

REVIEW

Open Access



Statistically inferred time warping: extending the cyclostationarity paradigm from regular to irregular statistical cyclicity in scientific data

William A. Gardner

Abstract

Statistically inferred time-warping functions are proposed for transforming data exhibiting irregular statistical cyclicity (ISC) into data exhibiting regular statistical cyclicity (RSC). This type of transformation enables the application of the theory of cyclostationarity (CS) and polyCS to be extended from data with RSC to data with ISC. The non-extended theory, introduced only a few decades ago, has led to the development of numerous data processing techniques/algorithms for statistical inference that outperform predecessors that are based on the theory of stationarity. So, the proposed extension to ISC data is expected to greatly broaden the already diverse applications of this theory and methodology to measurements/observations of RSC data throughout many fields of engineering and science. This extends the CS paradigm to data with inherent ISC, due to biological and other natural origins of irregular cyclicity. It also extends this paradigm to data with inherent regular cyclicity that has been rendered irregular by time warping due, for example, to sensor motion or other dynamics affecting the data.

Keywords: Cyclostationarity, Irregular cyclicity, Rhythmicity, Signals, Time-series, Time warping

1 One-sentence summary

Well-known data analysis benefits of cyclostationary signal-processing methodology are extended from regular to irregular statistical cyclicity in scientific data by using statistically inferred time-warping functions.

2 The cyclostationarity paradigm in science

2.1 Cyclicity is ubiquitous in scientific data

Many dynamical processes encountered in nature arise from periodic or cyclic phenomena. Such processes, although themselves not periodic functions of time, can produce random or erratic or otherwise unpredictable data whose statistical characteristics do vary periodically with time and are called *cyclostationary (CS) processes* [1–3]. For example, in telecommunications, telemetry, radar, and sonar systems, statistical periodicity or regular cyclicity in data is due to modulation, sampling, scanning, framing, multiplexing, and coding operations. In

these information-transmission systems, relative motion between transmitter or reflector and receiver essentially warps the time scale of the received data. Also, if the clock that controls the periodic operation on the data is irregular, the cyclicity of the data is irregular. In mechanical vibration monitoring and diagnosis, cyclicity is due, for example, to various rotating, revolving, or reciprocating parts of rotating machinery; and if the angular speed of motion varies with time, the cyclicity is irregular. However, as explained herein, irregular statistical cyclicity (ISC) due to time-varying RPM or clock timing is not equivalent to time-warped regular statistical cyclicity (RSC). In astrophysics, irregular cyclicity arises from electromagnetically induced revolution and/or rotation of planets, stars, and galaxies and from pulsation and other cyclic phenomena, such as magnetic reversals of planets and stars, and especially Birkeland currents (concentric shells of counter-rotating currents). In econometrics, cyclicity resulting from business cycles has various causes including seasonality and other less regular sources of cyclicity. In atmospheric science, cyclicity

Correspondence: wag@statsig.com
University of California, Davis, 18 Hahnemann Lane, Napa, CA 94558, USA



© The Author(s). 2018, corrected publication October/2018. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

is due to rotation and revolution of Earth and other cyclic phenomena affecting Earth, such as solar cycles. In the life sciences, such as biology, cyclicity is exhibited through various biorhythms, such as circadian, tidal, lunar, and gene oscillation rhythms. The study of how solar- and lunar-related rhythms are governed by living pacemakers within organisms constitutes the scientific discipline of *chronobiology*, which includes comparative anatomy, physiology, genetics, and molecular biology, as well as development, reproduction, ecology, and evolution. Cyclicity also arises in various other fields of study within the physical sciences, such as meteorology, climatology, oceanology, and hydrology. As a matter of fact, the cyclicity in all data is irregular because there are no perfectly regular clocks or pacemakers. But, when the degree of irregularity throughout time-integration intervals required for extracting statistics from data is sufficiently low, the data's cyclicity can be treated as regular.

