ is the usual Nyquist frequency for samples taken on the lattice $\{k\Delta\}_{k=-\infty}^{\infty}$. This frequency pops up in our problem: if the frequency cutoff Ω exceeds π/Δ then stable recovery of μ is possible. Indeed, motivated by the Fourier inversion formula

$$\alpha_k = \frac{\Delta}{2\pi}\int_{-\pi/\Delta}^{\pi/\Delta} e^{i\omega k\Delta}\hat{\mu}(\omega)d\omega \tag{2}$$

one is led immediately to the rule

$$\tilde{\alpha}_k = \frac{\Delta}{2\pi}\int_{-\pi/\Delta}^{\pi/\Delta} e^{i\omega k\Delta}y(\omega)d\omega. \tag{3}$$

Parseval's relation implies that the reconstruction formula $\tilde{\mu} = \sum_{k=-\infty}^{\infty}\tilde{\alpha}_k\delta_{k\Delta}$ has error

$$||\tilde{\mu}(y)-\mu||_2^2 = \frac{\Delta}{2\pi}\int_{-\pi/\Delta}^{\pi/\Delta}|z(\omega)|^2 d\omega \leq \frac{\Delta}{2\pi}\epsilon^2.$$

Consequently, if $\mathcal{M} = \mathcal{L}(\Delta)$ we get

$$E^*(\epsilon;\mathcal{M},\Omega) \leq \sqrt{\frac{\Delta}{2\pi}}\cdot\epsilon;$$

and an extra argument shows that in fact equality holds. Hence, in the case $\Omega \geq \pi/\Delta$, stable recovery is possible, and in fact optimal recovery requires only the simplest of linear reconstruction formulas.

The case $\Omega << \pi/\Delta$ is more interesting. In this case data on $\hat{\mu}(\omega)$ are not available on the whole range [−Nyquist, Nyquist], and formulas like (3) are not immediately applicable. Indeed, formula (3) suggests that, if $\Omega < \pi/\Delta$, reconstruction will require some sort of process of *extrapolation* of the noisy measurements inside $[-\Omega, \Omega]$ to produce quasi-measurements over the whole of the fundamental interval $[-\pi/\Delta, \pi/\Delta]$.

However, such extrapolation is evidently impossible in the absence of special prior information. Indeed, if μ can be any lattice measure, then (α_k) can be any square summable sequence. We could therefore let $\hat{\mu}(\omega)$ be a nonzero function, periodic of period $2\pi/\Delta$, belonging to L^2 on the fundamental interval $(-\pi/\Delta, \pi/\Delta)$, and vanishing on $[-\Omega, \Omega]$. The sequence (α_k) obtained from (2) would give a nonzero lattice measure μ whose transform agrees, over the low frequency band $|\omega| \leq \Omega$, with the zero measure. Even at noise level $\epsilon = 0$, our observations could not distinguish this μ from the zero measure, nor from its sign-reversal $-\mu$. Consequently, if $\Omega < \pi/\Delta$,

$$\begin{aligned} E^*(\epsilon, \mathcal{L}(\Delta), \Omega) &\geq \sup\{||\mu - \tilde{\mu}||_2 : \mu, \tilde{\mu} \in \mathcal{L}(\Delta), \hat{\mu}(\omega) = \hat{\tilde{\mu}}(\omega), |\omega| \leq \Omega\} \\ &\geq \sup\{||\mu||_2 : \mu \in \mathcal{L}(\Delta), \hat{\mu}(\omega) = 0, |\omega| \leq \Omega\} \\ &= +\infty. \end{aligned}$$

Stable recovery is not possible under the condition $\Omega < \pi/\Delta$ if all we know *a priori* is the lattice constraint $\mu \in \mathcal{L}(\Delta)$.

1.2 The Rayleigh threshold $\Omega \geq \pi/\Delta$

A mathematically equivalent reformulation of our problem occurs in the theory of optics. The frequency-domain data (1) are equivalent to the spatial-domain data

$$Y(t) = (K_\Omega \star \mu)(t) + Z(t) \quad t \in (-\infty, \infty) \tag{4}$$

where $K_\Omega(t)$ is the sinc-Kernel $\sin(\Omega t)/(\pi t)$ and $\star$ denotes convolution; and Z is a bandlimited noise with Fourier transform

$$\hat{Z}(\omega) = \begin{cases} z(\omega) & |\omega| \leq \Omega \\ 0 & \text{else} \end{cases}$$

Hence, one observes not the measure μ directly, but instead a noisy version which is blurred by convolution with the kernel

K_Ω. In this form, Y is a noisy diffraction-limited image of μ, a superposition of point-sources.

The study of diffraction-limited imaging for such superpositions of point sources goes back a long way. Lord Rayleigh studied it, and formulated a "resolution limit" [10, Pages 33-35]: if a measure μ consists of two point sources of equal strength separated by a distance Δ, a visual inspection of the $Y(t)$ curve will suggest the presence of two point sources provided $\Delta \geq 1.22\pi/\Omega$ and of one point source provided that $\Delta < 1.22\pi/\Omega$. Rayleigh's constant 1.22 is rather arbitrary, and Rayleigh's argument could, with minor modifications, yield instead the constant 1.0. This replacement would lead to the criterion: *pointlike sources separated by at least Δ can be resolved into separate sources, using data diffraction-limited by K_Ω, provided $\pi/\Delta \leq \Omega$.*

This modified Rayleigh limit coincides with the threshold $\pi/\Delta \leq \Omega$ for stable recovery mentioned earlier. In this sense, if we *were* able to recover stably the lattice measure μ from data satisfying $\Omega << \pi/\Delta$, we would have exceeded Rayleigh's resolution limit. Therefore, the problem of stably recovering μ from noisy data when the parameters in the range $\Omega << \pi/\Delta$ below the Rayleigh threshold may be called the problem of *super-resolution.*

1.3 Empirical super-resolution via sparsity

Despite the mathematical fact that stable recovery of the class $\mathcal{L}(\Delta)$ is impossible when the data satisfy $\Omega << \pi/\Delta$, there has been considerable effort to develop super-resolving algorithms for specific problems. The idea is essentially that additional *a priori* information about the support of the measure μ should be exploited in the recovery process.

1. Högbom [11] and others, working in radio astronomy, have developed the method CLEAN, which involves finding a small set of delta-functions $\tilde{\mu} = \sum \alpha_k \delta_{t_k}$ such that $K \star \tilde{\mu}$ (where K is a sinc-like kernel) nearly reproduces the original measurements. As Schwarz [27] says, "... some extra information about the brightness distribution must be used. The CLEAN method is designed for the case that the brightness distribution

contains only a few sources at well-separated, small regions, i.e. *the brightness distribution is essentially empty.*"

2. Papoulis and Chamzas [23] have proposed a nonlinear iterative method which assumes implicitly that the underlying measure is sparse, attempts to adaptively identify the regions where coefficients might be nonzero, and recover an object supported only in those regions. They describe an application in medical ultrasound [24]. They point out that the Rayleigh limit is exceeded, in some examples, by their method, and that the actual limit of resolution depends on noise and signal in some yet-to-be determined fashion.

3. Working in seismic prospecting, Levy and Fullagar [15], Santosa and Symes [26] and Walker and Ulrych [32] describe methods which attempt to exploit the fact that the underlying object is a "sparse spike train" to recover wideband data from measurements over a limited frequency range. The Levy-Fullagar and Santosa-Symes work exploits special support properties of l_1-norm penalized reconstruction – namely that for large values of the multiplier attached to the penalty, the algorithm tends to employ very few nonzero elements in the reconstruction. Walker and Ulrych exploit a method based on low-order autoregressive extrapolation of the Fourier data away from the measured frequency band. The low-order autoregressive model for the Fourier transform may be justified b an assumption that few elements in the spatial domain representation of the object are nonzero. Wang [33] recently introduced a method which constrains the reconstruction so that in any segment of a certain length there are only a few nonzero elements.

4. Working in Fourier transform spectroscopy, Kawata, Minami, and Minami [14, 20] and Mammone [18], also exploit parsimony. Kawata et al. use low-order autoregressive extrapolation away from the measured frequency band, and Mammone uses parametric linear programming to get reconstructions which nearly reproduce the data with minimal numbers of nonzero elements. In later work Minami et al. refined their technique by using the singular value decomposition to improve the choice of order in autoregressive extrapolation.

