

DESIGNING SIGNALS WITH GOOD CORRELATION AND DISTRIBUTION PROPERTIES

Arindam Bose*, Neshat Mohammadi and Mojtaba Soltanalian

Department of Electrical and Computer Engineering
University of Illinois at Chicago
Chicago, USA

ABSTRACT

Sequences with good correlation and distribution properties play a central role in various areas of signal processing. In this paper, we propose an efficient computational framework for designing sequences with two key properties: (i) an impulse-like auto-correlation, and (ii) a probability distribution of sequence entries which is uniform in nature; although the results can be easily extended to an arbitrary distribution. The proposed method is based on utilizing the Fast Fourier Transform (FFT) operations, and thus can generate very long sequences in small time frames. Several numerical examples are provided to exhibit the performance of the suggested construction framework.

Index Terms— Auto-correlation, integrated-sidelobe level (ISL), Parkinson’s disease, sequence design, uniform distribution

1. INTRODUCTION

Sequences with good correlation properties are necessary components in a wide range of signal and information processing applications including active sensing, spread spectrum communication systems, radar sensing, signal synchronization and cryptography [1–7]. For example, sequences with small auto-correlation improve the performance in pulse compression radars. In CDMA (Code-Division Multiple Access) systems it is also desirable to have small correlation for better synchronization purposes, as well as separation of multiple users. Due to such extensive range of applications, one can find a rich literature on the design of signals with small auto-correlation sidelobe [1, 3, 8–11]. However, until now there has been very little effort on designing sequences possessing good correlation and good distribution properties which has crucial potential applications in biomedical system identification [12]. In this paper, we focus on this critical aspect missing from the recent signal design approaches developed in the literature— i.e. the distribution properties of the signal itself. Some important applications such as biomedical signal processing often requires sequences with

good correlation properties as well as a user defined distribution properties for the purpose of system identification. Such an application in the context of Parkinson’s Disease (PD) diagnosis and treatment, has been discussed in Section 1.1.

In our approach, we will use the *Cyclic Algorithm-New* (CAN), a computational framework introduced in [11], in order to achieve low out-of-phase auto-correlation, and at the same time, achieve the desired distribution by incorporating a sort-based algorithm to form partitions for various distribution bins. Note that CAN is based on FFT operations and can be effectively used for values of sequence length N up to $N \sim 10^6$ or even larger.

The remainder of the paper is organized as follows. The application of signals with good correlation and good distribution properties in Parkinson’s Disease (PD) diagnosis and treatment has been put forward in Section 1.1. Section 2 talks about the necessary background required for the proposed algorithm which is discussed in Section 3. Numerical results are given in Section 4. Finally, Section 5 concludes the paper with a proper view of future research aspect.

Notation: We use bold lowercase letters for vectors and bold uppercase letters for matrices. $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote the vector/matrix transpose, the complex conjugate, and the Hermitian transpose, respectively. $\|\mathbf{x}\|_n$ or the l_n -norm of the vector \mathbf{x} is defined as $\{\sum_k |x_k|^n\}^{\frac{1}{n}}$ where x_k is the k -th entry of \mathbf{x} . $\mathbf{0}_N$ is the all-zero vector of order $N \times 1$. Also for a real number r , $[r]$ denotes the *integral part* of r , i.e. the greatest integer $\leq r$; also $r = r - [r]$ denotes the *fractional part* of r . Here we note that, the fractional part of any real number is contained in the unit interval $I = [0, 1)$.

1.1. An Application to Biomedical System Identification

Currently, the status of PD in a patient is evaluated through the Unified Parkinson’s Disease Rating Scale (UPDRS) which is very time consuming and is very time consuming and prone to human error [12]. Hence, it is of great interest to search for tools that facilitate a quick and objective quantification of the PD status. The new framework of eye tracking for quantifying (or *system identification* [2] of) the human smooth pursuit system (SPS) promises a solution to such difficulties with a revolutionary potential [13]. It has been shown that the SPS

* Corresponding author (e-mail: abose4@uic.edu)

is meaningfully affected by the PD and that the severity of the impairment is related to the progression of the disease [14–16]. Nonetheless, among the several methods for SPS quantification, eye tracking is the only non-invasive method suggested by the literature.

