

LEARNING SPARSE SYSTEMS AT SUB-NYQUIST RATES: A FREQUENCY-DOMAIN APPROACH

Martin McCormick, Yue M. Lu and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Switzerland

ABSTRACT

We propose a novel algorithm for sparse system identification in the frequency domain. Key to our result is the observation that the Fourier transform of the sparse impulse response is a simple sum of complex exponentials, whose parameters can be efficiently determined from only a narrow frequency band. From this perspective, we present a sub-Nyquist sampling scheme, and show that the original continuous-time system can be learned by considering an equivalent low-rate discrete system. The impulse response of that discrete system can then be adaptively obtained by a novel frequency-domain LMS filter, which exploits the parametric structure of the model. Numerical experiments confirm the effectiveness of the proposed scheme for sparse system identification tasks.

Index Terms— Sparse system identification, LMS, finite rate of innovation, sub-Nyquist sampling.

1. INTRODUCTION

In many practical applications, the impulse response of an unknown system can be modeled as a linear combination of several Diracs, *i.e.*,

$$h(t) = \sum_{k=1}^K c_k \delta(t - t_k), \quad (1)$$

where $\{c_k\}_{k=1}^K$ and $\{t_k\}_{k=1}^K$ are some unknown parameters. Examples of such sparse systems include certain wireless communication channels, acoustic room impulse responses, as well as those seen in ultra-wideband (UWB) ranging and line echo cancellation problems, to name a few. In this paper, we study the *adaptive* learning of the sparse system specified in (1), from a number of its input and output instances.

Linear system identification is a classical and well-studied subject. The common practice in learning a general impulse response $h(t)$ can be summarized as follows. Let $x(t)$ and $y(t)$ be the input and output signal of the unknown system, respectively. Suppose that the input $x(t)$ is bandlimited, we can then convert $x(t)$ and $y(t)$ into the discrete domain by sampling them at or above the Nyquist rate. Since the sampling process is alias-free, the resulting discrete signals—denoted by $x[n]$ and $y[n]$, respectively—are also related by a linear shift-invariant system, *i.e.*,

$$y[n] = (x * h_d)[n].$$

The discrete filter $h_d[n]$ above either completely determines the original continuous-time impulse response $h(t)$ [if the latter is bandlimited and has a bandwidth below that of $x(t)$], or is a lowpass approximation of $h(t)$. In practice, we usually approximate $h_d[n]$ by a finite impulse response (FIR) filter with N taps, where N is chosen to be large enough to retain most of the energy of $h_d[n]$. This finite-order system can then be learned from $x[n]$ and $y[n]$ by using standard adaptive filtering techniques, such as the least mean squares (LMS) algorithm [1].

Despite its widespread and successful use in practice for learning *general* linear time-invariant systems, the traditional approach described above does not exploit the special sparse structure of the unknown system in our problem. Intuitively, since the impulse response in (1) can be specified by only a small number of parameters ($2K$, to be exact), one should expect a much more efficient scheme in learning $h(t)$.

Note that in the traditional scheme for system identification, the input and output of the system must be sampled at or above the Nyquist Rate. In some cases (e.g. UWB ranging), this acquisition rate can be very high. Therefore, the first problem we want to address in this paper is as follows.

Problem 1: By exploiting the sparse structure of the system, can we reduce the sampling rate to below the Nyquist rate, but still recover the entire impulse response?

After sampling, the discrete filter $h_d[n]$ is approximated by an FIR model, whose length, N , must be large enough to capture the longest possible delay in the system. More specifically, $N \sim (\max_k t_k)/T$, where T is the sampling interval. This is generally much larger than the number of delays that make up the system itself. Hence, the second problem we want to address is on reducing the model order:

Problem 2: Can we learn the sparse system by an FIR filter who has roughly the same order as the sparsity level K ?

In this paper, we address the above two problems by proposing a new adaptive algorithm for sparse system identification, which combines a sub-Nyquist sampling step and a new frequency-domain LMS algorithm. Key to the proposed scheme is the observation that the Fourier transform of the sparse impulse response is a simple sum of complex exponentials. By borrowing tools from spectral analysis [2] and array signal processing, we can use only a small band of the Fourier transform to recover its frequencies, phases, and amplitude, the collection of which parameterizes the entire system.