5. Working in NMR spectroscopy, Barkhuisen et al. [1] use autoregressive extrapolation, combined with singular value decomposition (LPSVD); Newman [22] proposes the use of l_1-norm penalized reconstruction. Tang

and Norris [28] and Mazzeo et al. [19] divide the signal into segments and treat individual segments by sparsity-enhancing methods (e.g. LPSVD) mentioned above.

All these researchers seek to recover wideband objects from narrow-band data; all proceed by in some way imposing sparsity limitations on the recovered object; and all have achieved successes in certain computational experiments. Implicitly or explicitly, these successes amount to a claim that a certain sparsity of the unknown object enables recovery.

At first glance, the computational work just mentioned seems to conflict with the Rayleigh and Stability criteria developed above. In fact there is no conflict, since the Rayleigh and the stability criteria do not seek to describe the impact of sparsity.

1.4 Theoretical results

We now develop theory that sheds light on the possibilities, and difficulties, of super-resolution via sparsity constraints. We will show that, if the support of μ, though unknown, is known to be sufficiently sparse, then even in the case $\Omega << \pi/\Delta$, stable recovery is possible. On the other hand, the quantitative degree of stability might be disappointingly poor if we must recover objects which possess a high degree of complexity.

We do not exhibit a practical method for achieving stable recovery, but instead exhibit inequalities which show that, in the sense of the theory of optimal recovery, the object admits of stable reconstruction. Any stable reconstruction scheme is necessarily highly nonlinear. It would be interesting to know whether stability of the kind we establish below holds for the nonlinear methods [1, 11, 14, 15, 20, 22, 26, 32, 24] mentioned above.

Before developing stability results, we discuss uniqueness. Let S be a discrete set. Following Beurling [3], we define the upper uniform density

$$u.u.d.(S) = \lim_{r\to\infty} r^{-1} \sup_t \#(S \cap [t, t+r))$$

and the lower uniform density

$$l.u.d.(S) = \lim_{r\to\infty} r^{-1} \inf_t \#(S \cap [t, t+r)).$$

Both limits exist.

Theorem 1 *a. Let $\mathcal{M}_{<1}(\Delta)$ denote the class of finite signed lattice measures $\mu \in \mathcal{L}(\Delta)$ which have density $u.u.d.(supp(\mu)) < 1$. If $\Omega \geq 2\pi$, μ is uniquely characterized among $\mathcal{M}_{<1}(\Delta)$ by the transform $\hat{\mu}(\omega)$, $|\omega| \leq \Omega$.*

b. Let S_1 and S_2 be any two disjoint sets with $l.u.d.(S_i) > 1$, $i = 1, 2$. Let μ_1 be any finite signed measure supported on S_1. There exists μ_2 with support S_2 such that $\hat{\mu}_1(\omega) = \hat{\mu}_2(\omega)$ for $|\omega| \leq \pi$, yet $\mu_1 \neq \mu_2$.

c. There exist disjoint equispaced sets S_i, with $l.u.d.(S_i) = u.u.d.(S_i) = 1$, $i = 1, 2$, and measures μ_i supported in the S_i, such that $\hat{\mu}_1(\omega) = \hat{\mu}_2(\omega)$ for $|\omega| \leq 2\pi - \delta$, for $\delta > 0$ yet $\mu_1 \neq \mu_2$.

The proof, which relies on Beurling's theory of interpolation and balayage [3, 2], is given in section 7.

We conclude that $u.u.d. < 1$ and $\Omega \geq 2\pi$ ensures uniqueness; $l.u.d. > 1$ and $\Omega \leq \pi$ ensures nonuniqeness; and $l.u.d. = u.u.d. = 1$ and $\Omega < 2\pi$ may, in some cases, lead to nonuniqueness. Hence in searching for stability results, we confine attention to the case $u.u.d. < 1$ and $\Omega \geq 2\pi$.

By rescaling, this pair of conditions is equivalent to the single condition $u.u.d. < \Omega/2\pi$. Now compare this uniqueness condition with the modified Rayleigh criterion $\Delta > \pi/\Omega$. For a typical non-sparsely supported measure $\mu \in \mathcal{L}(\Delta)$, $u.u.d.(supp(\mu)) = \Delta^{-1}$; hence Rayleigh's criterion is comparable to $u.u.d. < \Omega/\pi$. Our uniqueness criterion therefore demands exactly twice the frequency-domain measurement band (or half the spatial domain density) as the Rayleigh criterion. Our stability estimates will demand at least four times as much as the Rayleigh criterion.

We now return to the scaling convention $u.u.d. < 1$.

Definition 1 *Let S be a discrete set of upper uniform density less than one. The* **Rayleigh index** *of S is*

$$R^*(S) = min\{R : R \geq \sup_t \#(S \cap [t, t+R))\}.$$

The class of lattice measures μ with $(\alpha_k) \in l_1$ and Rayleigh index $R^*(supp(\mu)) \leq R$ will be denoted $\mathcal{S}(R, \Delta)$.

The Rayleigh index measures, for sets which have on average, no more than one element per unit cell, the maximum number that can be clustered very closely together. We aim to show that the clustering of many elements together in one cell makes super-resolution difficult; we will see that the degree of clustering, as measured by the Rayleigh index, enters directly into our bounds on the stability coefficient.

Definition 2 *The* **modulus of continuity** *for the recovery of measures in $\mathcal{S}(R, \Delta)$ is*

$$\begin{aligned} \Lambda(\epsilon; \mathcal{S}(R, \Delta), \Omega) &= \sup\{||\mu_1 - \mu_2||_2 : \mu_i \in \mathcal{S}(R, \Delta), \\ & \quad ||\hat{\mu}_1 - \hat{\mu}_2||_{L_2[-\Omega,\Omega]} \leq \epsilon\}. \end{aligned}$$

This modulus of continuity measures the extent to which two lattice measures, both satisfying the sparsity condition $R^*(supp(\mu_i)) \leq R$, can differ, if the bandlimited data $\{\hat{\mu}_i(\omega), |\omega| \leq \Omega\}$ differ by at most ϵ in L_2-norm. Its relevance comes from

Lemma 1

$$E^*(\epsilon, \mathcal{S}(R, \Delta), \Omega) \leq \Lambda(2\epsilon, \mathcal{S}(R, \Delta), \Omega). \tag{5}$$

The proof is given in section 7. [For arguments relating a modulus of continuity to a minimax error in other contexts, see [21, 25, 29, 30, 6].]

Our main result bounds the modulus of continuity directly

Theorem 2 *Let $\Omega > 4\pi$.*

$$\Lambda(\epsilon, \mathcal{S}(R, \Delta), \Omega) \leq \Delta^{-4R-1} \cdot \beta(R, \Omega) \cdot \epsilon \tag{6}$$

for a positive finite constant β defined below.

The raw materials on which this result depends are developed in the body of the paper, sections 2 through 6 below. They are assembled to give a formal proof in section 7.

The following lower bound shows that our upper bound is nearly sharp. It is proved in section 7 below.

Theorem 3 *Let $\Delta_0 \in (0,1)$. If $\Delta < \Delta_0$ then*

$$\Lambda(\epsilon, \mathcal{S}(R,\Delta), \Omega) \geq \Delta^{-2R+1} \cdot b(R,\Omega,\Delta_0) \cdot \epsilon \tag{7}$$

for a positive finite constant $b(R,\Omega,\Delta_0)$ defined below.

1.5 Interpretation of the theory

To interpret these results, we introduce some terminology. We speak of the stability coefficient as the *noise amplification* factor in the relation

Reconstruction Error = Noise Amplification $\cdot$ Noise in Data.

We speak of the ratio Ω/Δ as the *super-resolution factor.* Our results indicate that *noise amplification increases polynomially with the super-resolution factor.* Hence, to achieve reconstructions with a fixed degree of reconstruction error requires data of increasingly low noise level as the super-resolution factor increases. Moreover, the rate at which the noise requirement imposes itself is directly tied to the Rayleigh index, and hence it may be extremely difficult to recover an object with a high degree of clumping or irregularity.

Hence super-resolution is possible if the object is known to contain, on average, less than one pointlike event per cell of size $4\pi/\Omega$; but this may require extremely precise data, particularly if we cannot rule out the possibility that a few cells contain many more than one pointlike event. These relations, rather than the Rayleigh criterion, determine the ultimate limits of resolution.

