# MINIMUM NUMBER OF POSSIBLY NON-CONTIGUOUS SAMPLES TO DISTINGUISH TWO PERIODS

Srikanth V. Tenneti and P. P. Vaidyanathan

California Institute of Technology Department of Electrical Engineering Pasadena, California, USA 91125 Email: stenneti@caltech.edu; ppvnath@systems.caltech.edu

# ABSTRACT

Given that a sequence x(n) is periodic with period P belonging to a known integer set  $\{P_1, P_2, \ldots, P_L\}$ , what is the minimum number of samples of x(n) required to find the period? For the special case where the samples of x(n) are constrained to be contiguous in time, this problem has recently been solved. More generally, when the samples are allowed to be non-contiguous, the problem is quite difficult. This paper provides the answer for the restricted situation where  $P \in \{P_1, P_2\}$ . With  $P_1 < P_2$ , the necessary and sufficient number of (possibly noncontiguous) samples for period estimation turns out to be (a)  $P_1$ , if  $P_1$  is not a divisor of  $P_2$ , and (b)  $P_2$  otherwise. While the proof is quite involved even in this restricted case, it is likely to form the basis for addressing the more general situation where  $P \in \{P_1, P_2, \ldots, P_L\}$ .

*Index Terms*— Period Estimation, Minimum Samples, Ramanujan Sums, Nested Periodic Matrices.

# 1. INTRODUCTION

Periodic signals arise in many applications of science and engineering [1, 2, 4, 5], and there are several techniques to identify the period based on Fourier and non-Fourier methods [12, 16, 9, 17, 6, 14]. In discrete time, we say that x(n) has period P if x(n) = x(n + P) for all n, and P is the smallest positive integer with this property. Most of the known techniques obtain an estimate of P from a larger number of samples, and one basic question has largely escaped attention in the past. Namely, what is the minimum number of samples of x(n) necessary to identify P, if P is known to belong to some integer set? Very recently this question has been addressed for the case where the samples used are required to be contiguous [13]. The result is as follows:

**Theorem 1.** Let x(n) be a periodic signal, whose period is known to lie in the integer set  $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$ . If one were to estimate the period using L consecutive samples, then, it is both necessary and sufficient for L to satisfy:

$$L \ge L_{min} = \max_{P_i, P_j \in \mathbb{P}} P_i + P_j - \gcd(P_i, P_j)$$
(1)

For the special case where  $P \in \{1, 2, \dots, P_{max}\}$ , the above number of samples becomes  $2P_{max} - 2$ .

Next, if the samples are allowed to be non-contiguous, can the period be estimated using a smaller number of samples? If so, what exactly is that number? In this paper we will address this question for the simple case where  $P \in \{P_1, P_2\}$ . While this is admittedly a very restricted special case, we consider it for two reasons. Firstly, even in this case the details are quite involved, and secondly, this might be the basis for future generalizations to more useful practical situations. Our main result in this paper is the following:

**Theorem 2.** Given a periodic signal x(n) whose period P lies in the set  $\mathbb{P} = \{P_1, P_2\}$ , where  $P_1 < P_2$ , the following number of samples is necessary and sufficient to identify P:

$$M_{min} = \begin{cases} P_2 & \text{if } P_1 \text{ divides } P_2 \\ P_1 & \text{otherwise} \end{cases}$$
(2)

Notice that when  $P_1$  divides  $P_2$ ,  $L_{min}$  (Eq. (1)) and  $M_{min}$  (Eq. (2)) are equal. So even if the samples are allowed to be non-contiguous, the period cannot be estimated using a smaller number of samples than  $L_{min}$  in this case. However, when  $P_1$  does not divide  $P_2$ ,  $M_{min}$  can be significantly smaller than  $L_{min}$ . For example, let us consider the case where  $P_1 = 8$  and  $P_2 = 50$ . For the contiguous samples case, the minimum number of samples is 56, and a set of such sample locations are:

$$\{0, 1, 2, \dots, 54, 55\}\tag{3}$$

For the non-contiguous case, the minimum number of samples required is only 8, and an example of the sample locations that work are:

$$\{0, 1, 50, 51, 100, 101, 150, 151\}$$
(4)

This work was supported by the ONR grant N00014-15-1-2118, and the California Institute of Technology.

Notice how the non-uniform sample locations occur in uniformly spaced bunches.

*Outline:* The sufficiency and necessity parts of Theorem 2 are proved in Sec. 2 and Sec. 3 respectively. In the presence of noise, having more samples than  $M_{min}$  improves the accuracy of period estimation. This is demonstrated in Sec. 4.

