Optimal Sub-Nyquist Nonuniform Sampling and Reconstruction for Multiband Signals

Raman Venkataramani and Yoram Bresler, *Fellow, IEEE*

***Abstract*—We study the problem of optimal sub-Nyquist sampling for perfect reconstruction of multiband signals. The signals are assumed to have a known spectral support $\mathcal{F}$ that does not tile under translation. Such signals admit perfect reconstruction from periodic nonuniform sampling at rates approaching Landau's lower bound equal to the measure of $\mathcal{F}$. For signals with sparse $\mathcal{F}$, this rate can be much smaller than the Nyquist rate. Unfortunately, the reduced sampling rates afforded by this scheme can be accompanied by increased error sensitivity. In a recent study, we derived bounds on the error due to mismodeling and sample additive noise. Adopting these bounds as performance measures, we consider the problems of optimizing the reconstruction sections of the system, choosing the optimal *base sampling rate*, and designing the nonuniform *sampling pattern*. We find that optimizing these parameters can improve system performance significantly. Furthermore, uniform sampling is optimal for signals with $\mathcal{F}$ that tiles under translation. For signals with nontiling $\mathcal{F}$, which are not amenable to efficient uniform sampling, the results reveal increased error sensitivities with sub-Nyquist sampling. However, these can be controlled by optimal design, demonstrating the potential for practical multifold reductions in sampling rate.**

***Index Terms*—Error bounds, Landau–Nyquist rate, matrix inequalities, multiband, nonuniform periodic sampling, optimal sampling and reconstruction.**

I. Introduction

THERE has been a long history of research [1]–[4] devoted to sampling theory, with perhaps the most fundamental and important piece of work in this area being the classical sampling theorem. Also known as the Whittaker–Kotelńikov–Shannon (WKS) theorem, it states that a lowpass signal bandlimited to the frequencies $(-f_0, f_0)$ can be reconstructed perfectly from its samples taken uniformly at no less than the Nyquist rate of $2f_0$ [5]. Another important result in sampling theory due to Landau is a lower bound on the sampling density required for any sampling scheme that allows perfect reconstruction [6]. For multiband signals, this fundamental lower bound is given by the total length (measure) of support of the Fourier transform of the signal. Landau's bound applies to an arbitrarily sampling scheme: uniform or not, and the minimum rate is not necessarily achievable except asymptotically. Landau's bound is often much lower than the corresponding Nyquist rate. This motivates the study of *sub-Nyquist sampling* of multiband signals and their perfect reconstruction, cf. [7]–[14]

From a practical viewpoint, sub-Nyquist sampling is very important in several Fourier imaging applications such as sensor array imaging, synthetic aperture radar (SAR), and magnetic resonance imaging (MRI), where the physics of the problem provides us samples of the unknown sparse object in its Fourier domain [15]–[18]. Our objective, then, is to reconstruct the object from the Fourier data. It is often expensive or physically impossible to collect many samples, and it becomes necessary to sample minimally and exploit the sparsity (i.e., multiband structure) in the object to form its image. These problems are, of course, duals to the problem considered here since the sparsity is in the spatial domain and sparse sampling in the frequency domain.

For a given signal $x(t)$, its *spectral support* $\mathcal{F}$ is defined as the set of frequencies where the Fourier transform $X(f)$ does not vanish, and the *spectral span* $[\mathcal{F}]$ is defined as the smallest interval containing $\mathcal{F}$. We consider here only spectral supports that can be expressed as a finite union of finite intervals called *bands*. The set of multiband signals bandlimited to $\mathcal{F}$ is denoted by $\mathcal{B}(\mathcal{F})$. Landau's lower bound for these signals is $\lambda(\mathcal{F})$, where $\lambda(\cdot)$ is the Lebesgue measure. However, in general, the Nyquist rate $f_{\rm nyq}$ for sampling $x(t) \in \mathcal{B}(\mathcal{F})$ without aliasing (overlap between translates of $\mathcal{F}$ by multiples of $f_{\rm nyq}$) is equal to the width of its spectral span $f_{\rm nyq} = [\mathcal{F}]$. Hence, for multiband signals with sparse spectral supports $[\mathcal{F}]$, the Nyquist rate can be much larger than the lower bound $\lambda(\mathcal{F})$.

A favorable case is when the widths of the bands and the gaps between them satisfy special relationships so that there is no overlap between uniform translates of $\mathcal{F}$ by multiples of a quantity $f_0 < \lambda([\mathcal{F}])$. In these cases, when the spectral support is *packable*, $f_{\rm nyq} = f_0 < \lambda([\mathcal{F}])$. The most favorable situation of these is when $\mathcal{F}$ tiles the real line under uniform translations, i.e., is packable without gaps, or "$\mathcal{F}$ is an explosion of the interval" [4]. In this (very special) case, $f_{\rm nyq} = \lambda(\mathcal{F})$, i.e., Landau's lower bound is achievable by uniform sampling.

Instead, the case of interest to us in this paper is the general case, with $\lambda(\mathcal{F}) < f_{\rm nyq} \leq \lambda([\mathcal{F}])$. Without loss of generality, we focus on the extreme (worst) case of *nonpackable* $\mathcal{F}$, such that $f_{\rm nyq} = \lambda([\mathcal{F}])$.[1] For such multiband signals, it has been shown that perfect reconstruction is possible from *nonuniformly* spaced samples taken at a sub-Nyquist average rate ap-

Manuscript received January 11, 2001; revised June 14, 2001. This work was supported in part by the Joint Services Electronic Program under Grant N00014-96-1-0129, the National Science Foundation under Grant MIP 97-07633, and DARPA under Contract F49620-98-1-0498. The associate editor coordinating the review of this paper and approving it for publication was Dr. Olivier Cappe.

The authors are with Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail raman@ifp.uiuc.edu; ybresler@uiuc.edu).

Publisher Item Identifier S 1053-587X(01)07767-4.

[1]Intermediate cases with $\lambda(\mathcal{F}) < f_{\rm nyq} < [\mathcal{F}]$ are reduced to this case by first sampling the signal at $f_{\rm nyq}$ and then considering the problem of further downsampling the discrete-time signal, which now has a nonpackable spectral support.

proaching the Landau lower bound [8]–[14].[2] In fact, the work in [19]–[22] typically addresses signals with lowpass spectral supports, which, being packable, are best sampled uniformly, as we show in this paper.

We note that the related problem of perfect reconstruction in filter banks (cf. [23]) is fundamentally different from the problem considered here. In the filterbank work, all samples at the Nyquist rate are assumed available, and the analysis stage can be designed together with the synthesis stage. In contrast, in our problem, Nyquist-rate acquisition is too expensive or even impossible, and only the minimum number of samples of the continuous time signal is acquired. As shown later, the filterbank interpretation for this is that the analysis filters are restricted to the form of z^{-k}, where $k \in \mathbb{Z}$.

Given the obvious advantages of such reduced sampling rates (by, e.g., a factor of 10 in one of the examples in this paper), one would expect extensive use and applications of these methods. However, a very high sensitivity to errors has been observed in some cases [9], [12]. In fact, it turns out that unless the sampling and reconstruction system is very carefully designed and optimized, the sensitivity to small errors can be so great that although perfect reconstruction is possible with perfect data, the signal is corrupted beyond recognition in most practical situations.

The goal of this paper is to explore these limitations and develop systematic design methods to minimize the sampling rate and, at the same time, minimize the error sensitivity of the system. This will provide the necessary tools for practical applications of minimum-rate sub-Nyquist sampling.

We consider the problem of periodic nonuniform sampling and reconstruction of multiband signals. We focus primarily on the results presented in [14], where we derived an explicit reconstruction formula and a multirate realization for a sampling scheme called *multicoset sampling* that allows us to approach the Landau minimum rate arbitrarily closely. As an important tool for systematic feasibility evaluation and design of the system, we derived estimates of the error resulting from signal mismodeling. More precisely, we derived bounds on a) the peak value and energy of the aliasing error resulting when the input signal $x(t)$ lies in $\mathcal{B}([\mathcal{F}])$, rather than $\mathcal{B}(\mathcal{F})$, and b) the output noise variance when the input samples contain additive white noise of variance σ^2.

In this paper, we use these various bounds to optimize the performance of the sub-Nyquist sampling and reconstruction system by minimizing the sensitivity bounds. It turns out that the reconstruction system that provides perfect reconstruction of signals $x(t) \in \mathcal{B}(\mathcal{F})$ in the modeled class has free parameters, which can be chosen to optimize the sensitivity bounds. We present closed-form or otherwise efficient approximation numerical algorithms to solve these optimization problems. Likewise, we use the bounds to determine the best sampling pattern among all patterns that achieve a given sampling rate for a given $\mathcal{F}$.

[2]This discussion and intentional use of nonuniform sampling are fundamentally distinct from other extensive recent work on nonuniform sampling in which it is usually regarded as a "necessary evil" imposed by sampling jitter or other physical limitations [19]–[22].

In addition, we solve the problem of an optimal choice of the base sampling frequency to minimize the average sampling rate achievable by a design with a given sampling period L. This allows us to minimize the sampling rate for a given system complexity, rather than asymptotically, with $L \to \infty$. This problem is related to the problem of pairing band edges of $\mathcal{F}$ [13]. We provide a simple algorithm to solve the problem, whether or not Landau's lower bound is attainable for the particular $\mathcal{F}$ and choice of L.

We derive additional relationships and bounds that allow us to quantify the performance loss in terms of increased error sensitivity due to nonpackability of the spectral support and compare uniform and nonuniform sampling patterns for packable spectra. Not surprisingly, perhaps, we find that uniform sampling is more suitable for packable spectra. Most importantly, however, we find that the sensitivity penalty for sub-Nyquist sampling of signals with nonpackable spectra can be controlled by optimal design and by backing off slightly from the minimum rate. The resulting low error sensitivities with multifold reductions from the Nyquist rate in our numerical examples suggest that these techniques have considerable practical potential.

II. Multicoset Sampling

A. Signals and Data

Let the class of continuous complex-valued signals of finite energy, bandlimited to a subset $\mathcal{F}$ of the real line (consisting of a finite union of bounded intervals), be denoted by $\mathcal{B}(\mathcal{F})$

$$\mathcal{B}(\mathcal{F}) = \{x(t) \in L^2(\mathbb{R}) \cap C(\mathbb{R})\colon X(f) = 0,\ f \notin \mathcal{F}\}$$

where

$$\mathcal{F} = \bigcup_{i=1}^{n} [a_i, b_i) \quad \text{and} \quad X(f) = \int_{\mathbb{R}} x(t) \exp(-j2\pi f t)\, dt. \tag{1}$$

The *span of* $\mathcal{F}$, which is denoted by $[\mathcal{F}]$, is the convex hull of $\mathcal{F}$, i.e., the smallest interval containing $\mathcal{F}$.

Definition 1: The spectral support $\mathcal{F}$ is said to be packable at rate f_0 if $\mathcal{F} \cap (\mathcal{F} \oplus nf_0) = \emptyset, \forall\, n \in \mathbb{Z}\backslash\{0\}$, where $\oplus$ is the translation operator defined by $f \oplus \mathcal{F} = \{f + x\colon x \in \mathcal{S}\}$ for any real set $\mathcal{F}$ and $f \in \mathbb{R}$.

In other words, $\mathcal{F}$ is packable at rate f_0 if signals with spectral support $\mathcal{F}$ can be sampled uniformly at rate f_0 without inducing aliasing. Hence, $\mathcal{F}$ is always packable at rate $\lambda([\mathcal{F}])$. We call $\mathcal{F}$ *nonpackable* if it is not packable at any rate smaller that $\lambda([\mathcal{F}])$. We assume that $\mathcal{F}$ is nonpackable and that $\inf \mathcal{F} = a_1 = 0$ and $\sup \mathcal{F} = b_n$. There is no loss of generality because any signal spectrum whose span $[\mathcal{F}]$ is known can be shifted to the origin by multiplication of the signal in the time domain by a suitable complex exponential. Since multiplication and sampling in the time domain commute, we are justified in making the assumption.

We now describe multicoset sampling. Given a bandlimited signal $x(t) \in \mathcal{B}(\mathcal{F})$, we obtain its samples taken on a periodic nonuniform grid consisting of the sampling locations $(nL + c_i)T$ for $n \in \mathbb{Z}$ and $i = 1, \ldots, p$, where $\{c_i\}$ is a set of p distinct integers in the set $\mathcal{L} = \{0, 1, \ldots, L-1\}$, and $1/T$

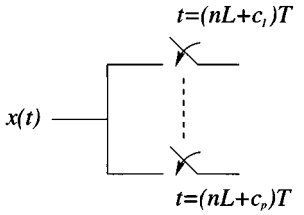

Fig. 1. Model for multicoset sampling.

is the "base frequency," which is at least equal to the Nyquist rate for $x(t)$: $1/T \geq \lambda([\mathcal{F}])$. This is illustrated in Fig. 1. In Section IV, we address the problems of selecting the optimal base frequency $1/T$ and sampling pattern $\{c_i\}$.