The relevance of the theory of cyclostationarity to many fields of time-series analysis was proposed in the mid-1980s in the seminal theoretical work and associated development of data processing methodology reported in [1–3], which established cyclostationarity as a new paradigm in data modeling and analysis, especially—at that time—in engineering fields and particularly in telecommunications signal processing where the signals typically exhibit RSC. More generally, the majority of the development of such data processing techniques that ensued up to the turn of the century was focused on statistical processing of data with RSC for engineering applications, such as telecommunications/telemetry/radar/sonar and, subsequently, mechanical vibrations of rotating machinery. But today—more than 30 years later—the literature reveals not only expanded engineering applications but also many diverse applications to measurements/observations of RSC data throughout the natural sciences (see Appendix), and it is to be expected there will be many more applications found in the natural sciences for which benefit will be derived from transforming ISC into RSC, and applying the now classical theory and methodology.

Wide-sense cyclostationary stochastic processes have autocorrelation functions that vary periodically with time. This function of time, under mild regularity conditions on its mathematical model, can be expanded in a Fourier series whose coefficients, referred to as *cyclic autocorrelation functions*, depend on the lag parameter; the Fourier frequencies, called *cycle frequencies*, are multiples of the reciprocal of the period of cyclostationarity [1–3].

More generally, if the frequencies of the (generalized) Fourier series expansion of the autocorrelation function are not commensurate, that is, if the autocorrelation function is an almost-periodic (in the mathematical sense) function of time, then the process is said to be

almost-cyclostationary [4]. This large class includes as subclasses the *polycyclostationary (polyCS) processes*, which exhibit only a finite number of incommensurate periods, and the cyclostationary processes which exhibit only one period. The (almost) periodicity property of the autocorrelation function is manifested in the frequency domain of the data as statistical dependence (e.g., correlation) between the spectral components of the data process that are separated in frequency by amounts equal to the cycle frequencies of the process and are shifted to any common spectral band for correlation measurement. In contrast to this, stationary (in the wide-sense) processes have joint moments (autocorrelation functions) that are independent of time, depending on only the lag parameter, and all spectral components at distinct frequencies are statistically independent (uncorrelated) with each other.

This subject has been further broadened by the generalization of the theory to *generalized almost-cyclostationary processes*, which exhibit cycle frequencies of the autocorrelation function that are dependent on the value of the lag variable, in [5].

As a simple means of assessing the current prevalence of the *cyclostationarity paradigm* in scientific data processing—that is, the concept of cyclostationarity and the associated body of data processing theory and method—in various fields of science and engineering, a web search using <https://scholar.google.com/> was performed in April 2018, as a refinement and update of the search performed during the writing of this paper in 2015. This latter search was based on just under 50 nearly distinct applications areas in science and engineering, and the search terms were chosen to yield only results involving cyclostationarity. By “nearly distinct”, it is meant that the search terms were also selected to minimize redundancy (multiple search application areas producing the same “hits”). The results are shown in Table 1 in Section 17, Appendix. The total number of hits was about 136,000. The hits grow from a trickle of 1 to 2 figures per decade in the 1960s/1970s to a total of 6 figures over the ensuing half century.

The same is true, with 5 figures, for a search performed on the single general search term “cyclostationary OR cyclostationarity”. Also, as shown in Table 2, another search was performed using just over 20 search terms that represent partially-redundant general subjects in science and engineering. The total number of hits was about 238,000. These hits also grew from a trickle of 1 to 2 figures per decade in the 1960s/1970s to a total of 6 figures over the ensuing half century.

Some analysis of Google Scholar's search results obtained using the terms shown in these tables suggest that this search engine's proprietary search algorithm is corrupting the logical “OR” operation and possibly the

“AND” operation. Further attempts will be made in an attempt to minimize the impact of this hypothesized corruption, and results obtained will be posted in [6]. Yet, there is good reason to believe that this body of theory and method and their applications would be even more pervasive if its utility could be extended from data with regular statistical cyclicity (RSC) to data with irregular statistical cyclicity (ISC).

The purpose of this paper is to enable an extension of the cyclostationarity paradigm from data exhibiting RSC to data exhibiting ISC. The approach taken, when performed in discrete time (as required when implemented digitally), can be classified as adaptive non-uniform resampling of data, and the adaptation proposed is performed blindly (requires no training data) using a property-restoral technique specifically designed to exploit cyclostationarity.