Notation:

- 1. D|P denotes that D is a divisor of P.  $D \nmid P$  denotes that D is not a divisor of P.
- 2. The greatest common divisor (gcd) of two numbers P and Q is denoted by (P, Q).
- 3.  $P \mod Q$  returns the remainder of dividing P by Q.
- 4.  $\phi(D)$ , the Euler-totient function of D, is equal to the number of positive integers smaller than and co-prime to D.

# 2. PROVING SUFFICIENCY

Let us start with the following lemma:

**Lemma 1.** Let x(n) be periodic with period P. Let y(n) be the M-fold decimated version of x(n), that is,

$$y(n) = x(Mn), \ \forall \ n \in \mathbb{Z}$$
(5)

Then, y(n) is also periodic, its period being a divisor of  $\frac{P}{(M,P)}$ .

Proof.

$$y\left(n + \frac{P}{(M,P)}\right) = x\left(Mn + M\frac{P}{(M,P)}\right)$$
$$= x(Mn)$$
$$= y(n)$$

So  $\frac{P}{(M,P)}$  is a repetition index of y(n). A repetition index of a signal must always be a multiple of its period (see Lemma 3 in [15]). This completes the proof.

When M is coprime to P in the above result, we can say something more:

**Lemma 2.** Let x(n) be periodic with period  $P \neq 1$ . Let M be an integer coprime to P, and let y(n) = x(Mn). Then, y(n) is periodic with period  $\neq 1$ .

*Proof.* From Lemma 1, it follows that y(n) must be periodic with period being a divisor of P. So y(n) can have period 1 if and only if all the entries in the following set are equal:

$$\mathbb{Y} = \{y(0), y(1), \dots, y(P-1)\}$$
(6)

Using (5), we can re-write  $\mathbb{Y}$  as:

$$\mathbb{Y} = \{x(0), x(M), \dots, x(MP - M)\}$$
(7)

Since x(n) is periodic with period P, we can re-write this further as:

$$\mathbb{Y} = \{x(0 \bmod P), x(M \bmod P), \dots, y(MP - M \bmod P)\}$$

It is a well known result in Number Theory that when (M, P) = 1, the following two sets are just permuted versions of each other:

$$\{(0 \bmod P), (M \bmod P), \dots, (MP - M \bmod P)\} = \{0, 1, \dots, P - 1\}$$

Hence the set  $\mathbb{Y}$  is in fact (a permuted version of) the following set:

$$\{x(0), x(1), \dots, x(P-1)\}$$
(8)

Since x(n) has period  $P \neq 1$ , all the elements in the above set cannot be equal to each other, due to which the period of y(n) cannot be 1.

We will use the above two lemmas to prove the sufficiency part of Theorem 2. We will do this by considering two distinct cases as follows:

## **2.1.** Sufficiency when $P_1|P_2$

In this case, we can directly use Theorem 1. When  $P_1|P_2$ ,  $P_1 + P_2 - (P_1, P_2) = P_2$ . Hence, using Theorem 1, we can conclude that  $P_2$  samples are sufficient to estimate the period.

#### **2.2.** Sufficiency when $P_1 \nmid P_2$

When  $P_1 \nmid P_2$ , we have to prove that  $P_1$  samples are sufficient to find the period. We will start with the case when  $P_1$  and  $P_2$  are coprime, and then extend the proof to the more general case.

Let  $y(n) = x(P_2n)$ . If x(n) had period  $P_2$ , then the period of y(n) is clearly 1. However, if x(n) had period  $P_1$ , and if  $(P_1, P_2) = 1$ , then y(n) cannot have period 1 (using Lemma 2). To check whether y(n) has period 1, we need at most  $P_1$  samples of y(n), since its period can only be a divisor of  $P_1$ . This proves that  $P_1$  samples are sufficient to find the period when  $(P_1, P_2) = 1$ .

Let us now consider the more general case when  $(P_1, P_2) = G$ , where G is not necessarily 1. We propose to use the setup shown in Fig. 1. If x(n) has period  $P_2$ , then clearly, all the outputs  $y_0(n), y_1(n), \ldots, y_{G-1}(n)$  will have period 1. However, if x(n) has period  $P_1$ , then at least one of those outputs will have period > 1.