Acknowledgements. The author would like to thank D. Dacunha-Castelle, E. Gassiat, J.C. Hoch, R.L. Parker, P.B. Stark, A. Stern, and Kong Wang for interesting discussions and correspondence.

2 Balanced (R, Δ)-Sets

Definition 3 *A* **balanced** (R, Δ)**-set** *is a countable set of points* $\{t_k\}$ *on the real line which may be obtained from the union of two bilateral sequences* $(u_i)_{i=-\infty}^{\infty}$ *and* $(v_i)_{i=-\infty}^{\infty}$ *satisfying these conditions:*

$$\begin{array}{rlll} \textit{(Internal Symmetry)} & u_i = i + \delta_i & i \in \mathbf{Z} & (8) \\ & v_i = i - \delta_i & i \in \mathbf{Z} & \\ \textit{(Uniform Density)} & |\delta_i| \leq R & i \in \mathbf{Z} & (9) \\ \textit{(Uniform Discreteness)} & |t_k - t_l| \geq \Delta & k \neq l & (10) \\ \textit{(Unicity)} & u_i \neq u_j & i \neq j & (11) \end{array}$$

We note that no assumption is made that $v_i \neq v_j$ for $i \neq j$. Nor is there an assumption that $u_i \neq v_j$ for every i, j. Consequently, the multiplicity

$$m_k = \#\{i : u_i = t_k\} + \#\{i : v_i = t_k\}$$

may be greater than 1; in fact as large as $2R + 2$. Also, no assumption is made that the points t_k belong to a lattice, although this is not excluded, either.

These four conditions describe a set which is allowed to be locally irregular, yet must be globally regular. A long interval, of length N, say, contains roughly $2 \cdot N$ elements of the set $\{t_k\}$, counting with multiplicities m_k. The internal symmetry condition is also important; it implies that even though the points are not equispaced; they may be arranged in pairs whose centers of gravity are equispaced.

Obviously, balanced (R, Δ)-sets are quite special, and do not occur "naturally"; our interest is in sets which can be "filled out", by the addition of new elements, to become balanced (R, Δ)-sets.

Definition 4 $\{s_k\}$ *is a pre-*(R, Δ)*-set if it is a subsequence of the* (u_i) *sequence associated to* (R, Δ) *set. A measure* μ *is an* (R, Δ)*-measure if its support* $\{s_k\}$ *is a pre-*(R, Δ)*-set.*

We prove the following in section 8.

Lemma 2 *The measures in $\mathcal{S}(R, \Delta)$ are all (R, Δ)-measures.*

Our introduction of balanced (R, Δ)-sets is geared to the development of certain interpolation schemes based on entire functions. Entire functions with real zeros have zeros which are roughly equispaced (compare [5, 12, 13]) and possess a certain symmetry (compare the discussion of Lindelöf's theorem in [13]). The conditions (8)-(11) serve to guarantee that there is an entire function of exponential type 2π with $\{t_k\}$ as its set of zeros. We prove the following in section 8.

Lemma 3 *Let $\{t_k\}$ be a balanced (R, Δ)-set. Define*

$$G_n(t) = \frac{\Pi_{-n}^{n}(t - u_i)(t - v_i)}{\Pi_{-n}^{n}(t - i)^2} \sin^2(\pi t).$$

Then (G_n) is a sequence of entire functions of exponential type 2π, uniformly bounded on the real axis. This sequence converges uniformly on compacts to a limit function G, entire of exponential type 2π. The $\{t_k\}$ are the zeros of the function G; the multiplicity of t_k is m_k.

The representation as a limit of the sequence (G_n) seems nonstandard; we use it for the purpose of bounding $||G||_\infty$ and related quantities. We record now several important bounds, all of which only involve the parameters R and Δ rather than the detailed properties of the set $\{t_k\}$. The proofs are given in section 8.

Lemma 4 *Let $\{t_k\}$ be a balanced (R, Δ)-set. Then*

$$\sup_t |G(t)| \leq A_1(R)$$

where $A_1(R)$ is defined below.

Lemma 5 *Let $\{t_k\}$ be a balanced (R, Δ)-set. Let G have a zero at t_0 of multiplicity m_0. Then*

$$\sup_{|t-t_0|\leq 1} \frac{|G(t)|}{|t - t_0|^{m_0}} \leq A_2(R)$$

where $A_2(R)$ is defined below.

Lemma 6 *Let $\{t_k\}$ be a balanced (R, Δ)-set. Let G have a zero of multiplicity m_k at t_k. Let*

$$g_k = \lim_{t \to t_k} G(t)/(t - t_k)^{m_k}.$$

Then

$$g_k \geq A_3(R)\Delta^{2R+1},$$

where $A_3(R)$ is a strictly positive constant defined below.

These bounds on the entire function G, although stated as lemmas, are in fact the "hard analysis" on which our main result depends; in succeeding sections we reduce the question of superresolution to these inequalities by the use of "soft analysis".

3 Super-resolution for (R, Δ) sets

Let $B_q(\Omega)$, $1 \leq q \leq \infty$, denote the space of entire functions of exponential type Ω belonging to L_q on the real axis [4, 5, 16, 17]. For a sigma-finite signed measure ν, define

$$||\nu||_{p,\Omega} = \sup\{\int f d\nu : f \in B_q(\Omega), ||f||_q \leq 1\}$$

where $1/p + 1/q = 1$ as usual. In case $p = 2$, Parseval's relation implies that

$$||\nu||_{2,\Omega}^2 = \frac{1}{2\pi}||\hat{\nu}||_{L_2[-\Omega,\Omega]}^2.$$

Also, for ν a discrete measure, let

$$||\nu||_p = (\sum_{t \in supp(\nu)} |\nu(\{t\})|^p)^{1/p}.$$

Thus, for example the case $p = 1$ gives the total variation norm of ν.

We call any inequality of the form

$$||\nu||_p \leq C_p(R, \Delta, \Omega)||\nu||_{p,\Omega}, \tag{12}$$

when valid for all finite signed (R, Δ)-measures ν, a *super-resolution inequality.* Besides the case $p = 2$, the main case of interest to us is the case $p = 1$, or

$$Variation(\nu) \leq C_1(R, \Delta, \Omega)||\nu||_{1,\Omega}.$$

The rationale for calling (12) a super-resolution inequality is that, ordinarily, bounds on the norms $||\nu||_p$ would seem to require knowledge of the transform $\hat{\nu}(\omega)$ at all frequencies up to the Nyquist π/Δ; but the norm $||\nu||_{p,\Omega}$ involves only knowledge of frequencies in the smaller band, since, by Parseval, we have

$$\int_{-\infty}^{\infty} f(t)d\nu(t) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} \hat{f}(\omega)\hat{\nu}(\omega)d\omega$$

for $f \in B_q(\Omega)$, $1 \leq q < \infty$.

4 Duality with Interpolation

Consider the following interpolation problem for the sequence $(u_i)_{i=-\infty}^{\infty}$ associated with a balanced (R, Δ)-set. *Given constants* $(c_i)_{i=-\infty}^{\infty}$, $(c_i) \in l_q$, *find a function* $f \in B_q(\Omega)$ *satisfying*

$$f(u_i) = c_i \quad i \in \mathbf{Z}. \tag{13}$$

Suppose that for *every* sequence $(c_i) \in l_q$, and *every* (u_i) associated to a balanced (R, Δ)-set, we have a solution to (13) satisfying

$$||f||_{L_q} \leq K_q(R, \Delta, \Omega)||c||_{l_q}, \tag{14}$$

where K_q does not depend on the details of the (u_i), but only on the parameters R and Δ. We claim that then, if p and q are conjugate indices $1/p + 1/q = 1$,

$$K_q(R, \Delta, \Omega) \geq C_p(R, \Delta, \Omega). \tag{15}$$

This expresses a certain duality between the superresolution inequality (12) and the interpolation problem (13).

