ESTIMATION OF PSEUDO-PERIODIC SIGNALS

Hoi Wong and William A. Sethares

Department of Electrical and Computer Engineering
University of Wisconsin-Madison

ABSTRACT

Many phenomena exhibit cyclical patterns without being periodic. This can be modeled using "pseudo-periodic" functions which enable cyclical patterns to be decomposed into a periodic signal along with a set of parameters that define the deviations of the pattern from true periodicity. This paper describes algorithms that estimate the parameters and the template function using the ρ-norm cost function. Implementation issues such as varying amplitude between pseudo-periods and problems using finite data records are also discussed. Applications to sunspot data explore practical uses and limitations of the method.

1. INTRODUCTION

A pseudo-periodic function $R(t)$ may be defined by

$$R(t) \equiv \sum_i m_i s(\alpha_i t + \beta_i) \qquad (1)$$

where $s(t)$ is a template function with compact support on $[0, 2\pi)$, α_i scales frequency, β_i translates in time, and m_i adjusts the amplitude. Earlier work [6] shows that if $m_i = 1$ and if $S(t)$ is the periodic extension of $s(t)$, then it is possible to achieve unbiased estimation of the α_i and β_i using the ρ-norm $|| \cdot ||_\rho$. This is defined to be the norm induced by the inner product [7]

$$\langle s, r \rangle_\rho \equiv \lim_{k \to \infty} \frac{1}{2k} \int_{-k}^{k} s(t) r(t) \, dt \qquad (2)$$

where $s(t)$ and $r(t)$ are periodic functions of periods T_s and T_r (which need not be the same). This paper tackles four issues that arise in the estimation of pseudo-periodic functions: (1) incorporation of nonunity amplitude weightings, (2) algorithms (LρNE and INRT) for estimating the unknown parameters, (3) a method to find a good template function $s(t)$ if it is unknown. Finally, (4) since any use of the method is necessarily in discrete time and data driven, some practical implementation issues are considered.

The mathematical setup is reminiscent of wavelet transforms: the template function plays a role analogous to that of the mother wavelet, while the frequency scaling parameter is analogous to the scale factor. However, wavelet scale factors are often constrained to specific values which insure that the wavelet basis functions are orthogonal, while the frequency scaling parameters assume arbitrary values and so the template functions need not be orthogonal. Hence, template functions do not form a basis, rather, they form a *frame* [1], a more-than-complete spanning set. A more fundamental difference is that the goal of the pseudo-periodic analysis is to directly provide information such as the time and frequency of repetitions of the template within a waveform.

2. SIGNAL VS. FUNCTION

The definitions (1) and (2) apply to functions that are defined on the whole real line. To apply these ideas to (finite) data records, suppose that the template function $s(t)$ is defined by w_s points in $[0, 2\pi)$. Then the ρ-inner product can be calculated as the ℓ_2-inner product normalized by the size of the support of the template. Thus

$$\langle \mathbf{s}_\alpha, \mathbf{r} \rangle_\rho \equiv \frac{\langle \mathbf{s}_\alpha, \mathbf{r} \rangle}{N_\alpha} \equiv \frac{1}{N_\alpha} \sum_{t=1}^{N_\alpha} s(\alpha t) r(t) \qquad (3)$$

where $N_\alpha = \omega_s / \alpha$ denotes the number of samples in the support. The goal of the estimation is to find the parameters m_i, α_i, and β_i so that the template best matches the ith pseudo-period of the data $R(t)$. It is possible to either hold $R(t)$ unchanged and compare it with $s(\alpha t + \beta)$, or to hold $s(t)$ unchanged and compare it with $R(\alpha't + \beta')$. Theorem 3.2 of [6] shows that these two operations are identical over $\Re$, so it is a matter of computational convenience which approach to use. In practice it is generally easier to stretch and compress the template.