To prove this, we re-draw Fig. 1 as shown in Fig. 2. If x(n) had period  $P_1$ , then at least one of  $u_0(n), u_1(n), \ldots$ 



**Fig. 1**. Finding the period of x(n) when  $(P_1, P_2) = G \neq P_1$ . See text for details.

 $u_{G-1}(n)$  must have period > 1. This is because, if all of them have period 1, then x(n) must satisfy:

$$x(n+G) = x(n) \ \forall \ n \in \mathbb{Z}$$
(9)

That is, G must be a repetition index, which then necessitates that  $P_1|G$  (since the period of a signal must always divide any repetition index). For  $P_1$  to divide its gcd with  $P_2$ ,  $P_1$  must divide  $P_2$ , which contradicts our assumption that  $P_1 \nmid P_2$ . This shows that at least one of  $u_0(n), u_1(n), \ldots, u_{G-1}(n)$ must have period > 1.

Let us assume that  $u_i(n)$  has period > 1. Because of Lemma 1, the period of  $u_i(n)$  must be a divisor of  $\frac{P_1}{G}$ . Further, since  $\frac{P_1}{G}$  and  $\frac{P_2}{G}$  are always co-prime, any divisor of  $\frac{P_1}{G}$ is also coprime to  $\frac{P_2}{G}$ . So using Lemma 2, we can conclude that the period of  $y_i(n)$  must be > 1.

So by checking whether or not all of  $y_0(n), y_1(n), \ldots$ ,  $y_{G-1}(n)$  have period 1, we can estimate the period of x(n). How many samples of each  $y_i(n)$  do we need to establish its period? Because of Lemma 1,  $u_0(n), u_1(n), \ldots, u_{G-1}(n)$ can have their periods as any divisors of  $\frac{P_1}{G}$ , and so the outputs  $y_0(n), y_1(n), \ldots, y_{G-1}(n)$  can have their periods as any divisors of  $\frac{P_1}{G}$ . So  $\frac{P_1}{G}$  samples of each output are sufficient to check if they have period 1. Since there are G such outputs, it follows that  $\frac{P_1}{G} \times G = P_1$  samples of x(n) are sufficient to determine whether the period of x(n) is  $P_1$  or  $P_2$ . This completes the proof of the sufficiency part of Theorem 2.

Notice that the  $P_1$  samples of x(n) required in the above technique occur in uniformly spaced bunches as shown in Eq. 4. There are  $P_1/G$  bunches, spaced  $P_2$  samples apart, and with G contiguous samples within each bunch.

#### **3. PROVING NECESSITY**

We will first show that at least  $P_1$  samples are necessary to find the period, irrespective of whether  $P_1|P_2$  or  $P_1 \nmid P_2$ .



Fig. 2. Re-drawing Fig. 1 for analysis. See text for details.

Later, we will show that when  $P_1|P_2$ ,  $P_2$  samples are necessary.

**Theorem 3.** Let  $P_1$  and  $P_2$  be positive integers such that  $P_1 < P_2$ . Then, given any set of  $L < P_1$  integers  $\mathbb{N}_T = \{n_1, n_2, \ldots, n_L\}$ , there exist periodic signals  $x_{P_1}(n)$  and  $x_{P_2}(n)$  with periods  $P_1$  and  $P_2$  respectively such that

$$x_{P_1}(n) = x_{P_2}(n) \ \forall \ n \in \mathbb{N}_T \tag{10}$$

*Proof.* Since  $L < P_1$ , there exists at least one integer in the set  $\{0, 1, \ldots, P_1 - 1\}$  that does not belong to the set  $\{(n_1 \mod P_1), (n_2 \mod P_1), \ldots, (n_L \mod P_1)\}$ . Let *m* be such an integer. We define  $x_{P_1}(n)$  as follows:

$$x_{P_1}(n) = \begin{cases} 0 & \text{if } n \mod P_1 = m\\ 1 & \text{otherwise} \end{cases}$$
(11)

It is easy to see that  $x_{P_1}(n)$  has period  $P_1$ . Notice that  $x_{P_1}(n) = 1 \forall n \in \mathbb{N}_T$ . In the same way, we can construct a period  $P_2$  signal  $x_{P_2}(n)$  that satisfies  $x_{P_2}(n) = 1 \forall n \in \mathbb{N}_T$ . Clearly, for these  $x_{P_1}(n)$  and  $x_{P_2}(n)$ , (10) is satisfied. This completes the proof.

We will now prove that when  $P_1|P_2$ , one needs at least  $P_2$  samples to estimate the period.

