

Data recovery with sub-Nyquist sampling: fundamental limit and a detection algorithm^{*#}

Xiqian LUO, Zhaoyang ZHANG^{†‡}

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

[†]E-mail: ning_ming@zju.edu.cn

Received June 28, 2019; Revision accepted Nov. 26, 2019; Crosschecked Oct. 20, 2020

Abstract: While the Nyquist rate serves as a lower bound to sample a general bandlimited signal with no information loss, the sub-Nyquist rate may also be sufficient for sampling and recovering signals under certain circumstances. Previous works on sub-Nyquist sampling achieved dimensionality reduction mainly by transforming the signal in certain ways. However, the underlying structure of the sub-Nyquist sampled signal has not yet been fully exploited. In this paper, we study the fundamental limit and the method for recovering data from the sub-Nyquist sample sequence of a linearly modulated baseband signal. In this context, the signal is not eligible for dimension reduction, which makes the information loss in sub-Nyquist sampling inevitable and turns the recovery into an under-determined linear problem. The performance limits and data recovery algorithms of two different sub-Nyquist sampling schemes are studied. First, the minimum normalized Euclidean distances for the two sampling schemes are calculated which indicate the performance upper bounds of each sampling scheme. Then, with the constraint of a finite alphabet set of the transmitted symbols, a modified time-variant Viterbi algorithm is presented for efficient data recovery from the sub-Nyquist samples. The simulated bit error rates (BERs) with different sub-Nyquist sampling schemes are compared with both their theoretical limits and their Nyquist sampling counterparts, which validates the excellent performance of the proposed data recovery algorithm.

Key words: Nyquist-Shannon sampling theorem; Sub-Nyquist sampling; Minimum Euclidean distance; Under-determined linear problem; Time-variant Viterbi algorithm

<https://doi.org/10.1631/FITEE.1900320>

CLC number: TN911.72

1 Introduction

The Nyquist-Shannon sampling theorem establishes a sufficient condition for a sample rate that permits a discrete sample sequence to capture all the information of a bandlimited continuous-time signal. With a sampling rate of no less than twice the


maximum frequency of the bandlimited signal, the samples capture all the information and the original analog signal can be perfectly reconstructed from the samples. The lower bound of the sampling rate, which equals twice the bandwidth, is called the Nyquist rate. However, it is also possible to sample and recover an analog signal with a certain time and/or frequency structure at a sub-Nyquist rate perfectly with no information loss. It is called sub-Nyquist sampling (SNS), and it is an efficient way to reduce the hardware cost on sampling, signal processing, and storage.

Sub-Nyquist sampling has been widely studied as one of the fundamental problems in signal processing. Early works mainly focused on nonuniform

[‡] Corresponding author

^{*} Project supported by the National Natural Science Foundation of China (Nos. 61725104 and 61631003) and Huawei Technologies Co., Ltd. (Nos. HF2017010003, YB2015040053, and YB2013120029)

[#] A preliminary version was presented at the 11th International Conference on Wireless Communications and Signal Processing, October 23–25, 2019, China

 ORCID: Xiqian LUO, <https://orcid.org/0000-0002-6219-5240>; Zhaoyang ZHANG, <https://orcid.org/0000-0003-2346-6228>

© Zhejiang University Press 2021

sampling of the multiband signal with known frequency support (Scoular and Fitzgerald, 1992; Venkataramani and Bresler, 2001). Landau (1967) showed that the average sampling rate of a multiband signal can be reduced to twice the occupied bandwidth; that is, the “Landau rate,” which is typically far smaller than the corresponding Nyquist rate, serves as the lower bound for sampling multiband signals.

Later, many researchers attempted to generalize the problem to multiband signals with unknown spectral support (Feng and Bresler, 1996; Herley and Wong, 1999). Multicoset sampling (Domínguez-Jiménez et al., 2012) and the modulated wideband converter (Mishali and Eldar, 2010) are two widely used sub-Nyquist sampling methods for multiband signals with unknown spectral support. Such a problem could be addressed well under a compressive sensing framework in some specific cases such as cognitive radio, where the spectral occupation of the signal is unknown but sparse (Mishali and Eldar, 2009; Tropp et al., 2010; Sun et al., 2013). In such a context, compressive sensing is introduced to sample signals with a rate much lower than the Nyquist rate. The sparsity of the signal in the frequency domain guarantees adequate detection accuracy. Of course, the multiband signal is not the only kind of signals that can be perfectly sampled and reconstructed with sub-Nyquist sampling. In the recent works from Lu and Do (2008) and Mishali et al. (2011), a model for the union of subspaces was proposed to reveal the underlying structure of the signals for which sub-Nyquist sampling works. Chen et al. (2013) studied the impact of sub-Nyquist sampling on channel capacity.

On the other hand, the early studies and recent revisits to faster-than-Nyquist signaling (Mazo, 1975; Mazo and Landau, 1988; Liveris and Georgiades, 2003; Fan et al., 2017) revealed that there exists some threshold of the signaling rate on the transmitter side, which is beyond the Nyquist rate and is called the “Mazo limit.” Within the limit, it is found that the minimum Euclidean distance between the transmitted signals does not decrease as the symbol rate increases beyond the Nyquist rate. This finding supports a potentially efficient signaling scheme to approach the constrained channel capacity for future communications (Anderson et al., 2013). What is more, it inspires us to consider whether its dual prob-

lem, i.e., sub-Nyquist sampling on the receiver side, has a similar property in information conservation.

The above problem of sub-Nyquist sampling of a general signal is a fundamental problem in signal processing. Unlike sub-Nyquist sampling in compressive sensing, sub-Nyquist sampling of a general signal without a special structure usually results in an under-determined linear problem, which is not sparse. Furthermore, it brings about energy loss as a result of the reduction of the number of available samples, which may in turn degrade the performance of data recovery. In this study, we take a closer look at the fundamental aspects of this problem. Specifically, we study the performance limit and the data recovery algorithm for data recovery from the sub-Nyquist sample sequence of a general baseband signal, which is linearly modulated with symbols from a finite alphabet. In the studied system, the transmitter transmits symbols at the Nyquist rate, while the receiver samples the received signal uniformly at a sub-Nyquist rate. Unlike sub-Nyquist sampling in compressive sensing, the general baseband signal is sparse neither in the time domain nor in the frequency domain. As a result, its degree of freedom is not reducible. We consider two practical sub-Nyquist sampling schemes, i.e., direct sub-Nyquist sampling (DSNS) and filtered sub-Nyquist sampling (FSNS). The former directly samples the modulated signal at some sub-Nyquist rate, while the latter conducts the sampling after a perfect anti-aliasing low-pass filtering. For both sampling schemes, the impacts of sub-Nyquist sampling on the error performance in data recovery are investigated, and a practical algorithm to resolve the very challenging data recovery problem is proposed.