The rest of the paper is organized as follows. We show in

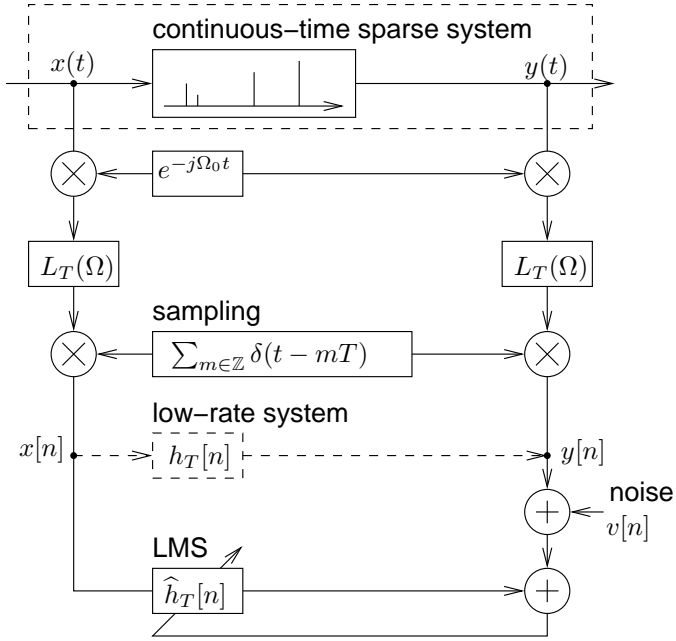


Fig. 1. The block diagram of the proposed sparse system identification algorithm.

Section 2 that the original continuous-time system can be *equivalently* determined by considering a sampled discrete-time system, whose sampling rate can vary freely below the Nyquist rate. Meanwhile, with sufficient subsampling, this equivalent discrete system can be learned by running LMS for an FIR filter with only $2K$ taps. By exploiting the prior knowledge of the complex exponential structure, we present in Section 3 a new frequency-domain LMS algorithm with significantly improved convergence performance. The parameters of the original continuous-time system can then be recovered by standard tools in harmonic retrieval, as outlined in Section 4. Numerical experiments in Section 5 confirm the effectiveness and efficiency of the proposed sub-Nyquist algorithm for sparse system identification.

2. A SUB-NYQUIST SAMPLING SCHEME FOR SPARSE SYSTEM IDENTIFICATION

Figure 1 shows the block diagram of the proposed algorithm for learning a continuous-time sparse impulse response $h(t)$ as specified in (1). We first apply a frequency modulation $e^{-j\Omega_0 t}$ to the input $x(t)$ and the output $y(t)$, each of which is then passed through a lowpass filter $L(\Omega)$, defined as

$$L(\Omega) = \mathbb{1}_{[-\pi/T, \pi/T]}(\Omega), \quad (2)$$

where $\mathbb{1}_{\mathcal{F}}(\cdot)$ denotes the indicator function defined on an interval \mathcal{F} . Subsequently, we sample the two filtered signals at rate $1/T$ and obtain

$$x_T[n] \stackrel{\text{def}}{=} x(nT) \quad \text{and} \quad y_T[n] \stackrel{\text{def}}{=} y(nT).$$

Note that the traditional setup for system identification as described in Section 1 can be regarded as a special case of the

above scheme, when we set the modulation frequency $\Omega_0 = 0$ and let the sampling rate $1/T$ equal to the Nyquist rate of the input signal $x(t)$. In what follows, we are interested in cases when the acquisition rate $1/T$ is lower than the Nyquist rate.

Lemma 1 *The discrete signals $x_T[n]$ and $y_T[n]$ are related by a discrete-time filtering operation*

$$y_T[n] = (x_T * h_T)[n],$$

where the filter $h_T[n]$ has a discrete-time Fourier transform

$$H_T(e^{j\omega}) = \sum_{k=1}^K c_k e^{-j\Omega_0 t_k} e^{-j\omega t_k / T}. \quad (3)$$

