Extension of the FOT-Probability Theory of CS and Poly-CS Time Series from Infinitely Long Data Records to Finite Segments

Consider the finite segment of data $\{z(t): t = 0,1,2,...,NT-1\}$, and define the discrete-time *Finite Synchronized Average by*

$$\tilde{z}_T(\tilde{t}) \triangleq \left\langle z(\tilde{t}+nT)\right\rangle_0^{N-1} \triangleq \frac{1}{N}\sum_{n=0}^{N-1} z[(\tilde{t}+nT)\,\mathrm{mod}(NT)], \qquad \tilde{t} = 0,1,2,...,T-1 \tag{1}$$

for positive integers N and T. For $N \gg 1$, the modulo operation has only a small impact in general, because it increases the data used by only the fraction

$$\frac{\tilde{t}/T}{N-\tilde{t}/T}, \quad \text{for } \tilde{t}/T < 1-1/T \tag{2}$$

which is smaller than $1/N$ for $N \gg 1$. But it is used here because it ensures that the same amount of data is used for every value of $\tilde{t}$, regardless of the overall segment length NT and because, for any *T*-periodic component in $\tilde{z}_T(t)$ that is extended beyond $t = NT$, the $\mathrm{mod}(NT)$ operation has no effect.

Now, define the *Periodic Extension* of the function $\tilde{z}_T(\tilde{t})$ by

$$\tilde{z}(t) \triangleq \sum_{k=-\infty}^{\infty} \tilde{z}_T(t+kT) \tag{3}$$

The function $\tilde{z}(t)$ is the *T*-periodic component of the finite segment $z(t)$ over $t \in \{0,1,2,...,NT-1\}$, extended to be periodic for all integer time. That is, if one uses (1) to calculate the periodic component of the residual $z(t)-\hat{z}(t)$ over $t \in \{0,1,2,...,NT-1\}$, the result will be zero. So, $\tilde{z}(t)$ is the entire *T*-periodic component of $z(t)$ over $t \in \{0,1,2,...,NT-1\}$, and is extended over all integer values of t.

Furthermore, the complex amplitude of the complex sine-wave component with frequency $\alpha = q/T$, for any integer q, of $z(t)$ over $t \in \{0,1,2,...,NT-1\}$ is given by

$$c^{\alpha} \triangleq \frac{1}{NT}\sum_{t=0}^{NT-1} z(t)\exp\{-i2\pi\alpha t\} \tag{4}$$

and it is identical to the sine-wave component of $\tilde{z}(t)$ over $t \in \{0,1,2,...,NT-1\}$; therefore, the sine-wave component of the residual $z(t)-\hat{z}(t)$ over $t \in \{0,1,2,...,NT-1\}$ is zero. In addition, each of these harmonically related sine-wave components is orthogonal to the others.

It follows that this periodic-component extraction operation and the associated sine-wave-component extraction operation for a finite-segment of data share all same properties as those defined for a persistent function $z(t)$ specified for all integer time. And, when these components for a finite-segment of data are extended over all integers as described above, they provide the basic statistics needed for defining a Cyclostationary data model in terms of FOT-Probability theory that is identical to that already defined for infinitely long data segments. That is, the same definitions of Periodic Cumulative FOT Probability Distributions (CDFs) apply, and they possess the same properties, like the representation of the periodic CDF in terms of its sinusoidal complex CDFs. Similarly, we have the same Fundamental Theorem of Sine-Wave Component Extraction and Fundamental Theorem of Periodic Component Extraction.

These CDFs also can be combined over multiple incommensurate periods with the same formulas derived for infinitely long data sequences, to obtain Poly-Cyclostationary data models in terms of poly-periodic CDFs.

However, here there is a difference that should be surfaced. For the purpose of simplifying the argument, the discussion shifts here from discrete time to continuous time. For continuous time, the time domain $t \in \{0, 1, 2, ..., NT-1\}$ for the data is replaced with $t \in [0, NT]$, and the formulas in (1) and (3) are replaced with the following

$$\tilde{z}_T(\tilde{t}) \triangleq \left\langle z(\tilde{t} + nT) \right\rangle_0^{NT} \triangleq \frac{1}{N} \sum_{n=0}^{N-1} z[(\tilde{t} + nT) \bmod (NT)], \qquad \tilde{t} \in [0, T] \tag{5}$$

and

$$c^{\alpha} \triangleq \frac{1}{NT} \int_0^{NT} z(t) \exp\{-i 2\pi \alpha t\} dt \tag{6}$$

