

ON THE NECESSARY DENSITY FOR SPECTRUM-BLIND NONUNIFORM SAMPLING SUBJECT TO QUANTIZATION

Michael Gastpar

Audio Visual Communications Laboratory
Swiss Federal Institute of Technology
Lausanne, Switzerland
Email: gastpar@lcvsun1.epfl.ch

Yoram Bresler

Coordinated Science Laboratory
Department of ECE
U. of Illinois, Urbana IL 61801
Email: ybresler@uiuc.edu

ABSTRACT

It is known that in the absence of distortion, the necessary sampling density for a multiband signal is given by its spectral occupancy. However, in general, the samples have to be acquired nonuniformly. There exist sampling patterns such that reconstruction is feasible even if the actual spectral support of the multiband signal is not known. If the samples are distorted, an increased sampling density may lead to a superior performance. In this paper, we consider the case of small distortion due to fine quantization of the samples, and we derive a necessary condition on the optimal sampling density.

1. INTRODUCTION

Band-limited signals have a spectrum that is supported (nonzero) on exactly one finite interval of the frequency domain. They may be represented without loss by periodic samples. The necessary sampling density is equal (by the sampling theorem) to the bandwidth of the signal. When the spectrum of a signal is supported on a finite union of multiple (finite) frequency intervals, it is called a multiband signal. The length of the union of these intervals is called the spectral occupancy. It has been shown [1, 2, 3, 4, 5] that a multiband signal can also be represented without loss by samples of an average density equal to the spectral occupancy, thus achieving the Landau lower bound [6] (also called the Landau-Nyquist density). However, this is not achieved by uniformly spaced samples, requiring nonuniform sampling patterns instead. As an important additional development, it has been shown that, for certain *universal sampling patterns*, the reconstruction can be performed without prior knowledge of the spectral support of the signal [4]. This so-called *spectrum-blind* reconstruction recovers both the spectral support information and the signal itself from the samples acquired on a universal pattern at the Landau-Nyquist density.

In this paper, we consider the representation of finite-length discrete-time multiband signals (sequences) by *distorted* samples, e.g. due to quantization noise. In general, both the spectral support reconstruction, and that of the signal

itself will then suffer distortion. Even if the spectral support is reconstructed perfectly (an event of positive probability, because of the finite number of possibilities involved), the reconstructed signal will be subject to some distortion due to noise (see [7]). However, erroneous support recovery will result in a large distortion in the final reconstruction. The question we address, therefore, is to determine the sampling density (higher than the Landau-Nyquist density) required to ensure a fixed probability of error in the reconstruction of the spectral support. In a rate-distortion sense, if a particular overall bit rate budget is available, it is generally not optimal to sample at the minimum possible density; rather, increasing the sampling density (and thus reducing the available rate per sample) may improve the overall performance. In this work, we compute a necessary condition on the optimal sampling density.

By the duality of the time and frequency domain, our results apply also to the case of signals that are sparse in the time domain while the samples are taken in the frequency domain. From this perspective, nonuniform sampling may be considered a compression technique for sparse sequences in which the location of the non-zeroes (the digital information) and their values (the analog information) are encoded jointly. The compression algorithm would include a discrete Fourier transform followed by nonuniform sampling. Moreover, in Fourier imaging systems, in cases where the image itself is sparse, nonuniform sampling allows to reduce the number of necessary samples [8]. This may be of interest for instance in magnetic resonance imaging (MRI) of moving objects.

2. NONUNIFORM SAMPLING

Consider a complex-valued length- N sequence x with discrete Fourier transform (DFT) X . For convenience, we use vector notation, i.e. $x, X \in \mathbb{C}^N$. In this paper, we use the unitary form of the DFT, i.e.

$$X(m) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) W^{nm}$$

where $W = e^{-j\frac{2\pi}{N}}$. Let x be a multiband sequence of spectral occupancy q/N , i.e. let X have (at most) q non-zero components in arbitrary locations, indexed by $\underline{K} =$

THIS WORK WAS SUPPORTED IN PART BY NSF GRANT NO. MIP 97-07633 AND DARPA CONTRACT F49620-98-1-0498, ADMINISTERED BY AFOSR.