Proof: The specification of the lowpass filter $L(\Omega)$ in (2) guarantees that there is no aliasing in the sampling process. Consequently, the discrete-time Fourier transform of $x_T[n]$ can be obtained as

$$X_T(e^{j\omega}) = \frac{1}{T} X\left(\frac{\omega}{T} + \Omega_0\right), \quad \text{for } |\omega| \leq \pi, \quad (4)$$

where $X(\Omega)$ denotes the continuous-time Fourier transform of $x(t)$. Similarly, for $y_T[n]$, we have

$$\begin{aligned} Y_T(e^{j\omega}) &= \frac{1}{T} Y\left(\frac{\omega}{T} + \Omega_0\right) \\ &= \frac{1}{T} X\left(\frac{\omega}{T} + \Omega_0\right) H\left(\frac{\omega}{T} + \Omega_0\right), \quad \text{for } |\omega| \leq \pi, \end{aligned} \quad (5)$$

where $H(\Omega)$ is the Fourier transform of $h(t)$. From the definition of $h(t)$ in (1),

$$H(\Omega) = \sum_{k=1}^K c_k e^{-j\Omega t_k}.$$

Substitute the above equality into (5) and compare with (4), we can verify that

$$Y_T(e^{j\omega}) = X_T(e^{j\omega}) H_T(e^{j\omega}),$$

where $H_T(e^{j\omega})$ is as defined in (3).

Proposition 1 *The sampled discrete-time system $h_T[n]$ specified in (3) completely determines the original continuous-time sparse system $h(t)$.*

Proof: On the one hand, the parameters $\{t_k\}_{k=1}^K$ and $\{c_k\}_{k=1}^K$ from (3) clearly determine $h(t)$. On the other hand, it is well-known from harmonic retrieval [2] that these parameters can be uniquely determined from at least $2K$ values of $H_T(e^{j\omega})$, with a variety of algorithms available for that purpose. See Section 4 for an outline of a concrete algorithm.

Remark 1 *In principle, Proposition 1 holds for arbitrary choices of T or Ω_0 . Thus, by increasing T , the sampling rate for the equivalent discrete-time system can be set to well below the Nyquist rate. Additionally, the modulation frequency Ω_0 can be used to select a frequency band of $x(t)$ with the highest signal to noise ratio (SNR). This is useful when either the input signal $x(t)$ or the noise is not white.*

3. LEARNING THE DISCRETE-TIME SYSTEM

In this section, we present a novel frequency-domain LMS algorithm for learning the equivalent low-rate discrete-time system specified in Proposition 1.

3.1. FIR Model Approximation

It follows from (3) that the discrete-time filter $h_T[n]$ has the form

$$h_T[n] = \sum_{k=1}^K c_k e^{-j\Omega_0 t_k} \frac{\sin(\pi(n - t_k/T))}{\pi(n - t_k/T)}.$$

In general, this filter is of infinite length, unless the t_k parameters are all integer multiples of T . In what follows, we approximate $h_T[n]$ by an FIR filter of length N , where

$$N \geq \left\lceil \frac{\max(t_i)}{T} \right\rceil + c. \quad (6)$$

for some constant c . We can show that, even for relatively small c , the above truncated FIR filter, denoted by $\tilde{h}_T[n]$, can capture most of the energy of $h_T[n]$.¹

Furthermore, we can show that the discrete Fourier transform (DFT) of the truncated filter, denoted by $H[m]$, can be well approximated by a discretized version of (3), *i.e.*,

$$H_m = \sum_{k=1}^K c_k e^{-j\Omega_0 t_k} e^{-j2\pi \frac{t_k m}{TN}}, \quad m = 0, 1, \dots, N-1. \quad (7)$$

By increasing the sampling interval T , the model order N can be decreased, subject to the constraint in (6). In our following discussions, however, we set the lower bound of N to be $2K + 1$, the reason of which will be clear in Section 4.

3.2. A New Frequency-Domain LMS Algorithm

Many types of adaptive filters can be used to learn the FIR filter $\tilde{h}_T[n]$, or equivalently its DFT, $\{H_m\}_{m=0}^{N-1}$. The frequency-domain LMS algorithm (FLMS) [1], in particular, learns the latter directly; at each iteration, an estimate $\hat{\mathbf{H}} = [\hat{H}_0, \dots, \hat{H}_{N-1}]^T$ is made by following a gradient descent step.

In this work, we propose an improved frequency-domain LMS algorithm that can exploit the special parametric structure of the filter given in (7). First, we form a Toeplitz matrix

$$\mathbf{A} = \begin{bmatrix} H_L & H_{L-1} & \dots & H_0 \\ H_{L+1} & H_L & \dots & H_1 \\ \vdots & \ddots & \ddots & \\ H_N & H_{N-1} & \dots & H_{N-L} \end{bmatrix}_{(N-L+1) \times (L+1)}, \quad (8)$$

where $L = \lfloor \frac{N}{2} \rfloor$. In the noiseless case, it can be shown that the rank of \mathbf{A} must be equal to K [3]. This property can be used

¹For example, $c = 5$ will include at least 99% of this sidelobe energy for the largest t_k in the worst case.

to accelerate the convergence of the FLMS algorithm, as summarized in the following procedure, which we call the Cadzow-FLMS.