To prove (15), let ν be any (R, Δ)-measure. Then, by definition, $supp(\nu) \subset \{u_i\}$ for some sequence $(u_i)_{i=-\infty}^{\infty}$ associated with a balanced (R, Δ)-set. Let $(c_i)_{i=-\infty}^{\infty}$ be aligned with $(\nu\{u_i\})$ in the usual sense that

$$c_i = \lambda \, \mathrm{sgn}(\nu\{u_i\}) \, |\nu\{u_i\}|^{p-1},$$

with the scalar λ chosen to make $||c||_{l_q} = 1$. Then

$$\begin{aligned} \sum_{i=-\infty}^{\infty} c_i \nu\{u_i\} &= (\sum |\nu\{u_i\}|^p)^{1/p} \\ &= ||\nu||_p, \end{aligned}$$

as $supp(\nu) \subset \{u_i\}$. Now, if f solves the interpolation problem (13), we have

$$\int f d\nu = \sum_{i=-\infty}^{\infty} c_i \nu\{u_i\} = ||\nu||_p;$$

and as $f \in B_q(\Omega)$,

$$\int f d\nu \leq ||f||_{L_q} ||\nu||_{p,\Omega}.$$

Thus, $||\nu||_p \leq ||f||_{L_q} ||\nu||_{p,\Omega}$, and so by (14),

$$||\nu||_p \leq K_q(R, \Delta, \Omega) ||\nu||_{p,\Omega}.$$

Relation (15) follows.

As a result of (15) we now turn our attention to problems of interpolation in $B_q(\Omega)$.

5 Pointwise Bounds on Interpolation

As indicated in section 2, a balanced (R, Δ)-set generates a function G, entire of exponential type 2π. By convention, G has a zero of multiplicity m_k at t_k; defining as before

$$g_k = \lim_{t \to t_k} G(t)/(t - t_k)^{m_k}, \tag{16}$$

the function

$$\xi_k(t) = \frac{G(t)}{g_k (t - t_k)^{m_k}} \tag{17}$$

satisfies formally

$$\xi_k(t_l) = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases} \tag{18}$$

Actually, ξ_k is a well-defined entire function of type 2π which belongs to L_2 on the real axis. Therefore, formally, the "Lagrange interpolation series"

$$f = \sum_k d_k \xi_k \tag{19}$$

gives a solution of the interpolation problem

$$f(t_k) = d_k, \quad k \in \mathbf{Z}. \tag{20}$$

However one would need regularity conditions on (ξ_k) in order to be sure that such sums converge and define elements of $B_q(\Omega)$.

(For discussion of Lagrange interpolation for sets where the corresponding function G has only simple zeros ($m_k \equiv 1$), see [3, 34]).

Regularity is easier to establish if we mollify the ξ_k. We record, without proof, the following essentially obvious technical fact.

Lemma 7 *Let $\eta > 0$. There exists a smooth function h of exponential decay at ∞ satisfying*

(a) $\hat{h}(\omega)$ is a smooth function supported in $(-\eta, \eta)$.

(b) $h \geq 0$.

(c) $h(0) = 1$.

(d) $h(x) \leq C(\eta)/(1 + x^2)$ for some positive finite constant $C(\eta)$ and all x.

The mollified functions $\tilde{\xi}_k = \xi_k(t)h(t - t_k)$ again formally satisfy the Lagrange interpolation conditions

$$\tilde{\xi}_k(t_l) = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

However, they also have good decay at ∞, and so belong to $B_1(2\pi + \eta)$.

The interpolation problem (20) at the (t_k) is somewhat more general than the problem (13) at the (u_i), so it is convenient to work with a subsequence of the $(\tilde{\xi}_k)$. By (11), each u_i occurs exactly once in the set $\{t_k\}$. Hence, a one-one mapping $k(i)$ exists so that $t_{k(i)} = u_i$. Define the function

$$\psi_i = \tilde{\xi}_{k(i)} = \xi_{k(i)} \cdot h(\cdot - t_{k(i)}).$$

Then $(\psi_i)_{i=-\infty}^{\infty}$ is a sequence of functions in $B_1(2\pi + \eta)$, satisfying the formal Lagrange interpolation conditions

$$\psi_i(u_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Let $\phi(t) = (1 + t^2)^{-1}$ and $\phi_i(t) = \phi(t - i)$. It turns out that the functions ψ_i are effectively not worse behaved than ϕ_i.

Theorem 4

$$|\psi_i(t)| \le A\phi_i(t) \;\; t \in (-\infty, \infty)$$

where

$$A(R, \Delta, \eta) = \Delta^{-2R-1}\alpha(R, \eta)$$

and $\alpha(R,\eta)$ is a positive, finite constant specified below.

The theorem follows from earlier estimates on G and g_k, on the mollifier h. To apply these estimates, put for short $k = k(i)$. Property (d) of the mollifier h gives $h(\cdot - u_i)/\phi(\cdot - u_i) \le C(\eta)$. Hence

$$\sup_t \frac{|\psi_i(t)|}{|\phi(t-u_i)|} \le C(\eta) \sup_t \frac{|G(t)|}{g_k|t-t_k|^{m_k}}.$$

Now

$$\begin{aligned} \sup_t \frac{|G(t)|}{g_k|t-t_k|^{m_k}} &\le g_k^{-1} \max[\sup_t |G(t)|, \sup_{|t-t_k|\le 1} \frac{|G(t)|}{|t-t_k|^{m_k}}] \\ &\le \Delta^{-2R-1} A_3(R)^{-1} \max[A_1(R), A_2(R)] \\ &= A_5(R)\Delta^{-2R-1}, \end{aligned}$$

say. It follows that

$$|\psi_i| \le C(\eta) A_5(R) \Delta^{-2R-1} \phi(\cdot - u_i).$$

Note that, by (9), $|u_i - i| \le R$. Simple algebra gives

Lemma 8 *If $|\delta| \le R$,*

$$\phi(t \pm \delta) \le A_4(R)\phi(t) \;\; t \in (-\infty, \infty)$$

where

$$A_4(R) = (R+1)^2.$$

It follows that $\phi(\cdot - u_i) \le A_4(R)\phi(\cdot - i)$. Hence

$$|\psi_i| \le \alpha(R,\eta)\Delta^{-2R-1}\phi_i,$$

say, where

$$\alpha(R,\eta) = C(\eta) \cdot A_5(R) \cdot A_4(R).$$

This completes the proof of the theorem.

6 Operator Norms for Interpolation

Let $\mathcal{I}(c) = \sum_{i=-\infty}^{\infty} c_i \psi_i$ denote the Lagrange interpolation operator formally "solving" the interpolation problem (13). We now show that this operator is a bounded linear operator from l_q into $L_q(\mathbf{R})$, and so $\mathcal{I}$ rigorously solves the interpolation problem.

Lemma 9 (1,1)-Boundedness. *Suppose that* $|\psi_i| \leq A\phi_i$, $i \in \mathbf{Z}$. *Then*

$$||\sum_i c_i \psi_i||_{L_1(\mathbf{R})} \leq A \cdot \pi \cdot ||c||_{l_1}.$$

Proof.

$$\begin{aligned} ||\sum_i c_i \psi_i||_{L_1(\mathbf{R})} &\leq \sum_i |c_i| \, ||\psi_i||_{L_1(\mathbf{R})} \\ &\leq \sum_i |c_i| \cdot \sup_i ||\psi_i||_{L_1(\mathbf{R})} \\ &\leq ||c||_{l_1} \, A \, ||\phi||_{L_1(\mathbf{R})} \end{aligned}$$

The result follows from $||\phi||_{L_1(\mathbf{R})} = \int_{-\infty}^{\infty} (1+t^2)^{-1} dt = \pi$.

Lemma 10 (∞,∞)-Boundedness. *Suppose that* $|\psi_i| \leq A\phi_i$, $i \in \mathbf{Z}$. *Then*

$$||\sum_i c_i \psi_i||_{L_\infty(\mathbf{R})} \leq A \left(\frac{2\pi}{1-e^{-2\pi}} \right) ||c||_{l_\infty}.$$

Proof. Using positivity of ϕ,

$$\begin{aligned} ||\sum_i c_i \psi_i||_{L_\infty(\mathbf{R})} &\leq \sup_t \sum_i |c_i| \, |\psi_i(t)| \\ &\leq A \sup_t \sum_i |c_i| \, \phi_i(t) \\ &\leq A \, ||c||_{l_\infty} \sup_t \sum_i \phi_i(t). \end{aligned}$$

Using the Poisson summation formula [34, Page 105], the formula $\hat{\phi}(\lambda) = \pi e^{-|\lambda|}$, and the relation $|\hat{\phi}_i(\omega)| = \hat{\phi}(\omega)$,

$$\sum_i \phi_i(t) = \sum_j \hat{\phi}_i(2\pi j)$$

$$
\begin{aligned}
&\leq \sum_j |\hat{\phi}_i(2\pi j)| = \sum_j \hat{\phi}(2\pi j) \\
&\leq 2\sum_0^\infty \pi e^{-2\pi j} = 2\pi/(1-e^{-2\pi})
\end{aligned}
$$