B. Definitions and Notation

This section is largely a collection of definitions and notations needed to describe the various error bounds derived in [14]. For any real set $\mathcal{S} \in \mathbb{R}$, denote its indicator function by

$$\chi(f \in \mathcal{S}) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } f \in \mathcal{S} \\ 0, & \text{otherwise} \end{cases}$$

and for the given spectral support $\mathcal{F} = \bigcup_i [a_i, b_i)$, define a finite set Γ as

$$\Gamma \stackrel{\text{def}}{=} \left\{ a_i - \frac{\lfloor LTa_i \rfloor}{LT} : 1 \leq i \leq n \right\} \cup \left\{ b_i - \frac{\lfloor LTb_i \rfloor}{LT} : 1 \leq i \leq n \right\} \tag{2}$$

where $\lfloor \cdot \rfloor$ is the floor function. Suppose we write $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_M\}$ as a set of $M \leq 2n$ elements of Γ arranged in increasing order; then, $\gamma_1 = 0$ as a consequence of $a_1 = 0$. Furthermore, defining $\gamma_{M+1} \stackrel{\text{def}}{=} 1/(LT)$, we see that $0 = \gamma_1 < \gamma_2 < \cdots < \gamma_{M+1} = (LT)^{-1}$, and a collection of intervals $\{\mathcal{G}_m\}$ that partitions the set $\mathcal{F}_0 = [0, 1/(LT))$ is given by

$$\mathcal{G}_m = [\gamma_m, \gamma_{m+1}), \qquad 1 \leq m \leq M.$$

The reason we partition $\mathcal{F}_0$ in this manner, as discussed in [14], is so that $\chi(f \in \mathcal{F})$ is constant (either 0 or 1) for $f \in \mathcal{G}_m \oplus r/(LT)$, and each pair of indices m and r. In other words, each "subcell" of the form $\mathcal{G}_m \oplus r/(LT)$, for $l \in \mathcal{L}$ and $1 \leq m \leq M$, is either disjoint from or fully contained in $\mathcal{F}$. Now, define *spectral index sets* $\mathcal{K}_m$ and their complements $\overline{\mathcal{K}}_m$ for $m \in \{1, 2, \ldots, M\}$ as follows:

$$\mathcal{K}_m \stackrel{\text{def}}{=} \left\{ r \in \mathcal{L} : \frac{r}{LT} \oplus \mathcal{G}_m \subset \mathcal{F} \right\}, \qquad \overline{\mathcal{K}}_m = \mathcal{L} \backslash \mathcal{K}_m.$$

The set $\mathcal{K}_m$ contains the indices of subcells in the collection $\{r/(LT) \oplus \mathcal{G}_m : r \in \mathcal{L}\}$ that are contained in $\mathcal{F}$. The following example illustrates the construction of these sets.

Example 1: Suppose we want to design sampling patterns for a class of bandlimited signals $\mathcal{B}(\mathcal{F})$ with

$$\mathcal{F} = \left[0, \frac{21}{400}\right) \cup \bigcup_{m=1}^{19} \left[\frac{21m}{400}, \frac{21m}{400} + \frac{1}{400}\right)$$

which is nonpackable with $[\mathcal{F}] = [0, 1)$ and $\Omega = 0.1$. Hence, the Nyquist rate equals $\lambda([\mathcal{F}]) = 1$, and the Landau minimum rate is $\lambda(\mathcal{F}) = 0.1$. For the choice $L = 20$, we find that $M = 19$ and $\Gamma = \{m/200 : m = 0, \ldots, 19\}$. Hence, the partitions of $\mathcal{F}_0 = [0, 1/20)$ and the spectral index sets are given by

$$\mathcal{G}_m = \begin{cases} \left[0, \frac{2}{400}\right), & \text{if } m = 1 \\ \left[\frac{m}{400}, \frac{m+1}{400}\right), & 2 \leq m \leq 19 \end{cases}$$

and

$$\mathcal{K}_m = \begin{cases} \{0, 1\}, & \text{if } m = 1 \\ \{0, m\}, & 2 \leq m \leq 19. \end{cases}$$

Denote the number of elements of $\mathcal{K}_m$ by $q_m = |\mathcal{K}_m|$. Then, $|\overline{\mathcal{K}}_m| = L - q_m$. Observe that

$$q(f) = q_m, \qquad \forall f \in \mathcal{G}_m$$

where

$$q(f) \stackrel{\text{def}}{=} \sum_{r=0}^{L-1} \chi\left(f + \frac{r}{LT} \in \mathcal{F}\right), \qquad f \in \mathcal{F}_0. \tag{3}$$

Next, define the following matrices for each m:

$$\begin{aligned} \boldsymbol{A}_m &= \boldsymbol{W}_L(\mathcal{C}, \mathcal{K}_m) \quad \text{of size: } p \times q_m \\ \boldsymbol{B}_m &= \boldsymbol{W}_L(\mathcal{C}, \overline{\mathcal{K}}_m) \quad \text{of size: } p \times (L - q_m) \end{aligned} \tag{4}$$

where $\boldsymbol{W}_L$ is the $L \times L$ unitary DFT matrix whose (m, n) entry is $\boldsymbol{W}_{L,mn} = 1/\sqrt{L} \exp(j2\pi mn/L)$, and $\boldsymbol{W}_L(\mathcal{C}, \mathcal{K})$ denotes the submatrix of $\boldsymbol{W}_L$ obtained by selecting its rows indexed by $\mathcal{C}$ and columns by $\mathcal{K}$. Observe that $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ satisfy

$$\boldsymbol{A}_m \boldsymbol{A}_m^* + \boldsymbol{B}_m \boldsymbol{B}_m^* = \boldsymbol{I}. \tag{5}$$

These matrices play in important role in the error bounds, as we will see later. A necessary and sufficient condition for reconstruction of every signal $x \in \mathcal{B}(\mathcal{F})$ is the existence of left inverses $\boldsymbol{A}_m^L$ for each $\boldsymbol{A}_m$ such that $\boldsymbol{A}_m^L \boldsymbol{A}_m = \boldsymbol{I}$. This, in turn, requires that $\boldsymbol{A}_m$ have full rank for each $1 \leq m \leq M$. This motivates the definition of two notions of goodness that characterize sampling patterns.

Definition 2: Given an index set $\mathcal{K}$ with $|\mathcal{K}| = q \leq p$, we call $\mathcal{C}$ a $\mathcal{K}$-reconstructive sampling pattern if the matrix $\boldsymbol{W}_L(\mathcal{C}, \mathcal{K})$ has full row rank.

Definition 3: A pattern $\mathcal{C}$ with $|\mathcal{C}| = p \geq q$ is (p, q) universal if the matrix $\boldsymbol{W}_L(\mathcal{C}, \mathcal{K})$ has full row rank for every index set $\mathcal{K}$ of q elements, i.e., whenever $|\mathcal{K}| = q$. A (p, p)-universal pattern is simply called *universal*.

The existence of a universal pattern is demonstrated by the *bunched sampling pattern* $\mathcal{C} = \{0, 1, \ldots, p-1\}$. It is universal because the resulting matrix $\boldsymbol{W}_L(\mathcal{C}, \mathcal{K})$ has a Vandermonde structure for any $|\mathcal{K}| = p$. A (p, q) universal pattern is $\mathcal{K}$-reconstructive for every $\mathcal{K}$ with $|\mathcal{K}| \leq q$ so that the second definition is stronger than the first for fixed p and q. Therefore, $\mathcal{C}$ has to be $\mathcal{K}_m$-reconstructive for each m [14] for perfect reconstruction. However, for simplicity, we assume in the rest of the paper that $\mathcal{C}$ is universal with $p \geq \max_m q_m$. This automatically satisfies the reconstruction condition. In view of (3), this

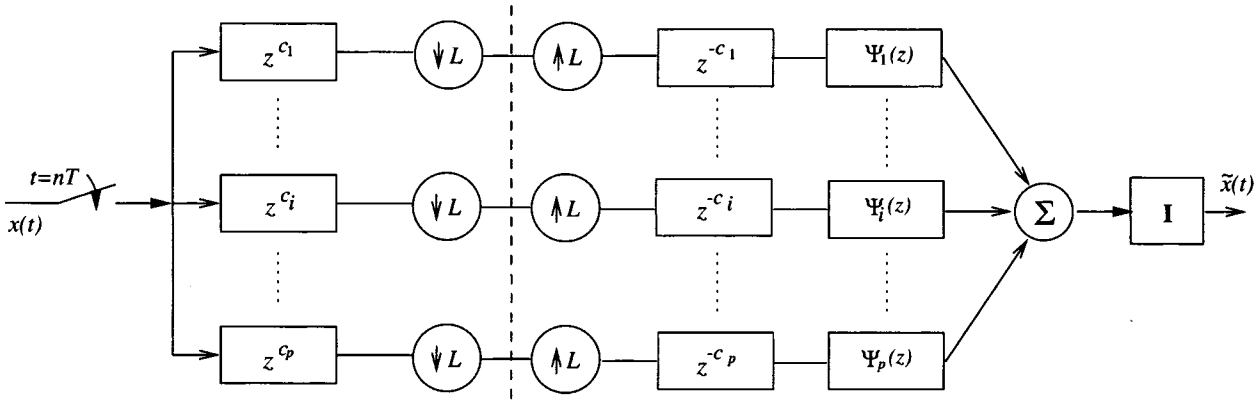

Fig. 2. Multicoset sampling and reconstruction. The block "I" is an ideal sinc interpolator.

condition reduces to $p \geq \max_f q(f)$. Hence, the average sampling rate f_{avg} satisfies

$$f_{\text{avg}} = \frac{p}{LT} \geq \frac{1}{LT} \max_{f \in \mathcal{F}_0} \sum_{r=0}^{L-1} \chi\left(f + \frac{r}{LT} \in \mathcal{F}\right). \quad (6)$$

The equality in (6) is achieved for universal sampling patterns $\mathcal{C}$ with $p = \max_m q_m$. The condition $p \geq \max_m q_m$ guarantees that $\boldsymbol{A}_m$ has full column rank for each m. Under this condition, we obtain the following explicit reconstruction formula for $x(t)$ from its multicoset samples:

$$x(t) = \sum_{i=1}^{p} \sum_{j=-\infty}^{\infty} x((c_i + Lj)T)\phi_i(t - (c_i + Lj)T) \quad (7)$$

where the functions $\phi_i(t)$, $i = 1, \ldots, p$ are the (nonunique) interpolation filters. These filters can be parameterized in terms of matrices $\boldsymbol{A}_m^L$ of size $q_m \times p$ and $\boldsymbol{C}_m$ of size $(L - q_m) \times p$, which are defined by

$$\boldsymbol{A}_m^L \boldsymbol{A}_m = \boldsymbol{I}_{q_m} \quad \text{and} \quad \boldsymbol{C}_m \boldsymbol{A}_m = \boldsymbol{0}. \quad (8)$$

It is clear that these matrices are nonunique if $p > q_m$, and this reflects the nonuniqueness of the interpolation filters. A multirate realization of the reconstruction scheme is illustrated in Fig. 2. The analysis part, to the left of the broken line, is a model for the multicoset sampling process of the continuous-time signal (see Fig. 1). In other words, the simple structure of the analysis part of the filterbank is dictated by the assumption that only samples on the multicoset grid are available, whereas the synthesis part has a fully general structure. The digital filters $\Psi_i(z)$, $i = 1, \ldots, p$ on the synthesis side are related to the interpolation filters by

$$\Psi_i(z) = \sum_{n \in \mathbb{Z}} \phi_i(nT) z^{-n}.$$

See [14] for the exact expressions for these interpolation filters in terms of $\boldsymbol{A}_m^L$ and $\boldsymbol{C}_m$.

We summarize the bounds below as they will be the basis for all the results in this paper. All our error bounds are expressed in terms of $\boldsymbol{A}_m^L$, $\boldsymbol{C}_m$, and $\boldsymbol{S}_m$ defined as

$$\boldsymbol{S}_m = \begin{pmatrix} \boldsymbol{A}_m^L \boldsymbol{B}_m \\ \boldsymbol{C}_m \boldsymbol{B}_m - \boldsymbol{I}_{q_m} \end{pmatrix}. \quad (9)$$

C. Error Bounds

Assume that $x(t)$ lies in $\mathcal{B}([\mathcal{F}])$ rather than in the class of signals for which it was designed, namely, $\mathcal{B}(\mathcal{F})$. Thus, $x(t)$ has out-of-band frequency components. This causes the reconstructed signal $\tilde{x}(t)$ to be in error, which, for simplicity, we call the *aliasing error*. The following are some error bounds derived in [14]. The peak value of $e(t)$ is bounded as

$$\sup_t |e(t)| \leq \psi_\infty \int_{[\mathcal{F}]\backslash\mathcal{F}} |X(f)|\, df$$

where

$$\psi_\infty = \max_m \|\boldsymbol{S}_m\|_1 \quad (10)$$

while the energy of $e(t)$ and an upper bound on it are given by

$$\int_{\mathbb{R}} |e(t)|^2\, dt = \sum_{m=1}^{M} \int_{\mathcal{G}_m} (\boldsymbol{x}_m^-(f))^* (\boldsymbol{S}_m^* \boldsymbol{S}_m) \boldsymbol{x}_m^-(f)\, df \quad (11)$$

$$\int_{\mathbb{R}} |e(t)|^2\, dt \leq \psi_2^2 \int_{[\mathcal{F}]\backslash\mathcal{F}} |X(f)|^2\, df, \quad \psi_2 = \max_m \|\boldsymbol{S}_m\|_2 \quad (12)$$

where $\boldsymbol{x}_m^-(f)$, which is defined for $f \in \mathcal{G}_m$, are vectors containing the out-of-band signal components

$$\boldsymbol{x}_m^-(f) = \left\{ X\left(f + \frac{k}{LT}\right) : \ k \in \overline{\mathcal{K}}_m \right\}, \qquad f \in \mathcal{G}_m.$$

Suppose the input samples $x(nT)$ contain additive white noise $w(n)$ with variance σ^2 representing, for instance, sensor noise or quantization error associated with sampling. Then, the corresponding output noise has average power equal to

$$\langle E|\tilde{w}(t)|^2 \rangle_t = \sigma^2 \psi_{\text{n}}$$

$$\psi_{\text{n}} = T \sum_{m=1}^{M} \lambda(\mathcal{G}_m) \left(\|\boldsymbol{A}_m^L\|_F^2 + \|\boldsymbol{C}_m\|_F^2 \right). \quad (13)$$

The bounds (12) and (13) are tight because there exist nonzero input signals that satisfy the bounds with equality; however, (10) need not be tight. All three bounds depend only on a measure of the out-of-band signal content, which is either the magnitute of the spectrum $|X(f)|$ or its square integrated over the out-of-band region $[\mathcal{F}]\backslash\mathcal{F}$.