In an eye tracking system, the visual stimulus consists of a moving circle whose trajectory is the signal to be designed, and the eye’s gaze direction is the output. One approach for the SPS analysis is the system theory approach where the SPS is mathematically modeled as a parameterized dynamical system correlating gaze direction to visual stimulus using the behavior of the SPS by imposing general restrictions on the model variables. In such a scenario, sequences with not only good correlation properties but also a well defined distribution is required in order to identify the system with high degree of accuracy [12].

It is widely known that a judicious design of the probing signals has a significant impact on the performance of identification. In particular, it was shown that signals with good correlation properties (i.e. with low out-of-phase auto-correlation lags) are influential for high performance SPS quantification, owing to their unique spectral properties [17]. The goal of this paper is thus not only to design and study signals possessing good correlation properties but also signals that follow a given distribution; preferably a uniform distribution, however can be converted to any other distributions.

2. PROBLEM FORMULATION

An efficient approach to sequence design for system identification is to seek for sequences with small *out-of-phase* auto-correlations, also referred to as *good correlation properties*. Let $\mathbf{x} \in \mathbb{C}^N$ be a sequence whose periodic (c_k) and aperiodic (r_k) auto-correlations are defined as

$$c_k \triangleq \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_{n+k(\text{mod } N)}^*, \quad (1)$$

$$r_k \triangleq \sum_{n=1}^{N-k} \mathbf{x}_n \mathbf{x}_{n+k}^* = r_{-k}^*, \quad (2)$$

for all $0 \leq k \leq N - 1$. The energy of \mathbf{x} is given as the inner product $\mathbf{x}^H \mathbf{x}$ which is equal to $c_0 = r_0$, and the out-of-phase lags are those with $k \neq 0$. The integrated sidelobe level (ISL) of the sequence \mathbf{x} is defined as,

$$\text{ISL} \triangleq \sum_{k=1}^{N-1} |r_k|^2. \quad (3)$$

The main focus of this paper is on the algorithms for minimizing the ISL or ISL-related metrics over a set of sequences. The significantly large application area of sequences with good correlation (in particular, with small ISL) has produced an active area of research in sequence design, and as a result, there is a rich (yet growing) literature on this topic.

On the other hand, a generic definition of (sequence \mathbf{x} with) uniform distribution can be given as follows [20]: The sequence $\mathbf{x} = \{x_n\}_{n=1}^N$ of real numbers is said to be uniformly distributed if for every pair a, b of real numbers with $0 \leq a < b \leq 1$ we have

$$\lim_{N \rightarrow \infty} \frac{\mathcal{C}([a, b]; \mathbf{x})}{N} = b - a \quad (4)$$

where, $\mathcal{C}(E; \mathbf{x})$ is the counting function defined as the number of values x_n ($1 \leq n \leq N$) for which $\{x_n\} \in E$.

In our proposed method, we achieve sequences with uniform distribution by partitioning the sequence entries into a number of range bins and populating each bin with (almost) same number of elements building a uniform histogram). In a discrete sense, for the sake of simplicity, we rewrite the definition of uniform distribution as follows:

Definition 2.1. A sequence $\mathbf{x} = \{x_n\}_{n=1}^N$ of real numbers partitioned into P equi-spaced range bins denoted as $\{p_i\}_{i=1}^P$, can be called uniformly distributed if the number of elements in each bin, denoted as $\mathcal{C}(p_i; \mathbf{x})$ follows:

$$\left| \mathcal{C}(p_i; \mathbf{x}) - \frac{N}{P} \right| \simeq 0, \quad i \in \{1, 2, \dots, P\}, \quad (5)$$

where the above expression is as small as possible, pertinent to cases where N is not perfectly divisible by P .