The main contributions of our work can be summarized as follows:

1. The fundamental performance limit of data recovery of a sub-Nyquist sampling system for general linearly modulated baseband signals is investigated. The minimum Euclidean distances between two different sample sequences for both DSNS and FSNS are derived as the indicators of the upper bounds of data recovery performances. The derived minimum Euclidean distances are proportional to the sampling rates within certain thresholds, and thus provide a bound of the performance loss resulting from sub-Nyquist sampling.
2. A low-complexity but efficient time-variant

Viterbi detection algorithm is proposed to recover data from the sub-Nyquist sampled data sequence. As the original signal is a general linearly modulated signal which lacks any inherent sparsity, data recovery is an under-determined problem. Using the constraint of a finite alphabet of transmitted symbols and the priori knowledge of the intrinsic interference structure of the sample sequence, the algorithm can reliably recover the data. Its bit error rate (BER) performance is simulated and compared with the theoretical limits and its Nyquist sampling counterpart, which validates the algorithm's excellent performance.

2 System model

In this study, we focus on the sub-Nyquist sampling problem for a linearly modulated baseband signal with the constraint of a finite symbol set.

As shown in Fig. 1, at the transmitter, the transmitted symbols $\{u_m\}$ are selected from a finite alphabet set and sent with period T . The transmitted symbol sequence $\sum_m u_m \sigma(t - mT)$ is then filtered with a pulse shaping filter $g(t)$ to limit the bandwidth. The transmitted signal can be denoted as

$$x(t) = \sqrt{E_s} \sum_m u_m g(t - mT), \quad (1)$$

where E_s is the average symbol energy.

Here, we assume that binary phase shift keying (BPSK) modulation and the sinc pulse shaping filter are used. Then, $u_m \in \{+1, -1\}$ and the signal is

$$x(t) = \sqrt{\frac{E_s}{T}} \sum_m u_m \operatorname{sinc}\left(\frac{t}{T} - m\right). \quad (2)$$

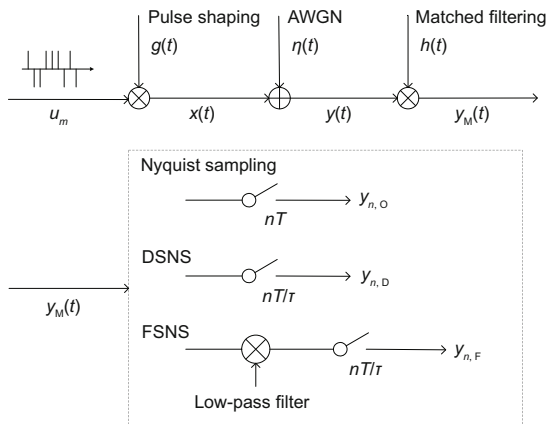


Fig. 1 System model. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE

The signal is bandlimited to $[-\pi/T, \pi/T]$ and its spectrum is

$$X(\omega) = \sqrt{E_s T} \left(\sum_m u_m e^{-j\omega m T} \right) R\left(-\frac{\pi}{T}, \frac{\pi}{T}\right), \quad (3)$$

where $R(-\pi/T, \pi/T)$ is the box function that equals 1 when $\omega \in (-\pi/T, \pi/T)$ and 0 otherwise. To demonstrate the spectrum change in the sampling process, we show the spectrum in Fig. 2a with a triangle, which is not the actual shape of the spectrum but can show the spectral overlapping clearly.

The signal is transmitted over an additive white Gaussian noise (AWGN) channel. So, the received signal is $y(t) = x(t) + \eta_0(t)$, where $\eta_0(t)$ is AWGN with power spectral density $N_0/2$.

At the receiver, the received signal is first matched filtered with the sinc filter as

$$y_M(t) = y(t) * \left(\frac{1}{T} \operatorname{sinc}\left(\frac{t}{T}\right) \right) = x(t) + \eta(t), \quad (4)$$

where $\eta(t)$ denotes the bandlimited Gaussian with spectral density $N_0/2$ within its bandwidth. The bandlimited transmitted signal $x(t)$ has no change but the noise becomes bandlimited.

When the filtered signal is sampled with the Nyquist rate $1/T$, the samples are orthogonal as

$$y_{n,O} = y_M(t)|_{t=nT} = \sqrt{\frac{E_s}{T}} u_n + \eta_{n,O}, \quad (5)$$

where $\eta_{n,O}$ for all n 's are independent and identically distributed as $\eta_{n,O} \sim N(0, N_0/2)$. The spectrum of the Nyquist sampled waveform $\sum_n y_{n,O} \delta(t - nT)$ can be demonstrated as Fig. 2b, where $\delta(t)$ is the unit impulse function. Spectrum $X(\omega)$ is replicated with interval $2\pi/T$ and there exists no spectrum overlapping.

If the filtered signal is directly sampled with the sub-Nyquist rate τ/T , where $\tau < 1$, the samples are the linear combinations of the transmitted symbols as

$$\begin{aligned} y_{n,D} &= [x(t) + \eta(t)]|_{t=nT/\tau} \\ &= \sqrt{\frac{E_s}{T}} \sum_m u_m \operatorname{sinc}\left(\frac{n}{\tau} - m\right) + \eta_{n,D}. \end{aligned} \quad (6)$$

Sub-Nyquist sampling without pre-filtering is called direct sub-Nyquist sampling in this study. The spectrum of the direct sub-Nyquist sampled waveform $\sum_n y_{n,D} \delta(t - nT/\tau)$ can be demonstrated as

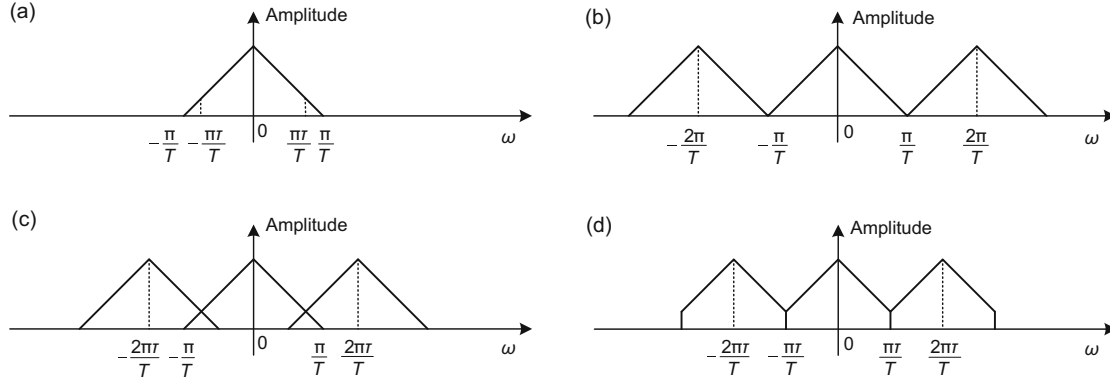


Fig. 2 Comparison of the spectra of the original signal (a), Nyquist sampled signal (b), DSNS signal (c), and FSNS signal (d). Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE

shown in Fig. 2c. Spectra $X(\omega)$ overlap on the high-frequency end.