These lemmas combine to prove that the operator $\mathcal{I}(c) = \sum_i c_i\psi_i$ gives a bounded mapping from l_q into $L_q(\mathbf{R})$ for each $q \in [1,\infty]$. Indeed, with $\theta = 1/q$ we get, from the Riesz-Thorin Interpolation theorem [35, Page 95],

$$
||\sum_i c_i\psi_i||_{L_q(\mathbf{R})} \leq (A\pi)^\theta \left(A\frac{2\pi}{1-e^{-2\pi}}\right)^{1-\theta} ||c||_{l_q}.
$$

Combining these bounds with theorem 4 gives

Theorem 5 *Let $\Omega > 2\pi + \eta$. Then for each $q \in [1,\infty]$ the interpolation problem (13) has a solution in $B_q(\Omega)$ satisfying the interpolation inequality (14) with*

$$
K_q(R,\Delta,\Omega) = \Delta^{-2R-1}\kappa(R,\Omega)
$$

where

$$
\kappa(R,\Omega) = \frac{2\pi}{1-e^{-2\pi}} \cdot \alpha(R,\eta).
$$

7 Proofs for Section 1

7.1 Main Result

Suppose that μ_1, μ_2 are lattice measures with support of Rayleigh index $\leq R$. For a set A, let $A/2$ denote the dilation $\{t : (2t) \in A\}$. Define the measure ν by

$$
\nu(A) = (\mu_1 - \mu_2)(A/2).
$$

Note that ν is supported in the lattice of span $\Delta' = \Delta/2$, and that

$$
\begin{aligned}
u.u.d.(supp(\nu)) &= \frac{1}{2}u.u.d.(supp(\mu_1) \cup supp(\mu_2)) \\
&\leq \frac{1}{2}(u.u.d.(supp(\mu_1)) + u.u.d.(supp(\mu_2))) < 1.
\end{aligned}
$$

Moreover, $R^*(supp(\nu)) \leq R^*(supp(\mu_1)) + R^*(supp(\mu_2))$. Hence, putting $R' = 2R$, ν is an (R', Δ')-measure.

Applying the super-resolution inequality (12), the interpolation inequality (14), and Theorem 5, we have for any $\Omega' > 2\pi$,

$$\begin{aligned} ||\nu||_2 &\leq C_2(R', \Delta', \Omega')||\nu||_{2,\Omega'} \\ &\leq K_2(R', \Delta', \Omega')||\nu||_{2,\Omega'} \\ &\leq (\Delta')^{-4R-1}\, \kappa(R', \Omega')||\nu||_{2,\Omega'}. \end{aligned}$$

Now we observe that $||\nu||_2 = ||\mu_1 - \mu_2||_2$. Also $\hat{\nu}(\omega) = (\hat{\mu}_1 - \hat{\mu}_2)(2\omega)$. Hence, making the particular choice $\Omega' = \Omega/2$ we have

$$\begin{aligned} ||\nu||_{2,\Omega'} &= \sqrt{2}||\mu_1 - \mu_2||_{2,\Omega} \\ &= \frac{1}{\sqrt{\pi}}||\hat{\mu}_1 - \hat{\mu}_2||_{L_2[-\Omega,\Omega]} \end{aligned}$$

Combining these relations, (6) follows once we set

$$\beta(R, \Omega) = 2^{4R+1} \cdot \kappa(2R, \Omega/2)/\sqrt{\pi}.$$

7.2 Proof of Theorem 1

Part a. Let $\mu_1 \neq \mu_2$ be two distinct measures, and s be a point of $S = supp(\mu_1) \cup supp(\mu_2)$ such that $d \equiv \mu_1\{s\} - \mu_2\{s\} \neq 0$. As the μ_i are lattice measures, S is a *separated* set: any two distinct elements of S differ by at least some strictly positive amount. Moreover, $u.u.d.(S) < 2$. Hence by Theorem 1, Page 352, of Beurling [3], for all sufficiently small $\delta > 0$, there exists a function in $B_\infty(2\pi - \delta)$ solving the interpolation problem

$$f(t) = \begin{cases} 0 & t \in S\backslash\{s\} \\ sgn(d) & t = s \end{cases}$$

Pick a mollifier $h \in B_1(\delta)$ satisfying (i) $h \geq 0$, (ii) $h(0) = 1$. Set

$$\psi(t) = f(t)\, h(t - s).$$

Then, by construction,

$$\int \psi\, d(\mu_1 - \mu_2) = |\mu_1\{s\} - \mu_2\{s\}| > 0.$$

On the other hand, as $\psi \in B_1(2\pi)$, $\hat{\psi}(\omega)$ exists, and by Parseval,

$$\int \psi \, d(\mu_1 - \mu_2) = \frac{1}{2\pi} \int_{-2\pi}^{2\pi} \hat{\psi}(\omega)(\hat{\mu}_1(\omega) - \hat{\mu}_2(\omega)) d\omega.$$

As the integrand of the right side is continuous and as the integral on the left side is strictly positive, we conclude that

$$\hat{\mu}_1(\omega) \neq \hat{\mu}_2(\omega) \text{ for some } \omega \in [-2\pi, 2\pi].$$

Part b. This is an application of Beurling's theory of Balayage [2]. Say that S *admits balayage* if the restriction to $[-\Omega, \Omega]$ of any Fourier transform of a finite signed measure ν supported in $\mathbf{R}$ can be represented as the restriction to $[-\Omega, \Omega]$ of the Fourier transform of a finite signed measure ν' supported entirely in S, and if the ratio $||\nu'||_1/||\nu||_1$ is bounded above independently of ν. Beurling shows that a necessary and sufficient condition for S to admit balayage is $l.u.d.(S) > \pi/\Omega$; see Theorem 5, page 346, [2].

To apply this, simply set $\nu = \mu_1$ and $S = S_2$. By assumption, $l.u.d.(S_2) > 1 > \pi/\Omega$ and so we may perform balayage. Let $\mu_2 = \nu'$ be the result. The measure μ_2 is supported on S_2, which by assumption is disjoint from S_1, and yet $\hat{\mu}_1(\omega) = \hat{\mu}_2(\omega)$ for $|\omega| \leq \pi$.

Part c. Let $\Delta = 1/2$ and let $f(\omega)$ be a nonzero, smooth function, periodic of period 4π, vanishing on $[-2\pi + \delta, 2\pi - \delta]$, and satisfying the hermitian symmetry $f(-\omega) = \overline{f}(\omega)$. Set

$$\alpha_k = \frac{\Delta}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} e^{i\omega k \Delta} f(\omega) d\omega$$

and define

$$\mu_1 = \sum_{k \text{ odd}} \alpha_k \delta_{k\Delta}$$

and

$$\mu_2 = -\sum_{k \text{ even}} \alpha_k \delta_{k\Delta}.$$

This defines a pair of finite signed measures; the first supported at the half-integers; the second at the integers. These measures have Fourier transforms and

$$(\hat{\mu}_1(\omega) - \hat{\mu}_2(\omega)) = f(\omega) = 0 \quad |\omega| \leq 2\pi - \delta$$

and so the two measures, although supported disjointly, have Fourier transforms which agree throughout the band $|\omega| \le 2\pi - \delta$. Note that $l.u.d.(supp(\mu_i)) = u.u.d.(supp(\mu_i)) = 1$.

7.3 Proof of Lemma 1

Let $N(y,\epsilon) = N(y,\epsilon; R, \Delta, \Omega) = \{\nu : ||\hat{\nu} - y||_{L_2[-\Omega,\Omega]} \le \epsilon, \nu \in \mathcal{S}(R,\Delta)\}$ denote the set of all lattice measures with Rayleigh index R which are within an ϵ-distance of y. This set contains μ by assumption. Hence it is nonempty. Consider any "feasible reconstruction" rule $\tilde{\mu}(y)$, i.e. any rule with

$$\tilde{\mu}(y) \in N(y,\epsilon).$$

Such a rule selects, from among all measures which could have generated the data, one which satisfies the assumed sparsity. (It is possible to select from the set $N(y,\epsilon)$ so that $\tilde{\mu}(y)$ is a measurable function of y in the topology generated from $L_2[-\Omega,\Omega]$-norm balls).