Now, sinusoids that are not harmonically related are not orthogonal on any finite interval of the data; they are only orthogonal over all real time. Therefore, the periodic extension of the periodic components to the whole real line is more important here than it is for harmonically related sinusoids. Furthermore, the complex strengths of non-harmonically related sinusoidal components will be nearly identical if the frequencies are close enough, relative to the reciprocals of the lengths of the data segments NT_i. This is generally not true for infinitely long data segments. That is, cycle resolution width is limited to the reciprocal of the data-segment length, which should be no surprise.

Other than these subtle differences, the relationships among the various component-extraction operators and the FOT-CDFs, such as the fundamental Theorem of Sine-Wave-Component Extraction, are completely equivalent for finite data-segments and infinitely long records of data. For this reason, the same notation can be used for component extraction on finite segments of data:

$$
\begin{aligned}
&1)\ E_T\{z(t)\} \equiv \langle z(t)\rangle_T \equiv \tilde{z}(t)\\
&2)\ E^{\alpha}\{z(t)\} \equiv \langle z(t)\rangle^{\alpha} \equiv c^{\alpha}\exp\{i2\pi\alpha t\}\\
&3)\ E^{\{\alpha\}}\{z(t)\} \equiv \langle z(t)\rangle^{\{\alpha\}} \equiv \sum_{\{\alpha\}} c^{\alpha}\exp\{i2\pi\alpha t\}\\
&4)\ E_{\{T\}}\{z(t)\} \equiv \langle z(t)\rangle_{\{T\}} \equiv \langle z(t)\rangle + \sum_{\{T\}}\left[\langle z(t)\rangle_{\{T\}} - \langle z(t)\rangle\right]
\end{aligned}
\tag{7}
$$

where 1) is the *Periodic Component Extraction Operation,* 2) is the *Sine-Wave Component Extraction Operation,* 3) is the *Poly-Periodic Component Extraction Operation*, which includes the *Periodic Component Extraction Operation* as a special case, and 4) is the alternative representation of the *Poly-Periodic Component Extraction Operation.* In (7) the left and center members of each identity comprise the original notation **[Bk2]**, **[JP34]** and the right members comprise the notation used in this presentation.

In summary, for the case treated here, of finite data-segments, the fundamental statistics and their relationships with each other that define the *Purely Empirical* FOT-Probabilistic Models of Cyclostationarity and Poly-Cyclostationarity are exactly as presented in the seminal publications **[Bk2]**, **[JP34]**. The only difference is that here we do not consider Almost Cyclostationary models that are not Poly-Cyclostationary because infinitely many incommensurate periods is not an empirical concept. However, despite the equivalence of the *models*, the *theories* based on these models differ significantly. The impact of signal processing operations on the fundamental statistics are more complicated in the empirical theory, and the properties of these statistics also are more complicated. Simplifications that are so useful conceptually occur only in the limit as the amount of data approaches infinity. So, in practice, both the idealized and empirical non-stochastic theories are valuable tools.

CONCLUSIONS

- As explained above, there exist entirely empirical FOT probabilistic models of stationary, cyclostationary and poly-cyclostationary times series.
- All quantities occurring in these models can be calculated from physically measured/observed time series data on finite intervals.
- This theory should appeal to practitioners who analyze and process empirical data.
- CAVEATS: Drawbacks of the finite-time FOT probabilistic models are:
  - *Cumulant Selectivity* is not exact; it is only approximate; the more data, the more nearly exact it is.
  - *Signal separability* with FRESH filtering is limited. The degree of spectral redundancy decreases with decreasing amounts of data.
  - *The spectral resolution* of spectral correlation functions is limited

    - *Sinewave generation* is only approximately measurable since spectral features narrower than the reciprocal of the length of the data segment are not resolvable.

- PRAGMATIC IMPACT: In practice, one never has more than a finite amount of data; so, the above caveats exist in practice, regardless of the theory used. The Empirical theory is consistent with what is actually realizable in practice, whereas the more idealistic theory reveals what could be achieved with unlimited amounts of data. The theoretical simplifications of the idealized theory are definitely of conceptual value. But, as always, one must beware that these simplifications are not realizable exactly.