To avoid spectral overlapping, a low-pass filter is often added before sampling. The sub-Nyquist sampling with pre-filtering is called filtered sub-Nyquist sampling in this study. Suppose the ideal low-pass filter $\tau/T \cdot \text{sinc}(t\tau/T)$ is used. Then the low-pass filtered signal is

$$\begin{aligned} y_F(t) &= [x(t) + \eta(t)] * \left(\frac{\tau}{T} \cdot \text{sinc} \left(\frac{t\tau}{T} \right) \right) \\ &= \sqrt{\frac{E_s}{T}} \tau \sum_m u_m \text{sinc} \left(\frac{\tau}{T} t - m\tau \right) + \eta_F(t), \end{aligned} \quad (7)$$

where $\eta_F(t)$ is the bandlimited Gaussian noise with a plain power density $N_0/2$ within its band $[-\pi\tau/T, \pi\tau/T]$. The spectrum of the signal without noise is the low-pass filtered $X(\omega)$ as

$$X_F(\omega) = X(\omega)R \left(-\frac{\pi\tau}{T}, \frac{\pi\tau}{T} \right), \quad (8)$$

where $R(-\pi\tau/T, \pi\tau/T)$ is the box function that equals 1 when $\omega \in (-\pi\tau/T, \pi\tau/T)$ and 0 otherwise.

When the filtered signal $y_F(t)$ is uniformly sampled with intervals of T/τ , the samples are

$$\begin{aligned} y_{n,F} &= y_F(t)|_{t=nT/\tau} \\ &= \sqrt{\frac{E_s}{T}} \tau \sum_m u_m \text{sinc}(n - m\tau) + \eta_{n,F}, \end{aligned} \quad (9)$$

where $\eta_{n,F}$ for all n 's are independent and identically distributed as $N(0, N_0/2)$. The spectrum of the sampled signal $\sum_m y_n \delta(t - nT/\tau)$ is the repetition of $X_F(\omega)$ with intervals $2\pi\tau/T$ (Fig. 2d). There is no overlapping as a result of low-pass filtering.

3 Minimum Euclidean distance

Pairwise error probability (PEP) is the error probability that when signal \mathbf{u}_1 is transmitted, the receiver detects the distorted version \mathbf{u}_2 instead of \mathbf{u}_1 . Suppose \mathbf{y}_1 and \mathbf{y}_2 are the noiseless sample sequences corresponding to the transmitted symbol sequences \mathbf{u}_1 and \mathbf{u}_2 respectively, and \mathbf{y} is the sample sequence with noise when \mathbf{u}_1 is transmitted. A detection error occurs when \mathbf{y} is closer to \mathbf{y}_2 than to \mathbf{y}_1 , that is,

$$P(\mathbf{u}_1 \rightarrow \mathbf{u}_2 | \mathbf{y}) = P(d^2(\mathbf{y}, \mathbf{y}_1) > d^2(\mathbf{y}, \mathbf{y}_2)). \quad (10)$$

As a result, the error probability above depends on the noise power normalized Euclidean distance between \mathbf{y}_1 and \mathbf{y}_2 . The PEP of the system depends on the minimum normalized Euclidean distance for all possible \mathbf{u}_1 's and \mathbf{u}_2 's.

In this section, the minimum noise power normalized Euclidean distances of DSNS and FSNS under different sampling rates are derived. Compared to the minimum distance with Nyquist sampling, the derived minimum Euclidean distances of DSNS and FSNS show how the detection performances degrade with the decreasing sampling rates beyond the Nyquist rate.

In Nyquist sampling, DSNS, and FSNS, the samples are linear combinations of the transmitted symbols. For a linear system, the Euclidean distance between \mathbf{y}_1 and \mathbf{y}_2 is the function of the error pattern $\mathbf{e} = \mathbf{u}_1 - \mathbf{u}_2$:

$$d^2 = |\mathbf{y}_1 - \mathbf{y}_2|^2 = |\mathbf{H}\mathbf{x}_1 - \mathbf{H}\mathbf{x}_2|^2 = |\mathbf{H}\mathbf{e}|^2. \quad (11)$$

The difference between the sample sequences, which is defined as $\mathbf{y}_e \doteq \mathbf{y}_1 - \mathbf{y}_2 = \mathbf{H}\mathbf{e}$, is

equivalent to the sample sequence corresponding to the transmitted symbol sequence \mathbf{e} . Calculating the minimum Euclidean distance is to search over all possible nonzero \mathbf{e} 's to find $\min_{\mathbf{e}} d^2 = \min_{\mathbf{e}} |\mathbf{H}\mathbf{e}|^2$. With this fact, we can calculate the minimum normalized Euclidean distances for DSNS and FSNS with BPSK modulation and sinc pulse shaping.

As for BPSK modulation, the elements in the error pattern $\mathbf{e} = \mathbf{u}_1 - \mathbf{u}_2$ have three possible values: 2, -2, and 0. Define half of the error pattern as $\mathbf{b} = \frac{1}{2}\mathbf{e}$, the elements of which satisfy $b_m \in \{1, -1, 0\}$.

To calculate the Euclidean distance between two distinct sample sequences, the following lemma is first introduced:

Lemma 1 Suppose that $x(t)$ is a continuous-time signal which is sampled uniformly with interval T , and the corresponding sample sequence is $\{x_n\}$. Denote the spectrum of $x(t)$ as $X_C(j\omega)$. The total energy of the sample sequence can be calculated by integrating its power spectrum as

$$\sum_n |x_n|^2 = \frac{1}{2\pi T} \int_{2\pi/T} \left| \sum_k X_C \left(j \left(\omega - \frac{2\pi k}{T} \right) \right) \right|^2 d\omega. \quad (12)$$

Proof Denote the discrete Fourier transform of the sample sequence $\{x_n\}$ as $X_D(e^{j\omega})$. According to Parseval's theorem for discrete-time signals, the total energy of all the samples $\{x_n\}$ is equal to the integration of its power spectrum over one period:

$$\sum_{n=-\infty}^{\infty} |x_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X_D(e^{j\omega})|^2 d\omega. \quad (13)$$

The relationship between $X_D(e^{j\omega})$ and $X_C(\omega)$ is

$$X_D(e^{j\omega}) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} X_C \left(\frac{\omega}{T} - \frac{2\pi n}{T} \right). \quad (14)$$

Combining Eqs. (13) and (14), Eq. (12) is proved.

For Nyquist sampling in Eq. (5), obviously the minimum Euclidean distance occurs when there is only one nonzero element in \mathbf{b} , and the minimum normalized distance is $d_{O,\min}^2 = 8E_s/N_0$. In the following, we present two basic results for the minimum Euclidean distances of the DSNS and FSNS systems, respectively.

