Single Channel RF Signal Recovery for Nyquist Folding Receiver

F. Uysal, J. C. Martin, N. A. Goodman

1,2,3 Advanced Radar Research Center, University of Oklahoma, Norman, OK, USA
1 Microwave Sensing, Systems and Signals (MS3), Delft University of Technology, Delft, Netherlands
2,3 School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA
1 f.uysal@tudelft.nl, 2 james.c.martin@ou.edu, 3 goodman@ou.edu

Keywords: Nyquist folding receiver (NYFR), analog-to-information (A2I), compressed sensing (CS)

Abstract

This paper presents a method for exploiting wideband spectral information of real-valued radio frequency (RF) signals using the Nyquist Folding Receiver architecture. A new system model based on a symmetric modulation matrix is introduced so that the frequency band of the real input signals can be estimated without in-phase and quadrature reception and processing. To recover the original frequency of the input RF signal, we use the parameter-free sparse learning via iterative minimization (SLIM) method. Finally, the proposed model and the success of the recovery algorithm are demonstrated with data collected from an experimental testbed.

1 Introduction

Continuous spectrum monitoring is important for tracking spectrum occupancy as well as detection and identification of hostile or unauthorized transmission activities. Meanwhile, improvements in Radio Frequency (RF) hardware technologies allow modern radars to operate up to terahertz frequencies, which increases the difficulty of wideband spectrum monitoring. The recently proposed Nyquist Folding Receiver (NYFR) folds a broadband signal into a narrow sampling bandwidth prior to digitization by a narrowband analog-to-digital converter (ADC) [2]. The captured data has a time-varying frequency modulation that varies with Nyquist zone, such that the embedded information can be used to identify a signal's original RF band.

A basic NYFR architecture is depicted in Figure 1. In the NYFR, the local oscillator (LO) signal is modulated such that its frequency varies with time. Harmonics of the LO signal have increased amounts of frequency modulation. Therefore, RF signals at increasingly higher frequencies are demodulated with harmonics of the primary LO signal, and these harmonic signals have unique modulation strengths that vary with harmonic number [3]. Then, the baseband output of the NYFR has a unique Nyquist-zone-dependent frequency modulation that can be exploited to recover the original input signal.

Different approaches have been proposed in the literature to exploit the information that is embedded in the received signal. X-Gram processing, which depends on demodulating the input signal for all Nyquist zones of interest, is investigated in [2]. It is reported that X-Gram is computationally intense since it computes all Nyquist zone of interest. Moreover, it needs a 2D peak detection algorithm to identify the input frequencies. The measurement of induced modulation through time-frequency analysis such as the spectrogram, wavelet transform, and Wigner–Ville transform are studied in [2] and [3]. In the spectrogram approach, the tradeoff between time and frequency resolution makes it difficult to detect modulation information [2]. A Wigner-Ville transform can be used to achieve good time and frequency resolution; however, it is computationally intense due to the second order terms involved in the processing.

In this paper, we first revisit the system model and modify it so that real-valued signals can be represented without the need of explicit in-phase and quadrature receiver paths. Next, we investigate different compressive sensing (CS) methods to recover real-valued signals. We propose a solution for information recovery based on the sparse learning via iterative minimization (SLIM) method, which unlike many other methods in the literature, can identify the original RF band without any user-dependent parameter selection. Finally, we demonstrate the application of the proposed method with data from an experimental testbed.

2 System Model and Derivation

Let us assume that the input to the LO port of a harmonic mixer is a frequency-modulated tone centered at \( \omega_{s1} = 2\pi F_{s1} \), such that

\[
s_{LO}(t) = \sin(\varphi(t))
\]

where \( \varphi(t) = \omega_{s1}t + \theta(t) \). The LO frequency should be continuously modulated to introduce a corresponding time-varying and RF-dependent modulation on signals at the output. Thus, one may define the information modulation of the LO signal as

\[
\theta(t) = \frac{F_A}{F_m} \sin(2\pi F_m t + \alpha),
\]

where \( F_A \) is the LO frequency deviation, \( F_m \) is the frequency of the LO modulations (rate of the change of the frequency deviations) and \( \alpha \) is the known initial phase of modulation.
In a harmonic mixer, narrow pulses are produced at the zero crossings of the LO signal, \( s_{LO}(t) \) (in our case, positive-slope zero crossings). In other words, the zero crossings occur when \( \phi(t) = 2\pi k \). Because of the frequency modulation of the LO signal, the spacing of these narrow 'sampling' pulses are non-uniform, and the time-varying rate is

\[
F_s(t) = \frac{1}{2\pi} \phi'(t) = \frac{1}{2\pi} \frac{d}{dt} (\omega_c t + \theta(t)) = F_s + \theta'(t)
\]  

(3)

crossings per second.