For what follows, readers would benefit from some basic knowledge of the concept of cyclostationarity—the periodic time variation of probabilistic (mathematical) or statistical (empirical) parameters of time-series data, which are sometimes called *signals*. These parameters are most notably the joint probability density function for the signal’s amplitude at multiple points in time, or moments of these density functions. A *polyperiodic* function of time (typically not the signal itself) is defined by its characteristic of being able to be expressed as a finite sum of periodic functions with multiple incommensurate periods (these are periods whose ratios are all irrational numbers. Polyperiodic time variation of probabilistic/statistical parameters characterizes polycyclostationary signals. The frequencies of the individual harmonics associated with each period of a cyclostationary/polycyclostationary signal are called the *cycle frequencies*. Polycyclostationary signals were originally [4], and are still [1–3, 7], most frequently called almost-cyclostationary (particularly by mathematically oriented authors) because polyperiodic functions are examples of *almost periodic* (in the mathematical sense) functions; such functions need not have only a finite number of incommensurate periods.

Tutorial treatments of cyclostationarity theory and method are available in the books and journal articles [1–5, 7, 8] and references therein; for the highest fidelity treatments in the literature, all of which share a common terminology and a self-consistent foundational theory (over the last three decades), readers are referred to those authored by the first author of [1–4] and [7–11], the originator of the cyclostationarity paradigm in signal processing, whose publications on this topic date back to the early 1970s [9]. Also recommended are the more recent publications by the originator of several extensions and generalizations of cyclostationarity, the author of [5], who uses

terminology and develops theory that are (for the most part) consistent with that in the foundational literature.

There exists a duality between two alternative theories of polycyclostationarity: (1) *the traditional theory*, introduced in 1978 [4], which is more abstract and is based on the stochastic-process model (introduced in the 1940s by Kolmogorov) and the associated probabilistic expectation operation, and (2) *the empiricist’s alternative*, introduced during 1985–1991 [1, 2, 8, 10], which is recommended for scientists working with empirical data and is based on *fraction-of-time probability* and *the sine-wave-extraction operation* (introduced during that same period [1, 2, 8, 10]). For tutorial treatments of the concepts underlying this duality, these four originating publications plus [3, 11] are recommended, particularly for analytically inclined practitioners but also for any reader seeking a treatment that starts from basics and proceeds step by step to build advanced concepts and theory. For deeper mathematical treatments of fundamentals, the primary publications to date are [5, 12]. (Some of the references cited herein, such as out-of-print books, authored by the author of this paper are accessible to all for free as downloadable PDF documents at the webpage [9]; in the future, the primary source on cyclostationarity is expected to be the website (presently under construction) with domain name *cyclostationarity* followed by any of the domain extensions *.com*, *.org*, *.info*, *.net*, and *.us* [6].) The presently inadequate Wikipedia article entitled *Cyclostationary_process* (and several inadequate/redundant articles at other Wikipedia sites) requires major upgrading.

One simple example of a CS signal is described here to illustrate that what is here called *regular statistical cyclicity* for time-series data can represent extremely erratic behavior relative to a periodic time series. Consider a long train of pulses or bursts of arbitrary complexity and identical functional form and assume that, for each individual pulse the shape parameters, such as amplitude, width (or time expansion/contraction factor), time-position, and center frequency, are all random variables—their values change unpredictably from one pulse in the train to the next. If these multiple sequences of random parameters associated with the sequence of pulses in the train are jointly stationary random sequences, then the signal is CS and therefore exhibits regular statistical cyclicity, regardless of the fact that the pulse train can be far from anything resembling a periodic signal. As another example, any broadband noise process with center frequency and/or amplitude and/or time scale that is varied periodically is CS. Thus, exactly regular statistical cyclicity can be quite subtle and even unrecognizable to the casual observer. This is reflected in the frequent usage in recent times of CS models for time-series data from natural phenomena of many distinct

origins (see [Appendix](#)). Yet, there are many ways in which a time-series of even exactly periodic data can be affected by some dynamical process of cycle-time expansion and/or contraction in a manner that renders its statistical cyclicity irregular: not CS or polyCS. The particular type of dynamic process of interest in this contribution is *time warping*.

3 Time warping

3.1 Time warping is a dynamical process of cycle-time expansion and/or contraction

Let $x(t)$ be a wide-sense CS, or wide-sense polyCS signal or time series of data with at least one cycle frequency α , and let $y(t)$ be a *time-warped* version

$$y(t) = x(\psi(t)) \tag{1}$$

in which the time-warping function $\psi(t)$ represents a causal data transformation, meaning warped time never pauses or reverses direction: if $t_2 > t_1$, then $\psi(t_2) > \psi(t_1)$. In this case, the warping function is an invertible function. The notation $t = \psi^{-1}(s)$ is used herein to denote the inverse of $s = \psi(t)$.