**Theorem 4.** Let  $P_1|P_2$ . Then, given any set of  $L < P_2$ integers  $\mathbb{N}_T = \{n_1, n_2, \dots, n_L\}$ , and any period  $P_1$  signal  $x_{P_1}(n)$ , there exists a period  $P_2$  signal  $x_{P_2}(n)$  such that

$$x_{P_1}(n) = x_{P_2}(n) \ \forall \ n \in \mathbb{N}_T \tag{12}$$

*Proof.* Let  $x_{P_2}(n)$  be defined to be equal to  $x_{P_1}(n)$  for all  $n \in \mathbb{N}_T$ . Is it possible that doing so violates the following condition:

$$x_{P_2}(n+P_2) = x_{P_2}(n) \tag{13}$$



**Fig. 3**. Error rate vs. number of samples.  $P_1 = 9$ ,  $P_2 = 13$ , SNR = 0dB. See Sec. 4 for details.

for some  $n, n + P_2 \in \mathbb{N}_T$ ? Luckily no, since  $n + P_2$  can be written as  $n + kP_1$  for some integer k, and so  $x_{P_1}(n)$  will satisfy:

$$x_{P_1}(n+P_2) = x_{P_1}(n) \tag{14}$$

Further, since  $L < P_2$ , there exists at least one integer in the set  $\{0, 1, \ldots, P_2 - 1\}$  that does not belong to the set  $\{(n_1 \mod P_2), (n_2 \mod P_2), \ldots, (n_L \mod P_2)\}$ . Let *m* be such an integer. Moreover, let *u* and *v* be integers such that  $u > \max_n x_{P_1}(n)$  and  $v < \max_n x_{P_1}(n)$ . We define  $x_{P_2}(n)$ as follows:

$$x_{P_2}(n) = \begin{cases} x_{P_1}(n) & \text{if } n \mod P_2 \in \mathbb{N}_T \\ u & n \mod P_2 = m \\ v & \text{otherwise} \end{cases}$$
(15)

It is easy to see that  $x_{P_2}(n)$  has period  $P_2$  (*u* occurs only once every  $P_2$  samples). Hence, we have constructed a period  $P_2$ signal  $x_{P_2}(n)$  satisfying the conditions of the theorem.  $\Box$ 

This completes the proof of Theorem 2.

## 4. SIMULATIONS UNDER NOISE

We can easily adapt the period estimation techniques of Sec. 2 to deal with noisy inputs. Recall that the period estimation techniques in Sec. 2.2 involve, apart from downsampling, checking whether certain signals have period 1. When there is noise, we can compute the sample variance of a signal to check whether its period is 1. If the variance is less than a suitably chosen threshold, we may hypothesize that the period is 1.

In the following experiments, we chose  $P_1 = 9$  and  $P_2 = 13$ . Let x(n) denote the input, and  $y(n) = x(P_2n)$ . If x(n) had period  $P_2$ , then the period of y(n) is 1. However, if x(n) had period  $P_1$ , then y(n) cannot have period 1, and is in fact a permutation of the input itself (Lemma 2). Let us now assume that x(n) was contaminated by an independent AWGN process s(n) with sample variance  $\sigma_n^2$ . Further, suppose that x(n) was itself a randomly generated periodic signal, that is,



**Fig. 4.** Error rate vs. SNR for the minimum samples case.  $P_1 = 9$ ,  $P_2 = 13$ , number of samples  $= P_1 = 9$ . See Sec. 4 for details.

with sample values in one period randomly generated with sample variance  $\sigma_x^2$ .

In this case, y(n) would have sample variance  $\sigma_y^2$  given by:

$$\sigma_y^2 = \begin{cases} \sigma_x^2 + \sigma_n^2 & \text{if } x(n) \text{ had period } P_1 \\ \sigma_n^2 & \text{if } x(n) \text{ had period } P_2 \end{cases}$$
(16)

So we may choose a threshold parameter  $T = \sigma_n^2 + \frac{\sigma_x^2}{2}$ , and predict the input period as  $P_1$  if the observed variance of y(n) is larger than T, and as  $P_2$  otherwise.

Using this technique, the following two experiments were performed. First, we study the accuracy of the period estimate as a function of the number of samples of y(n) used for computing  $\sigma_y^2$ . The SNR was 0dB. The minimum number of samples needed in this case, as given by Theorem 2, is 9. For each value of 'number of samples', we generated 50000 signals with periods randomly chosen from the set  $\{P_1, P_2\}$ . In Fig. 3, the fraction of times the period was incorrectly estimated is plotted as the error rate. It is intuitive that the error rate decreases as we have more samples. In our second experiment, we fixed the number of samples to be  $P_1$ , and plotted the error rate for various values of SNR (Fig. 4). Once again, as is consistent with intuition, we observed that the error rate decreases as the SNR increases.