By definition, a Dirac comb is constructed from a train of Dirac delta functions according to

\[
\Delta(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT),
\]  

(4)

for some given period \( T \). Multiplying any function by a Dirac comb transforms it into a train of weighted impulses at the nodes of the comb. If a real, narrowband RF input signal \( x(t) \in \mathbb{R} \) centered at \( \omega_c = 2\pi F_c \) with phase \( \phi \) is defined as

\[
x(t) = \cos(\omega_c t + \phi),
\]  

(5)

then the output of the harmonic mixer can be expressed as

\[
y(t) = x(t)\Delta(t) = \sum_k x(t)\delta(t - tk).
\]  

(6)

The conversion from zero crossings to impulses happens inside the harmonic mixer when \( t = tk \), which occurs when \( \phi(t)|_{t=\frac{tk}{2\pi}} = 2\pi k \). According to the Dirac scaling property, the non-uniform pulse train takes the form \([1, 4]\)

\[
\tilde{\Delta}(t) = \phi'(t) \sum_{k=-\infty}^{\infty} 2\pi \delta(\phi(t)-kT).
\]  

(7)

Using the identity \([1, 5]\)

\[
2\pi \sum_k \delta(v - 2\pi k) = \sum_k e^{jvk}
\]  

(8)

with \( v = \phi(t) = \omega_c t + \theta(t) \), we can rewrite (7) as

\[
\tilde{\Delta}(t) = \phi'(t) \sum_{k=-\infty}^{\infty} e^{j(k\omega_c t + \theta(t))}.
\]  

(9)

Because the frequency modulation is narrowband and satisfies \( \omega_c \gg \max|\theta'(t)| \), it is possible to further simplify to

\[
\tilde{\Delta}(t) \approx \omega_c \sum_{k=-\infty}^{\infty} e^{j(k\omega_c t + \theta(t))}.
\]  

(10)

Substituting (5) and (10) into (6) results in

\[
y(t) \approx \sum_k \cos(\omega_c t + \phi) e^{j(k\omega_c t + \theta(t))},
\]  

(11)

which can be expanded further to

\[
y(t) = \frac{1}{2} \sum_k \left( e^{j(k\omega_c + \phi)} + e^{-j(k\omega_c + \phi)} \right) e^{j(k\omega_c t + \theta(t))}
\]  

\[
= \frac{1}{2} \sum_k e^{j(\omega_c k + \phi) + \theta(k\omega_c t)}
\]  

\[
+ e^{-j(\omega_c k + \phi) + \theta(k\omega_c t)}
\]  

\[
\frac{1}{2} \sum_{k \neq k_H} \left( e^{j(\omega_c k + \phi) + \theta(k\omega_c t)} + e^{-j(\omega_c k + \phi) + \theta(k\omega_c t)} \right).
\]  

(12)

Let \( k_H \) be the harmonic in the Fourier series of the impulse train that satisfies \( |\omega_c - k_H\omega_s| \leq \frac{1}{2} \omega_s \), then \( y(t) \) can be written as

\[
y(t) = \frac{1}{2} \left( e^{j(\omega_c - k_H\omega_s) + \phi + k_H\theta(t)} + e^{-j(\omega_c - k_H\omega_s) + \phi - k_H\theta(t)} \right)
\]  

\[
+ \frac{1}{2} \sum_{k \neq k_H} \left( e^{j(\omega_c k + \phi) + \theta(k\omega_c t)} + e^{-j(\omega_c k + \phi) + \theta(k\omega_c t)} \right).
\]  