For each iteration of FLMS, let $\hat{\mathbf{H}}_n$ denote the current estimate of the DFT.

1. Find the standard update vector of FLMS, and call it \mathbf{S} . In normal FLMS, the next estimate would be $\hat{\mathbf{H}}_{n+1} = \hat{\mathbf{H}}_n + \mathbf{S}$.
2. Form the Toeplitz matrix \mathbf{A} from $\hat{\mathbf{H}}_{n+1}$.
3. To enforce the rank condition, compute the SVD $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Build a diagonal matrix \mathbf{S}' from \mathbf{S} by zeroing all terms less than the K most significant ones and produce $\mathbf{A}' = \mathbf{U}\mathbf{S}'\mathbf{V}^T$.
4. Compute \mathbf{C} , the denoised version of $\hat{\mathbf{H}}_{n+1}$, by averaging the diagonals of \mathbf{A}' .
5. Update $\hat{\mathbf{H}}_{n+1} = \hat{\mathbf{H}}_n + \mathbf{S} + (\mathbf{I} - \frac{\mathbf{S}\mathbf{S}^T}{\mathbf{S}^T\mathbf{S}})(\mathbf{C} - \hat{\mathbf{H}}_n)$.

Intuitively, the SVD “denoising” step brings the current estimate closer to the model assumption in (7). Meanwhile, the matrix $(\mathbf{I} - \frac{\mathbf{S}\mathbf{S}^T}{\mathbf{S}^T\mathbf{S}})$ in the last step of the update procedure ensures that the inner product between \mathbf{S} and the “denoised” update \mathbf{C} is always positive, which is important in guaranteeing the convergence of the above algorithm. Due to space limitations, we omit further discussions on this convergence property.

Numerical experiments in Section 5 show that the proposed algorithm described above can dramatically improve the convergence rate and the steady state error of the FLMS with sparse system identification.

4. RECOVERING THE ORIGINAL CONTINUOUS-TIME SPARSE SYSTEM

After finding the FIR model $H[m]$ as in (7), the final step in our algorithm is to estimate the parameters $\{c_k\}$ and $\{t_k\}$, which are then used to reconstruct the original continuous-time sparse impulse response $h(t)$.

Many algorithms are available for this parameter estimation task. See [2] for an extensive review. In what follows, we briefly outline the Prony’s method, which is used in our experiments.

1. From the Fourier coefficients $\{\hat{H}_m\}_{m=0}^{N-1}$ learned from the Cadzow-FLMS algorithm, form the Toeplitz matrix \mathbf{A} as in (8). This time, set the parameter $L = K + 1$.
2. Perform the SVD of $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.
3. Take the smallest right singular vector, $\mathbf{a} = [a_0, a_1, \dots, a_k]^T$, *i.e.*, the column of \mathbf{V} with the smallest singular value. Form a K th-order polynomial whose coefficients are elements of \mathbf{a} . The K zeros of this polynomial lie directly on the complex exponential frequencies. Taking the complex angles of these roots then gives a list of 2π -normalized frequencies.

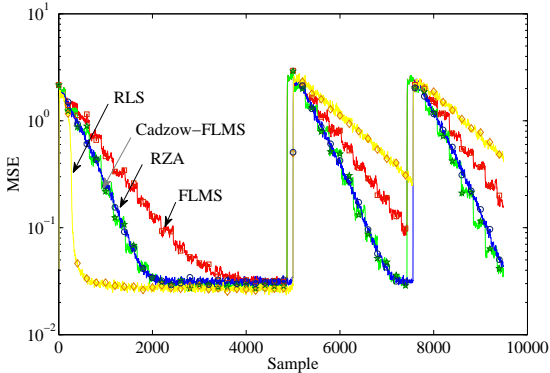


Fig. 2. The mean-square output error for sparse system identification.

4. Multiplying this list by $\frac{NT}{2\pi}$ gives the values $\{t_k\}_{k=1}^K$.
5. The parameters $\{c_k\}_{k=1}^K$ are found from the least-squares solution to $\hat{H}_m = \sum_{k=1}^K c_k e^{-j\Omega_0 t_k} e^{-j2\pi \frac{t_k m}{TN}}$.