#### 5. CONCLUSION AND FUTURE WORK

This paper derived the minimum necessary and sufficient number of samples to identify the period of a signal from a set of size 2. This is a simple step in the generalization of the recent results of [13], to non-contiguous samples. Further, it was shown through simulations that in the presence of noise, having more samples than the minimum number improves the accuracy of period estimation.

It will be interesting to extend these results to the case when the period of x(n) is known a priori to belong to an arbitrarily sized set  $\mathbb{P} = \{P_1, P_2, \dots, P_L\}$ . We are also working towards extending these results to the case when the input is a mixture of multiple periodic signals, and to the case of multidimensional signals.

## 6. REFERENCES

- M. A. Andrade, C. P.-Iratxeta and C. P. Ponting, "Protein repeats: Structures, functions, and evolution", J. Structur. Biol., vol. 134, pp.117 -131 2001.
- [2] A. Grover, A. Veenu and P. C. Sharma, "Searching Microsatellites in DNA Sequences: Approaches Used and Tools Developed.", Physiology and molecular biology of plants: an international journal of functional plant biology, 18.1 (2012): 1119. PMC. Web. 28 July 2015.
- [3] G. H. Hardy and E. M. Wright, "An introduction to the theory of numbers", Oxford University Press, Inc., 2008.
- [4] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott and R. A. Collins, "Observation of a rapidly pulsating radio source", Nature 217, 709 - 713, Feb 1968.
- [5] L. T. Mainardi, M. Bertinelli, and R. Sassi, Analysis of T-wave alternans using the Ramanujan Transform", Computers in Cardiology, vol. 35 pp. 605-608, 2008.
- [6] D. D. Muresan and T. W. Parks, "Orthogonal, Exactly Periodic Subspace Decomposition", IEEE Transactions on Signal Processing, Vol. 51, no. 9, September 2003.
- [7] M. Nakashizuka, "A Sparse Decomposition for Periodic Signal Mixtures", Proc. of the 15th Intl. Conf. on Digital Sig. Proc., 2007.
- [8] B. Santhanam and P. Maragos, "Harmonic analysis and restoration of separation methods for periodic signal mixtures: Algebraic separation vs comb filtering," Signal Process., vol. 69, no. 1, pp. 8191, 1998.
- [9] W. A. Sethares and T. W. Staley, "Periodicity transforms," IEEE Trans. on Sig. Proc., vol. 47, pp. 2953 -2964, Nov. 1999.
- [10] S. V. Tenneti and P. P. Vaidyanathan, "Ramanujan Filter Banks for Estimation and Tracking of Periodicities", Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc., Brisbane, Australia, 2015.
- [11] S. V. Tenneti and P. P. Vaidyanathan, "Nested Periodic Matrices and Dictionaries: New Signal Representations for Period Estimation," IEEE Transactions on Signal Processing, vol. 63, no. 14, pp. 3736-3750, July, 2015.
- [12] S. V. Tenneti and P. P. Vaidyanathan, "A Unified Theory of Union of Subspaces Representations for Period Estimation," IEEE Trans. on Sig. Proc., vol. 64, no. 20, pp. 5217-31, Oct, 2016.
- [13] S. V. Tenneti and P. P. Vaidyanathan, "Critical Datalength for Period Estimation", Proc. IEEE Int. Symp. on Circuits and Systems, Canada, 2016.

- [14] P.P. Vaidyanathan and S.V. Tenneti, "Properties of Ramanujan Filter Banks", Proc. European Signal Processing Conference, France, August 2015.
- [15] P. P. Vaidyanathan, "Ramanujan sums in the context of signal processing: Part I: fundamentals" IEEE Trans. on Signal Processing, Vol. 62, No. 16, pp. 4145 - 4157, August 2014.
- [16] P. P. Vaidyanathan, "Ramanujan sums in the context of signal processing: Part II: FIR representations and applications" IEEE Trans. on Signal Processing, Vol. 62, No. 16, pp. 4158 - 4172, Augut 2014.
- [17] J. D. Wise, J. Caprio and T. W. Parks, "Maximum Likelihood Pitch Estimation" IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. assp-24, no. 5, October 1976.
- [18] M. Zou, Z. Chai and R. Unbehauen, "Separation of periodic signals by using an algebraic method," IEEE International Symposium on Circuits and Systems, vol.5, pp.2427-2430, June 1991.