The output of the harmonic mixer \( y(t) \) is filtered with a lowpass (LP) anti-aliasing filter before digitization. For now, assuming an ideal filter with a cutoff frequency \( \frac{1}{2} \omega_s \), only the terms with RF frequency in the \( k_H \) harmonic’s Nyquist zone pass the filtering because \( |\omega_c - k_H\omega_s| \leq \frac{1}{2} \omega_s \). Any other terms \( k \neq k_H \) are rejected by the LP filter. Thus, the output of the anti-aliasing filter is

\[
z(t) = \frac{1}{2} \left( e^{j(\omega_c + k_H\omega_s) + \phi} + e^{-j(\omega_c + k_H\omega_s) + \phi} \right) e^{j2\pi k \theta(t)}.
\]  

(14)

Note that the first two exponential terms in \( z(t) \) are the time-domain representations of any signals (positive and negative frequency of the spectrum) that passed through the LP filter. The Nyquist-zone-dependent frequency modulation impressed on \( z(t) \) is present in the last term \( e^{j2\pi k \theta(t)} \).

### 2.1 NYFR Compressive Sensing (CS) Model

It is possible to express the input-output relationship of the NYFR as a CS model to separate and recover the input signal.
Let $X = [x_0, x_1, \cdots, x_{N-1}]$ be the length-$N$ Discrete Fourier Transform (DFT) of the Nyquist-rate sampled input signal $x(t)$. Because we previously defined $x(t)$ as a real signal (which is always true for physical RF signals), its DFT $X$ has symmetry. Moreover, $X$ is highly compressible (sparse) in the frequency domain due to the narrowband assumption. By defining a sensing matrix $H$, the system model can be written in compact form such that

$$z = HX,$$

where $z$ is the $K \times 1$ measurement vector (real-valued ADC samples). In this formulation, $X$ consist of $Z$ Nyquist zones (folds) each of length $K$ ($Z = N/K$).

As seen from Figure 2, the $K \times N$ sensing matrix can be explicitly written as

$$H = RS\Psi,$$

where $\Psi$ is the block diagonal matrix comprising a modified inverse Discrete Fourier Transform (IDFT) matrix $D$, such that

$$\Psi = I_Z \otimes D.$$

The $Z \times Z$ identity matrix is represented by $I_Z$, and $\otimes$ denotes the Kronecker product. The modified Inverse Discrete Fourier Transform matrix $D$ transforms the positive and negative frequency halves of each sub-Nyquist zone separately to a time-domain signal in the form of (14). The traditional IDFT matrix is defined by $W_{m,n} = e^{j2\pi mn/K}$, which can be separated into two halves that cover the positive and negative frequency of the spectrum as

$$W = \begin{bmatrix} W_+ & W_- \end{bmatrix}.$$  

The $2K \times K$ modified IDFT matrix $D$ is formed from the subparts of the regular $K \times K$ IDFT matrix $W$ according to

$$D = \begin{bmatrix} W_+ & 0_{K,K/2} \\ 0_{K,K/2} & W_- \end{bmatrix}.$$  

where $0_{m,n}$ represents the $m \times n$ all-zero matrix.

The induced sample modulation matrix $S$ is a conjugate-symmetric $2N \times 2N$ diagonal matrix whose entries are partitioned into $2Z$ sub-blocks,

$$S_p = e^{j\pi \theta(s)} I_K$$

where $p = [0, 1, \ldots, Z]_N$, $I_n$ is discrete-time at the ADC sample rate, and $k_H = \text{round} (\omega_s/\omega_s)$ which is in the form of $k_H = [0, 1, 2, 3, \ldots, Z/2, Z/2]$.

The projection matrix $R$ that folds the $Z$ Nyquist bands onto baseband is the horizontal concatenation of $2Z$ identity matrices, each of size $K \times K$,

$$R = J_{1,2Z} \otimes I_K$$

where $J_{m,n}$ is the $m \times n$ unit matrix (all-ones matrix).

3 Information Recovery

It is possible to recover (estimate) the original input RF signal from the received signal by solving the system (15) introduced in the previous section. The sensing matrix $H$ is wider than it is tall ($N = ZK > K$), which means the system has more unknowns than observations. This kind of system is known as an under-determined system and has infinitely many solutions under the assumption that $HH^*$ is invertible (where $(\cdot)^*$ denotes conjugate transpose). It is possible to use the linear least-squares approach ($\ell_2$ norm) to solve this system of equations [6]. However, the system in (15) has special sparsity properties that can be further exploited by sparse solvers.