Since both the template $s(t)$ and the data $R(t)$ are represented by a finite collection of points, the domain over which the sum (3) is carried out may be too short (when $\alpha > 1$) or too long (when $\alpha < 1$). This can be handled in two ways: (a) wrap the data in the template when it is too short, or (b) zero pad when it is too short. Both are roughly equivalent, though there will be slight discrepancies in the behavior of the resulting algorithms, especially in terms of

stopping conditions. For the wrap-shift method, $S(t)$ is the 2π-periodic extension of $s(t)$, and $R(t)$ is compared with $s(\alpha t + \beta)$. This may be most useful for signals (such as those from musical analyses [8]) where the phase of the local period varies. The weakness of this setting is that discontinuity at junctions between local pseudo-periods may complicate the interpolation. The pad zero method[1] assumes the deadband is filled with zeros, that is $\{S(t) = 0\}|_{t \in (-\infty, 0) \cup [2\pi, \infty)}$, $\{S(t) = s(t)\}|_{t \in [0, 2\pi)}$. This may be more useful for signals (such as heartbeat or sunspot data) where there are silences (deadband) between pseudo-periods, that is, where the template may be assumed to be approximately zero at either end of its domain.

3. STEEPEST DESCENT - LEAST ρ-NORM ERROR

One way to estimate the unknown parameters is with a gradient method. Two different cost functions were suggested in [6]: minimize the ρ-norm of error (LρNE), or maximize the inner product (MaxCorr). LρNE is implemented with the gradient method in this section. MaxCorr is discussed in section 5. If the template contains inadequate information (for example, is too short), the estimation is not meaningful (A degenerate case would be a template with only a single point.) It is the user's responsibility to choose a good template, and this choice is discussed in section 6.

3.1. Estimation Procedure (Approximate Gradient)

At each local period, $s(\alpha t)$ is known. The algorithm extracts $r(t)$ from $R(t)$ (with an appropriate translation) and then estimates the three parameters m_i, α_i and β_i. In implementation, it is easier to use the parameter δ_i in place of β_i where the two are related by

$$\beta_i \equiv \sum_{j < i} \frac{2\pi}{\alpha_j} + \delta_i. \tag{4}$$

With appropriate $s(\alpha t)$ and $r(t)$, the cost function at the ith pseudo-period can be computed:

$$J(\alpha, \delta, m) = \|\mathbf{s}(\alpha t + \beta) - m\mathbf{r}(t)\|_\rho^2.$$

The parameters are updated by:

$$\alpha_{k+1} = \alpha_k - \mu_\alpha \frac{dJ}{d\alpha} \approx \alpha_k - \mu_\alpha \frac{J(\alpha_k) - J(\alpha_k + \Delta\alpha)}{\Delta\alpha}$$

$$\delta_{k+1} = \delta_k - \mu_\delta \frac{dJ}{d\delta} \approx \delta_k - \mu_\delta \frac{J(\delta_k) - J(\delta_k + \Delta\delta)}{\Delta\delta} \quad (5)$$

$$m_{k+1} = m_k - \mu_m \frac{dJ}{dm} \approx m_k - \mu_m \frac{J(m_k) - J(m_k + \Delta m)}{\Delta m}$$

---

[1] The estimation procedure need not, in reality, fill a vector with zero values. Instead, shifting can be performed by sliding the window that extracts $r(t)$ from $R(t)$ according to β.

The algorithm will stop (and consider the current parameters satisfactory) when all three derivatives change sign and:

$$\left|\frac{dJ}{dm}\right| < \epsilon_m \qquad \left|\frac{dJ}{d\alpha}\right| < \epsilon_\alpha \qquad \left|\frac{dJ}{d\delta}\right| < \epsilon_d$$

where ϵ_m, ϵ_α and ϵ_δ are user specified values. Because the cost function is nonlinear, it should be expected that the cost surface might have various local minima. Since multiple parameters are adapting simultaneously, their rate of convergence (related to step-sizes) must be carefully set. Good sets of μ_α, μ_δ and μ_m were found by trial and error.