Theorem 1 The minimum normalized Euclidean distance between two distinct sample sequences in DSNS with respect to that in the Nyquist sampling

(orthogonal transmission) case is given by

$$\min_{\mathbf{b}} \frac{d_D^2}{d_{O,\min}^2} = \min_{b_m \in \{\pm 1, 0\}} \left\{ \int_0^{\tau_1} \left| \sum_m b_m e^{-j\pi\omega m} \right|^2 d\omega + \frac{1}{2} \int_{\tau_1}^{\tau} \left| \sum_m b_m e^{-j\pi\omega m} (1 + e^{j2\pi\tau m}) \right|^2 d\omega \right\}. \quad (15)$$

Proof Obviously, the sample sequence and the original transmitted data sequence constitute a linear system, with the transmitted data sequence as input and the sample sequence as output. As a result, for two distinct original data sequences, the difference between the two corresponding sample sequences is equal to the sample sequence corresponding to the original data sequence of the error pattern $2\mathbf{b}$. Given the original transmitted symbol sequence $\{2b_m\}$, the total energy of the corresponding samples is calculated as the Euclidean distance between sample sequences.

After matched filtering at the receiver, the continuous-time signal before sampling is bandlimited to $[-\pi/T, \pi/T]$ as

$$X_M(\omega) = \begin{cases} 2\sqrt{E_s T} \sum_m b_m e^{-j\omega m T}, & |\omega| < \pi/T, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where the noise is also bandlimited AWGN with a two-sided spectral density of $N_0/2$.

In DSNS, the signal is directly under-sampled uniformly with time interval T/τ , which is larger than its symbol interval T . Define the partition point as $\tau_1 = 2\tau - 1$. As shown in Fig. 2c, the higher frequency bands beyond the partition point $\tau_1\pi/T < |\omega| < \pi/T$ of the spectrum of the sampled signal $X(\omega)$ will overlap, while the lower band $|\omega| < \tau_1\pi/T$ stays the same as the spectrum of the original signal. The spectrum of the discrete sample sequence can be represented by the spectrum of the continuous-time signal according to Eq. (14) as

$$X_D(e^{j\omega}) = \frac{\tau}{T} \sum_{n=-\infty}^{+\infty} X_M \left(\frac{\tau}{T} (\omega - 2\pi n) \right). \quad (17)$$

The noise after sampling is no longer white. Its power density is $N_0/2$ in the lower band and N_0 in the higher band. Thus, a noise whitening filter should be used as

$$W_D(e^{j\omega}) = \begin{cases} 1, & |\omega| < \pi\tau_1/\tau, \\ \sqrt{1/2}, & \pi\tau_1/\tau < |\omega| < \pi. \end{cases} \quad (18)$$

After whitening filtering, the whitened noise has the power of $N_0\tau/(2T)$ and the spectrum of the signal is $X_D(e^{j\omega}) \cdot W_D(e^{j\omega})$.

With whitened noise, the energy of the sample sequence \mathbf{y}_e normalized by the noise power can be calculated according to Parseval's theorem in Eq. (13) as

$$\begin{aligned} \frac{\sum_n |y_{n,e}|^2}{N_0\tau/(2T)} &= \frac{T}{\pi N_0\tau} \int_{2\pi} |X_D(e^{j\omega}) \cdot W_D(e^{j\omega})|^2 d\omega \\ &= \frac{8E_s}{N_0} \left[\int_0^{\tau_1} \left| \sum_m b_m e^{-j\pi\omega m} \right|^2 d\omega \right. \\ &\quad \left. + \frac{1}{2} \int_{\tau_1}^{\tau} \left| \sum_m b_m e^{-j\pi\omega m} (1 + e^{j2\pi\tau m}) \right|^2 d\omega \right]. \end{aligned} \quad (19)$$

Compared to the minimum Euclidean distance $8E_s/N_0$ of orthogonal transmission, the normalized Euclidean distance of DSNS is given by

$$\begin{aligned} \frac{d_D^2}{d_{O,\min}^2} &= \int_0^{\tau_1} \left| \sum_m b_m e^{-j\pi\omega m} \right|^2 d\omega \\ &\quad + \frac{1}{2} \int_{\tau_1}^{\tau} \left| \sum_m b_m e^{-j\pi\omega m} (1 + e^{j2\pi\tau m}) \right|^2 d\omega. \end{aligned} \quad (20)$$

In Eq. (20), the first term comes from the lower-frequency band where there is no overlapping, and the second term comes from the overlapping higher-frequency band. The minimum of the normalized Euclidean distance of DSNS in Eq. (20) is obtained by searching all nonzero \mathbf{b} 's.

Similarly, we have the following result for the FSNS system:

Theorem 2 The minimum normalized Euclidean distance of FSNS with respect to that of the Nyquist sampling (orthogonal transmission) case is given by

$$\min_{\mathbf{b}} \frac{d_F^2}{d_{O,\min}^2} = \min_{b_m \in \{\pm 1, 0\}} \int_0^{\tau} \left| \sum_m b_m e^{-j\pi m \omega} \right|^2 d\omega. \quad (21)$$

Proof In FSNS, the received signal is first matched filtered and then low-pass filtered with an ideal low-pass filter $\tau/T \cdot \text{sinc}(t\tau/T)$. The spectrum of the signal after low-pass filtering is

$$X_F(\omega) = \begin{cases} 2\sqrt{E_s T} \sum_m b_m e^{-j\omega m T}, & |\omega| < \tau\pi/T, \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

Then the low-pass filtered signal is sampled with interval T/τ . As shown in Fig. 2d, there is no overlapping in the frequency domain of the sampled signal.

Besides, the sampled noise remains to be AWGN with a power spectrum density of $N_0/2$.

Normalized by the noise power $N_0\tau/(2T)$, the total energy of the sample sequence can be calculated according to Eq. (12) as

$$\begin{aligned} \frac{\sum_n |y_{n,e}|^2}{N_0\tau/(2T)} &= \frac{1}{\pi N_0} \int_{2\pi\tau/T} \left| \sum_k X_F \left(j \left(\omega - \frac{2\pi\tau k}{T} \right) \right) \right|^2 d\omega \\ &= \frac{8E_s}{N_0} \int_0^{\tau} \left| \sum_m b_m e^{-j\pi m \omega} \right|^2 d\omega. \end{aligned} \quad (23)$$

Compared to the minimum Euclidean distance $8E_s/N_0$ of Nyquist sampling, the normalized distance of FSNS is

$$\frac{d_F^2}{d_{O,\min}^2} = \int_0^{\tau} \left| \sum_m b_m e^{-j\pi m \omega} \right|^2 d\omega. \quad (24)$$

The minimum of the normalized Euclidean distance of FSNS in Eq. (24) is obtained by searching all nonzero \mathbf{b} 's.

We search the sequence \mathbf{b} for the minimum Euclidean distances $d_{D,\min}^2 \triangleq \min_{\mathbf{b}} d_D^2$ and $d_{F,\min}^2 \triangleq \min_{\mathbf{b}} d_F^2$ under different sampling rates τ/T . The upper bounds and minimum normalized Euclidean distances $d_{D,\min}^2/d_{O,\min}^2$ and $d_{F,\min}^2/d_{O,\min}^2$ versus τ are shown in Fig. 3. We can see that at the same sampling rate, the minimum normalized Euclidean distance of DSNS is much lower than that of FSNS, indicating the negative effect of frequency overlapping in DSNS on the performance.