Surprisingly, defining a sequence based on its elements’ membership to the range bins will also provide an extra control to modify the distribution of the sequence itself at a later point.

3. THE PROPOSED METHOD

In the following, we describe in detail our proposed design approach for generating highly uncorrelated sequences with a uniform distribution.

3.1. Construction with Low Correlation

The CAN algorithm in [11] provides an efficient mathematical formalism confirming our intuitive observation that a sequence with small out-of-phase periodic correlation has a flat spectrum in the frequency domain. Following such observation, the periodic out-of-phase correlations of a sequence \mathbf{x} of length N can be optimized conveniently by following minimization problem:

$$\min_{\mathbf{x}, \mathbf{v}} \|\mathbf{A}^H \mathbf{x} - \mathbf{v}\|_2^2 \quad (6)$$

where \mathbf{A} denotes the $N \times N$ (inverse) DFT matrix, whose elements can be given as,

$$[A]_{k,l} = \frac{1}{\sqrt{N}} \exp \{j2\pi kl/N\}, \quad k, l = 1, \dots, N \quad (7)$$

Data: \mathbf{h}, N, P
Result: $\hat{\mathbf{h}} \triangleq \{\hat{h}_n\}_{n=1}^N$
initialize $n = 1$;
maxnum $\leftarrow \text{floor}(N/P)$;
while $n \leq N$ **do**
 bin_index $\leftarrow \text{ceil}(n/\text{maxnum})$;
 if $h_n < \text{lower_edge}(\text{bin_index})$ **then**
 $\hat{h}_n \leftarrow \text{lower_edge}(\text{bin_index})$;
 else if $h_n > \text{upper_edge}(\text{bin_index})$ **then**
 $\hat{h}_n \leftarrow \text{upper_edge}(\text{bin_index})$;
 else
 $\hat{h}_n \leftarrow h_n$;
 end
 $n \leftarrow n + 1$;
end

Algorithm 1: Optimal modification of \mathbf{h} according to the edge values, and their corresponding range bins.

Table 1. Algorithm for Construction of Uniformly Distributed Sequences with Low Auto-Correlation

Input parameters: sequence length N , number of range bins P .

- Step 0:** Initialize \mathbf{x} using a randomly generated sequence.
Step 1: Compute $\hat{\mathbf{v}}$ using (10).
Step 2: Compute $\hat{\mathbf{x}}$ using (12).
Step 3: Compute \mathbf{h} using (13) and preserve the index of each elements of original sequence $\hat{\mathbf{x}}$ in $\mathcal{I}_a(\hat{\mathbf{x}})$.
Step 4: Partition \mathbf{h} into P bins of equal length and compute the edges of each bin.
Step 5: Modify the elements of \mathbf{h} using Algorithm 1.
Step 6: Compute $\tilde{\mathbf{x}}$ by restoring the index order previously stored in step 3.
Step 7: Let $d = \|\tilde{\mathbf{A}}^H \tilde{\mathbf{x}} - \tilde{\mathbf{v}}\|_2$. Repeat steps 1–6 until $|d^{(s)} - d^{(s-1)}| \leq 10^{-6}$ where s denotes the iteration number.
-

where $j^2 = -1$ and \mathbf{v} is the representation of \mathbf{x} in Fourier domain. Here, \mathbf{x} is constrained to have given distribution, as in (5). Observe that, the aperiodic correlations of \mathbf{x} can be given in terms of periodic correlations of the sequence $\tilde{\mathbf{x}} = [\mathbf{x} \mathbf{0}_N]^T$.

Hence, in the aperiodic case, we can consider the following frequency-domain design problem to minimize the aperiodic out-of-phase correlations of \mathbf{x} :

$$\min_{\tilde{\mathbf{x}}, \tilde{\mathbf{v}}} \|\tilde{\mathbf{A}}^H \tilde{\mathbf{x}} - \tilde{\mathbf{v}}\|_2^2 \quad (8)$$

in which $\tilde{\mathbf{A}}$ denotes the $2N \times 2N$ (inverse) DFT matrix, and also \mathbf{x} is constrained to have uniform distribution as previously described.