4. AMPLITUDE MISMATCH

In addition to translation and frequency scaling, real-life pseudo-periodic signals usually have amplitude mismatch. In contrast, the theoretical development in [6] assumed $m_i = 1, \forall i$. This section shows that, depending on the cost function chosen, the amplitude mismatch might or might not bias the estimation. As intuition suggests, MaxCorr is immune to amplitude mismatch, while ignoring amplitude mismatch when using LρNE may result in biased estimates. The cost function of MaxCorr is:

$$\tilde{J} = \langle \mathbf{r}(t), m\mathbf{s}(\alpha t) \rangle_p. \tag{6}$$

Theorem 1 *With MaxCorr (6) as the cost function, the amplitude mismatch factor m need not be taken into account in order to estimate α and δ. Moreover, m cannot be estimated using the MaxCorr cost.*

Proof: Let $\tilde{J}$ be the cost function with amplitude mismatch:

$$\tilde{J} = \langle \mathbf{r}(t), m\mathbf{s}(\alpha t) \rangle_\rho = m\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle_\rho = m \cdot J$$

When the match is perfect

$$\frac{\partial \tilde{J}}{\partial \alpha} = m\frac{\partial \tilde{J}}{\partial \alpha} = m \cdot 0 = 0$$

The above proof also shows that it is impossible to estimate the amplitude mismatch factor m using MaxCorr as cost function because the derivative at the correct coordinate (α^*, δ^*) will be zero, regardless of m.

Theorem 2 *With LρNE as the cost function, the amplitude mismatch factor m must be taken into account in order to estimate α and δ. Ignoring m may lead to biased estimation.*

Proof: Let m be the amplitude factor, and

$$J = \|\mathbf{r}(t) - m\mathbf{s}(\alpha t)\|_\rho^2. \tag{7}$$

The objective is to show that if $m \neq 1$, then J does not achieve its minimum at $\alpha = \alpha^*$. Suppose the template

$s(\alpha t)$ matches the data $\mathbf{r}(t)$ at $\alpha = \alpha^*$, i.e. $\mathbf{r}(t) = s(\alpha^* t)$. Expressing the ρ-norm in terms of the L_2-norm yields

$$J = \frac{\|s(\alpha^* t) - m s(\alpha t)\|^2}{\omega_s / \alpha}.$$

For the estimation to be unbiased,

$$\left. \frac{\partial J}{\partial \alpha} \right|_{\alpha = \alpha^*} = 0. \tag{8}$$

In general

$$\left. \frac{\partial s(at)}{\partial \alpha} \right|_{\alpha = \alpha^*} \neq 0.$$

Hence

$$\frac{\partial J}{\partial \alpha} = \frac{\alpha}{\omega_s} \frac{\partial \|s(\alpha^* t) - m s(at)\|^2}{\partial \alpha} + \frac{\|s(\alpha^* t) - m s(at)\|^2}{\omega_s}$$

$$\frac{\partial J}{\partial \alpha} = \frac{\|s(\alpha^* t) - m s(at)\|^2}{\omega_s} - 2 m \alpha \|s(\alpha^* t) - m s(at)\| \frac{\partial s(\alpha t)}{\partial \alpha}$$

$$\left. \frac{\partial J}{\partial \alpha} \right|_{\alpha^*} = \frac{(1-m)^2 \|s(\alpha^* t)\|^2}{\omega_s} + 2(m^2 - m)\alpha^* \|s(\alpha^* t)\| \left. \frac{\partial s(\alpha t)}{\partial \alpha} \right|_{\alpha^*}$$

$$\left. \frac{\partial J}{\partial \alpha} \right|_{\alpha = \alpha^*} \neq 0 \qquad \forall m \neq 1$$

By violating (8), the estimation is, in general, biased when $m \neq 1$.

5. THE MAXCORR ALTERNATIVE

As a result of using MaxCorr (6) as the cost function for gradient estimation, a relationship between α and the template is found. This relationship can be exploited to create a faster method of estimating α_i.