In general, if the time warping is not an affine transformation, $\psi(t) = at + b$, or some periodic or polyperiodic generalization thereof, such as $\psi(t) = at + b(t)$, in which $b(t)$ is a periodic or polyperiodic function, then any cyclicity in $x(t)$ is absent in $y(t)$: the signal $y(t)$ is not CS or polyCS. Nevertheless, by *de-warping* time in $y(t)$, $x(t)$ is recovered and therefore cyclicity is restored.

The periodically (or polyperiodically) time-varying autocorrelation function for $x(t)$ is given by

$$R_{xx^{(*)}}(t, \tau) = \sum_{\alpha} R_{xx^{(*)}}^{\alpha}(\tau) \exp(j2\pi\alpha t) \tag{2}$$

where the cyclic autocorrelations $R_{xx^{(*)}}^{\alpha}(\tau)$ are defined in the usual manner [1–3] in terms of either sinusoidally weighted time averages of lag products or same for probabilistic expected values of lag products. For CS $x(t)$ with period T_1 , we have $\{\alpha\} = \{h/T_1; h = \text{some integers}\}$; and, for polyCS $x(t)$, we have

$$\{\alpha\} = \left\{ \begin{array}{l} h_p/T_p; p = 1, 2, \dots, P < \infty; \text{ and,} \\ \text{for each } p, h_p = \text{some integers and} \\ T_p = \text{one of } P \text{ incommensurate periods} \end{array} \right\}.$$

For example, the expected value of the lag product with lag τ is given by

$$\begin{aligned} R_{xx^{(*)}}(t, \tau) &\triangleq E\{x(t + \tau), x^{(*)}(t)\} \\ &= \sum_{\alpha} R_{xx^{(*)}}^{\alpha}(\tau) \exp(j2\pi\alpha t), \end{aligned} \tag{3}$$

where the superscript $(*)$ denotes optional conjugation [1–3] of data (such as the baseband complex-valued representations of real-valued bandpass signals) that is represented in terms of complex data values, and the cyclic autocorrelation is given (ideally) by the limit of the sinusoidally weighted time average of the lag product, as the averaging time T (ideally, for polyCS $x(t)$) approaches infinity:

$$\begin{aligned} R_{xx^{(*)}}^{\alpha}(\tau) &= \lim_{T \rightarrow \infty} \left\langle E\{x(t + \tau)x^{(*)}(t)\} \right. \\ &\quad \left. \times \exp(-j2\pi\alpha t) \right\rangle_T. \end{aligned} \tag{4}$$

The usual estimate of this statistical function, obtained from the data $x(t)$, is given by

$$\hat{R}_{xx^{(*)}}^{\alpha}(\tau) = \left\langle x(t + \tau)x^{(*)}(t) \exp(-j2\pi\alpha t) \right\rangle_T \tag{5}$$

where T is the finite length of the time-averaging interval used. Readers are warned that the presentation here follows the convention in [5] for which the asymmetric lag product $x(t + \tau)x^{(*)}(t)$ is adopted, whereas the earlier treatments in all references herein by the Author follow the alternative convention based on the symmetric lag product $x(t + \tau/2)x^{(*)}(t - \tau/2)$. Because of this difference in convention, the asymmetric lag cyclic autocorrelations in [5] and herein differ from the symmetric lag autocorrelations in the original work [1–3] in that the former equals the latter multiplied by the lag-dependent phase-shifting factor $\exp(j\pi\alpha\tau)$. The choice of convention used herein was dictated by the benefit gained by avoiding the need for translations back and forth between this paper and the important complementary source [5] and the directly related upcoming publications by the author of [5], and also because because $\tau/2$ is generally undefined for discrete time.

In the one special case for which $x(t)$ is CS (not polyCS), and the probabilistic expectation operation is used as in Eqs. 3 and 4, the theory reveals that a finite averaging time equal to the period of CS, $T = T_1$ (or any non-zero integer multiple thereof) suffices in Eq. 4. There is no need for an infinite amount of time-averaging to obtain the idealized result. However, if the probabilistic expectation is not used, then (ideally) infinite averaging time is required to obtain the mathematically idealized cyclic autocorrelations.