Sparsity-based signal reconstruction algorithms can be divided into two categories. The first includes greedy algorithms including Matching Pursuit (MP) [7] and Orthogonal Matching Pursuit (OMP) [8]. The application of greedy algorithms to the NYFR architecture and their weakness are discussed in [2].
The second category includes convex optimization algorithms such as basis pursuit [9].

One may define a convex optimization problem to solve (15) as

$$\arg\min_X \lambda \|X\|_1$$

such that $z = HX$  \hspace{1cm} (22)

where $\|X\|_1$ is the $\ell_1$-norm of vector $X$.\(^1\) The optimization defined in (22) is known as the basis pursuit (BP) problem and is usually applied in cases where there is an under-determined system of linear equations that must be exactly satisfied and the sparsest solution in the $\ell_1$ sense is desired one [9].

Note that in many cases, especially in real applications, the observation (received signal) is noisy,

$$z = HX + \sigma \mathcal{N},$$

and it does not make sense to solve (22) exactly. Instead, it is desirable to trade off exact congruence in exchange for a sparser estimation. In these cases, an approximate solution can be found by minimizing the cost function,

$$\arg\min_X \|z - HX\|_2^2 + \lambda \|X\|_1$$

which is known as the basis pursuit denoising (BPDN) problem. In this formulation, $\lambda$ controls the trade-off between sparsity and reconstruction fidelity and needs to be selected appropriately to achieve a sparse solution.

Different approaches have been proposed in the literature to solve basis pursuit (22) and basis pursuit denoising (24) optimization problems such as FISTA [10], SpaRSA\(^2\) [11], SALSA [12], etc. There is a tradeoff between BP and BPDN approaches. For instance, BP (22) preserves the input signal; however, it is not suitable for real data because it does not address the noise issue. On the other hand, BPDN needs an appropriate $\lambda$ value to recover the input signal in noisy environments. It is possible to estimate a good value for $\lambda$ when $\mathcal{N}$ is standard white Gaussian noise and the noise level $\sigma$ is known [9], but $\lambda$ also needs to be adjusted according to the input signal level and signal sparsity.

In addition to the $\ell_1$-norm approach, there are parameter-free $\ell_q$-norm approaches available in the literature (for $0 < q \leq 1$), such as Sparse Learning via Iterative Minimization (SLIM) [13, 14]. SLIM is a regularized minimization approach with an $\ell_q$-norm constraint and can also be regarded as a natural extension to $\ell_1$-norm based approaches [14, 15]. SLIM considers the regularized minimization algorithm for sparse recovery as

$$\min_{X, \eta} g_q(X, \eta),$$

where

$$g_q(X, \eta) = K \log \eta + \frac{1}{\eta} \|z - HX\|_2^2 + \sum_{n=1}^{N} \frac{2}{q} (|X_n|^q - 1)$$

\hspace{1cm} (26)

\(^1\)For a length $N$ signal $u$, the $\ell_1$-norm is denoted by $\|u\|_1 = \sum_{n=0}^{N-1} |u_n|$ and the "sum of squares" of $u$ is denoted by $\|u\|_2^2 = \sum_{n=0}^{N-1} |u_n|^2$.

\(^2\)The application of SparSA to the NYFR architecture is discussed in [2].

4 Experimental Setup

An experimental testbed, shown in Figure 3a, was assembled to verify the proposed NYFR architecture for real signal recovery. The testbed consisted of an arbitrary waveform generator (AWG) that generated the frequency-modulated LO input signal $s_{LO}(t)$, at least two RF signal generators with an RF combiner that provide the input test signal, Nyquist Folding Receiver (actual hardware), a workstation with ADC for data acquisition, and a pulse generator for triggering.

The top view of the NYFR box is shown in Figure 3b while its high-level schematic is depicted in Figure 4. In its current configuration, the ADC is set to operate at 1.5 GSPS. Thus, the cutoff frequency of the low-pass filters after the harmonic mixer and after the amplifier are set to 750 MHz.