Theorem 3 *The frequency scaling factor α is determined by the ratio between the energy of the template and energy of frequency scaled (by α_i) version of itself:*

$$\alpha^* = \frac{\|\mathbf{s}(t)\|^2}{\|\mathbf{s}(\alpha^* t)\|^2}. \tag{9}$$

Proof: Writing (6) in terms of the L_2 norm yields

$$J = \langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle_\rho = \frac{\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle}{N_\alpha}$$

where

$$\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle = \sum_{i=1}^{N_\alpha} r_i(t) s_i(\alpha t). \tag{10}$$

At match, $\alpha = \alpha^*$ and $\mathbf{r}(t) = s(\alpha^* t)$. Thus

$$\left. \frac{\partial J}{\partial \alpha} \right|_{\alpha = \alpha^*} = \left. \frac{\partial (\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle \frac{\alpha}{\omega_s})}{\partial \alpha} \right|_{\alpha = \alpha^*} = 0$$

which implies that

$$\left. \frac{1}{\omega_s} \left(\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle + \alpha \frac{\partial \langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle}{\partial \alpha} \right) \right|_{\alpha = \alpha^*} = 0.$$

The right hand side is complicated since α appears in the template $s(\alpha t)$ and in the upper limit of the sum. This is in the form of an ODE

$$\left\{ \frac{\partial \langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle}{\langle \mathbf{r}(t), \mathbf{s}(\alpha t) \rangle} = -\frac{\partial \alpha}{\alpha} \right\}_{\alpha = \alpha^*}$$

which can be solved

$$ln\langle \mathbf{r}, \mathbf{s}(\alpha^* t) \rangle = -ln(\alpha^*) + K.$$

The matching condition $\alpha = \alpha^*$ implies $\mathbf{r}(t) = \mathbf{s}(\alpha^* t)$, and

$$ln\|\mathbf{s}(\alpha^* t)\|^2 = -ln(\alpha^*) + K.$$

Realizing that the template has $\alpha^* = 1$ shows that $ln\|\mathbf{s}(t)\|^2 = K$. Hence

$$ln\|\mathbf{s}(\alpha^* t)\|^2 = -ln(\alpha^*) + ln\|\mathbf{s}(t)\|^2$$

which can be rearranged to show (9).

5.1. Application: Independent Norm Ratio Test (INRT)

If $\mathbf{s}(\alpha t)$ on the right hand side of (9) is replaced by $\mathbf{r}(t)$, the following equation is obtained

$$\alpha^* = \frac{\|\mathbf{s}(t)\|^2}{\|\mathbf{r}(t)\|^2}. \tag{11}$$

The expression on the right hand side will be evaluated and compared with the current value of α to be tested. If they are equal, then $\|r(t)\| = \|s(at)\|$. If the template function correctly describes the data, then $\mathbf{r}(t) = \mathbf{s}(\alpha t)$. From the equation above, INRT does not give information about translation parameters. This is obvious for the 'wrap-shift' method discussed earlier where the norm of $\mathbf{r}(t)$ is independent of any shifts.[2] In the 'pad zeros' method, incorrect translation introduces undesired signals into $\|\mathbf{r}(t)\|$, causing $\frac{\|\mathbf{s}(t)\|^2}{\|\mathbf{r}(t)\|^2}$ to deviate from α. With other parameters fixed, δ_i can be estimated using a simple one-dimensional search. INRT is based on the error surface $|\alpha^* - \|s(t)\|^2 / \|r(t)\|^2|$. Since this is not very smooth, it may have many local minima. On the other hand, since INRT is numerically very fast, it may be useful when computation is at a premium. Even small errors in the norm, however, may lead to large changes in α, thus making INRT sensitive to noise and to template mismatch.

---

[2]This occurs because, for finite series, numbers can be added in different order without affecting the sum. [2]

6. APPLICATION TO SUNSPOT DATA

The raw sunspot data from [3] is smoothed and called the 'original data' $R(t)$ in Fig. 1. An initial template was chosen (shown in Fig 2) to be equal to the data in the first pseudo-period, and all values were normalized so that the initial template has height $m = 1$ and width $\alpha = 1$. The parameters α_i, β_i, and m_i were then estimated using the LρNE gradient method of (5).
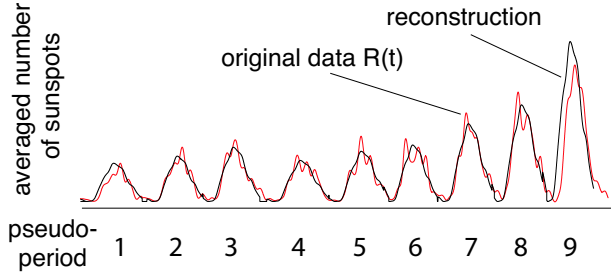