III. Optimal Reconstruction

The multiplying constants ψ_∞, ψ_2 in the bounds (10) and (12) are nondecreasing functions of the 1- and 2-norms, respectively, of the matrix $\boldsymbol{S}_m$ for each m. Hence, our problem is to make $\boldsymbol{S}_m$ have the smallest norm possible for each m, and this is clearly a collection of mutually independent problems. Therefore, for the sake of readability, we drop the index m everywhere from now on with the understanding that the solution needs to be applied to each m. Given sampling pattern $\mathcal{C}$ and a spectral index set $\mathcal{K}$, our objective is to pick appropriate matrices $\boldsymbol{A}^L$ and $\boldsymbol{C}$ that satisfy $\boldsymbol{A}^L\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{CA} = \boldsymbol{0}$ [viz. (8)] and minimize the norm of

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{A}^L\boldsymbol{B} \\ \boldsymbol{CB} - \boldsymbol{I} \end{pmatrix} \tag{14}$$

where "norm" means either $||\cdot||_2$ (spectral norm) or $||\cdot||_1$ (maximum-column-sum norm). The other possibility is to minimize the output noise power in (13). A close look at this expression reveals that we need to minimize (for each fixed m) the quantity $||\boldsymbol{A}^L||_F^2 + ||\boldsymbol{C}||_F^2$ over all valid matrices $\boldsymbol{A}^L$ and $\boldsymbol{C}$.

Note that if $p = q$, the matrix $\boldsymbol{A}$ is square, and hence, the left-inverse $\boldsymbol{A}^L$ is unique. In addition, the only matrix $\boldsymbol{C}$ that would satisfy $\boldsymbol{CA} = \boldsymbol{0}$ is trivial in this case, namely, $\boldsymbol{C} = \boldsymbol{0}$. In other words, there are no free parameters when $p = q$. Therefore, the reconstruction system only needs to be optimized when $p > q$. We assume that $p > q_m$ in the rest of the section. The other point to note is that the optimization needs to be carried out for each value of the index m, the subscript of which we have chosen to omit. We will see in a moment that the selection of the best $\boldsymbol{A}$ and $\boldsymbol{C}$ to minimize a) the spectral norm of $\boldsymbol{S}$ and b) the output noise power in (13) can be solved analytically. Minimizing the quantity $||\boldsymbol{S}||_1$ is a little harder, however, requiring the use of numerical methods.

A. Minimizing the Aliasing Error Energy

The following lemmas (whose proofs can be found in the Appendix) address the problem of minimizing the bounding constant ψ_2 for the aliasing error energy (12).

Lemma 1: The solution $\boldsymbol{X}_\star$ to the problem $\min_{\boldsymbol{X}} ||\boldsymbol{R} + \boldsymbol{XQ}||_2$, for given $\boldsymbol{R}$ and $\boldsymbol{Q}$ with compatible dimensions, is $\boldsymbol{X}_\star = -\boldsymbol{RQ}^\dagger$, where $\boldsymbol{Q}^\dagger$ is the pseudo-inverse of $\boldsymbol{Q}$. Furthermore, $\boldsymbol{X}_\star$ simultaneously minimizes *all* singular values of $\boldsymbol{R} + \boldsymbol{XQ}$.

Lemma 2: Let $(\boldsymbol{A}\ \boldsymbol{B})$ be a $p \times L$ submatrix of the $L \times L$ DFT matrix $\boldsymbol{W}$ whose entries are $\boldsymbol{W}_{mn} = \exp(-j2\pi mn/L)$, with possible rearrangements of the columns. Suppose $\boldsymbol{A}$ with $q \leq p$ columns has full rank. Then, the minimization of the largest singular value of $\boldsymbol{S}$

$$\min ||\boldsymbol{S}||_2 \equiv \min \left\| \begin{pmatrix} \boldsymbol{A}^L\boldsymbol{B} \\ \boldsymbol{CB} - \boldsymbol{I} \end{pmatrix} \right\| \tag{15}$$

performed over all matrices $\boldsymbol{A}^L$ and $\boldsymbol{C}$ that satisfy $\boldsymbol{A}^L\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{CA} = \boldsymbol{0}$ has the solution

$$\boldsymbol{A}_\star^L = \boldsymbol{A}^\dagger = (\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^* \quad \text{and} \quad \boldsymbol{C}_\star = \boldsymbol{B}^*(\boldsymbol{I} - \boldsymbol{AA}^\dagger) \tag{16}$$

and the corresponding minimum value of the objective function is

$$||\boldsymbol{S}_\star||_2 = \begin{cases} \sqrt{\lambda_{\max}((\boldsymbol{A}^*\boldsymbol{A})^{-1})}, & \text{if } p < L \\ 0, & \text{if } p = L. \end{cases} \tag{17}$$

Furthermore, the solution (16) simultaneously minimizes all singular values of $\boldsymbol{S}$ and, therefore, its Frobenius norm as well.

Next, we address the problem of optimizing the actual reconstruction in terms of the aliasing error energy.

Theorem 1: The choice of optimal matrices $\boldsymbol{A}_\star^L$ and $\boldsymbol{C}_\star$ in Lemma 2 minimizes the *actual aliasing error energy* for each $x(t) \in \mathcal{B}([\mathcal{F}])$ as well as the aliasing error bound (12).

Proof: Equation (16) clearly minimizes the constant ψ_2 in the bound (12), or equivalently the bound itself. Moreover, Lemma 2 says that (16) minimizes all the eigenvalues of $\boldsymbol{S}^*\boldsymbol{S}$. Hence, $\boldsymbol{S}^*\boldsymbol{S} - \boldsymbol{S}_\star^*\boldsymbol{S}_\star$ is non-negative definite for any feasible $\boldsymbol{S}$. Examining the expression for the actual aliasing error energy [see (11)], we see that (16) is indeed the best solution. □

Remark 1: Suppose that $p < L$, and we let $\boldsymbol{A}^L = \boldsymbol{A}^\dagger$ and $\boldsymbol{C} = \boldsymbol{0}$. Then, (14) yields $\boldsymbol{S} = \left(\begin{smallmatrix} \boldsymbol{A}^\dagger\boldsymbol{B} \\ -\boldsymbol{I} \end{smallmatrix}\right)$, whose spectral norm can be shown to be

$$\begin{aligned} ||\boldsymbol{S}||_2 &= \sqrt{\lambda_{\max}(\boldsymbol{A}^\dagger\boldsymbol{B}(\boldsymbol{A}^\dagger\boldsymbol{B})^* + \boldsymbol{I})} = \sqrt{\lambda_{\max}(\boldsymbol{A}^\dagger(\boldsymbol{A}^\dagger)^*)} \\ &= \sqrt{\lambda_{\max}((\boldsymbol{A}^*\boldsymbol{A})^{-1})} \equiv ||\boldsymbol{S}_\star||_2 \end{aligned}$$

using (5). Thus, the pair $(\boldsymbol{A}^L, \boldsymbol{C}) = (\boldsymbol{A}^\dagger, \boldsymbol{0})$ produces precisely the optimal constant multiplier ψ_2 in the bound (12) but does not minimize the actual aliasing error energy. In other words, these matrices produce the same worst case but not the same case-by-case performance as the optimal matrices in (16).

B. Minimizing the Output Noise Power

We seek the optimal matrices $\boldsymbol{A}^L$ and $\boldsymbol{C}$ that satisfy $\boldsymbol{A}^L\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{CA} = \boldsymbol{0}$ and minimize ψ_{n} in (13), or equivalently, $||\boldsymbol{A}^L||_F^2 + ||\boldsymbol{C}||_F^2$. This is a fairly easy problem because $\boldsymbol{C}$ and $\boldsymbol{A}^L$ are independent of each other, and the objective function is separable. Therefore, we need to minimize $\min_{\boldsymbol{A}^L} ||\boldsymbol{A}^L||_F^2$ and $\min_{\boldsymbol{C}} ||\boldsymbol{C}||_F^2$ individually. For the second term, $\boldsymbol{C}_\star = \boldsymbol{0}$ is clearly the unique solution, whereas for the first term, we use the representation $\boldsymbol{A}^L \equiv \boldsymbol{A}^\dagger + \boldsymbol{XP}$, where $\boldsymbol{P}$ is the projection operator onto the null space of $\boldsymbol{A}^*$, namely, $\boldsymbol{P} \stackrel{\text{def}}{=} (\boldsymbol{I} - \boldsymbol{AA}^\dagger) = (\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*)$. This representation is justified in the proof of Lemma 2 in the Appendix. This yields the following minimization:

$$\min_{\boldsymbol{A}^L} ||\boldsymbol{A}^L||_F^2 = \min_{\boldsymbol{X}} ||\boldsymbol{A}^\dagger + \boldsymbol{XP}||_F^2.$$

Applying Lemma 1, we obtain the minimizing solution $\boldsymbol{X}_\star = -\boldsymbol{A}^\dagger\boldsymbol{P}^\dagger \equiv -\boldsymbol{A}^\dagger\boldsymbol{P} = \boldsymbol{0}$. Therefore, $\boldsymbol{A}_\star^L = \boldsymbol{A}^\dagger$. Finally

$$||\boldsymbol{A}^\dagger||_F = \left[\mathrm{tr}(\boldsymbol{A}^\dagger(\boldsymbol{A}^\dagger)^*)\right]^{1/2} = \sqrt{\mathrm{tr}((\boldsymbol{A}^*\boldsymbol{A})^{-1})}. \tag{18}$$

C. Minimizing the Peak Aliasing Error

The relevant quantity to minimize in order to obtain the tightest bound in (10) is the 1-norm of the matrix $\boldsymbol{S}$ defined in (14):

$$\min_{\boldsymbol{C},\boldsymbol{A}^L} \left\| \begin{pmatrix} \boldsymbol{A}^L\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{B}-\boldsymbol{I} \end{pmatrix} \right\|_1 \quad \text{subject to} \quad \begin{pmatrix} \boldsymbol{A}^L \\ \boldsymbol{C} \end{pmatrix} \boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{pmatrix}. \quad (19)$$

The problem of choosing $\boldsymbol{A}^L$ and $\boldsymbol{C}$ to minimize $||\boldsymbol{S}||_1$, unlike the spectral or Frobenius norms of $\boldsymbol{S}$, cannot be solved analytically. We resort to numerical methods instead. As shown in the proof of Lemma 2, we can represent $\boldsymbol{A}^L$ and $\boldsymbol{C}$ as follows:

$$\boldsymbol{A}^L = \boldsymbol{A}^\dagger + \boldsymbol{X}_1\boldsymbol{P} \quad \text{and} \quad \boldsymbol{C} = \boldsymbol{X}_2\boldsymbol{P} \quad (20)$$

where $\boldsymbol{X}_1$ is a $q \times p$ matrix, and $\boldsymbol{X}_2$ a $(L-q) \times p$ matrix. We can now rewrite (19) in an unconstrained form as

$$\min_{\boldsymbol{X}_1,\boldsymbol{X}_2} \left\| \begin{pmatrix} \boldsymbol{A}^\dagger\boldsymbol{B} \\ -\boldsymbol{I} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \boldsymbol{P}\boldsymbol{B} \right\|_1 \equiv \min_{\boldsymbol{X}} ||\boldsymbol{S}_0 + \boldsymbol{X}\boldsymbol{M}||_1 \quad (21)$$

where $\boldsymbol{S}_0 = \binom{\boldsymbol{A}^\dagger\boldsymbol{B}}{-\boldsymbol{I}}$, $\boldsymbol{X} = \binom{\boldsymbol{X}_1}{\boldsymbol{X}_2}$, and $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{B}$. This problem may be solved approximately using a linear program. We first rewrite (21) as follows:

$$\min d \quad \text{sub. to } d \geq ||\boldsymbol{z}_i||_1, \quad \boldsymbol{Z} = \boldsymbol{S}_0 + \boldsymbol{X}\boldsymbol{M}.$$

Each constraint $||z||_1 \leq d$ can be approximated by the set of linear inequalities:

$$\sum_{k=1}^{K} u_{mik} \exp(-j2\pi k/K) = z_{mi}, \qquad \forall m, i$$

$$\sum_{m=1}^{L}\sum_{k=1}^{K} u_{mik} \leq d, \qquad \forall i$$

$$u_{mik} \geq 0, \qquad \forall m, i, k.$$

This region is clearly a subset of $||\boldsymbol{z}_i|| \leq d$, but it can be verified easily that it is a superset of the region $||\boldsymbol{z}_i|| \leq d/\cos(\pi/K)$. Hence, the approximate LP produces an answer that is accurate to within a factor of $\cos(\pi/K)$ of the correct answer. Hence, the normalized error is bounded by $1 - \cos(\pi/K) \leq \pi^2/(2K^2)$, and the approximation is quite good for moderately large K. This is essentially like approximating circles by K-sided polygons. The optimization (21) can also be solved using semi-infinite programming [24].