During the course of the experiment, it was observed that unintended high-order harmonics generated by the AWG operating in loopback mode were mixed with the input RF signal to yield shadows of the input RF signals with different modulation factors. To overcome this issue, a low-pass filter was placed between the AWG and mixer LO port as shown in Figure 4.
The LO frequency deviation $F_\Delta$ was set to $\pm 4$ MHz whereas the frequency of the LO modulations $F_m$ is set to 5 MHz. The center frequency of the LO signal $F_s$ was set to 1.5 GHz to match the ADC sampling rate as well as the cutoff frequency of the LP filters after the harmonic mixer. The initial phase the LO modulation $\alpha$ was maintained at a constant value by triggering the LO signal and data acquisition (ADC) simultaneously ($\alpha \approx 0$).

A single-tone test signal was generated according to (5) where the center frequency $F_c$ was set to 3.2 GHz. Figure 5 shows the comparison of the spectrum of the ideal system response and the collected data. As seen from the figure, the spectrum of the received signal (red line) corresponds to the proposed system model (blue line). Although the input signal was a narrowband tone, the received signal is modulated over a broader frequency band due to the modulation introduced by the LO signal.

Figure 6 shows the spectrogram of the received signal for a test signal that includes two tones at 2.5 and 9.4 GHz (with same power). It can be seen in the figure that the induced modulation has a 2 $\mu$sec period corresponding to the frequency of the LO modulation $F_m = 5$ MHz ($F_m = 1/T_m$). The tone at 2.5 GHz (lower sideband of the 3-GHz harmonic) is folded to 500 MHz; Similarly, the tone at 9.4 GHz (upper sideband of the 9-GHz harmonic) is folded to 400 MHz. A 180° phase shift is observed in the modulation of the upper and lower sideband signals, which fits the theoretical derivation in (14). The amount of frequency modulation is defined by the folding zone (corresponding harmonic in the Fourier series), which is $k_H = 2$ and $k_H = 6$ for the 2.5- and 9.4-GHz input signals, respectively. Thus, there is a factor of $6/2 = 3$ difference in the modulation between the two received signals (see in Figure 6).

Next, we set the frequency grid resolution (accuracy of the estimation) to 5 MHz and apply SLIM for signal estimation. The recovered signal $\hat{X}$ for the two-tone example is shown in Figure 7. SLIM algorithm estimates two tones exactly at 2.5 and 9.4 GHz without any user inputs (parameter).

5 Discussions

In Figure 7, it should be noted that the magnitude of the estimated signals are different even though the input powers were set to be the same. To investigate the power differences, we empirically measured the frequency response of the NYFR system, with the results shown in Figure 8. From the figure, the difference between the frequency responses at 2.5 and 9.4 GHz RF signal.
and 9.4 GHz is approximately 8.74 dB (factor of $\approx 2.74$ in magnitude). The peak magnitudes are observed at 179.8 and 709.3 for 9.4 and 2.5 GHz respectively. Therefore, the factor of 2.74 is not quite sufficient to equalize the output powers $(2.74 \times 179.3 = 491.28)$, but accounts for some of the difference. Additional system calibration and characterization is needed to improve the equalization across a wide RF spectrum.

The rolloff of the anti-aliasing LP filters is seen to produce notches at every 750 MHz difference from an LO harmonic. This notching causes blind zones around $k$ notches at every 750 MHz difference from an LO harmonic. The rolloff of the anti-aliasing LP filters is seen to produce additional aliasing. Additional system calibration and characterization is needed to improve the equalization across a wide RF spectrum.

6 Conclusion

We present the implementation of a Nyquist-Folding Receiver architecture for recovery of wideband spectral information. A sensing model is defined that exploits the symmetry of the Fourier transform of real signals so that the frequency band (upper or lower sideband) of the input RF signals can be resolvable without in-phase and quadrature processing. We investigate different CS formulations and discussed their pros and cons, while demonstrating the use of the parameter-free Sparse Learning via Iterative Minimization to recover the origin of the input RF signal. The proposed model was demonstrated and verified through collected data with real hardware, and the success of the recovery algorithm was demonstrated via an experimental testbed.

Acknowledgment

This work is partially supported by Kansas City National Security Campus operated by Honeywell FM&T, Contract DE-NA0002839. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the agency that provided support for the project.

References