The triangle inequality implies that any such $\tilde{\mu}$ formally has

$$||\hat{\mu} - \hat{\tilde{\mu}}(y)||_{L_2[-\Omega,\Omega]} \le ||\hat{\mu} - y||_{L_2[-\Omega,\Omega]} + ||y - \hat{\tilde{\mu}}(y)||_{L_2[-\Omega,\Omega]} \le 2 \cdot \epsilon.$$

Hence, by definition of Λ,

$$||\mu - \tilde{\mu}(y)||_2 \;\; \le \;\; \Lambda(2 \cdot \epsilon).$$

The lemma follows.

Remark. The idea of "feasible reconstruction", i.e. of selecting any reconstruction matching known constraints on signal and noise, while of practical value in other contexts [31] is not necessarily practical here, because "projection" onto the set of sparse objects is not a contraction, and so certifiably convergent iterative algorithms are lacking; notwithstanding [23].

7.4 Proof of Theorem 3

Theorem 3 follows by a simple computation. Let $\nu_{r,h} = \sum_{k=0}^{r}(-1)^k C(r,k)\delta_{kh}$, where $C(r,k)$ is the standard combinatorial factor

$$C(r,k) = \frac{r!}{k!(r-k)!}.$$

Then $||\nu_{r,h}||_2 = (\sum_{k=0}^{r} C(r,k)^2)^{1/2}$ independent of h. On the other hand, we recognize that $\nu_{r,h} \star f = D_h^r f$, where D_h^r is the r-th order finite difference operator of span h. From the fact that

$$h^{-r} D_h^r f \to_{L_2(\mathbf{R})} f^{(r)} \quad \text{as} \quad h \to 0,$$

for every smooth f of compact support, we get

$$h^{-2r} \int_{-\Omega}^{\Omega} |\hat{\nu}_{r,h}(\omega)|^2 d\omega \to \int_{-\Omega}^{\Omega} \omega^{2r} d\omega = \frac{\Omega^{2r+1}}{r+1/2}.$$

Hence, as $h \to 0$,

$$||\nu_{r,h}||_{2,\Omega} \sim h^r \frac{\Omega^{r+1/2}}{\sqrt{\pi\,(2r+1)}}. \tag{21}$$

Let now

$$\mu_1 = \eta \sum_{\substack{k \text{ odd} \\ 0 < k < 2R}} C(2R-1,k)\delta_{k\Delta}$$

and

$$\mu_2 = \eta \sum_{\substack{k \text{ even} \\ 0 \le k < 2R}} C(2R-1,k)\delta_{k\Delta}.$$

Then the μ_i belong to $\mathcal{S}(R,\Delta)$ and $\mu_1 - \mu_2 = \eta\, \nu_{2R-1,\Delta}$. Hence if we choose η so that $||\mu_1 - \mu_2||_{2,\Omega} = \epsilon$, then

$$\Lambda(\epsilon) \ge ||\mu_1 - \mu_2||_2. \tag{22}$$

Now, evidently,

$$||\mu_1 - \mu_2||_2 = \frac{||\nu_{2R-1,\Delta}||_2}{||\nu_{2R-1,\Delta}||_{2,\Omega}} ||\mu_1 - \mu_2||_{2,\Omega}.$$

Define

$$b(R, \Omega, \Delta_0) = \inf_{\Delta < \Delta_0} \frac{||\nu_{2R-1,\Delta}||_2}{||\nu_{2R-1,\Delta}||_{2,\Omega}} \Delta^{2R-1}$$

By (21), $b(R, \Omega, \Delta_0) > 0$ and, as $\Delta_0 \to 0$,

$$b(R, \Omega, \Delta_0) \to \frac{\Omega^{2R-1/2}}{\sqrt{\pi (4R-1)}}.$$

Hence if $\Delta < \Delta_0$,

$$||\mu_1 - \mu_2||_2 \geq \Delta^{-2R+1} b(R, \Omega, \Delta_0) ||\mu_1 - \mu_2||_{2,\Omega}$$

and so, by (22) the Theorem follows.

8 Proofs for Section 2

8.1 Proof of Lemma 2

Let $S = supp(\mu) = \{s_k\}$. Partition the line into intervals $I_m = [mR+1/2, (m+1)R+1/2)$, $m = \ldots, -1, 0, 1, \ldots$. In I_m there are by assumption $r_m \leq R$ elements of $supp(\mu)$. We pair up these elements of $S \cap I_m$ with integer elements of I_m from right to left (say). For an integer i paired with an element $s \in S$ in this way, set $u_i = s$. Some integer elements of I_m may remain unpaired after this step; we simply set u_i to be the closest element of the lattice $\{k\Delta\}$ to i which has not been previously assigned to a u_i and which does belong to I_m. In this way we get a bilateral sequence $(u_i)_{i=-\infty}^{\infty}$.

Defining $v_i = 2i - u_i$, gives (8). Since for $i \in I_m$, $u_i \in I_m$, we have $|u_i - i| \leq R$; hence (9). Now any two distinct points in $\{u_i\} \cup \{v_i\}$ are separated by at least the lattice span Δ; hence (10). Also, by our pairing convention, we never assign a u_i to a value which has been previously assigned. Thus all four properties (8)-(11) are verified.

8.2 Proof of Lemma 3

If $f \in B_\infty(\Omega)$ has a zero at t_0, then $g(t) = f(t)\frac{t-s_0}{t-t_0}$ is again in $B_\infty(\Omega)$. Compare, for example, [34, Pages 126-129]. Now, $\sin^2(\pi t) \in B_\infty(2\pi)$; hence we have $G_n \in B_\infty(2\pi)$, for all $n \geq 1$. Moreover, as $\sin^2(\pi z)$ has only real zeros, $G_n(z)$ has only real zeros.

We show that $\{G_n\}$ forms a Normal Family. Indeed, by the Theorem on page 47 of Koosis [13], an entire function of exponential type with only real zeros satisfies, for $\Im(z) > 0$

$$\log|f(z)| = A_+\Im(z) + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\Im(z)}{|z-t|^2}\log|f(t)|dt, \tag{23}$$

where

$$A_+ = \limsup_{y\to\infty} \frac{\log|f(iy)|}{y}.$$

A parallel relation, employing

$$A_- = \limsup_{y\to-\infty} \frac{\log|f(iy)|}{|y|}$$

holds in $\Im(z) < 0$. We note that for $f = G_n$, $A_+ = A_- = 2\pi$. Indeed, $G_n(z) = Q_n(z)\sin^2(\pi z)$, where Q_n is a rational function of degree (n, n) which satisfies

$$Q_n(z) \to 1 \quad \text{as} \quad |z| \to \infty$$

because the numerator and denominator polynomials both have coefficient 1 on the highest-order term z^n. Moreover,

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{|\Im(z)|}{|z-t|^2}dt = 1$$

for each nonreal z, and so the second term on the right side of (23) is never larger than $\log||f||_{L_\infty(\mathbf{R})}$. Combining these facts we have

$$\log|G_n(z)| \leq 2\pi|z| + \log||G_n||_{L_\infty(\mathbf{R})}. \tag{24}$$

The proof of Lemma 3 below shows that $||G_n||_{L_\infty(\mathbf{R})} \leq A_1(R)$ for all n. Hence (24) shows that $\{G_n(z)\}$ is uniformly bounded in each bounded region of the complex plane. Hence, $\{G_n\}$ is a normal family of entire functions. By Montel's Theorem, an entire function G exists as a cluster point of $\{G_n\}$; it must obey, because of (24),

$$\log |G(z)| \leq 2\pi |z| + \log A_1(R)$$

and so G must be entire of exponential type 2π. However, a calculation based on the definition of G_n shows that for each fixed t, $G_n(t)$ converges to a definite limit as $n \to \infty$ so there can only be one cluster point of G_n, and hence the sequence actually has G as a proper limit. The limit function G must satisfy the same relation (23) as the G_n, so G has only real zeros. The following lemma allows us to see that these zeros are at the points t_k and have the same multiplicity as the m_k.

Lemma 11

$$(n!)^4 \cdot \frac{\sin^2(\pi t)}{\prod_{i=-n}^{n}(t-i)^2} \to \pi^2 \quad as \quad n \to \infty$$

uniformly on compact sets of t in $(-\infty, \infty)$.