In contrast to $x(t)$, the cyclic autocorrelations of the time-warped data $y(t)$, whether defined with or without (cf. [10]) the probabilistic expectation operation, are generally zero, $R_{yy^{(*)}}^\alpha(\tau) = 0$, for all the values of α in Eq. 2, except possibly $\alpha = 0$, and for all other non-zero values of α .

Before proceeding, it is clarified here that, although the notation used does not reveal any dependence of the time-warping function $\psi(\cdot)$ on the phenomenon characterized by $x(\cdot)$, there is nothing in the theory or method presented here that prohibits such dependence, with one exception that is explained in Section 6. The only limitation on the nature of such dependence is that, to be physically viable, the dependence must (according to generally—but not unanimously—accepted principles of cosmology) be causal, meaning that the dependence $\psi(t) = \psi(t; \{x(\nu) : \nu < t\})$ is possible, but ν in this expression cannot be allowed to exceed t . This clarification can be summarized as follow:

The manner in which present cyclicity of a phenomenon departs from being regular can depend on past behavior of the phenomenon.

(The mathematical question of whether or not this mathematical model should be modified to allow ν to equal t is not addressed here.)

As a final introductory remark on time warping and cyclicity, let us take into account the fact that data obtained from measurement/observation of physical phenomena cannot, in reality, exhibit exact RSC. This property is a mathematical idealization of physical reality. The extent to which data departs from exact RSC sets an upper limit on how long sinusoidally weighted time averages (with sinusoid frequencies approximately equal to the data's cycle repetition frequency and its harmonics) of the data, and/or time-invariant nonlinear transformations of the data, can be integrated (without dividing the integral by the integration time to produce an average value) before the magnitude of the result stops growing with a linear trend with increasing integration time. This upper limit is here referred to as the *cycle coherence time (CCT)*—the maximum length of time over which the cycle frequency is stable. (This is distinct from the *cyclic coherence time*, which is defined to be the width of the cyclic autocorrelation function of the data—the maximum length of time separation (lag) for which time samples are cyclically correlated.)

The objective of the data processing methods presented herein is to increase the cycle coherence time (CCT) of data—render the data's statistical cyclicity more regular or less irregular (increase RSC

or decrease ISC) – enough to be able to achieve coherent cyclic processing gain sufficient for the information-extraction task at hand.

Generally speaking, when unintentional time warping of data exhibiting some level of inherent regularity of statistical cyclicity decreases the data's CCT (renders it less RSC or more ISC), the purpose of the de-warping described herein is to recover the longer CCT. But the same de-warping methods (but possibly better interpreted in this case as *warping methods*) can produce a substantial increase in CCT when the cyclicity is inherently irregular even though the unprocessed data has not been subjected to any time warping; in fact, useful levels of CCT can be obtained in some cases even when the original data exhibits such highly irregular statistical cyclicity that its CCT is negligible to start with.

4 De-warping to restore cyclostationarity

If $y(t)$ is *de-warped* using the inverse of the warping function that has transformed RSC in the data into ISC, the regular statistical cyclicity present in $x(t)$ is recovered:

$$y(\psi^{-1}(s)) = x(s) \tag{6a}$$

or, changing the variable's label from s to t , we obtain

$$y(\psi^{-1}(t)) = x(t). \tag{6b}$$

More generally, assuming that $\psi^{-1}(t)$ is completely or, at least, partially unknown, it is in principle possible to estimate it from the observed data $y(t)$ by searching for the particular function $\phi(t)$ (an estimate of $\psi^{-1}(t)$) that maximizes the strength of a measurement of some cyclic feature, such as the cyclic autocorrelation function, for the candidate de-warped data

$$y(\phi(t)) = x_\phi(t) \tag{7}$$

at one or more values of lag τ and cycle frequency α , where $x_\phi(t) = y(\phi(t)) = x(\psi([\phi(t)]) = x(t)$ for $\phi(t) = \psi^{-1}(t)$. In some applications, doing this jointly for appropriate multiple values of τ and α can improve the quality of the estimate. In such cases, the most appropriate values for τ may be different from one value of α to another. In practice, such values may be determinable only by experimentation with trial values. Nevertheless, it is stated here, on the basis of decades of experience, that more often than not the magnitude of $R_{xx^{(*)}}^\alpha(\tau)$ peaks at $\tau = 0$ for physically realistic models of $x(t)$.