D. Lower Bounds on ψ_∞, ψ_2 and ψ_{n}

The choice of sampling pattern $\mathcal{C}$ that minimizes the optimal constants ψ_∞, ψ_2, and ψ_{n} can be a difficult problem. It is therefore useful to know, even before attempting such a design, how small these constants can be made. In this section, we present lower bounds on these "error gain constants" $||\boldsymbol{S}||_2$, $||\boldsymbol{A}^L||_F$, and $||\boldsymbol{S}||_1$. These are the relevant quantities that affect the bounds (10) and (12) and the noise power (13). All bounds presented here are independent of the sampling pattern and only depend on L, p, and q.

Suppose we denote the (real positive) eigenvalues of the $q \times q$ matrix $\boldsymbol{A}^*\boldsymbol{A}$ by $\lambda_1, \lambda_2, \ldots, \lambda_q$ in decreasing order. Then, we have from (17) that

$$||\boldsymbol{S}||_2^2 \geq ||\boldsymbol{S}_\star||_2^2 = \lambda_{\max}\left((\boldsymbol{A}^*\boldsymbol{A})^{-1}\right) = \frac{1}{\lambda_q} \quad (22)$$

where $\boldsymbol{S}_\star$ is optimized to obtain the lowest aliasing error energy. Of course, $\lambda_q > 0$ since $\boldsymbol{A}$ is assumed to have full rank. Notice that we can use the fact that $|\boldsymbol{A}_{ij}| = 1/\sqrt{L}$ to bound the average eigenvalue $\overline{\lambda}_{\mathrm{am}} = \mathrm{tr}(\boldsymbol{A}^*\boldsymbol{A})/q$:

$$\overline{\lambda}_{\mathrm{am}} = \frac{1}{q}\sum_{i=1}^{q}\sum_{j=1}^{p} |\boldsymbol{A}_{ij}|^2 = \frac{p}{L}. \quad (23)$$

This result along with (22) and $\lambda_q \leq \overline{\lambda}_{\mathrm{am}}$ yields

$$||\boldsymbol{S}||_2 \geq \sqrt{\frac{L}{p}}. \quad (24)$$

Our next bound on $||\boldsymbol{A}^\dagger||_F$ provides an estimate for the worst-case output noise power. Equation (18) implies that

$$||\boldsymbol{A}^\dagger||_F^2 = \mathrm{tr}\left((\boldsymbol{A}^*\boldsymbol{A})^{-1}\right) = \sum_{i=1}^{q} \lambda_i^{-1} = q/\overline{\lambda}_{\mathrm{hm}}$$

where $\overline{\lambda}_{\mathrm{hm}}$ is the harmonic mean of the eigenvalues. Using the standard inequality $\overline{\lambda}_{\mathrm{hm}} \leq \overline{\lambda}_{\mathrm{am}}$ and (23), we immediately obtain

$$||\boldsymbol{A}^\dagger||_F^2 = \mathrm{tr}\left((\boldsymbol{A}^*\boldsymbol{A})^{-1}\right) \geq \frac{qL}{p}. \quad (25)$$

See [25] for some stronger bounds on the eigenvalues of these matrices, but they hold only for special sampling patterns and bandpass spectral supports. Our next theorem concerns the tightness of the bounds (24) and (25).

Theorem 2: The bounds (24) and (25) are tight in that they hold when $\mathcal{C}$ is a uniform sampling pattern and $\mathcal{F}$ is packable corresponding to $\mathcal{C}$.

Proof: Assume that $\mathcal{C} = \{0, a, 2a, \ldots, (p-1)a\}$, i.e., $c_i = (i-1)a$, where $a = L/p$ is an integer. This corresponds to subsampling $x(nT)$ by a factor of a. Let $\mathcal{G}_m$ and $\mathcal{K}_m$ denote the spectral subcells and index sets corresponding to $\mathcal{F}$. Let m be a fixed index, and let $\mathcal{K}_m = \{k_1, k_2, \ldots, k_{q_m}\}$. We see that packablility of $\mathcal{F}$ implies that $k_i - k_j$ cannot be a multiple of p for $i \neq j$ for otherwise, the subcells $\mathcal{G}_m \oplus (k_i/LT)$ and $\mathcal{G}_m \oplus (k_j/LT)$ will overlap in the spectrum of $x(nT)$ when subsampled by $a = L/p$. Therefore, the (i, l) entry of $\boldsymbol{A}_m$ is

$$\boldsymbol{A}_{m,il} = \frac{1}{\sqrt{L}} \exp\left(j2\pi\frac{(i-1)ak_l}{L}\right) = \frac{1}{\sqrt{ap}} \exp\left(j2\pi\frac{(i-1)\kappa_l}{p}\right)$$

where $\kappa_l \in \{0, 1, \ldots, p-1\}$ is such that $\kappa_l \equiv k_l \pmod{p}$. Evidently, $\{\kappa_l\}$ are distinct because $\mathcal{F}$ is packable. We see that $\sqrt{a}\boldsymbol{A}_m$ is a submatrix of the $p \times p$ DFT matrix consisting of *all* its rows and q of its columns. It immediately follows that $\lambda_{\max}(\boldsymbol{A}_m^*\boldsymbol{A}_m) = 1/a = L/p$ and $\mathrm{tr}(\boldsymbol{A}_m^*\boldsymbol{A}_m) = q_m/a =$

$q_m L/p$. Hence, the optimal matrices satisfy $||\boldsymbol{S}_{m\star}||_2 = \sqrt{L/p}$ and $||\boldsymbol{A}^\dagger||_F^2 = q_m L/p$. □

Finally, we provide a lower bound on $||\boldsymbol{S}||_1$ just as we did in the case of the spectral norm. From the definition of the 1-norm, we have

$$||\boldsymbol{S}||_1 = \max_s \sum_{r=1}^{L} |\boldsymbol{S}_{rs}| \geq \max_s \left(\sum_{r=1}^{L} |\boldsymbol{S}_{rs}|^2 \right)^{1/2}$$

where the last step follows from the positivity of terms in the sum. Therefore

$$\begin{aligned} ||\boldsymbol{S}||_1 &\geq \left(\max_s \sum_{r=1}^{L} |\boldsymbol{S}_{rs}|^2 \right)^{1/2} \\ &\geq \left(\frac{1}{L-q} \sum_{s=1}^{L-q} \sum_{r=1}^{L} |\boldsymbol{S}_{rs}|^2 \right)^{1/2} = \frac{1}{\sqrt{L-q}} ||\boldsymbol{S}||_F. \quad (26) \end{aligned}$$

Observe that

$$\begin{aligned} ||\boldsymbol{S}||_F^2 &= ||\boldsymbol{A}^L \boldsymbol{B}||_F^2 + ||\boldsymbol{C}\boldsymbol{B} - \boldsymbol{I}||_F^2 \\ &\geq ||\boldsymbol{A}^\dagger \boldsymbol{B}||_F^2 + ||\boldsymbol{B}^* \boldsymbol{P} \boldsymbol{B} - \boldsymbol{I}||_F^2 \end{aligned}$$

which follows from the fact that the solution (16) minimizes the Frobenius norm of $\boldsymbol{S}$. Therefore

$$\begin{aligned} ||\boldsymbol{S}||_F^2 &\geq \text{tr}(\boldsymbol{A}^\dagger \boldsymbol{B}\boldsymbol{B}^* (\boldsymbol{A}^\dagger)) + \text{tr}\left((\boldsymbol{B}^* \boldsymbol{P} \boldsymbol{B} - \boldsymbol{I})^2 \right) \\ &= \text{tr}\left((\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{I} \right) + \text{tr}(\boldsymbol{I} - \boldsymbol{B}^* \boldsymbol{P} \boldsymbol{B}). \end{aligned}$$

The expressions in the last step were obtained using the facts that $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^*$ and that $(\boldsymbol{I} - \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B})$ is a projection operator. These are justified in the proof of Lemma 2. Therefore, using $\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{P}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^*) = \boldsymbol{P}$ and $\text{tr}(\boldsymbol{P}) = \text{tr}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^\dagger) = p - \text{tr}(\boldsymbol{A}^\dagger \boldsymbol{A}) = p - q$, we conclude that

$$\begin{aligned} ||\boldsymbol{S}||_F^2 &\geq \text{tr}\left((\boldsymbol{A}^* \boldsymbol{A})^{-1} \right) - q + (L-q) - \text{tr}(\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^*) \\ &= \text{tr}\left((\boldsymbol{A}^* \boldsymbol{A})^{-1} \right) - q + L - p. \end{aligned}$$

Combining the last inequality with (26) and (25), we finally obtain

$$\begin{aligned} ||\boldsymbol{S}||_1 &\geq \frac{1}{\sqrt{(L-q)}} \sqrt{\frac{qL}{p} - q + L - p} \\ &= \sqrt{\frac{(L-p)(q+p)}{(L-q)p}}. \quad (27) \end{aligned}$$

We can obtain a stronger bound (27) when $p = q$. In this case, we see that the matrices $\boldsymbol{A}^L$ and $\boldsymbol{C} = \boldsymbol{0}$ are unique. Therefore

$$||\boldsymbol{S}||_1 = \left\| \begin{pmatrix} \boldsymbol{A}^L \boldsymbol{B} \\ -\boldsymbol{I} \end{pmatrix} \right\|_1 = 1 + ||\boldsymbol{A}^L \boldsymbol{B}||_1 \geq 1 + \frac{||\boldsymbol{B}||_1}{||\boldsymbol{A}||_1} = 2. \quad (28)$$

This bound, although applicable only for $p = q$, is indeed stronger than (27), whose right-hand side evaluates to $\sqrt{2}$ for $p = q$.

E. Which Criterion to Optimize?

As seen in Sections III-A–C, each of the three criteria leads to a different optimal choice for $\boldsymbol{A}^L$ and $\boldsymbol{C}$. Furthermore, the computation of those matrices is more difficult in one of the cases (minimizing the peak error). Therefore, the following question arises: Suppose that $\boldsymbol{A}^L$ and $\boldsymbol{C}$ are chosen to optimize one of the criteria. Then, how far from the optimum would the other criteria be for this choice? First of all, it is not fair to compare the optimum matrices corresponding to the criteria ψ_2 and ψ_{n} because the two problems have different underlying settings: The error energy is due to input signal mismodeling, whereas the output noise power is due to additive sample noise. Yet we have seen that the "least-squares" solution $\boldsymbol{A}^\dagger$ is the optimal choice for $\boldsymbol{A}^L$ for both criteria, although the optimal choices for $\boldsymbol{C}$ are different. However, it is more meaningful to compare the optimal matrices for the criteria ψ_2 and ψ_∞ because the two problems are similar: The imperfections in the input signal in both cases are due to mismodeling. It is therefore reasonable to expect the optimal matrices that minimize the 1-norm of $\boldsymbol{S}$ to be close to those that minimize its 2-norm for which, of course, we have the analytical least-squares solution (16). A question that springs into mind is whether or not the least-squares solution is a good enough approximation to the exact solution for the 1-norm problem. The answer partially lies in the following result, which is a bound on the improvement factor that the solution to (21) can offer over the least-squares solution (16). Observe that (21) can be rewritten in a slightly different constrained form as

$$\min ||\boldsymbol{S}||_1 \quad \text{subject to } \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{S}_{\mathrm{ls}}\boldsymbol{\xi}, \quad \forall \boldsymbol{\xi} \text{ such that } \boldsymbol{P}\boldsymbol{B}\boldsymbol{\xi} = \boldsymbol{0}$$

$$\text{where } \boldsymbol{S}_{\mathrm{ls}} = \begin{pmatrix} \boldsymbol{A}^\dagger \boldsymbol{B} \\ \boldsymbol{B}^* \boldsymbol{P} \boldsymbol{B} - \boldsymbol{I} \end{pmatrix} \quad (29)$$

since the difference $(\boldsymbol{S} - \boldsymbol{S}_{\mathrm{ls}})$ can be expressed as $\boldsymbol{X}\boldsymbol{P}\boldsymbol{B}$ for a suitable $\boldsymbol{X}$. Recall that $\boldsymbol{A}^L = \boldsymbol{A}^\dagger + \boldsymbol{X}_1 \boldsymbol{P}$ and $\boldsymbol{C} = \boldsymbol{X}_2 \boldsymbol{P}$ for suitable $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Note that $\boldsymbol{S}_{\mathrm{ls}}$ is the optimal matrix for the 2-norm minimization. Therefore, (29) yields