5. NUMERICAL EXPERIMENTS

To examine the performance of the proposed Cadzow-FLMS for sparse system identification, we compare it to normal FLMS, Recursive Least Squares (RLS), and another effective sparse LMS method: the reweighted zero-attracting method (RZA) proposed in [4].

For fair comparison, the parameters for each algorithm are made equal when possible. We imitate the test setup in [4] which identifies an unknown FIR system with 256 taps, containing 28 nonzero coefficients. The parameters used for RZA are from [4]. White Gaussian noise (WGN) of unit variance is used as input. The output is observed with SNR of 30dB. FLMS and Cadzow-FLMS each operate on overlapping blocks of 512 samples. The unknown system taps are perturbed slightly at sample 5000 to show the ability of the algorithms to track a changing system. A larger shift in the system occurs at sample 7500². The improvement from applying Cadzow to FLMS is clearly observed in the output MSE plot in Figure 2.

In comparing RZA to Cadzow-FLMS, it appears that the convergence rate, final MSE, and adaptability are quite similar. In Figure 2, Cadzow-FLMS is only barely distinguishable from RZA by its periodic jumps (updates are made only once per block for FLMS methods). Each of the methods require some parameters, RZA being somewhat sensitive its to values of ρ , and ϵ , and Cadzow FLMS requiring an (over)estimate of K .³

In terms of computational complexity, Cadzow FLMS must perform an SVD per block of N samples—this requires $O(\frac{N^3}{N}) = O(N^2)$ computation time as opposed to $O(N)$ for RZA or $O(\log N)$ for general FLMS. On the other hand, Cadzow FLMS proves more versatile in that sparsity can be

²In particular, RLS is seen to not adapt well to these changes, even when a small “forgetting factor” is applied.

³ K can be estimated as the number of singular values of \mathbf{A} above a threshold.

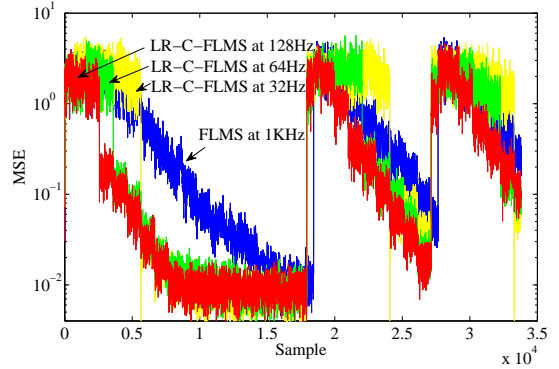


Fig. 3. Comparing MSE of low rate Cadzow LMS.

modeled off of the sampling grid—a situation that RZA does not handle well. In addition, Cadzow FLMS can operate at a lower sampling rate than RZA or others—effectively reducing its computational complexity to $O(K^2)$, where K is the system sparsity.

To illustrate this point, a second setup with various sampling rates of Cadzow-FLMS is constructed. A 512Hz WGN signal of unit variance passes through a system with 5 different reflections (t_k) between 0 and 1 second. The output is observed at an SNR of 40dB after corruption by WGN. Low-rate Cadzow-FLMS (LR-C-FLMS) is performed at 128Hz, 64Hz, and 32Hz—using 12, 6, and 3 times the minimum parameters ($2 \times 5 + 1 = 11$), respectively. These low rate versions are compared to FLMS running at the full rate (1KHz) with block size $N = 1024$.

It is clear from the simulations that all versions of LR-C-FLMS converge at about the same rate. Initial MSE is high for a longer period for lower rates because a higher model SNR is needed to properly reconstruct from fewer parameters.

6. CONCLUSION

We proposed a novel algorithm for sparse system identification in the frequency domain. The parametric structure of sparse impulse responses allows for a reduction in the sampling rate and the model order, along with the freedom to choose a band with higher SNR. Numerical experiments indicate the advantages of the proposed algorithm over standard system identification methods (e.g., LMS and RLS) as well as other sparsity-exploiting methods [4, 5] operating in the time domain.

7. REFERENCES

- [1] A. H. Sayed, *Adaptive Filters*, Wiley-IEEE Press, Hoboken, NJ, 2008.
- [2] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1997.
- [3] J. A. Cadzow, “Signal enhancement—A composite property mapping algorithm,” *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.
- [4] Y. Chen, Y. Gu, and A. O. Hero, “Sparse LMS for system identification,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Taipei, 2009.
- [5] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, “Exploiting sparsity in adaptive filters,” *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1883–1894, Aug. 2002.