For any valid cycle frequency α and lag value τ for which $R_{xx^{(*)}}^\alpha(\tau)$ is not zero and not negligibly small, the property-restoral optimization proposed here is

$$\max_{\varphi(t)} \left\{ \left| \hat{R}_{x_\varphi}^\alpha(\tau) \right|^2 \right\} \quad (8)$$

where $\hat{R}_{x_\varphi}^\alpha(\tau)$ is a *shorthand notation* (the double subscript used in Section 3 is replaced with a single subscript from this point forward) for a measurement (estimate) of the cyclic autocorrelation of $x_\varphi(t)$ obtained from a finite time-averaging interval (and, of course, no expectation operation):

$$\hat{R}_{x_\varphi}^\alpha(\tau) = \left\langle x_\varphi(t + \tau)x_\varphi^*(t) \exp(-j2\pi\alpha t) \right\rangle_T. \quad (9)$$

If no valid cycle frequency α for $x(t)$ is known, then this parameter also must be searched over in the optimization Eq. 8. One possibility for initializing the estimate of α is described in Section 11, where there also is described a possibility for initializing φ .

The values of cycle frequency α for which Eq. 8 is a valid objective function for de-warping include any/all cycle frequencies for which the cyclic correlation coefficient for $x(t)$ is non-negligible (not much less than unity in magnitude). An alternative to the *single-cycle objective function* Eq. 8 is a *multi-cycle objective function* which can be either a sum over cycle frequencies of squared magnitudes of cyclic autocorrelation functions or a sum of the complex values of cyclic autocorrelations. The latter may perform best, but it may be impractical in many cases because of the need for equalizing the phases of the signal component in each term in order to obtain coherent addition (cf. literature on *maximum-likelihood multi-cycle detectors*). Another alternative is to sum squared magnitudes of cyclic autocorrelations over multiple lag values τ (for either one or multiple values of α). As illustrated in the example presented in Section 14, multiple harmonically related values of α could be useful as could a range of lag values τ centered at $\tau = 0$. However, in that example, the strongest cyclic autocorrelation value occurs at the first harmonic and at a lag of zero.

Substituting Eq. 7 into Eq. 9 yields the measured statistic whose squared magnitude is the performance functional to be maximized w.r.t the candidate de-warping function φ :

$$\hat{R}_{x_\varphi}^\alpha(\tau) = \left\langle y(\varphi[t + \tau])y^*(\varphi[t]) \exp(-j2\pi\alpha t) \right\rangle_T. \quad (10)$$

It is important to recognize that, although the above concept and method is presented as if the observed data arose from the time warping of other data that exhibited regular cyclicity prior to time

warping, there is in fact no need for this conceptual model—no need for there to be an underlying physical mechanism exhibiting regular statistical cyclicity that is then transformed into irregular statistical cyclicity by some actual time-warping process. Direct sources of ISC (e.g., an EKG from a beating heart or many other biological functions that naturally produce ISC, or some long-term climate, geological, or celestial data, etc.) can, in principle, be de-warped in many cases even if it were not warped to start with. However, in this case, we should say it can be “warped” not “de-warped” since there is no original warping to be removed. To summarize:

The objective addressed by the theory and method presented here is twofold:

- (i) To convert naturally occurring ISC in data into RSC (or at least increase the data’s CCT) by time warping—thereby rendering the converted data (more) amenable to CS and/or polyCS data processing techniques, algorithms, and theory;
- (ii) To de-warp time in data that exhibited RSC prior to having been subjected to time warping—thereby increasing the data’s CCT, rendering it more amenable to CS and/or polyCS data processing techniques, algorithms, and theory.

The optimization method based on Eq. 8 is an example of a *property-restoral method* for *blind adaptation* (learning without training data). Introductions to cyclostationarity restoral for blind adaptive spatial filtering and frequency-shift spectral filtering for suppression of additive noise and interfering signals, and to joint cyclostationarity restoral for time-difference-of-arrival estimation in the presence of additive noise and interfering signals are presented in [1–3, 13, 14].