$\{k_1, \dots, k_q\}$, where $k_i \in \{0, \dots, L-1\}$. Define the vector $x_{\underline{c}}$ containing only p of the N components of x , at locations indexed by $\underline{c} = \{c_1, \dots, c_p\}$. $x_{\underline{c}}$ are the nonuniform samples.

The data vector x has to be reconstructed from the nonuniform samples $x_{\underline{c}}$ in a spectrum blind fashion, that is, without prior knowledge of \underline{K} . It has been shown that for $p > q$, there exist sets \underline{c} for which this is feasible [4]. In particular, if the nonzero components of X are distributed according to an absolutely continuous probability measure then this is true w.p.1 for $p = q + 1$ [7].

Notice that for $(p, q, N) \rightarrow \infty$ at fixed ratio $\Omega = q/N$, this means that the necessary sampling density is again equal to the spectral occupancy of the signal. As mentioned, this is achieved by certain universal nonuniform sampling patterns. The efficient design of such patterns that perform optimally in the presence of distortion in the samples is a problem by itself, which has not been completely solved to date.

For convenience, we introduce a matrix notation. In particular, we will use submatrices of the inverse DFT matrix. It follows directly from the definitions that we may write $x_{\underline{c}} = A_{\underline{c}}X$, where $A_{\underline{c}}$ is the inverse DFT-matrix with only the rows corresponding to the indices in \underline{c} retained. Let S denote the vector containing only the q nonzero entries in X , and let $A_{\underline{c}, \underline{K}}$ denote the submatrix of the inverse DFT matrix that is obtained by only retaining the rows with indices in \underline{c} and the columns with indices in \underline{K} . With these definitions, we will write $x_{\underline{c}} = A_{\underline{c}, \underline{K}}S$.

In this paper, we are interested in the case of distorted samples. Therefore, we introduce the additive noise z and define the noisy samples as

$$y_{\underline{c}} = x_{\underline{c}} + z = A_{\underline{c}, \underline{K}}S + z. \quad (1)$$

For our further arguments, we will use this representation of nonuniform sampling.

As an aside, we can also clarify now the definition of a *universal sampling pattern*. A sampling pattern \underline{c} is universal, if every p columns of $A_{\underline{c}}$ are linearly independent. For any L there exist many universal patterns [4]. For example, one of them is the bunched pattern, $\underline{c} = \{0, 1, \dots, p-1\}$. Its universality follows immediately from the Vandermonde structure of $A_{\underline{c}}$.

3. OPTIMUM SAMPLING DENSITY

In this paper, we compute the required sampling density $\rho = p/N$ in a statistical sense: we find the necessary sampling density to achieve a specified probability of error in recovering \underline{K} from the data. Therefore, we need a statistical description of the multiband signals to be sampled. Let the location of the q nonzero components of X be distributed uniformly over all possibilities, and let their (complex) values be distributed as circularly normal, $S \sim \mathcal{N}_c(0, \sigma_S^2 I_q)$. Moreover, let the additive distortion be (complex) white Gaussian noise, $z \sim \mathcal{N}_c(0, \sigma^2 I_p)$. We define the signal-to-noise ratio (SNR) $\beta = \sigma_S^2 / \sigma^2$.

In our approach, we consider mutual informations:

$$\begin{aligned} I(x_{\underline{c}}; y_{\underline{c}}) &= I((S, \underline{K}); y_{\underline{c}}) \\ &= I(\underline{K}; y_{\underline{c}}) + I(S; y_{\underline{c}} | \underline{K}), \end{aligned}$$

where the first equality follows from the data processing lemma and the fact that S and \underline{K} uniquely determine $x_{\underline{c}}$, and the second equality from the chain rule for mutual informations. Therefore,

$$\max I(x_{\underline{c}}; y_{\underline{c}}) \geq I(\underline{K}; y_{\underline{c}}) + I(S; y_{\underline{c}} | \underline{K}), \quad (2)$$

where the max is taken over all distributions of $x_{\underline{c}}$ with second moment equal to the one of the our considered source.