Note that for a given $\tilde{\mathbf{x}}$, the minimization of (8) with respect to $\tilde{\mathbf{v}}$ is straightforward. Let

$$\mathbf{f} \triangleq \{f_n\}_{n=1}^{2N} = \tilde{\mathbf{A}}^H \tilde{\mathbf{x}} \quad (9)$$

denote the FFT of $\tilde{\mathbf{x}}$. The optimum $\tilde{\mathbf{v}}$, denoted as $\hat{\mathbf{v}}$ can be obtained as

$$\hat{\mathbf{v}} = \frac{1}{\sqrt{2}} [e^{j\psi_1} \dots e^{j\psi_{2N}}]^T \quad (10)$$

where $\psi_n = \arg\{f_n\}$, for $n = 1, \dots, 2N$. Similarly, for fixed $\tilde{\mathbf{v}}$, we denote IFFT of $\tilde{\mathbf{v}}$ as, $\mathbf{g} \triangleq \{g_n\}_{n=1}^{2N} = \tilde{\mathbf{A}} \tilde{\mathbf{v}}$.

Observe that,

$$\|\tilde{\mathbf{A}}^H \tilde{\mathbf{x}} - \tilde{\mathbf{v}}\|_2^2 = \|\tilde{\mathbf{x}} - \tilde{\mathbf{A}} \tilde{\mathbf{v}}\|_2^2. \quad (11)$$

It follows from the above that the minimizer \mathbf{x} of (8), denoted as $\hat{\mathbf{x}} \triangleq \{\hat{x}_n\}_{n=1}^N$ is simply given by

$$\hat{x}_n = g_n, \quad n = 1, \dots, N. \quad (12)$$

3.2. Construction with Uniform Distribution

In this section, we extend our previous formulations to enforce a uniform distribution for the sequences obtained from the above framework, and particularly (11). This goal will be accomplished by finding the minimizer \mathbf{x} of (8) that has a uniform distribution. In other words, the *nearest-vector* problem in (11) is to be solved by a projection on the set of sequences with a uniform distribution. Therefore, we will revisit the concept of histogram equalization.

It is not difficult to verify that the aforementioned projection can be computed by segmenting the sequence entries into smaller range bins, and aiming to achieve a uniform distribution inside these smaller bins so that by controlling the smaller range bins, a sequence with a given distribution can be generated. The algorithm is as follows: we first sort the sequence entries in an ascending order to form

$$\mathbf{h} = \mathcal{S}_a(\hat{\mathbf{x}}) \quad (13)$$

where $\mathcal{S}_a(\cdot)$ denotes a sorting operation (in ascending order) on the vector argument. Along with sorting, we also preserve the index of each element in the original sequence $\hat{\mathbf{x}}$ in an index array $\mathcal{I}_a(\hat{\mathbf{x}})$.

Next, we partition \mathbf{h} into P equi-sized range bins, where $1 \leq P \leq N$. Note that, $P = 1$ resembles the whole sequence in one single bin, whereas $P = N$ suggests each bin p_i contains only one element \hat{x}_i . Naturally, both situations are not desirable for optimum distribution induction. Therefore, we choose P such that the sequence length N is divisible (or closely divisible, if N is prime) by P and also large enough to achieve a smooth transition between two consecutive bins. Finally, we compare the values of each element with the corresponding bin's edge values and modify those elements according to Algorithm 1. Once we have modified $\hat{\mathbf{h}}$, we are