$$\begin{aligned} ||\boldsymbol{S}||_1 ||\boldsymbol{\xi}||_1 \geq ||\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{\xi}||_1 \quad \forall \boldsymbol{\xi} \in N(\boldsymbol{P}\boldsymbol{B}) &\Longrightarrow ||\boldsymbol{S}||_1 \\ &\geq \max_{\substack{\boldsymbol{\xi} \in N(\boldsymbol{P}\boldsymbol{B}) \\ \boldsymbol{\xi} \neq \boldsymbol{0}}} \frac{||\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{\xi}||_1}{||\boldsymbol{\xi}||_1} \quad (30) \end{aligned}$$

where $N(\cdot)$ denotes the null space. In the proof of Lemma 2, we show that $N(\boldsymbol{P}\boldsymbol{B}) = R(\boldsymbol{K})$, where $\boldsymbol{K} \stackrel{\text{def}}{=} (\boldsymbol{I} - \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B})$ is an orthonormal projector. Furthermore, $\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{K} = \boldsymbol{S}_{\mathrm{ls}}$ is easy to check. Hence, for ease of computation, we can weaken the bound (30) by maximizing over the set consisting of the $L - q$ columns of $\boldsymbol{K}$, rather than the subspace $N(\boldsymbol{P}\boldsymbol{B}) = R(\boldsymbol{K})$. Thus, letting $\{\boldsymbol{k}_i\}$ denote the columns of $\boldsymbol{K}$, we obtain

$$\begin{aligned} ||\boldsymbol{S}||_1 \geq \max_i \frac{||\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{k}_i||_1}{||\boldsymbol{k}_i||_1} \geq \frac{\max_i ||\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{k}_i||_1}{\max_i ||\boldsymbol{k}_i||_1} &= \frac{||\boldsymbol{S}_{\mathrm{ls}}\boldsymbol{K}||_1}{||\boldsymbol{K}||_1} \\ = \frac{||\boldsymbol{S}_{\mathrm{ls}}||_1}{||\boldsymbol{K}||_1} \Longrightarrow ||\boldsymbol{S}_{\mathrm{ls}}||_1 \leq ||\boldsymbol{S}||_1 ||\boldsymbol{K}||_1. \quad (31) \end{aligned}$$

In other words, using the least-squares solution (16) instead of the solution to (21) cannot amplify $||\boldsymbol{S}||_1$ by a factor of more than $||\boldsymbol{K}||_1 \geq 1$.

We can apply these results to the norms of $\boldsymbol{S}_m$ and $\boldsymbol{A}_m^L$ for each m to obtain lower bounds on a) the constants present in (10) and (12) and b) the average output noise power in (13). Without explicitly deriving them, we summarize the bounds below in terms of Ω, L, p, and $q' = \max_m q_m$:

$$\|e\|_\infty \le \psi_\infty \int_{[\mathcal{F}]\setminus\mathcal{F}} |X(f)|\, df$$

$$\|e(t)\|_2 \le \psi_2 \left(\int_{[\mathcal{F}]\setminus\mathcal{F}} |X(f)|^2\, df \right)^{1/2}$$

$$\langle E|\tilde{w}(t)|^2\rangle_t = \psi_{\mathrm{n}}\sigma^2$$

where

$$\psi_\infty = \max_m \|\boldsymbol{S}_m\|_1 \ge \sqrt{\frac{(L-p)(q'+p)}{(L-q')p}} \tag{32}$$

$$\psi_2 = \max_m \|\boldsymbol{S}_m\|_2 \ge \sqrt{\frac{L}{p}} \tag{33}$$

$$\psi_{\mathrm{n}} = T \sum_m \lambda(\mathcal{G}_m)\|\boldsymbol{A}_m^\dagger\|_F^2 \ge \Omega\, \frac{L}{p}. \tag{34}$$

In each of the first two equations, we have a lower bound and an upper bound, which cannot be combined. However, note that the bounds for $\|e(t)\|_\infty$ or $\|e(t)\|_2$ are tight. Hence, the lower bounds on ψ_∞ and ψ_2 tell us how large the aliasing error can be in the worst cases for the corresponding bounds. However, they do not tell how large or small the errors are in other cases. In addition, note that the constants ψ_ω, $(\omega = 2, \mathrm{n}, \infty)$ decrease as expected when p is increased for a fixed $\{q_m\}$. If we increase L, p, and q_m in such a way that p/L and q/L remain constant (as would happen if one attempts to approach the Landau rate by increasing L), we find that none of the bounds above get worse. These bounds represent errors inherent to any sampling procedure, whether they are uniform or not, for any multiband signals, whether they are packable or not. In Section V, we study the increase in error sensitivity incurred for nonpackable signals that are sampled at a sub-Nyquist rate.

IV. Optimal Sampling

In this section, we discuss two important issues pertaining to optimal sampling, namely, determining the optimal base sampling frequency and sampling pattern design. The first one is concerned with choosing the period T that produces the lowest average sampling rate for a given $\mathcal{F}$ and L. The second issue is concerned with finding good sampling patterns $\mathcal{C}$ that optimize the aliasing error bounds or sensitivity to noise. Our study of this latter problem is somewhat numerical in nature.

A. Optimizing the Choice of T

In all our previous analysis, we set the base sampling frequency $1/T = f_{\mathrm{nyq}}$ for convenience. This corresponds to sampling the original signal uniformly at the Nyquist rate before some samples are dropped. Clearly, we could have chosen a slightly larger rate than the Nyquist rate and still obtained similar results. In this section, we examine the problem of choosing, for given L and $\mathcal{F}$, the optimal value $T_\star$ for T that minimizes the average sampling rate. We also provide a polynomial-time algorithm to find $T_\star$, which may be larger than $1/f_{\mathrm{nyq}}$, and we show this by example. This problem is related to that of pairing band edges [13] that Herley and Wong suggested. They provide a necessary and sufficient condition on the band edge frequencies a_i and b_i of the spectral support $\mathcal{F}$ for achieving the minimum rate. They also show that for sufficiently large L, with LT fixed, the minimum rate can be approached arbitrarily closely. In our case, however, we fix L and the corresponding optimal base sampling frequency $1/T$ because L determines the complexity of the reconstruction.

All the results derived for the specific case $1/T = f_{\mathrm{nyq}}$ (the Nyquist rate) extend to the general case $1/T \ge f_{\mathrm{nyq}}$, provided we replace the spectral span $[\mathcal{F}]$ everywhere by $[0,\, 1/T)$. We already know that for given L and T, the smallest average sampling rate [equal to $\max_m q_m/(LT)$] is given by the right-hand side of (6):

$$f_{\mathrm{avg}}(T, L) = \frac{1}{LT} \max_{f\in[0,(1/LT))} \sum_{r=0}^{L-1} \chi\left(f + \frac{r}{LT} \in \mathcal{F}\right).$$

Therefore, we seek the solution to

$$T_\star = \arg \min_{0<T\le T_0} f_{\mathrm{avg}}(T, L) \quad \text{where } T_0 = \frac{1}{f_{\mathrm{nyq}}} \tag{35}$$

for a fixed L. For spectral supports expressible as $\mathcal{F} = \bigcup_{i=1}^n [a_i, b_i)$, we have seen that the construction of the sets $\mathcal{G}_m$ and $\mathcal{K}_m$ can be done in polynomial time. Therefore, the computation of $f_{\mathrm{avg}}(T, L)$ requires polynomial time as well. The minimization (35), however, is over a continuous parameter T. The following theorem allows us to transform the minimization problem to an exhaustive search over a finite set of values for T.

Theorem 3: The optimum value for $T_\star$ in (35) for a spectral support of the form $\mathcal{F} = \bigcup_{i=1}^n [a_i, b_i)$ must satisfy

$$\frac{1}{LT_\star} = \frac{b_j - a_i}{k}$$

for some integers i, j, and k such that $1 \le i \le j \le n$ and $0 < k \le LT_0(b_j - a_i) \le L$.

The proof can be found in the Appendix. Theorem 3 enables us to generate no more than $n(n+1)(L-1)/2$ possible candidates for $T_\star$, which may then be used to minimize the average sampling rate. It turns out that $T_\star = T_0$ is optimal for the spectral support $\mathcal{F}$ of [14, Ex. 1]. However, as the following example shows, spectral supports exist for which $T = T_0$ is highly suboptimal.

Example 2: For a given L, consider the spectral support

$$\mathcal{F} = \bigcup_{r=0}^{L-1} [r,\, r+\epsilon) \cup [L-\epsilon,\, L)$$

where $\epsilon < 1/(L+1)$ is a small positive number. For the choice $T = T_0 = 1/L$, we have $1/(LT) = 1$ and

$$\sum_{r=0}^{L-1} \chi\left(f + \frac{r}{LT} \in \mathcal{F}\right) = \begin{cases} L, & \text{if } 0 \le f < \epsilon \\ 0, & \text{if } \epsilon \le f < 1-\epsilon \\ 1, & \text{if } 1-\epsilon \le f < 1. \end{cases}$$

Since L pieces of the spectrum overlap, we require $p = L$, and this makes the average sampling rate equal to $p/(LT) =$

TABLE I
DESIGN CRITERIA EVALUATED AT THE CANDIDATE OPTIMAL SAMPLING PATTERNS: $\mathcal{C}_a = \{0, 1, 2, 13, 16\}, \mathcal{C}_b = \{0, 4, 7, 14, 15\}$ AND $\mathcal{C}_c = \{0, 1, 2, 8, 17\}$. THE BOXED ENTRIES ARE OPTIMAL IN THEIR RESPECTIVE ROWS

Criterion	$\mathcal{C} = \mathcal{C}_a$	$\mathcal{C} = \mathcal{C}_b$	$\mathcal{C} = \mathcal{C}_c$
ψ_2	2.9032	3.0641	3.5241
ψ_{n}	0.4811	0.4769	0.4918
ψ_∞	2.3929	2.6479	2.2583

L. Next, for the optimal choice $T = T_\star = (L(1+\epsilon))^{-1}$, we have

$$\sum_{r=0}^{L-1} \chi\left(f + \frac{r}{LT} \in \mathcal{F}\right) = \begin{cases} 1 & \text{if } f \in [0, \epsilon) \cup [1 - L\epsilon, 1 - L\epsilon + \epsilon) \\ & \quad \cup [1 - L\epsilon + 2\epsilon, 1 + \epsilon) \\ 0 & \text{if } f \in [\epsilon, 1 - L\epsilon) \\ & \quad \cup [1 - L\epsilon, +\epsilon 1 - L\epsilon + 2\epsilon). \end{cases}$$

Therefore, $p_\star = 1$ is required, and the average sampling rate $p_\star/(LT) = (1+\epsilon)$. Hence, the choice $T = T_\star$ can improve the sampling rate over the choice $T = T_0$ by a factor of $L/(1+\epsilon)$, which is nearly as large as it can get because this factor never exceeds L

$$\frac{p_\star}{LT_\star} \geq \frac{p_\star}{LT_0} \geq \frac{1}{L}\left(\frac{p}{LT_0}\right) \quad \text{since } T_\star \leq T_0 \quad \text{and}$$
$$p_\star \geq 1 \geq \frac{p}{L}.$$

Although Example 2 is an extreme situation, it shows that optimizing T may be of significance, with the largest gains occurring typically for signals with sparse spectral supports.

B. Sampling Pattern Design

We now examine the problem of designing good sampling patterns. We propose to use the following minimizations as empirical design criteria:

$$\mathcal{C}_\star = \arg\min_{\mathcal{C}} \psi_\omega(\mathcal{C}, \{\mathcal{K}_m\}_{m=1}^M) \quad \text{for} \quad \omega = 2, \infty, \mathrm{n}$$

where $\psi_\omega(\mathcal{C}, \{\mathcal{K}_m\})$ for $\omega = 1, 2, \mathrm{n}$ are $\max_m \|\boldsymbol{S}_{\star m}\|_2$, $\max_m \|\boldsymbol{S}_{\star m}\|_1$, and $T\sum_m \lambda(\mathcal{G}_m)\|\boldsymbol{A}_m^\dagger\|_F^2$, respectively, and the subscript "$\star$" denotes optimality of $\boldsymbol{S}_m$ with respect to the matrices $\boldsymbol{A}_m^L$ and $\boldsymbol{C}_m$, which is discussed in the previous section. The functions $\psi_\omega(\mathcal{C}, \{\mathcal{K}_m\})$ are invariant under cyclic shifts of $\mathcal{C}$ in $\{0, 1, \ldots, L-1\}$, and this follows easily from the definitions of these functions.