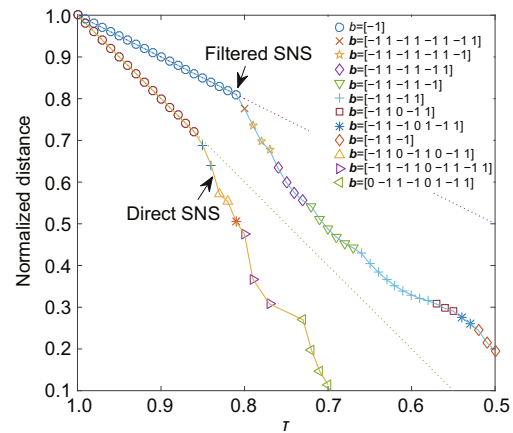


Fig. 3 Minimum of the normalized distance of two distinct sample sequences in DSNS and FSNS. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE

Based on the above two theorems and also as indicated by the curves in Fig. 3, we have the following two important observations:

Theorem 3 For FSNS, the minimum normalized Euclidean distance satisfies $d_{F,\min}^2/d_{O,\min}^2 = \tau$ when the sampling rate decreases below the Nyquist rate within the range $\tau \in [0.802, 1]$, and it is less than τ when $\tau < 0.802$.

Proof The equation for the minimum normalized Euclidean distance of FSNS in Theorem 2 is similar to the minimum normalized Euclidean distance of faster-than-Nyquist (FTN) signaling in Mazo (1975). For BPSK modulation and sinc pulse shaping, the minimum Euclidean distance of FTN signaling with symbol rate $1/(\tau T)$ is

$$\min_{\mathbf{b}} d_{FTN}^2 = \min_{b_m \in \{\pm 1, 0\}} \frac{1}{\tau} \int_0^{\tau} \left| \sum_m b_m e^{-j\pi m \omega} \right|^2 d\omega. \quad (25)$$

Comparing the Euclidean distance of the FTN signaling above with the minimum Euclidean distance of FSNS with a sampling rate τ/T in Eq. (21) yields

$$\frac{d_{F,\min}^2}{d_{FTN,\min}^2} = \tau. \quad (26)$$

The minimum Euclidean distance of FTN signaling in Eq. (25) was calculated in Mazo and Landau (1988) and Hajela (1990). It is obtained by searching the possible error patterns \mathbf{b} . To reduce the search complexity, several constraints such as patterns and the number of nonzero blocks, are made to \mathbf{b} . FTN signaling has a limit on the signaling rate called the Mazo limit. When increasing the signaling rate within the Mazo limit, the minimum normalized Euclidean distance between two different signals will not decrease as the signaling rate increases. For BPSK and sinc pulse shaping, the Mazo limit is $\tau = 0.802$.

Based on the relationship between FTN and FSNS in Eq. (26), the minimum normalized Euclidean distance of FSNS is equal to τ within the Mazo limit $\tau \in (0.802, 1]$, and will drop below τ as the sampling rate decreases beyond the Mazo limit.

Theorem 4 For DSNS, the minimum normalized Euclidean distance satisfies $d_{D,\min}^2/d_{O,\min}^2 = 2\tau - 1$ when $\tau \in [0.855, 1]$, and it is less than $2\tau - 1$ when $\tau < 0.855$.

Proof As shown in Eq. (20), the normalized Euclidean distance for DSNS is composed of two terms, the first term from the non-overlapping frequency

band and the second term from the overlapping frequency band. The first part $\int_0^{\tau_1} \left| \sum_m b_m e^{-j\pi \omega m} \right|^2 d\omega$ is the same as the Euclidean distance of FSNS, except that the integral limit is $\tau_1 = 2\tau - 1$ instead of τ . According to Theorem 3, the minimum value of the first term is no more than $2\tau - 1$.

Next, if we choose the error pattern $\mathbf{b} = [\dots 0 \dots 0 \quad -1 \quad 0 \dots 0 \dots]$ with the m^{th} element as nonzero, the non-overlapping part of the Euclidean distance in Eq. (20) just equals $2\tau - 1$. By choosing the position m to satisfy $2\tau m = 2k + 1$, where k is an arbitrary integer, the overlapping part turns out to be zero. Thus, the minimum Euclidean distance for DSNS is $2\tau - 1$.

When the sampling rate τ begins to decrease from 1, the minimum distance remains as $2\tau - 1$, corresponding to some error pattern \mathbf{b} with only one nonzero element at some position. The minimum sampling rate τ at which the minimum distance remains as $2\tau - 1$ can also be obtained by search. The combinatorial complexity for search can be greatly reduced by considering only those \mathbf{b} that make the second part no more than $2\tau - 1$ since the minimum distance never exceeds $2\tau - 1$. The minimum sampling rate for keeping the linear distance $2\tau - 1$ is $\tau = 0.855$.

In conclusion, the upper bound of the minimum Euclidean distance decreases proportionally to the sampling rate within some thresholds for both FSNS and DSNS.

4 Time-variant Viterbi algorithm for data recovery

4.1 Challenges in data recovery

As shown in Eqs. (6) and (9), the sub-Nyquist sampling forms an under-determined linear system which has a time-variant system impulse response. With BPSK modulation, the transmitted symbols are constrained to be chosen from $\{+1, -1\}$. So, the data recovery problem from the sub-Nyquist sampled sequences in Eqs. (6) and (9) can be formulated as

$$\min_{\mathbf{u} \in \{+1, -1\}^M} \|\mathbf{y} - \mathbf{H}\mathbf{u}\|_2, \quad (27)$$

where $\|\cdot\|_2$ denotes the ℓ^2 norm of a vector, and \mathbf{H} is the equivalent channel response matrix which is of full row rank with size $N \times M$.

As an under-determined problem that has more inputs than outputs, linear equalization methods are incapable for data recovery in this case. The conventional sparsity detection algorithms, such as Bases Pursuit and the greedy algorithm used in compressed sensing, are also unable to work since there is no sparsity in the sampled signal. Although in theory, the complicated maximum likelihood sequence detection approach (Goldsmith, 2005) may be applicable, the computational complexity becomes too high in practice because the search space becomes extremely large as the data length increases and, to make matters worse, a much severer inter-symbol interference (ISI) arises when the sampling is not at the Nyquist rate.