8.3 Proof of Lemma 3

We now arrive at the key estimates. It is enough to show that for large n, $||G_n||_\infty \leq A_1(R)$, as G is the limit of the G_n. We may assume, without loss of generality, that $t \in [-1/2, 1/2]$ and $n >> R$.

$$|G_n(t)| = \frac{\prod_{i=-n}^{n} |i + \delta_i - t| \, |i - \delta_i - t|}{\prod_{i=-n}^{n} |i - t|^2} \sin^2(\pi t).$$

Now, for $|i| > R$, the inequality of the Arithmetic-Geometric mean gives

$$|i + \delta_i - t| \, |i - \delta_i - t| \leq |i - t|^2.$$

On the other hand, for $|i| \leq R$,

$$|i + \delta_i - t| \, |i - \delta_i - t| \leq (2R + 1)^2.$$

We also note that

$$\sup_{t\in[-1/2,1/2]} \prod_{\substack{|i| \leq R \\ i \neq 0}} |t-i|^{-2} = \prod_{i=1}^{R} |i^2 - 1/4|^{-2}$$

Combining these estimates,

$$\begin{aligned}
|G_n(t)| &\leq \frac{\prod_{i=-R}^{R} |i+\delta_i - t|\,|i-\delta_i - t|}{\prod_{i=-R}^{R} |i-t|^2} \sin^2(\pi t) \\
&\leq ((2R+1)^2)^{2R+1} \sup_{t\in[-1/2,1/2]} \prod_{\substack{|i| \leq R \\ i \neq 0}} |t-i|^{-2} \sup_t \frac{\sin^2(\pi t)}{t^2} \\
&= ((2R+1)^2)^{2R+1} \prod_{i=1}^{R} |i^2 - 1/4|^{-2}\, \pi^2 \\
&= A_1(R).
\end{aligned}$$

8.4 Proof of Lemma 4

Without loss of generality let $k = 0$, let $t_0 \in [-1/2, 1/2)$ and let $i_0 = i_0(t)$ denote an integer closest to t. Let $a_i = 1$ if $u_i \neq t_0$, and 0 otherwise; and let $b_i = 1$ if $v_i \neq t_0$, and 0 otherwise. Pick $n >> 2R$.

$$\frac{|G_n(t)|}{|t-t_0|^{m_0}} = \frac{\prod_{i=-n}^{n} |i+\delta_i - t|^{a_i}\,|i-\delta_i - t|^{b_i}}{\prod_{i=-n}^{n} |i-t|^2} \sin^2(\pi t).$$

Note that if $|i| > 2R$ then necessarily $a_i = b_i = 1$. We again have

$$|i+\delta_i - t|\,|i-\delta_i-t| \leq |i-t|^2 \quad \text{for} \quad |i| > 2R.$$

and so

$$\frac{|G_n(t)|}{|t-t_0|^{m_0}} \leq \frac{\prod_{|i|\leq 2R} |i+\delta_i - t|^{a_i}\,|i-\delta_i - t|^{b_i}}{\prod_{\substack{|i| \leq 2R \\ i \neq i_0}} |t-i|^{-2}} \cdot \frac{\sin^2(\pi t)}{|t-i_0|^2}.$$

Now

$$\sup_{t\in[-1/2,1/2]} \prod_{\substack{|i|\le 2R \\ i\ne i_0}} |t-i|^{-2} \le \prod_{i=1}^{2R-1} |i^2-1/4|^{-2}$$

and also,

$$|i+\delta_i-t|^{a_i}\,|i-\delta_i-t|^{b_i} \le (3R+1)^2, \quad \text{for} \quad |i|\le 2R$$

Hence

$$\begin{aligned} \frac{|G_n(t)|}{|t-t_0|^{m_0}} &\le ((3R+1)^2)^{4R+1} \prod_{i=1}^{2R-1} |i^2-1/4|^{-2}\,\pi^2 \\ &= A_2(R). \end{aligned}$$

8.5 Proof of Lemma 5

Without loss of generality, let $k=0$, and suppose that $|t_0|\le 1/2$. Set

$$G_n(t) = \prod_{-n}^{n}(t-u_i)(t-v_i)\,H_n(t);$$

By Lemma 11, $(n!)^4 H_n(t_0) \to \pi^2$ as $n\to\infty$. Let n be so large that $(n!)^4 H_n(t_0) \ge \pi^2/2$. Then set $x_i = |u_i - t_0|$ and $y_i = |v_i - t_0|$. Define a_i and b_i as in the previous section. The convention $0^0 = 1$ will simplify notation below.

Now

$$|g_{0,n}| = |H_n(t_0)| \cdot \prod_{-n}^{n} x_i^{a_i} y_i^{b_i}$$

and so, by our choice of n,

$$|g_{0,n}| \ge \frac{\pi^2}{2} \cdot \frac{\prod_{-n}^{n} x_i^{a_i} y_i^{b_i}}{\prod_{i=1}^{n} i^4}. \tag{25}$$

We need two estimates. First, for a constant $E_0(R)$,

$$\prod_{i=-R}^{R} x_i^{a_i} y_i^{b_i} \ge E_0(R)\,\Delta^{2R+1}. \tag{26}$$

To see this, note that at most $2R+2$ of the x_i and y_i can be less than $1/2$. Indeed, if $i \neq 0$ at most one member of each pair (u_i, v_i) can be in the interval $[-1,1]$; and if $|i| > R$, neither member of a pair can be in the interval. On the other hand, by hypothesis, at least one of the (x_i) or (y_i) must be zero, as t_0 belongs to $\{u_i\} \cup \{v_i\}$. Therefore, the product on the left side of (26) contains at most $2R+1$ terms of size less than $1/2$. Such terms, by (10), must be at least Δ in size. Hence,

$$\prod_{i=-R}^{R} x_i^{a_i} y_i^{b_i} \geq \Delta^{2R+1}(1/2)^{2R+1}. \tag{27}$$

Thus (26) holds, with $E_0(R) = 2^{-2R-1}$. [Much larger values for E_0 may established with more effort.]

Our second estimate concerns the terms omitted by (26). We wish to show that

$$\prod_{R+1}^{n} \frac{x_i^{a_i} y_i^{b_i} x_{-i}^{a_{-i}} y_{-i}^{b_{-i}}}{i^4} \geq E_1(R) > 0. \tag{28}$$

The key point is that for $|i| > R$, $a_i = b_i = 1$. Now

$$\begin{aligned}
\frac{x_i y_i x_{-i} y_{-i}}{i^4} &= \frac{|i+\delta_i - t_0|\,|i-\delta_i-t_0|\,|-i+\delta_{-i}-t_0|\,|-i-\delta_{-i}-t_0|}{i^4} \\
&= \frac{|(i-t_0)^2-\delta_i^2|\,|(-i-t_0)^2-\delta_{-i}^2|}{i^4} \\
&= |(1-\frac{t_0}{i})^2-(\frac{\delta_i}{i})^2|\,|(1+\frac{t_0}{i})^2-(\frac{\delta_{-i}}{i})^2| \\
&\geq |(1-\frac{1}{2i})^2-(\frac{R}{i})^2|\,|(1+\frac{1}{2i})^2-(\frac{R}{i})^2| \\
&= e_i,
\end{aligned}$$

say. The inequality step is justified by additional calculations, which we omit. Now $e_i \geq [(\frac{2R+1}{2R+2})^2 - (\frac{R}{R+1})^2]^2 > 0$. A little algebra shows that for $i > 2R$, $e_i > (1-(\frac{2R}{i})^2)$. Hence, defining

$$E_1(R) = \prod_{i=R+1}^{\infty} e_i$$

we have

$$E_1(R) \geq \prod_{i=R+1}^{2R} e_i \cdot \prod_{i=2R+1}^{\infty} (1 - (\frac{2R}{i})^2) > 0$$

and (28) holds. Again, this is only a very crude estimate; much larger values of $E_1(R)$ can be established.

Combining the inequalities (25), (26) and (28), We have

$$|g_{0,n}| \geq \Delta^{2R+1} \cdot \frac{\pi^2}{2} \cdot E_0(R) \cdot E_1(R) \cdot (R!)^{-4}$$

which, by a limiting process, yields the sought-for inequality

$$|g_0| \geq \Delta^{2R+1} \cdot A_3(R)$$

with the (very crude) value

$$A_3(R) = \frac{\pi^2}{2} \cdot E_0(R) \cdot E_1(R) \cdot (R!)^{-4}.$$

9 Discussion

A few final remarks may clarify issues raised by the above.