First, consider the term on the left in Eqn. (2). There is a (complex) additive white Gaussian noise (AWGN) channel between $x_{\underline{c}}$ and $y_{\underline{c}}$. The signal power is $\Omega \sigma_S^2$ and the noise power is σ^2 . Therefore,

$$\max I(x_{\underline{c}}; y_{\underline{c}}) = p C_{AWGN} = p \log_2(1 + \Omega \beta),$$

where C_{AWGN} denotes the channel capacity.

Second, consider the first term on the right in Eqn. (2). This is the mutual information across the digital channel from \underline{K} to $y_{\underline{c}}$. Recall that \underline{K} is the location of the nonzero values in the spectrum X . Erroneous recovery will result in a large distortion in the final reconstruction. For a desired probability of error P_e in the recovery of \underline{K} , a lower bound on the mutual information follows from Fano's inequality:

$$I(\underline{K}; y_{\underline{c}}) \geq H(\underline{K}) - H_b(P_e) - P_e \log_2 \left(\binom{N}{q} - 1 \right) \quad (3)$$

Last, consider the second term on the right in Eqn. (2). It is the mutual information across the channel between S and $y_{\underline{c}}$, $I(S; y_{\underline{c}} | \underline{K})$. This is also a Gaussian channel,

$$y_{\underline{c}} = A_{\underline{c}, \underline{K}}S + z. \quad (4)$$

In our scenario, the covariance matrix Σ_S of S is known. This determines the mutual information that is achieved across the channel; in particular, we do not achieve capacity. As stated above, we assume $\Sigma_S = \sigma_S^2 I_q$. For given value $\underline{K} = \underline{k}$, the mutual information is

$$I(S; y_{\underline{c}} | \underline{K} = \underline{k}) = \log_2 \det \left(I + \beta A_{\underline{c}, \underline{k}}^H A_{\underline{c}, \underline{k}} \right). \quad (5)$$

Then, $I(S; y_{\underline{c}} | \underline{K})$ is found taking the expectation over all instances of \underline{K} . This proves the following statement:

Lemma 1 (Necessary Condition). *The sampling density $\rho = p/N$ has to satisfy*

$$\begin{aligned} &\rho \log_2(1 + \beta \Omega) \\ &\geq \frac{1}{N} \left(\log_2 \binom{N}{q} - H_b(P_e) - P_e \log_2 \left(\binom{N}{q} - 1 \right) \right) \\ &\quad + \frac{1}{N} E_{\underline{K}} \log_2 \det \left(I + \beta A_{\underline{c}, \underline{K}}^H A_{\underline{c}, \underline{K}} \right), \end{aligned} \quad (6)$$

where $\beta = \frac{\sigma_S^2}{\sigma^2}$ is the SNR with σ_S^2 and σ^2 the variances of the signal and noise, respectively, and P_e the probability of erroneous spectral support recovery.

For fixed signal and noise parameters and sampling pattern \underline{c} , this expectation can be numerically evaluated by summing over all possible instances of the random variable \underline{K} . However, we are particularly interested in the case $N \rightarrow \infty$, where numerical evaluation is not an option.

In order to gain insight, let us make a particular choice of the sampling pattern. First, note that $I(S; y_{\underline{c}} | \underline{K})$ is a

function of the sampling pattern \underline{c} . The bound in the lemma is tighter for \underline{c} with large $I(S; y_{\underline{c}}|\underline{K})$ (since all other terms do not depend on \underline{c}). Therefore, it makes sense to consider a pattern that makes this mutual information as large as possible. This will yield an upper bound on the optimal ρ . We consider the idealized situation where \underline{c} has been chosen such that the matrix $A_{\underline{c}, \underline{k}}$ has orthogonal columns for all \underline{k} . Note that such a pattern does not exist in general. Then,

$$\begin{aligned} I(S; y_{\underline{c}}|\underline{K} = \underline{k}) &= \log_2 \det (I + \beta A_{\underline{c}, \underline{k}}^H A_{\underline{c}, \underline{k}}) \\ &= \log_2 \det \left(I + \beta \frac{p}{N} I \right) \\ &= q \log_2 (1 + \beta \rho) \end{aligned} \quad (7)$$