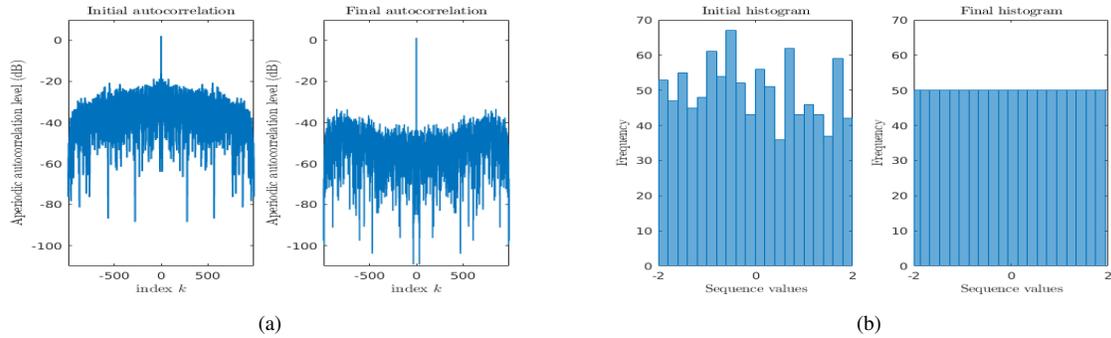


Fig. 1. The initial and final (a) normalized aperiodic auto-correlation, and (b) histogram of constructed sequence of length $N = 10^3$ and $P = 20$ using the proposed algorithm.

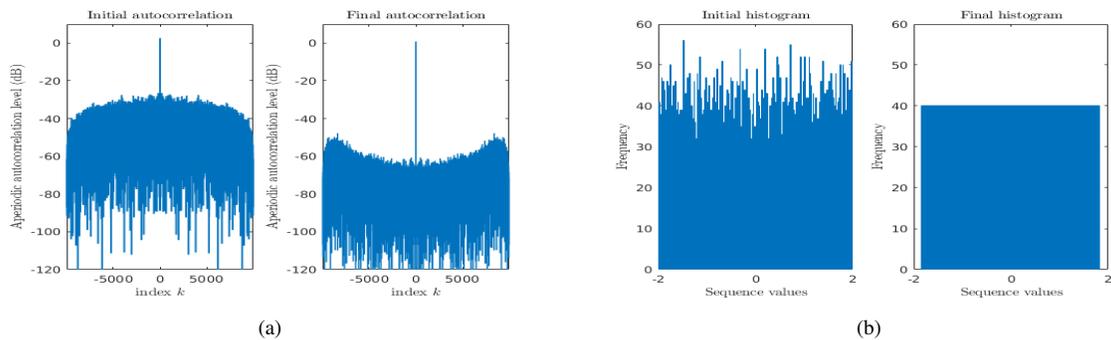


Fig. 2. The initial and final (a) normalized aperiodic auto-correlation, and (b) histogram of constructed sequence of length $N = 10^4$ and $P = 250$ using the proposed algorithm.

easily back to our optimal sequence \bar{x} by employing the index array of $\mathcal{I}_a(\hat{x})$. Note that the optimality of projection can be shown easily but is omitted herein for the sake of brevity.

The iterative algorithm for construction of the desired sequences with good correlation and good distribution properties is summarized in Table 1.

4. NUMERICAL RESULTS

We provide several numerical examples to investigate the performance of the proposed method. We use the proposed approach to design uniformly distributed sequences of length (i) $N = 10^3$ with number of partitions $P = 20$, and (ii) $N = 10^4$ with partition $P = 250$ using `rand` function in MATLAB. Although `rand` provides fairly uniformly distributed random sequences but is not at all uniform inside a range bin. For both cases the initial and final normalized aperiodic auto-correlation level (NAPC) = $20 \log_{10} |r_k/r_0|$ in dB.

are presented in Fig. 1(a) and 2(a). We also present the initial and final histogram in Fig. 1(b) and 2(b). A significant improvement in terms of both auto-correlation and distribution can be observed in both cases. It should also be noted that the generation of sequences using the proposed method was relatively fast in terms of computa-

tional time. Particularly, it took 1.56 secs and 57.39 secs on a standard PC to accomplish the sequence design for the first and second cases described above, respectively.