Fig. 1. The sunspot data in the top plot is superimposed over the estimates which use the final template and the parameter values given in the Table.
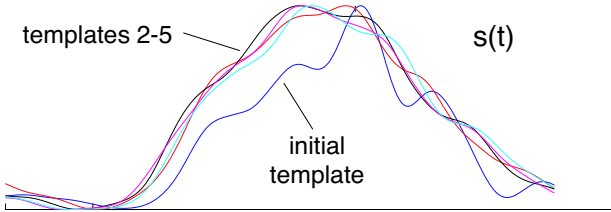

Fig. 2. The initial template is chosen to be equal to the data in the first pseudo-period of Fig 1. Successive templates are then chosen according to the bootstrap method 1-5.

Because the choice of template is important, we propose the following "bootstrap method" of finding and improving templates:

1. Begin with a nominal template. This must be made sensibly, though it need not be exact. In the above example we suggest that a reasonable (though not infallible) method is to choose the data in the first pseudo-period as an initial guess at the template.

2. Run the parameter estimation algorithm to find appropriate amplitude, scale, and translation parameters for the complete data record.

3. Parse the data record into its normalized pseudo-periods $s_i(t)$.

4. Create a new template $s(t)$ by averaging over all the $s_i(t)$, i.e., $s(t) = \frac{1}{|i|} \sum_i s_i(t)$.

5. Iterate (return to step 2) until the template ceases to change significantly.

Following this procedure with the sunspot data, the template changed from its initial version (the first pseudo-period) to the successive templates shown in Fig 2. Evidently, the template 'converged' after one iteration, and continues to wiggle around the cluster of curves. After five iterations, the resulting template $s(t)$ was used, along with the corresponding parameters m_i, α_i, δ_i to reconstruct the original data. The reconstruction is shown in Fig 1. The final converged values of the parameters are estimated for eight pseudo-periods using the fifth template.

i	2	3	4	5	6	7	8	9
m_i	1.19	1.42	1.07	1.31	1.48	2.03	2.53	4.18
α_i	1.10	1.08	0.99	1.03	1.16	1.10	1.13	1.22
δ_i	0.48	-0.01	0.70	-0.01	-0.51	0.07	-0.24	-0.42

7. REFERENCES

[1] C. S. Burrus, R. A. Gopinath, and H. Guo, *Wavelets and Wavelet Transforms,* Prentice Hall, NJ 1998.

[2] F. Dangello and M. Seyfried, *Introductory Real Analysis*, Houghton Mifflin Company, 2000.

[3] S. L. Marple, *Digital Spectral Analysis with Applications*, Englewood Cliffs, N.J., Prentice-Hall, 1987.

[4] P. Polotti and G. Evangelista, "Analysis and synthesis of pseudo-periodic 1/f-like noise by means of wavelets with applications to digital audio," EURASIP JASP 2001:1, 2003.

[5] P. Polotti and G. Evangelista, "Harmonic-band wavelet coefficient modeling for pseudo-periodic sound processing," *Proceedings of the COST G-6 Conference on Digital Audio Effects,* pp. 103-108, Verona, Italy, Dec. 2000.

[6] W. A. Sethares "Repetition and pseudo-periodicity," Tatra Mountains Mathematical Publications, Publication 23, 2001.

[7] W. A. Sethares and T. Staley, "Periodicity transforms," *IEEE Trans. Signal Processing*, Vol. 47, No. 11, Nov. 1999.

[8] W. A. Sethares and T. W. Staley, "Meter and periodicity in musical performance," *Journal of New Music Research,* Vol. 30, No. 2, June 2001.