Using this in Lemma 1, we find

$$\begin{aligned} \rho \log_2 (1 + \beta \Omega) &\geq \frac{1}{N} (\log_2 \binom{N}{q} - H_b(P_e) - P_e \log_2 (\binom{N}{q} - 1)) \\ &\quad + \frac{q}{N} \log_2 (1 + \beta \rho) \end{aligned} \quad (8)$$

Letting $N \rightarrow \infty$, we obtain

$$\rho \log_2 (1 + \beta \Omega) \geq (1 - P_e) H_b(\Omega) + \Omega \log_2 (1 + \beta \rho) \quad (9)$$

In this idealized framework, the optimal sampling density is the one that achieves equality in Eqn. (9).

4. MARGINAL-BASED QUANTIZATION OF THE SAMPLES

Let us now consider the case where the distortion of the samples is due to quantization. We study the case of a quantizer that does not exploit the correlation between the samples. We call this marginal-based quantization. (Note that this does *not* imply scalar quantization.) In order to apply the results derived above, we will model the distortion due to this quantization by additive white Gaussian noise. Therefore, our discussion only applies to high rate quantization. The marginal density of the samples (components of $x_{\underline{c}}$) in our scenario is Gaussian with zero mean and variance $\sigma_S^2 \Omega$. Let the rate available per symbol of the original source x be R . Then, the rate to encode the length N block x at hand is NR , and the rate available to encode one sample is $RN/p = R/\rho$. Thus, using the rate-distortion function for a complex Gaussian iid source the distortion per sample due to quantization, and hence the variance of the equivalent additive noise becomes

$$\sigma^2 = \sigma_S^2 \Omega 2^{-R/\rho},$$

from which we find that the SNR $\beta = 2^{R/\rho}/\Omega$. Plugging this into Eqn. (9), we obtain

$$\begin{aligned} \rho \log_2 (1 + 2^{R/\rho}) &\geq (1 - P_e) H_b(\Omega) + \Omega \log_2 \left(1 + \frac{\rho}{\Omega} 2^{R/\rho} \right) \end{aligned} \quad (10)$$

This is a transcendental inequality. However, we may numerically determine the ρ that achieves equality. This is

illustrated in Fig. 1 for a particular choice of the parameters. Within the assumptions of this paper, the figure indicates the optimum sampling redundancy, by which we mean the difference between the sampling density and the spectral occupancy, for various coding rates. As expected, for very high rates, the optimum sampling redundancy goes to zero.

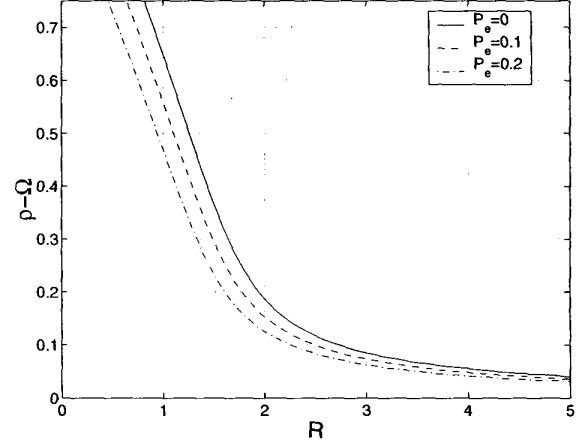


Figure 1: Sampling redundancy $\rho - \Omega$ versus the coding rate R for various P_e for $\Omega = 3/14$.

5. THE DISTORTION

In this section, we compute an approximation to the distortion achieved by a particular choice of the parameters. To this end, we consider the events $\{|\hat{K} \cap K| = m\}$. We assume that first, \hat{K} is determined from the data, whereafter S is recovered using the pseudo-inverse. This is a suboptimal reconstruction procedure, but it allows the consideration of the incurred distortion. Denoting by $U_{\underline{K}}$ the matrix such that $U_{\underline{K}} S$ is a vector of length N containing the q entries of S in the locations indexed by \underline{K} , we can write