Example 1 (Continued): For the chosen spectral support and $L = 20$, we have $q_m = 2$ for all m. Hence, $p \geq 2$ is necessary and sufficient for perfect reconstruction from the multicoset samples. For example, for $p = 5$, an exhaustive search over all sampling patterns yields

$$\psi_{2\star} = \min_{\mathcal{C}} \psi_2(\mathcal{C}, \{\mathcal{K}_m\}) = 2.9032 \text{ at } \mathcal{C}_a = \{0, 1, 2, 13, 16\}$$
$$\psi_{\mathrm{n}\star} = \min_{\mathcal{C}} \psi_{\mathrm{n}}(\mathcal{C}, \{\mathcal{K}_m\}) = 0.4769 \text{ at } \mathcal{C}_b = \{0, 4, 7, 14, 15\}$$
$$\psi_{\infty\star} = \min_{\mathcal{C}} \psi_\infty(\mathcal{C}, \{\mathcal{K}_m\}) = 2.2583 \text{ at } \mathcal{C}_c = \{0, 1, 2, 8, 17\}$$

where $\min_{\mathcal{C}} \psi_\infty(\mathcal{C}, \{\mathcal{K}_m\})$ was computed using the approximate LP method with $K = 12$ and, hence, cannot be claimed to be truly optimal. In this example, the three design criteria produce different optimal sampling patterns. Table I shows the three objective functions evaluated for each of these three candidate optimal patterns. It is evident that none of the three candidate sampling patterns is simultaneously optimal for all three design criteria. However, all three candidate solutions are close to optimal for each of the criteria, and we lose little in terms of optimality by restricting our attention to any one of the criteria. For example, we can pick ψ_{n}, which is the easiest to compute.

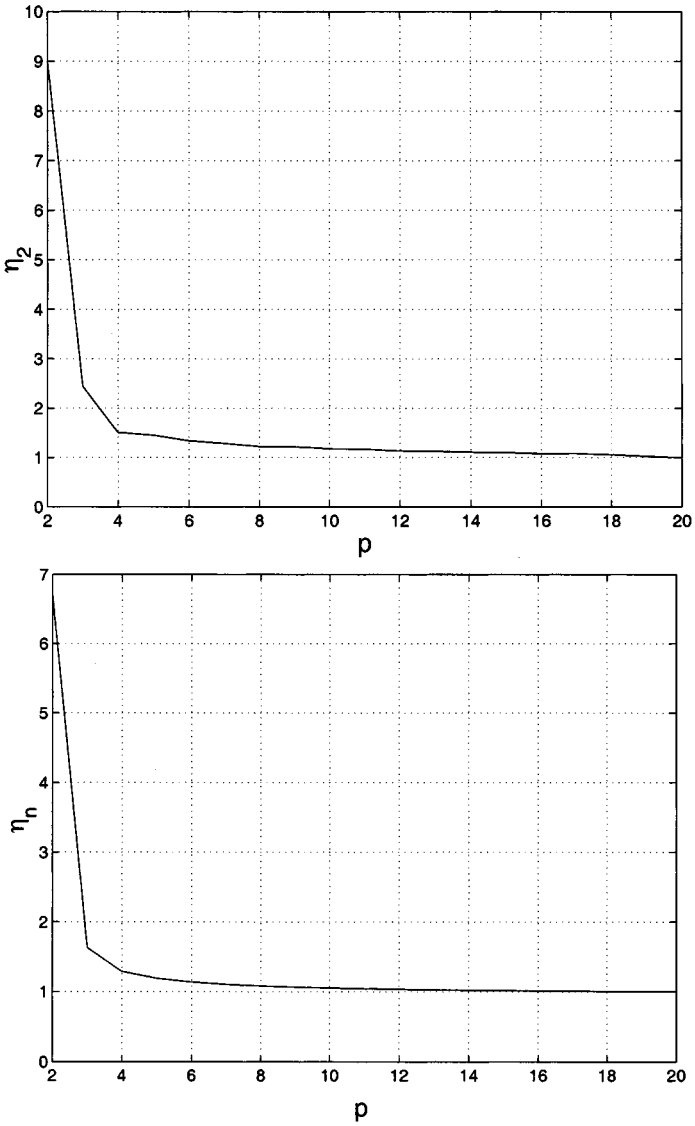

Fig. 3. Normalized error gain constants η_2 and η_{n} shown as functions of the number of multicoset samples p in each period.

Let $\psi_{2\star} = \min_{\mathcal{C}} \psi_2(\mathcal{C}, \{\mathcal{K}_m\})$ and $\psi_{\mathrm{n}\star} = \min_{\mathcal{C}} \psi_{\mathrm{n}}(\mathcal{C}, \{\mathcal{K}_m\})$ denote the optimal constants for a given p. Define the following quantities:

$$\eta_2 = \psi_{2\star}\sqrt{\frac{p}{L}}, \quad \eta_{\mathrm{n}} = \psi_{\mathrm{n}\star}\frac{p}{\Omega L}$$
$$\eta_\infty = \psi_{\infty\star}\sqrt{\frac{(L-q')p}{(L-p)(q'+p)}}$$

which are obtained by normalizing the quantities $\psi_{2\star}$ and $\psi_{\mathrm{n}\star}$ and $\psi_{\infty\star}$ by their lower bounds (32)–(34). Fig. 3 illustrates (for $2 \leq p \leq 20$) the behavior of η_2 and η_{n}. These were computed

TABLE II
NORMALIZED ERROR GAIN CONSTANT η_∞ FOR $2 \le p \le 5$

p	2	3	4	5
η_∞	9.8256	3.3764	2.3049	2.0908

by first computing $\psi_{2\star}$ and $\psi_{\mathrm{n}\star}$ by exhaustively searching over all sampling patterns of size p, followed by normalization. Due to its higher computational complexity, we computed η_∞ only for $2 \le p \le 5$, which is summarized in Table II. The two plots in Fig. 3 show that the optimal error gain constants η_2 and η_{n} approach their lower bounds rather quickly when p is increased. This behavior suggests the following design recommendation. For given $\mathcal{F}$ with occupancy Ω and given L, choose $p = L\Omega+1$ or $p = L\Omega+2$. This results in sampling rate p/L slightly larger than Ω but provides significant reduction in error sensitivities. Table II shows that η_∞ also approaches its lower bound, only more slowly.

V. COMPARISONS

The goal of this section is to compare uniform and nonuniform sampling below the Nyquist rate for packable and nonpackable supports $\mathcal{F}$. Some of the comparisons are numerical examples. We will examine the following two questions: i) For packable spectra, both uniform and nonuniform sampling at the same average rate is possible. How does the reconstruction error compare between the two options? ii) What is the penalty for nonpackability: Given $\mathcal{F}^{(1)}$ packable and $\mathcal{F}^{(2)}$ nonpackable, such that $\lambda(\mathcal{F}^{(1)}) = \lambda(\mathcal{F}^{(2)})$, which are sampled at the same rate (uniformly and nonuniformly, respectively), how do the error bounds compare?

A. Uniform Sampling Versus Nonuniform Sampling for Packable $\mathcal{F}$

Recall that a signal $x(t) \in \mathcal{B}(\mathcal{F})$ is packable at rate $1/T' < \lambda([\mathcal{F}])$ if $\mathcal{F} \cap (\mathcal{F} \oplus n/T') = \emptyset$ for all $n \neq 0$. Dodson and Silva [26] proved the following sampling theorem for packable signals.

Theorem 4: Suppose $\mathcal{F}$ is packable at rate $1/T'$ for some $T' > 0$; then, $x(t) \in \mathcal{B}(\mathcal{F})$ has the sampling representation

$$x(t) = \sum_{m \in Z} x(mT')\varphi(t - mT';\ \mathcal{F}) \tag{36}$$

with the sum converging uniformly and absolutely for $t \in \mathbb{R}$.

The following theorem due to Beaty and Higgins [27] is a bound on the peak value of the aliasing error for signals with finite energy and packable $\mathcal{F}$.

Theorem 5: Suppose that $x(t) \in L^2(\mathbb{R}) \cap C(\mathbb{R})$ with $X(f) \in L^1(\mathbb{R})$ is sampled at the rate $1/T'$ that satisfies $\mathcal{F} \cap ((n/T') \oplus \mathcal{F}) = \emptyset, \forall\, n \neq 0$. Then, the aliasing error $e(t)$ satisfies

$$|e(t)| \le 2 \int_{\mathbb{R} \setminus \mathcal{F}} |X(f)|\, df. \tag{37}$$

Incidentally, this equation is identical to the bound for the aliasing error in classical Shannon sampling of lowpass signals. Although valid for all $x(t) \in F$, (37) will only be used for the class signals $x(t) \in \mathcal{B}([\mathcal{F}])$ in order to compare with the bound (10) that applies to signals in $\mathcal{B}([\mathcal{F}])$. It turns out that uniform patterns are indeed the best patterns suited for packable spectra because the aliasing error bounds in (10) and (37) have identical forms, with the only difference being the premultiplying constants. The constant is 2 in (37), but $\max_m \|\boldsymbol{S}_m\|_1$ in (10). Since the lower bound (28) implies that $\psi_\infty = \max_m \|\boldsymbol{S}_m\|_1 \ge 2$ for $p = \max q_m$, we see that for packable signals, uniform sampling is most appropriate. Theorem 2 verifies this claim for the other two performance criteria because the constants ψ_2 and ψ_{n} attain their respective lower bounds for uniform sampling. In summary, for packable signals, uniform sampling is the best.

B. Penalty for Nonpackability

What is the penalty in error sensitivity for nonpackability? Unfortunately, the answer to this question is not easy to deduce analytically. From various numerical computations, it seems that there is a price to pay for nonpackability of $\mathcal{F}$. We provide an example here to support the conjecture.

Example 2: Consider spectral supports $\mathcal{F}^{(1)}$, $\mathcal{F}^{(2)}$, and $\mathcal{F}^{(3)}$

$$\mathcal{F}^{(1)} = [0,\ 7)$$
$$\mathcal{F}^{(2)} = [0,\ 3) \cup [5,\ 6) \cup [12,\ 13) \cup [14,\ 16)$$
$$\mathcal{F}^{(3)} = [0,\ 3) \cup [6,\ 8) \cup [13,\ 16).$$

For $L = 16$ and base sampling rate $1/T = 16$, they all have $M = 1$. The corresponding index sets are $\mathcal{K}_1^{(1)} = \{0 \to 7\}$, $\mathcal{K}_1^{(2)} = \{0 \to 2,\ 5,\ 11,\ 12,\ 14,\ 15\}$ and $\mathcal{K}_1^{(3)} = \{0 \to 2,\ 6,\ 7,\ 13 \to 15\}$. The spectral supports have the same measure, and the $\mathcal{K}$ sets have $q = 8$ elements each. Although $\mathcal{K}^{(1)}$ and $\mathcal{K}^{(2)}$ are packable, $\mathcal{K}^{(3)}$ is not. Upon minimizing the quantity $\psi_\omega(\mathcal{C}, \mathcal{K}^{(i)})$ for $i = 1,\ 2,\ 3$ and $\omega = \mathrm{n},\ 2,\ \infty$ using a forward selection greedy algorithm, we obtain the following optimal sampling patterns $\mathcal{C}_{\star\omega}^{(i)}$ with $p = 8$ (half the Nyquist rate) and objective functions $\psi_{\star\omega}^{(i)} \stackrel{\mathrm{def}}{=} \psi_{\star\omega}(\mathcal{C}_{\star\omega}^{(i)}, \{\mathcal{K}_m^{(i)}\})$.

- For $i = 1,\ 2$, we obtain $\mathcal{C}_{\star\mathrm{n}}^{(i)} = \mathcal{C}_{\star 2}^{(i)} = \mathcal{C}_{\star\infty}^{(i)} = \{1,\ 3,\ 5,\ 7,\ 9,\ 11,\ 13,\ 15\}$ with the corresponding objective functions being $\psi_{\star\mathrm{n}}^{(i)} = 1.0$, $\psi_{\star 2}^{(i)} = 1.4142$, and $\psi_{\star\infty}^{(i)} = 2.0$. Note that these values agree with the results of Theorem 2.
- For $i = 3$, we obtain $\mathcal{C}_{\star\mathrm{n}}^{(3)} = \mathcal{C}_{\star 2}^{(3)} = \{2,\ 4,\ 5,\ 6,\ 9,\ 12,\ 14,\ 15\}$ and $\mathcal{C}_{\star\infty}^{(3)} = \{1,\ 4,\ 5,\ 6,\ 9,\ 12,\ 13,\ 14\}$. The objective functions take the values $\psi_{\star\mathrm{n}}^{(1)} = 1.9291$, $\psi_{\star 2}^{(i)} = 3.3598$, and $\psi_{\star\infty}^{(i)} = 4.8284$ at optimality.

Hence, the price to pay to sample a signal with a nonpackable spectrum at the Landau rate manifests itself in the output noise and aliasing error bounds. They are larger for nonpackable spectra.

Remark 2: Example 2 also illustrates the point made in the last subsection that uniform sampling is, in general, better suited for packable signals than nonuniform patterns. The fact that the sampling patterns for the packable spectra $\mathcal{C}_{\star\omega}^{(1)}$ and $\mathcal{C}_{\star\omega}^{(2)}$ turn out to be uniform clearly supports the claim.

VI. CONCLUSION

We presented solutions to the problems of optimal sub-Nyquist sampling and reconstruction. We showed how to determine optimal matrices to obtain the best performance

in terms of the aliasing error bounds or the noise sensitivity (our measures of performance) that were derived in [14]. We provided explicit solutions for most of these problems.