5. CONCLUSIONS AND FUTURE WORK

Signals with both good auto-correlation and good distribution properties are required in eye tracking for Parkinson's Disease diagnosis and treatment. We have presented a new framework to design such signals based on the CAN computational framework. The proposed method is computationally efficient and can design very long sequences (of lengths up to $N \sim 10^6$ and even more) in relatively short time frames. The designed sequences using the proposed algorithm show significant enhancement in terms of out-of-phase auto-correlation as well as distribution properties. While the numerical examples showed promising results, as a future research direction, it would be of great interest to study the behavior of ISL or other correlation related metrics and particularly their relationship with given marginal distributions.

6. REFERENCES

- [1] H. He, P. Stoica, and J. Li, "Designing unimodular sequence sets with good correlations -including an application to MIMO radar," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4391–4405, Nov 2009.
- [2] T. Soderstrom and P. Stoica, *System identification*, Prentice Hall, Inc., New York, 1989.
- [3] H. He, J. Li, and P. Stoica, *Waveform Design for Active Sensing Systems: A Computational Approach*, Cambridge University Press, 2012.
- [4] F. Gini, A. De Maio, and L. Patton, *Waveform Design and Diversity for Advanced Radar Systems*, Electromagnetics and Radar Series. Institution of Engineering and Technology, 2012.
- [5] E. J. Hughes, *Radar Waveform Optimisation as a Many-Objective Application Benchmark*, pp. 700–714, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [6] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, Jan 2006.
- [7] N. Levanon and E. Mozeson, *Radar Signals*, Wiley - IEEE. Wiley, 2004.
- [8] M. Soltanalian and P. Stoica, "Computational design of sequences with good correlation properties," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2180–2193, May 2012.
- [9] J. Jedwab and K. Yoshida, "The peak sidelobe level of families of binary sequences," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2247–2254, May 2006.
- [10] P. Stoica, H. He, and J. Li, "On designing sequences with impulse-like periodic correlation," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 703–706, Aug 2009.
- [11] P. Stoica, H. He, and J. Li, "New algorithms for designing unimodular sequences with good correlation properties," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1415–1425, Apr 2009.
- [12] D. Jansson, *Mathematical modeling of the human smooth pursuit system*, Ph.D. thesis, Uppsala University, Division of Systems and Control, Automatic control, 2014.
- [13] L. Itti, "New eye-tracking techniques may revolutionize mental health screening," *Neuron*, vol. 88, no. 3, pp. 442 – 444, 2015.
- [14] S. Marino, E. Sessa, G. Di Lorenzo, P. Lanzafame, G. Scullica, A. Bramanti, F. La Rosa, G. Iannizzotto, P. Bramanti, and P. Di Bella, "Quantitative analysis of pursuit ocular movements in Parkinson's Disease by using a video-based eye tracking system," *European Neurology*, vol. 58, no. 4, pp. 193–197, Nov 2007.
- [15] T. Nakamura, R. Kanayama, R. Sano, M. Ohki, Y. Kimura, M. Aoyagi, and Y. Koike, "Quantitative analysis of ocular movements in Parkinson's Disease," *Acta Oto-Laryngologica*, vol. 111, no. sup481, pp. 559–562, 1991.
- [16] O. B. White, J. A. Saint-Cyr, R. D. Tomlinson, and J. A. Sharpe, "Ocular motor deficits in parkinson's disease: Ii. control of the saccadic and smooth pursuit systems," *Brain*, vol. 106, no. 3, pp. 571, 1983.
- [17] P. Stoica and R.L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.
- [18] M. Soltanalian and P. Stoica, "Design of perfect phase-quantized sequences with low peak-to-average-power ratio," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 2576–2580.
- [19] M. Soltanalian, P. Stoica, M. M. Naghsh, and A. De Maio, "Design of piecewise linear polyphase sequences with good correlation properties," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, Sep 2014, pp. 1297–1301.
- [20] H. Niederreiter L. Kuipers, *Uniform distribution of Sequences, (Pure and Applied Mathematics)*, John Wiley and Sons Inc., 1974.