The Viterbi algorithm (VA) (Forney, 1973) is a simplified realization of maximum likelihood sequence estimation. It can recover data from the outputs of an ISI channel with low computational complexity. However, in the case of SNS and as far as VA is concerned, the ISI structure varies from sample to sample, so does the state space. Let us elaborate more on these. Note that each output sample is associated with either one or two dominating input symbols (i.e., they constitute the principal components of that sample), as indicated by the red dotted lines in Fig. 4. As a matter of fact, the number of dominating symbols changes every several samples, which also makes the state transitions of the trellis graph varying (see Fig. 4 for illustration). The interferences contained in each sample come from

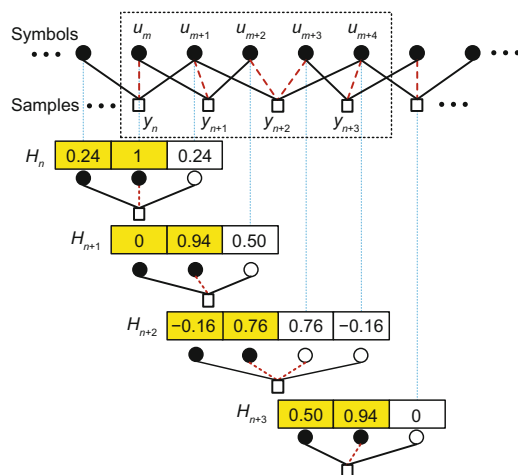


Fig. 4 Structure of the time-variant inter-symbol interference with $\tau = 0.8$ and $2L = 2$. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE

the neighboring symbols after being weighted by the equivalent channel impulse responses. As a result of non-integer (fractional) sub-Nyquist sampling, the equivalent channel impulse responses change from sample to sample, as shown in Fig. 4, resulting in a time-variant interference structure and a relatively large and dynamic constraint length as well as state space. Due to the above reasons, the conventional VA does not apply to this case.

4.2 Algorithm description

Here, we propose an efficient time-variant VA to recover data from the sub-Nyquist sampled sequence.

First, the overall equivalent channel impulse response is truncated to reduce the number of states. Let L be a tunable parameter denoting the number of neighboring interfering symbols at each side of the dominating symbol(s) associated with a sample after the truncation. The number of dominating symbol(s) could be one or two as shown in Fig. 4; thus, the length of the truncated impulse response is either $2L + 1$ or $2L + 2$. Obviously, the truncation length L provides a tradeoff between the performance and the computational complexity.

Next, the dynamic trellis graph is constructed, over which the new VA is conducted. For truncated impulse response with length $2L + 1$, the constraint length is normally $2L + 1$ (or equivalently the memory is $2L$); thus, the resultant number of states is 2^{2L} . When a new sample is obtained, the Euclidean measure is calculated along the paths which start from the current states and go to the next states. Note that for a certain sample that is associated with two dominating symbols, the constraint length is $2L + 2$; thus, there will be four branches (corresponding to two unknown single-bit symbols) starting from each of the current 2^{2L} states and ending to four of the next 2^{2L} states (Fig. 5a).

Without loss of generality, let us take FSNS with sampling rate $\tau = 0.8$ for illustration. With the sampling rate $\tau = 0.8 = 4/5$, four samples are generated within the duration of five symbols (Fig. 4). As described above, each sample is associated to one or two dominating symbols. As depicted in the upper dotted-line block in Fig. 4, each of the 1st, 2nd, and 4th samples y_n, y_{n+1} , and y_{n+3} has one dominating symbol, while the 3rd one y_{n+2} has two. Alongside the dominating symbols of each sample, there are two interfering symbols, one for each side.

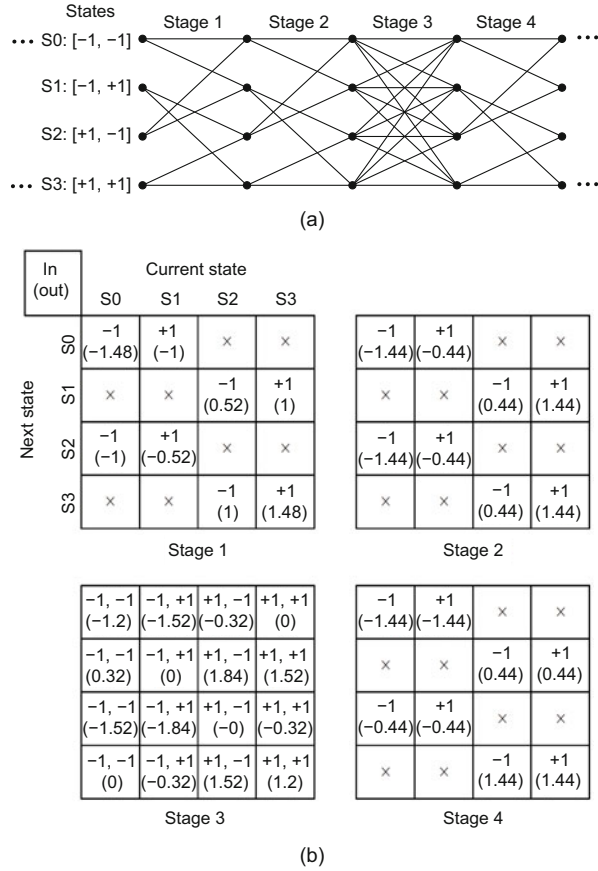


Fig. 5 The trellis graph (a) (reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE) and state transition (b) of a time-variant VA with $\tau = 0.8$ and $2L = 2$. In (b), each term in the form represents a trellis branch connecting the current state and the next state, the inputs and outputs denote the input symbol values and the predicted sample values respectively, and the “x” in the form stands for an impossible state transition

Therefore, here $2L = 2$ and the number of states is four. The truncated time-variant impulse responses H_n, H_{n+1}, H_{n+2} , and H_{n+3} , for the four successive samples y_n, y_{n+1}, y_{n+2} , and y_{n+3} , respectively, are shown in Fig. 4. As a new sample comes, the VA detector enters a new stage. At each stage, the first two symbols constitute the current state space, shown as the black disks. The other one or two white disks at each stage correspond to the unknown input symbols which lead to state transitions. Note that the state transition of the third stage is different, at which there are four branches starting from each of the current four states and ending at the next four states (Fig. 5a).

At each stage, the path metric is calculated as follows. Given the current state and the correspond-

ing two symbols, each state transition corresponds to one or two specific input symbols. All the above symbols are summed up after being weighted by the truncated channel coefficients to obtain the predicted sample value. Then the square error between the predicted sample value and the real sample value serves as the branch metric. Finally, the path metric can be achieved by accumulating the branch metrics along that path. The calculation of the metrics for each state of each stage is illustrated in Fig. 5b. Note that at stage 3 the states are driven by two input symbols and there are four next states for each current state; thus, four symbols are used to predict the sample value. At the other three stages, only three symbols are used to obtain the prediction. Finally, with the trellis dynamically growing, the path metric calculation and decision-making are conducted in the same way as in the conventional VA. The path with the minimum accumulated metric survives and the corresponding input symbols along that path will be the final detection output.

Note that the computational complexity of the time-variant VA is of the same order as that of the conventional VA, with only a slight difference in the treatment of the non-regular stages where there are two dominating input symbols.

4.3 Simulation results

In this subsection, the bit error rate (BER) performances of the above time-variant VA detection algorithm for both the DSNS and FSNS systems are simulated. The simulation results are then compared with both their theoretical limits and the benchmarks of their Nyquist sampling counterparts.