$$\begin{aligned} D_m &= (1/N) E \| U_{\hat{K}} A_{\hat{K}}^\dagger (A_{\underline{K}} S + z) - U_{\underline{K}} S \|^2 \\ &= (1/N) \left(E \| U_{\hat{K}} A_{\hat{K}}^\dagger A_{\underline{K}} S - U_{\underline{K}} S \|^2 + E \| A_{\hat{K}}^\dagger z \|^2 \right), \end{aligned}$$

where we have used the simplified notation $A_{\underline{K}} = A_{\underline{c}, \underline{K}}$ since in this discussion, the sampling pattern \underline{c} is fixed. The average distortion can thus be written as $D = E[D_m]$, where the expectation is over the events defined above. In this paper, we only consider an upper bound on the distortion, $D \leq (1 - P_e) D_q + P_e D_{max}$. For $\Omega \leq 1/2$, $D_{max} = D_0$. However, for $\Omega > 1/2$, the event $\{|\hat{K} \cap K| = 0\}$ cannot occur. In fact, for $\Omega > 1/2$, $D_{max} = D_{(2\Omega-1)N}$. We therefore need to compute the distortion when $|\hat{K} \cap K| = m$. Denote by $\underline{K}' = \underline{K} \setminus (\underline{K} \cap \hat{K})$ the elements in \underline{K} that do not occur in the reconstruction, and correspondingly $\hat{K}' = \hat{K} \setminus (\underline{K} \cap \hat{K})$. Then, we may write

$$D_m = (1/N) ((q - m) \sigma_S^2 + E \| A_{\hat{K}'}^\dagger A_{\underline{K}'} S \|^2 + E \| A_{\hat{K}'}^\dagger z \|^2).$$

To gain insight, let us again invoke the assumption that all the columns of the matrix $A_{\underline{K}}$ are mutually orthogonal. Then,

$$\begin{aligned}
E\|A_{\underline{K}}^\dagger A_{\underline{K}'} S\|^2 &= E\|A_{\underline{K}}^\dagger A_{\underline{K}'}\|_F^2 \sigma_S^2 \\
&\stackrel{(a)}{=} E\left\|\frac{N}{p} A_{\underline{K}}^H A_{\underline{K}'}\right\|_F^2 \sigma_S^2 \\
&= \sigma_S^2 \frac{N^2}{p^2} \frac{1}{N^2} E \sum_{i,j} \left|W^{c_1(k'_i - \hat{k}'_j)} + \dots + W^{c_p(k'_i - \hat{k}'_j)}\right|^2 \\
&\stackrel{(b)}{=} \sigma_S^2 \frac{N^2}{p^2} \frac{1}{N^2} (q-m)^2 E_{d \neq 0} |W^{c_1 d} + \dots + W^{c_p d}|^2 \\
&\stackrel{(c)}{=} \sigma_S^2 \frac{1}{p^2} (q-m)^2 \frac{1}{N-1} \sum_{d=1}^{N-1} |W^{c_1 d} + \dots + W^{c_p d}|^2 \\
&= \sigma_S^2 \frac{(q-m)^2}{N-1} \frac{1}{p^2} (Np - p^2) = \sigma_S^2 \frac{(q-m)^2}{N-1} \left(\frac{1}{\rho} - 1\right)
\end{aligned}$$

where, for (a), we used the orthogonal column assumption, for (b), we substituted $d_{ij} = k'_i - \hat{k}'_j$ and assumed that the d_{ij} are identically distributed, and for (c), we assumed that all d are equally likely. Note that under these assumptions, negative values of d need not be considered explicitly. The sum may be evaluated by noting its similarity to the magnitude-squared of the inverse DFT of a length- N sequence with ones at locations indexed by \underline{c} and zeroes otherwise. Then, the result follows from Parseval's formula. To compute the suggested upper bound on D , it remains to find

$$D_q = \frac{1}{N} E\|A_{\underline{K}}^\dagger z\|^2 = \frac{1}{N} E\|A_{\underline{K}}^\dagger\|_F^2 \sigma^2.$$