9.1 The Optimal Exponent

Theorems 2 and 3 together suggest that for a certain exponent $e(R)$ we have, for $\Omega > \Omega_0$,

$$E^*(\epsilon, R, \Delta, \Omega) \asymp \Delta^{-e(R)} \cdot Const \cdot \epsilon.$$

If such a relation holds, we must have, by Theorems 2 and 3,

$$2R - 1 \leq e(R) \leq 4R + 1.$$

What is the value of $e(R)$? We suspect that it is roughly $2R$; perhaps exactly $2R + 1$. We believe that the innocent-looking passage from $\mathcal{S}(R, \Delta)$ to (R, Δ)-measures, as given by Lemma 2, introduces an unnecessary factor of 2 slack in our approach. Certainly, we know of many examples of $\mathcal{S}(R, \Delta)$ measures which are in fact $(R/2, \Delta)$-measures.

9.2 Relation to Other Work

A number of papers treat the problem of recovering a signal from data which are missing information about a whole band of frequencies, by exploiting support limitations: compare results in [26, 7, 8, 25].

The closest result in those papers to the present one may be stated as follows [8, Section 6]. Suppose that (x_t) is a discrete-time signal, and that we have noisy information $\hat{y}(\omega) = \hat{x}(\omega) + \hat{z}(\omega)$ about the Fourier transform of x; only now the *high* frequencies $|\omega| \in [\frac{\pi}{m}, \pi]$ are observed, m an integer greater than 1. Then, despite the missing information about the *low* frequency band $[-\frac{\pi}{m}, \frac{\pi}{m}]$, one can stably recover (x_t) – provided that in every interval of length m, a fraction less than $1/\pi$ of the samples are nonzero.

The present paper covers the complementary case, where information for the *low* frequency interval $[-\frac{\pi}{m}, \frac{\pi}{m}]$ would be observed, and information for the *high* frequency band $|\omega| \in [\frac{\pi}{m}, \pi]$ would be missing. For stable recovery of (x_t), Theorem 3 would require that for each interval of length $4mR$, fewer than R samples are nonzero.

This considerably more restrictive sparsity condition expresses the fact that the problem of missing *high* frequencies is much more ill-posed than the problem of missing *low* frequencies.

In another direction, one might compare the interpolation results developed here with other work on interpolation of entire functions. Both in Boas [4] and in Duffin and Schaeffer [9] there is discussion of sets which are uniformly discrete and of uniform density 1, so that conditions (9) and (10) hold. However, the internal symmetry condition (8) seems different from earlier work and figures significantly in the key Lemmas 2, 3, and 4. It would be interesting to know whether this internal symmetry condition is in some sense necessary.

9.3 Removing the Lattice Constraint

We consider it plausible that a more general family of superresolution inequalities holds, in which the lattice constraint $\mu \in \mathcal{L}(\Delta)$ is removed. Such an inequality would be of the form

$$||\nu||_{p,\lambda\Omega} \leq A_p(\lambda, \Omega, R)||\nu||_{p,\Omega}$$

where $\lambda > 1$ is the superresolution factor, and the inequality is suppposed valid for all finite signed measures ν with support of Rayleigh index R. Such an inequality would express superresolution by the fact that the norm on a larger frequency band would be controlled by the norm on a smaller frequency band. For comparison, if $\mu \in \mathcal{L}(\Delta)$, and $\lambda\Omega = \frac{\pi}{\Delta}$, then

$$||\nu||^2_{2,\lambda\Omega} = \frac{1}{2\pi}||\nu_2||^2_2$$

so we get the inequality proved in this paper as a special case. Evidently λ plays the role of Δ^{-1}; it seems plausible that the stability coefficient A_p would grow with λ like $\lambda^{e(R)}$, where $e(R)$ is the optimal exponent in the Lattice case.

References

[1] H. Barkhuisen, R. de Beer, W.M.M.J. Bovée, and D. van Ormondt. Retrieval of Frequencies, Amplitudes, Damping Factors, and Phases from Time-Domain Signals Using a Linear Least-Squares Procedure. *Journal of Magnetic Resonance* **61**, 465-481. 1985.

[2] A. Beurling. IV. Balayage of Fourier-Stieltjes transforms. Mittag-Leffler Lectures on Harmonic Analysis 1977-78. Published in *Collected Works of Arne Beurling: Volume 2 Harmonic Analysis*, pp. 343-350. Birkhauser: Boston. 1989.

[3] A. Beurling. V. Interpolation for an interval on R^1. 1. A density theorem. Mittag-Leffler Lectures on Harmonic Analysis 1977-78. Published in *Collected Works of Arne Beurling: Volume 2 Harmonic Analysis*, pp. 351-359. Birkhauser: Boston. 1989.

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

[5] R.P. Boas, Jr. *Entire Functions.* Academic Press, New York. 1952.

[6] D.L. Donoho. Statistical Estimation and Optimal Recovery. Technical Report 213, Department of Statistics, University of California at Berkeley.

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

[8] D.L. Donoho and B.F. Logan. Signal recovery and the large sieve. To appear, *SIAM J. Appl. Math.*, 1991.

[9] R.J. Duffin and A.C. Schaeffer. Power series with bounded coefficients. *Am. J. Math.* **67**, 141-154. 1945.

[10] E.G. Steward. *Fourier Optics: an introduction* J. Wiley and Sons: New York. 1983.

[11] J.A. Högbom. Aperture Synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.*, **15**, 417-426. 1974.

[12] A.S.B. Holland. *Introduction to the Theory of Entire Functions.* Academic Press: New York. 1973.

[13] Paul Koosis. *The logarithmic integral.* Cambridge University Press. 1988.

[14] S. Kawata, K. Minami, and S. Minami. Superresolution of Fourier transform spectroscopy data by the maximum entropy method. *Applied Optics* **22**, 3593-3606. 1983.

[15] 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.

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

[17] 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.

[18] R.J. Mammone. Spectral Extrapolation of Constrained Signals. *J. Opt. Soc. Am.* **73**, 1476.

[19] A.R. Mazzeo, M.A. Delsuc, A. Kumar, and G.C. Levy. Generalized maximum entropy deconvolution of spectral segments. *Journ. Magnetic Resonance* **81**, 512-519, 1989.

[20] K. Minami, S. Kawata, and S. Minami. Superresolution of Fourier Transform spectra by autoregressive model fitting with singular value decomposition. *Applied Optics* **24**, 162-167. 1985.

[21] C.A. Micchelli and T.J. Rivlin. A survey of optimal recovery. pp. 1-54 in *Optimal Estimation in Approximation Theory* (C.A. Micchelli and T.J. Rivlin., Eds.) Plenum Press, N.Y. 1977.

[22] R.H. Newman. Maximization of Entropy and Minimization of Area as criteria for NMR Signal Processing. *Journal of Magnetic Resonance* **79**, 448-460.

[23] A. Papoulis and C. Chamzas. Detection of Hidden Periodicities by Adaptive Extrapolation. *IEEE Trans. Acoustics, Speech, and Signal Processing.* **ASSP-27**, 492-500. 1979.

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

[25] E.R. Pike, J.G. McWhirter, M. Bertero, and C. De Mol. Generalized information theory for inverse problems in signal processing. *IEE Proceedings* **131**, 660, 1984.

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

[27] U.J. Schwarz. Mathematical-statistical description of the iterative beam-removing technique (Method CLEAN). *Astronomy and Astrophysics* **65**, 345-356.

[28] J. Tang and J.R. Norris. LP-ZOOM, a linear-prediction method for local spectral analysis of NMR signals. *Journal of Magnetic Resonance* **79**, 190-196, 1988.

[29] J.H. Traub, G. Wasilkowski and H. Woźniakowski. *Information, Uncertainty, Complexity.* Academic Press. 1981.

[30] J.H. Traub, G. Wasilkowski and H. Woźniakowski. *Information-Based Complexity.* Academic Press. 1988.

[31] H.J. Trussel and M.R. Civanlar. The feasible solution in signal restoration. *IEEE Trans. Acoustics, Speech, and Signal Processing* **ASSP-32**, 201-212. 1984.

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

[33] K. Wang. Ph. D. Thesis, University of California, Berkeley.

[34] R.M. Young. *An Introduction to Nonharmonic Fourier Series.* Academic Press: New York. 1980.

[35] A. Zygmund. *Trigonometric series.* Vol. II. Cambridge University Press. Second edition. 1977.