Suppose the block length of the transmitted symbols is $M = 200$. At a sampling rate τ , the number of samples is $N = \lceil \tau \cdot M \rceil$ ($\lceil \cdot \rceil$ represents the rounding up operation). The equivalent channel impulse response is truncated to length $2L + 1$ or $2L + 2$, where L is first set to four. Figs. 6 and 7 show the performance curves of BER versus the sampling rate in DSNS and FSNS with sinc pulse shaping, respectively. Compared to the case of Nyquist sampling, the BER performance for both SNS systems degrades as the sampling rate decreases. FSNS always has a lower BER than DSNS probably at the same sampling rate because FSNS has anti-aliasing filtering.

For DSNS with $\tau = 0.9, 0.8$, and 0.7 , the square

normalized Euclidean distances are $d^2 = 0.8, 0.5,$ and $0.1,$ respectively, while for FSNS with $\tau = 0.9, 0.8,$ and $0.7,$ the square normalized Euclidean distances are $d^2 = 0.9, 0.8,$ and $0.5,$ respectively. Comparing the theoretical performance and the simulated BERs in Figs. 6 and 7, there is a trend that the BER performance approaches the theoretical limit as the sampling rate increases. However, there are gaps between the theoretical performances and the simulated ones. On one hand, the truncation of the

equivalent channel impulse response degrades the performance, especially when the signal-to-noise ratio (SNR) becomes high. On the other hand, the performance loss is due to the non-optimality of time-variant VA detection.

Next we show the impact of the truncation of the channel impulse response, which causes additional interference and degrades the performance. Fig. 8 presents the BER performances of FSNS with truncation length $2L + 1 = 5, 7, 9,$ with sinc pulse

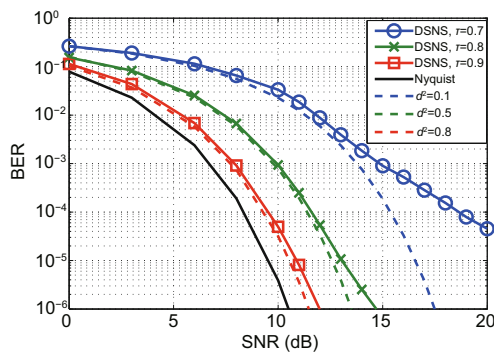


Fig. 6 BER of DSNS with sinc pulse shaping. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE. The solid lines show the simulated BERs, and the dotted lines are the theoretical BERs calculated based on the minimum normalized Euclidean distances in Fig. 3

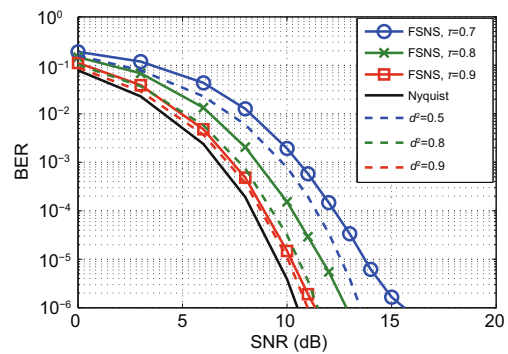


Fig. 7 BER of FSNS with sinc pulse shaping. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE. The solid lines show the simulated BERs, and the dotted lines are the theoretical BERs calculated based on the minimum normalized Euclidean distances in Fig. 3

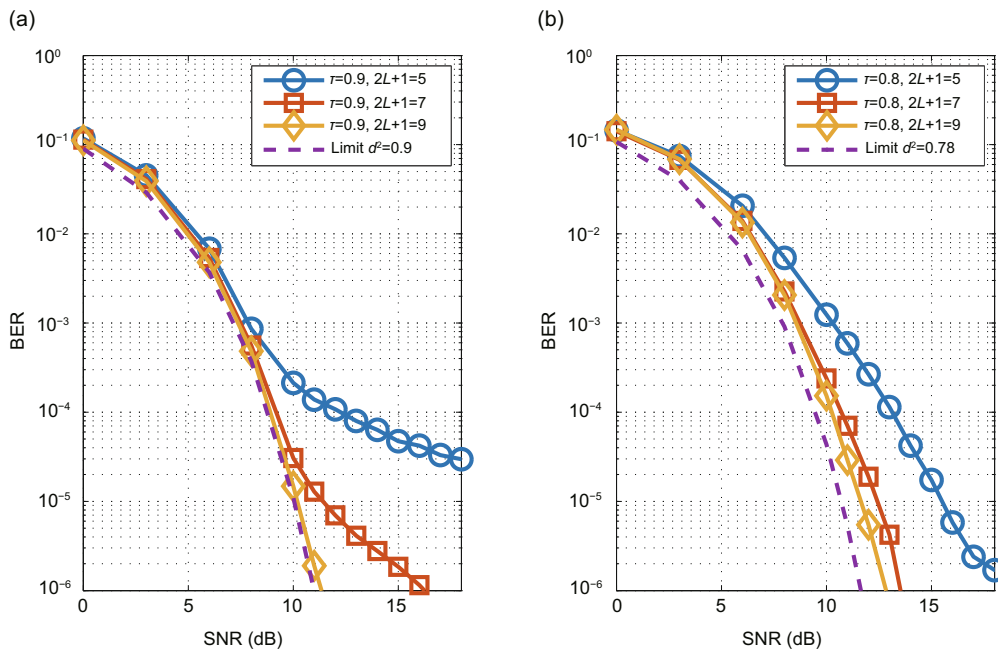


Fig. 8 BER of FSNS with different truncation lengths under sinc pulse shaping with $\tau = 0.9$ (a) and $\tau = 0.8$ (b)

shaping and sampling rate $\tau = 0.9$ and 0.8 . We can see that the effect of truncation is more obvious when the SNR becomes higher. As the truncation length increases, the residual interference is reduced and the error floor decreases as well, and more importantly, the BER performance gradually approaches the theoretical limit given by the minimum Euclidean distance.

To alleviate the negative effect of truncation, we test the performance when the root raised cosine (RRC) filter, instead of the sinc pulse filter, is used for shaping. In the case of RRC shaping, the ISI decays more rapidly, so that the truncation length can be much smaller than that of sinc pulse shaping, which greatly reduces the computational complexity. Fig. 9 shows the BER of DSNS and FSNS with RRC shaping with roll-off factor $\alpha = 0.3$ and single-side truncation length $L = 3$. There is no error floor in the observation window caused by the truncation. With alleviated effect of truncation, it shows the good performance of the time-variant Viterbi algorithm.