With the assumption of orthogonal columns, we obtain

$$D_q = \frac{q}{p} \sigma^2 = \frac{\Omega}{\rho} \sigma^2.$$

Thus, for $\Omega \leq 1/2$, $D_{max} = D_0$ and the output noise-to-signal ratio for $N \rightarrow \infty$ is

$$NSR := \frac{D}{\Omega \sigma_S^2} \leq \frac{1}{\rho \beta} + \left(1 + \Omega \left(\frac{1}{\rho} - 1\right)\right) P_e,$$

where β is the input signal-to-noise ratio, as defined earlier. For $\Omega > 1/2$, $D_{max} = D_{(2\Omega-1)N}$ and the output noise-to-signal ratio for $N \rightarrow \infty$ is

$$\frac{D}{\Omega \sigma_S^2} \leq \frac{1}{\rho \beta} + \left(1 + (1 - \Omega) \left(\frac{1}{\rho} - 1\right)\right) \frac{1 - \Omega}{\Omega} P_e,$$

We may thus derive an expression for P_e in terms of NSR , and substitute this in Eqn. (9). Furthermore, we can also substitute $\beta = 2^{R/\rho}/\Omega$ and replace P_e in Eqn. (10). This derivation allows therefore to specify a desired output noise-to-signal ratio rather than a desired resulting probability of error P_e in the spectral support recovery. Fig. 2 illustrates Eqn. (10) with P_e replaced by NSR according to the last two equations for a particular choice of the parameters. For fixed rate, the figure shows the optimum sampling redundancy for various Ω . For signals that are spectrally very sparse, we observe for instance that (nearly) vanishing sampling redundancy is optimum. However, for $\Omega \rightarrow 1$, the graph should be interpreted with caution: our setup explicitly excludes sampling rates $\rho > 1$ (see Eqn. (1)).

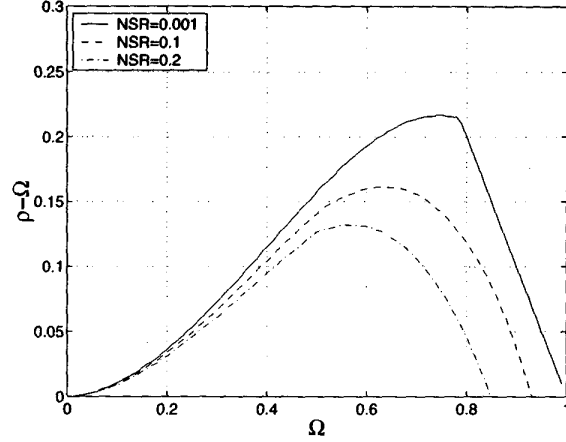


Figure 2: Sampling redundancy $\rho - \Omega$ for various NSR for $R = 5$.

6. REFERENCES

- [1] R. E. Kahn and B. Liu, "Sampling representations and the optimum reconstruction of signals," *IEEE Trans. Info. Theory*, vol. 11, pp. 339-347, 1965.
- [2] K. Cheung and R. Marks, "Image sampling below the Nyquist density without aliasing," *J. Opt. Soc. Am. A*, vol. 7, pp. 92-105, 1990.
- [3] P. Vaidyanathan and V. Liu, "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Trans. ASSP*, vol. 38, pp. 1927-1936, Nov. 1990.
- [4] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, Atlanta, GA, May 1996, vol. 3, pp. 1689-1692.
- [5] C. Herley and P. W. Wong, "Minimum rate sampling of signals with arbitrary frequency support," in *Proc. IEEE Int. Conf. Image Proc.*, (Lausanne, Switzerland), IEEE, Sept. 1996.
- [6] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, pp. 37-52, 1967.
- [7] R. Venkataramani and Y. Bresler, "Sub-Nyquist sampling of multiband signals: Perfect reconstruction and bounds on aliasing error," in *Proc. IEEE Int. Conf. Acoust. Speech, Sig. Proc.*, Seattle, Apr. 1998, vol. 3, pp. 1633-1636. (Also, submitted to *IEEE Trans. Infor. Theory*.)
- [8] Y. Bresler, M. Gastpar and R. Venkataramani, "Image compression on-the-fly by universal sampling in Fourier imaging systems," in *Proc. 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification, and Imaging*, Santa-Fe, NM, p. 48, Feb. 1999.