The error bounds reveal a dependence on the sampling pattern $\mathcal{C}$. We examined the problem of designing sampling patterns that optimize any of our performance measures. We used an exhaustive search algorithm in one example and a forward selection greedy algorithm in another to pick optimal sampling patterns for a few design examples. An exhaustive search over all sampling patterns is computationally very expensive for even moderately large L, whereas a greedy search is not guaranteed to produce the best pattern. Nevertheless, the greedy algorithm did produce very good results. The problem of designing sampling patterns efficiently is still an open problem. We also showed how to choose the optimal base sampling frequency that minimizes the average sampling rate for a given sampling period L. This is an important issue because sampling at a base frequency equal to the Nyquist rate may be severely suboptimal for certain spectral supports.

We made comparisons to determine whether nonuniform sampling is appropriate for packable signals. Our findings are that a) for packable spectral supports, a uniform sampling pattern yields a better performance than a general nonuniform sampling pattern, and b) for nonpackable signals, where uniform sampling is not applicable, there is a penalty associated with nonuniform sampling. The error bounds are larger for this case relative to uniform sampling of a packable signal of the same occupancy. We find that the sensitivity penalties (error gain constants) for sub-Nyquist sampling of signals with nonpackable spectra can be controlled by optimal design and by backing off slightly from the minimum rate. The resulting low error sensitivities, and the significant reduction in the sampling rate over the Nyquist rate of our numerical examples, suggest that these techniques have considerable practical potential. Most of the results presented for 1-D signals should be extensible to two and higher dimensions with little difficulty. In contrast, determining the optimal base sampling lattice in higher dimensions would be a harder problem.

APPENDIX

A. Proof of Lemma 1

Let $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{R}\boldsymbol{Q}^{\dagger}$. Rewriting in terms of $\boldsymbol{Y}$ followed by squaring, the objective function gives us an equivalent problem $\min_{\boldsymbol{Y}} \|\boldsymbol{R} + (\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{Q}^{\dagger})\boldsymbol{Q}\|_2^2 = \min_{\boldsymbol{Y}} \|\boldsymbol{R}\boldsymbol{P}_{\boldsymbol{Q}}^{\perp} + \boldsymbol{Y}\boldsymbol{Q}\|_2^2$, where $\boldsymbol{P}_{\boldsymbol{Q}}^{\perp} = \boldsymbol{I} - \boldsymbol{Q}^{\dagger}\boldsymbol{Q}$ is the orthogonal projection operator onto $R(\boldsymbol{Q})^{\perp}$. Hence, using $\boldsymbol{P}_{\boldsymbol{Q}}^{\perp}\boldsymbol{Q} = \boldsymbol{0}$ transforms the problem to

$$\min_{\boldsymbol{Y}} \lambda_{\max}\left((\boldsymbol{R}\boldsymbol{P}_{\boldsymbol{Q}}^{\perp} + \boldsymbol{Y}\boldsymbol{Q})(\boldsymbol{R}\boldsymbol{P}_{\boldsymbol{Q}}^{\perp} + \boldsymbol{Y}\boldsymbol{Q})^*\right) \equiv \min_{\boldsymbol{Y}} \lambda_{\max}\left(\boldsymbol{R}\boldsymbol{P}_{\boldsymbol{Q}\boldsymbol{Q}^*\perp}\boldsymbol{R}^* + \boldsymbol{Y}\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{Y}^*\right).$$

The choice $\boldsymbol{Y} = \boldsymbol{0}$ (or $\boldsymbol{X} = -\boldsymbol{R}\boldsymbol{Q}^{\dagger}$) is clearly optimal since $\boldsymbol{Y}\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{Y}^*$ is a positive semidefinite perturbation. In fact, all singular values of $(\boldsymbol{R} + \boldsymbol{X}\boldsymbol{Q})$ are simultaneously minimized.

B. Proof of Lemma 2

The proof of Lemma 2 relies on the following result.

Lemma 3: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same number of rows that satisfy $\boldsymbol{A}\boldsymbol{A}^* + \boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}$, and define $\boldsymbol{P} \stackrel{\text{def}}{=} \boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*$. Then, the pseudo-inverse of $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{B}$ is $\boldsymbol{M}^{\dagger} = \boldsymbol{B}^*\boldsymbol{P}$.

Proof: It suffices to check that the asserted pseudo-inverse $\boldsymbol{M}^{\dagger}$ satisfies a) $\boldsymbol{M}^{\dagger}\boldsymbol{M} = \boldsymbol{P}_{R(\boldsymbol{M}^{\dagger})}$ and b) $\boldsymbol{M}\boldsymbol{M}^{\dagger} = \boldsymbol{P}_{R(\boldsymbol{M})}$, where $\boldsymbol{P}_{R(\boldsymbol{U})}$ denotes the orthogonal projection operator onto the range space of any matrix $\boldsymbol{U}$. Note that $\boldsymbol{P}$, being the orthogonal projection matrix onto the null space of $\boldsymbol{A}^*$, satisfies the following properties, which are easy to verify directly: $\boldsymbol{P}^* = \boldsymbol{P}^2 = \boldsymbol{P}^{\dagger} = \boldsymbol{P}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{0}$ and $\boldsymbol{A}^*\boldsymbol{P} = \boldsymbol{0} = \boldsymbol{A}^{\dagger}\boldsymbol{P} = \boldsymbol{0}$. We use these properties without explicitly stating. To verify a), let $\boldsymbol{Q} = \boldsymbol{M}^{\dagger}\boldsymbol{M} = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{P}\boldsymbol{B} = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B}$. A standard way to check that $\boldsymbol{Q}$ is an orthonormal projection operator is to verify that $\boldsymbol{Q}^2 = \boldsymbol{Q}$ and $\boldsymbol{Q}^* = \boldsymbol{Q}$. It is evident from its definition that $\boldsymbol{Q}$ is Hermitian. Next, examine the quantity $\boldsymbol{Q}^2$:

$$\begin{aligned}\boldsymbol{Q}^2 &= (\boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B})^2 = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B} = \boldsymbol{B}^*\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^*)\boldsymbol{P}\boldsymbol{B} \\ &= \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B} = \boldsymbol{Q}\end{aligned}$$

which follows from $\boldsymbol{A}\boldsymbol{A}^* + \boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}$ and the properties of $\boldsymbol{P}$. The equation $\boldsymbol{Q} = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B} = \boldsymbol{M}^{\dagger}\boldsymbol{B}$ yields $R(\boldsymbol{Q}) \subset R(\boldsymbol{M}^{\dagger})$. Conversely, $\boldsymbol{Q}\boldsymbol{B}^*\boldsymbol{P} = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^*\boldsymbol{P} = \boldsymbol{B}^*\boldsymbol{P} = \boldsymbol{M}^{\dagger}$ implies that $R(\boldsymbol{M}^{\dagger}) \subset R(\boldsymbol{Q})$. This proves that $R(\boldsymbol{Q}) = R(\boldsymbol{M}^{\dagger})$ and, therefore, completes the verification of a). We may check b) similarly. Let $\overline{\boldsymbol{Q}} = \boldsymbol{M}\boldsymbol{M}^{\dagger} = \boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^*\boldsymbol{P} = \boldsymbol{P}$. Then, $\overline{\boldsymbol{Q}}$ is clearly a projection operator. As before, we have $R(\boldsymbol{M}) \subseteq R(\boldsymbol{P})$ since $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{B}$, whereas $R(\boldsymbol{P}) \subseteq R(\boldsymbol{M})$ since $\boldsymbol{M}\boldsymbol{M}^{\dagger} = \boldsymbol{P}$. This proves b) and, hence, the lemma. □

Proof: Any matrix $\boldsymbol{C}$ that satisfies $\boldsymbol{C}\boldsymbol{A} = \boldsymbol{0}$ can be expressed as $\boldsymbol{C} = \boldsymbol{Y}\boldsymbol{P}$ for a suitable matrix $\boldsymbol{Y}$ of the same size as $\boldsymbol{C}$. Conversely, the matrix $\boldsymbol{Y}\boldsymbol{P}$ is a valid "$\boldsymbol{C}$ matrix" since $\boldsymbol{Y}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{0}$. The upshot is that we can replace $\boldsymbol{C}$ with $\boldsymbol{Y}\boldsymbol{P}$ in the minimization, thereby eliminating the constraint $\boldsymbol{C}\boldsymbol{A} = \boldsymbol{0}$ altogether. By a similar argument, we can replace $\boldsymbol{A}^L$ by $\boldsymbol{A}^{\dagger} + \boldsymbol{X}\boldsymbol{P}$. The sizes of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $q \times (L - q)$ and $(L - q) \times (L - q)$, respectively. Therefore, (15) transforms to

$$\min_{\boldsymbol{X},\boldsymbol{Y}} \left\| \begin{pmatrix} (\boldsymbol{A}^{\dagger} + \boldsymbol{X}\boldsymbol{P})\boldsymbol{B} \\ \boldsymbol{Y}\boldsymbol{P}\boldsymbol{B} - \boldsymbol{I} \end{pmatrix} \right\|_2 \equiv \min_{\boldsymbol{X},\boldsymbol{Y}} \left\| \begin{pmatrix} \boldsymbol{A}^{\dagger}\boldsymbol{B} \\ -\boldsymbol{I} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \boldsymbol{P}\boldsymbol{B} \right\|_2.$$

We can now apply Lemma 1 to the above problem to obtain the minimizing solution

$$\begin{pmatrix} \boldsymbol{X}_{\star} \\ \boldsymbol{Y}_{\star} \end{pmatrix} = -\begin{pmatrix} \boldsymbol{A}^{\dagger}\boldsymbol{B} \\ -\boldsymbol{I} \end{pmatrix} (\boldsymbol{P}\boldsymbol{B})^{\dagger}. \tag{B.1}$$

Next, observe that $(\boldsymbol{A}\ \boldsymbol{B})$ is a $q \times L$ submatrix of the $L \times L$ DFT matrix with possible rearrangements of columns. Therefore, its rows are orthonormal, and it satisfies $\boldsymbol{A}\boldsymbol{A}^* + \boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}$. Combining Lemma 3 with (B.1) yields the minimizing solution $\boldsymbol{X}_{\star} = \boldsymbol{A}^{\dagger}\boldsymbol{B}\boldsymbol{B}^*\boldsymbol{P}$ and $\boldsymbol{Y}_{\star} = -\boldsymbol{B}^*\boldsymbol{P}$. Again, using $\boldsymbol{A}\boldsymbol{A}^* + \boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}$ and the properties of $\boldsymbol{P}$, we find that $\boldsymbol{X}_{\star} = \boldsymbol{A}^{\dagger}\boldsymbol{B}\boldsymbol{B}^*\boldsymbol{P}$ simplifies to zero. Hence, the choice $\boldsymbol{A}_{\star}^L = \boldsymbol{A}^{\dagger}$ and $\boldsymbol{C}_{\star} =$

$\boldsymbol{B}^*\boldsymbol{P}^2 \equiv \boldsymbol{B}^*\boldsymbol{P}$ minimizes $||\boldsymbol{S}||_2$. Note that this solution also simultaneously minimizes all the singular values of $\boldsymbol{S}$ and, therefore, its Frobenius norm as well. We now compute the minimum norm of $\boldsymbol{S}$ at optimality, namely, $||\boldsymbol{S}_\star||_2$, where

$$\boldsymbol{S}_\star = \begin{pmatrix} \boldsymbol{D}_\star \\ \boldsymbol{F}_\star \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^\dagger\boldsymbol{B} \\ \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B} - \boldsymbol{I} \end{pmatrix}. \tag{B.2}$$

From the proof of Lemma 3, we know that $\boldsymbol{Q} = \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B}$ is a projection operator. Therefore, $-\boldsymbol{F}_\star = \boldsymbol{I} - \boldsymbol{Q}$ is also an orthogonal projection operator. The case $p = L$ is trivial, and it is easy to check that $\boldsymbol{S}_\star = \boldsymbol{0}$ for this case. Moreover, $\boldsymbol{F}_\star$ is nonzero if and only if $p < L$. We assume that $p < L$ (and hence $\boldsymbol{F}_\star \neq \boldsymbol{0}$) in the rest of the proof. The spectral norm of $\boldsymbol{S}_\star$ can be bounded from above as follows:

$$\begin{aligned} ||\boldsymbol{S}_\star||_2^2 &= \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star + \boldsymbol{F}_\star^*\boldsymbol{F}_\star) \\ &\leq \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star) + \lambda_{\max}(\boldsymbol{F}_\star^*\boldsymbol{F}_\star) \\ &= 1 + \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star). \end{aligned} \tag{B.3}$$

Now, observe that $-\boldsymbol{F}_\star^*\boldsymbol{D}_\star^* = \boldsymbol{D}_\star^* - \boldsymbol{B}^*\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^*(\boldsymbol{A}^\dagger)^* = \boldsymbol{D}_\star^* - \boldsymbol{B}^*\boldsymbol{P}(\boldsymbol{A}^\dagger)^* = \boldsymbol{D}_\star^*$. Therefore, $R(\boldsymbol{D}_\star^*) \subset R(\boldsymbol{F}_\star^*)$, or equivalently, $R(\boldsymbol{D}_\star^*\boldsymbol{D}_\star) \subset R(\boldsymbol{F}_\star^*\boldsymbol{F}_\star)$. Let $\boldsymbol{\xi}$ be an eigenvector corresponding to the largest eigenvalue of $\boldsymbol{D}_\star^*\boldsymbol{D}_\star$, i.e., $\boldsymbol{D}_\star^*\boldsymbol{D}_\star\boldsymbol{\xi} = \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star)\boldsymbol{\xi}$. Then, $\boldsymbol{F}_\star^*\boldsymbol{F}_\star\boldsymbol{\xi} = \boldsymbol{\xi}$ because $\boldsymbol{\xi} \in R(\boldsymbol{F}_\star^*\boldsymbol{F}_\star)$. Therefore, $(\boldsymbol{D}_\star^*\boldsymbol{D}_\star + \boldsymbol{F}_\star^*\boldsymbol{F}_\star)\boldsymbol{\xi} = (1 + \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star))\boldsymbol{\xi}$, and we have $||\boldsymbol{S}_\star||_2^2 \geq 1 + \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star)$. Combining this with (B.3), we obtain

$$||\boldsymbol{S}_\star||_2^2 = 1 + \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star). \tag{B.4}$$