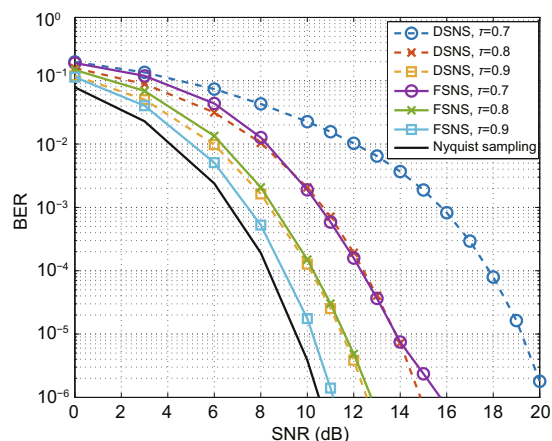


Fig. 9 BER of DSNS and FSNS with RRC pulse shaping. Reprinted from Luo and Zhang (2019), Copyright 2019, with permission from IEEE

5 Conclusions

In this paper, the sub-Nyquist sampling of a linearly modulated baseband signal has been investigated. While FTN signaling increases the symbol rate beyond the Nyquist rate at the transmitter, SNS reduces the sampling rate below the Nyquist rate at the receiver. Direct SNS and low-pass filtered SNS

have been proposed. The performance loss has been estimated by the minimum Euclidean distances of the sample sequences for DSNS and FSNS. While DSNS suffers from frequency overlapping, FSNS has better performance due to its anti-aliasing filtering. For FSNS with BPSK modulation and sinc pulse shaping, the minimum distance remains proportional to the sampling rate within the $\tau = 0.802$ limit, and it drops more rapidly beyond this limit. The Viterbi algorithm is adapted for detection in this under-determined time-variant linear system. The proposed time-variant Viterbi algorithm is able to recover the transmitted signal from the sub-Nyquist samples, and its performance approaches the theoretical limit as its truncation length grows.

Contributors

Xiqian LUO and Zhaoyang ZHANG designed the research. Xiqian LUO did the simulation and drafted the manuscript. Zhaoyang ZHANG revised the manuscript. Xiqian LUO revised and finalized the paper.

Compliance with ethics guidelines

Xiqian LUO and Zhaoyang ZHANG declare that they have no conflict of interest.

References

- Anderson JB, Rusek F, Öwall V, 2013. Faster-than-Nyquist signaling. *Proc IEEE*, 101(8):1817-1830. <https://doi.org/10.1109/JPROC.2012.2233451>
- Chen YX, Eldar YC, Goldsmith AJ, 2013. Shannon meets Nyquist: capacity of sampled Gaussian channels. *IEEE Trans Inform Theory*, 59(8):4889-4914. <https://doi.org/10.1109/TIT.2013.2254171>
- Domínguez-Jiménez ME, González-Prelcic N, Vazquez-Vilar G, et al., 2012. Design of universal multicore sampling patterns for compressed sensing of multiband sparse signals. *Proc IEEE Int Conf on Acoustics, Speech and Signal Processing*, p.3337-3340. <https://doi.org/10.1109/ICASSP.2012.6288630>
- Fan JC, Guo SJ, Zhou XW, et al., 2017. Faster-than-Nyquist signaling: an overview. *IEEE Access*, 5:1925-1940. <https://doi.org/10.1109/ACCESS.2017.2657599>
- Feng P, Bresler Y, 1996. Spectrum-blind minimum-rate sampling and reconstruction of multiband signals. *Proc IEEE Int Conf on Acoustics, Speech, and Signal Processing*, 3:1688-1691. <https://doi.org/10.1109/ICASSP.1996.544131>
- Forney GD, 1973. The Viterbi algorithm. *Proc IEEE*, 61(3): 268-278. <https://doi.org/10.1109/PROC.1973.9030>
- Goldsmith A, 2005. *Wireless Communications*. Cambridge University Press, Cambridge, USA, p.327-340. <https://doi.org/10.1017/CBO9780511841224>
- Hajela D, 1990. On computing the minimum distance for faster than Nyquist signaling. *IEEE Trans Inform Theory*, 36(2):289-295. <https://doi.org/10.1109/18.52475>

- Herley C, Wong PW, 1999. Minimum rate sampling and reconstruction of signals with arbitrary frequency support. *IEEE Trans Inform Theory*, 45(5):1555-1564. <https://doi.org/10.1109/18.771158>
- Landau HJ, 1967. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math*, 117(1):37-52. <https://doi.org/10.1007/BF02395039>
- Liveris AD, Georghiades CN, 2003. Exploiting faster-than-Nyquist signaling. *IEEE Trans Commun*, 51(9):1502-1511. <https://doi.org/10.1109/TCOMM.2003.816943>
- Lu YM, Do MN, 2008. A theory for sampling signals from a union of subspaces. *IEEE Trans Signal Process*, 56(6):2334-2345. <https://doi.org/10.1109/TSP.2007.914346>
- Luo X, Zhang Z, 2019. Data recovery from sub-Nyquist sampled signals: fundamental limit and detection algorithm. Proc 11th Int Conf on Wireless Communications and Signal Processing, p.1-6. <https://doi.org/10.1109/WCSP.2019.8927914>
- Mazo JE, 1975. Faster-than-Nyquist signaling. *Bell Syst Techn J*, 54(8):1451-1462. <https://doi.org/10.1002/j.1538-7305.1975.tb02043.x>
- Mazo JE, Landau HJ, 1988. On the minimum distance problem for faster-than-Nyquist signaling. *IEEE Trans Inform Theory*, 34(6):1420-1427. <https://doi.org/10.1109/18.21281>
- Mishali M, Eldar YC, 2009. Blind multiband signal reconstruction: compressed sensing for analog signals. *IEEE Trans Signal Process*, 57(3):993-1009. <https://doi.org/10.1109/TSP.2009.2012791>
- Mishali M, Eldar YC, 2010. From theory to practice: sub-Nyquist sampling of sparse wideband analog signals. *IEEE J Sel Top Signal Process*, 4(2):375-391. <https://doi.org/10.1109/JSTSP.2010.2042414>
- Mishali M, Eldar YC, Elron AJ, 2011. Xampling: signal acquisition and processing in union of subspaces. *IEEE Trans Signal Process*, 59(10):4719-4734. <https://doi.org/10.1109/TSP.2011.2161472>
- Scoular SC, Fitzgerald WJ, 1992. Periodic nonuniform sampling of multiband signals. *Signal Process*, 28(2):195-200. [https://doi.org/10.1016/0165-1684\(92\)90035-U](https://doi.org/10.1016/0165-1684(92)90035-U)
- Sun HJ, Nallanathan A, Wang CX, et al., 2013. Wideband spectrum sensing for cognitive radio networks: a survey. *IEEE Wirel Commun*, 20(2):74-81. <https://doi.org/10.1109/MWC.2013.6507397>
- Tropp JA, Laska JN, Duarte MF, et al., 2010. Beyond Nyquist: efficient sampling of sparse bandlimited signals. *IEEE Trans Inform Theory*, 56(1):520-544. <https://doi.org/10.1109/TIT.2009.2034811>
- Venkataramani R, Bresler Y, 2001. Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals. *IEEE Trans Signal Process*, 49(10):2301-2313. <https://doi.org/10.1109/78.950786>