Finally, a simple calculation reveals that

$$\begin{aligned} \boldsymbol{D}_\star\boldsymbol{D}_\star^* &= \boldsymbol{A}^\dagger\boldsymbol{B}\boldsymbol{B}^*(\boldsymbol{A}^\dagger)^* = \boldsymbol{A}^\dagger(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^*)(\boldsymbol{A}^\dagger)^* = \boldsymbol{A}^\dagger(\boldsymbol{A}^\dagger)^* - \boldsymbol{I} \\ &= (\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{A})^{-1} - \boldsymbol{I} = (\boldsymbol{A}^*\boldsymbol{A})^{-1} - \boldsymbol{I}. \end{aligned} \tag{B.5}$$

Therefore, $\lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star) = \lambda_{\max}(\boldsymbol{D}_\star^*\boldsymbol{D}_\star) = \lambda_{\max}((\boldsymbol{A}^*\boldsymbol{A})^{-1}) - 1$, and hence, (B.4) yields $||\boldsymbol{S}_\star||_2 = \sqrt{\lambda_{\max}((\boldsymbol{A}^*\boldsymbol{A})^{-1})}$ for $p < L$. □

C. Proof of Theorem 3

We can rewrite the optimization as

$$\min_{\alpha \geq \alpha_0} \left(\alpha \max_{f \in [0,\alpha)} \sum_{r=0}^{L-1} \nu(f + r\alpha) \right) \tag{C.1}$$

where $\nu(f) = \chi(f \in \mathcal{F})$, $\alpha = 1/LT$, and $\alpha_0 = 1/LT_0$. The set $\mathcal{F} = \bigcup_i [a_i, b_i)$ is assumed to satisfy $0 = a_1 < b_1 < a_2 < \ldots, < s_n < b_n = 1/T_0$. The function $\nu(f)$ is right continuous and has jump discontinuities of $+1$ and -1 at a_i and b_i, respectively. Hence, we obtain the following properties of $\nu(f)$: a) $\nexists a_i \in (f_1, f_2] \Longrightarrow \nu(f_1) \geq \nu(f_2)$, and b) $\nexists b_i \in (f_1, f_2] \Longrightarrow \nu(f_1) \leq \nu(f_2)$.We use these properties without explicitly stating them. Now, suppose that the optimum value $\alpha_\star$ does not satisfy $k\alpha_\star = b_j - a_i$ for any $i \leq j$ and $k \geq 0$. We will show that the choice $\alpha = \alpha_\star - \delta$ yields a smaller objective function than $\alpha = \alpha_\star$ for a sufficiently small δ, which is a contradiction. In particular, we prove that $\overline{q} \leq q$, where q and $\overline{q}$ are defined as

$$q \stackrel{\text{def}}{=} \max_{f \in [0,\alpha_\star)} \sum_{r=0}^{L-1} \nu(f + r\alpha_\star)$$

and

$$\overline{q} \stackrel{\text{def}}{=} \max_{f \in [0,\alpha_\star - \delta)} \sum_{r=0}^{L-1} \nu(f + r(\alpha_\star - \delta)) \tag{C.2}$$

for an appropriately chosen δ, and this implies that $(\alpha_\star - \delta)\overline{q} \leq \alpha_\star q$. Hence, the optimality of $\alpha_\star$ is contradicted. We start by choosing δ:

$$0 < \delta < \frac{1}{L} \min\{|b_j - a_i - k\alpha_\star| : i \leq j,\ 0 < k \leq L\}. \tag{C.3}$$

Since $\delta < L^{-1}|b_n - a_1 - L\alpha_\star|$, we see that $\alpha \stackrel{\text{def}}{=} \alpha_\star - \delta \geq (b_n - a_1)/L \equiv \alpha_0$ is feasible to (C.1). Let $f \in [0, \alpha_\star - \delta)$ be fixed. If there does not exist a b_j such that $b_j \in (f + r(\alpha_\star - \delta), f + r\alpha_\star]$ for any $r \in \mathcal{L} = \{0, \ldots, L-1\}$, then it follows that $\nu(f + r(\alpha_\star - \delta)) \leq \nu(f + r\alpha_\star)$ holds for every $r \in \mathcal{L}$, and hence, $\overline{q} \leq q$ follows from (C.2). Otherwise, let r_0 be the largest integer in $\mathcal{L}$ such that

$$f + r_0(\alpha_\star - \delta) < b_{j_0} \leq f + r_0\alpha_\star \tag{C.4}$$

for some b_{j_0}. We now claim that there does not exist an $r \in \{0, \ldots, r_0 - 1\}$ and a_i such that

$$f + r\alpha_\star - r\delta \geq a_i > f + r\alpha_\star - r_0\delta. \tag{C.5}$$

This is easily justified because subtracting (C.5) from (C.4) gives

$$\begin{aligned} &(r_0 - r)\alpha_\star + (r - r_0)\delta < b_{j_0} - a_i < (r_0 - r)\alpha_\star + r_0\delta \\ &\quad \Longrightarrow (r - r_0)\delta < b_{j_0} - a_i - (r_0 - r)\alpha_\star < r_0\delta \\ &\quad \Longrightarrow |b_{j_0} - a_i - (r_0 - r)\alpha_\star| < \max(r_0, r_0 - r)\delta < L\delta \end{aligned}$$

which contradicts the choice of δ in (C.3). A consequence of this is that

$$\nu(f + r\alpha_\star - r\delta) \leq \nu(f + r\alpha_\star - r_0\delta), \qquad 0 \leq r < r_0. \tag{C.6}$$

Next, for $r_0 \leq r \leq L - 1$, we know from the definition of r_0 that there does not exist a $b_j \in (f + r(\alpha_\star - \delta), f + r\alpha_\star] \supset (f + r(\alpha_\star - \delta), f + r\alpha_\star - r_0\delta]$. Hence

$$\nu(f + r\alpha_\star - r\delta) \leq \nu(f + r\alpha_\star - r_0\delta), \qquad r_0 \leq r \leq L - 1. \tag{C.7}$$

Summing (C.6) and (C.7) over their respective ranges and adding together yields

$$\sum_{r=0}^{L-1} \nu(f + r\alpha_\star - r\delta) \leq \sum_{r=0}^{L-1} \nu(f' + r\alpha_\star) \leq q \tag{C.8}$$

where $f' = f - r_0\delta$. The second inequality in (C.8) follows from the definition of q. Of course, we need to verify that $f' \geq 0$. This is true because, otherwise, if $f' = f - r_0\delta < 0$, the choice $r = 0$, $a_i = a_1 = 0$ would serve as a counterexample to the claim. Since (C.8) holds for each $f \in [0, \alpha_\star - \delta)$, we obtain $\overline{q} \leq q$. This proves the original statement that $1/LT_\star = (b_j - a_i)/k$

for some $1 \leq i \leq j \leq n$ and $0 < k \leq L$. Furthermore, the condition $T \leq T_0$ restricts k to $0 < k \leq T_0 L(b_j - a_i) \leq L$ for given indices i and j.

REFERENCES

[1] A. J. Jerri, "The Shannon sampling theorem—its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.

[2] R. J. Marks, II, Ed., *Advanced Topics in Shannon Sampling and Interpolation Theory*. New York: Springer-Verlag, 1993.

[3] A. Zayed, *Advances in Shannon's Sampling Theory*. Boca Raton, FL: CRC, 1993.

[4] J. R. Higgins, *Sampling Theory in Fourier and Signals Analysis Foundations*. Oxford, U.K.: Oxford, 1996.

[5] C. E. Shannon, "Communication in the presence of noise," in *Proc. IRE*, vol. 37, Jan. 1949, pp. 10–21.

[6] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, 1967.

[7] R. G. Shenoy, "Nonuniform sampling of signals and applications," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1994, vol. 2, pp. 181–184.

[8] R. E. Kahn and B. Liu, "Sampling representations and the optimum reconstruction of signals," *IEEE Trans. Inform. Theory*, vol. 1, 1965.

[9] K. Cheung and R. Marks, "Image sampling below the Nyquist density without aliasing," *J. Opt. Soc. Amer. A*, Jan. 1990.

[10] P. Vaidyanathan and V. Liu, "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1927–1936, Nov. 1990.

[11] K. Cheung, "A multidimensional extension of Papoulis' generalized sampling expansion with the application in minimum density sampling," in *Advanced Topics in Shannon Sampling and Interpolation Theory*, R. J. Marks, II, Ed. New York: Springer-Verlag, 1993, ch. 3.

[12] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multi-band signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Atlanta, GA, May 1996.

[13] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1555–1564, July 1999.

[14] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulae and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inform. Theory.*, vol. 46, pp. 2173–2183, Sept. 2000.

[15] R. N. Bracewell, *Two Dimensional Imaging*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[16] Z. P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging*. Piscataway, NJ: IEEE, 1999.

[17] M. Soumekh, *Synthetic Aperture Radar Signal Processing with MATLAB Algorithms*. New York: Wiley, 1999.

[18] R. J. Korzig and S. A. Kassam, "Synthetic aperture pulse-echo imaging with rectangular bounded arrays," *IEEE Trans. Image Processing*, vol. 2, pp. 68–79, Jan. 1993.

[19] H. G. Feichtinger and K. Grochenig, "Theory and practice of irregular sampling," in *Wavelets: Mathematics and Applications*. Boca Raton, FL: CRC, 1994, ch. 8.

[20] K. Grochenig, "Reconstruction algorithms in irregular sampling theory," *Comput. Math.*, vol. 59, pp. 181–194, 1992.

[21] J. J. Benedetto, "Irregular sampling and frames," in *Wavelets—A Tutorial in Theory and Applications*. New York: Academic, 1992, ch. 6.

[22] Y. C. Eldar and A. V. Oppenheim, "Filter bank reconstruction of bandlimited signals from nonuniform and generalized sampling," *IEEE Trans. Signal Processing*, vol. 48, pp. 2864–2875, Oct. 2000.

[23] K. Gosse and P. Duhamel, "Perfect reconstruction versus MMSE filter banks in source coding," *IEEE Trans. Signal Processing*, vol. 45, pp. 2188–2202, Sept. 1997.

[24] E. J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces*. New York: Wiley, 1987.

[25] P. J. S. G. Ferreira, "The eigenvalues of matrices that occur in certain interpolation problems," *IEEE Trans. Signal Processing*, vol. 45, pp. 2115–2120, Aug. 1997.

[26] M. M. Dodson and A. M. Silva, "Fourier analysis and the sampling theorem," *Proc. R. Irish Acad.*, vol. 85, pp. 81–108, 1985.

[27] M. G. Beaty and J. R. Higgins, "Aliasing and Poisson summation in the sampling theory of Paley–Wiener spaces," *J. Fourier Anal. Appl.*, vol. 1, pp. 67–85, 1994.

Raman Venkataramani received the B.Tech. degree from the Indian Institute of Technology, Madras, in 1995 and the M.S. degree from Johns Hopkins University, Baltimore, MD, in 1997. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Illinois at Urbana-Champaign.

His interests include statistical signal processing, signal sampling and interpolation, and inverse problems.

Yoram Bresler (F'99) received the B.Sc. (cum laude) and M.Sc. degrees from the Technion, Israel Institute of Technology, Haifa, in 1974 and 1981, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 1986, all in electrical engineering.

From 1974 to 1979, he served as an electronics engineer in the Israeli Defense Force. From 1985 to 1987, he was a Research Associate with the Information Systems Laboratory, Stanford University, working on sensor array processing and medical imaging. In 1987, he joined the University of Illinois at Urbana-Champaign, where he is currently a Professor with the Department of Electrical and Computer Engineering and the Bioengineering Program and a Research Professor with the Coordinated Science Laboratory. From 1995 to 1996, he spent a sabbatical leave at the Technion. His current research interests include multidimensional and statistical signal processing and their applications to inverse problems in imaging. He is is currently on the editorial board of *Machine Vision and Applications*.

Dr. Bresler was an Associate Editor for the IEEE TRANSACTIONS FOR IMAGE PROCESSING from 1992 to 1993 and a member of the IEEE Image and Multidimensional Signal Processing Technical Committee from 1994 to 1998. In 1988 and 1989, he received the Senior Paper Awards from the IEEE Acoustics, Speech, and Signal Processing Society. He is the recipient of a 1991 NSF Presidential Young Investigator Award, the Technion Fellowship in 1995, and the Xerox Senior Award for Faculty Research in 1998. He was named a University of Illinois Scholar in 1999 and was appointed Associate at the Center for Advanced Study of the University for 2001–2002.