5 Warping compensation instead of de-warping

It is shown here that the search for an optimum de-warping function $\varphi \equiv \hat{\psi}^{-1} \cong \psi^{-1}$, by the method described in Section 4, can be transformed into an equivalent search for an optimum *warping compensation* function $\varphi^{-1} \equiv \hat{\psi} \cong \psi$. Such a function, once found, can then be inverted if it is desired to de-warp the data. By using the definition

$$\Delta_\varphi^\tau[\varphi^{-1}(u)] \triangleq \varphi[\varphi^{-1}(u) + \tau] - \varphi[\varphi^{-1}(u)] \quad (11)$$

the measured statistic Eq. 10 to be used for optimization of φ can be re-expressed as follows (using the change of variables $u = \varphi(t)$):

$$\begin{aligned}
\hat{R}_{x_\varphi}^\alpha(\tau) &= \left\langle y(\varphi[t+\tau])y^{(*)}(\varphi[t])\exp(-j2\pi\alpha t) \right\rangle_T \\
&= \left\langle y(\varphi[t+\Delta_\varphi^\tau(t)])y^{(*)}(\varphi[t])\exp(-j2\pi\alpha t) \right\rangle_T \\
&= \frac{|\varphi(T)|}{|T|} \left\langle y(u+\Delta_\varphi^\tau[\varphi^{-1}(u)])y^{(*)}(u) \right. \\
&\quad \times \exp[-j2\pi\alpha\varphi^{-1}(u)]\dot{\varphi}^{-1}(u) \left. \right\rangle_{\varphi(T)} \\
&\cong \frac{|\varphi(T)|}{|T|} \left\langle y(u+\tau/\dot{\varphi}^{-1}(u))y^{(*)}(u) \right. \\
&\quad \times \exp[-j2\pi\alpha\varphi^{-1}(u)]\dot{\varphi}^{-1}(u) \left. \right\rangle_{\varphi(T)}
\end{aligned} \tag{12}$$

where $|T|$ denotes the length of the averaging interval $T \triangleq [t_o, t_o + T]$ (with some abuse of notation) and, similarly, for the de-warped averaging interval:

$$\begin{aligned}
\varphi(T) &\triangleq [\varphi(t_o), \varphi(t_o + T)] \\
|\varphi(T)| &\triangleq \varphi(t_o + T) - \varphi(t_o)
\end{aligned}$$

and where the approximation in the last line of Eq. 12 is

$$\begin{aligned}
\Delta_\varphi^\tau[\varphi^{-1}(u)] &\cong \tau[d\varphi/dt]_{t=\varphi^{-1}(u)} \\
&= \tau[1/d\varphi^{-1}(u)/du] \\
&\triangleq \tau[1/\dot{\varphi}^{-1}(u)].
\end{aligned}$$

This approximation is accurate when $\varphi(t)$ is accurately approximated as linear over intervals no longer than the width of the function $R_x^\alpha(\cdot) \equiv R_{xx^{(*)}}^\alpha(\cdot)$, the cyclic coherence time.

The inverse $\varphi^{-1}(t)$ of the candidate de-warping function is a candidate warping-compensation function. Equation 12 indicates that an estimate of the warping function $\varphi^{-1} \equiv \hat{\psi} \cong \psi$, from which its derivative $\dot{\varphi}^{-1}(u) \cong \dot{\psi}(u)$ can be obtained, can be used to compensate for warping in the data by time warping the sinusoids and scaling, in a time-varying manner, their amplitudes and the lags used in the data to compute the cyclic autocorrelations. One can use Eq. 12 in Eq. 8 and search over $\varphi^{-1} \cong \psi$ instead of $\varphi \cong \psi^{-1}$, to directly find the warping-compensation function $\varphi^{-1} \cong \psi$; or, one can use Eq. 10 in Eq. 8 to search directly for the data-de-warping function $\varphi \cong \psi^{-1}$. The relative advantages and disadvantages of these two theoretically approximately equivalent approaches are expected to involve somewhat complicated tradeoffs between algorithmic efficiency and estimation accuracy. For an iterative search algorithm of the sort described in Section 10, the efficiency depends on computational complexity and storage requirements per iteration, and the number of iterations required for convergence. There are tradeoffs among these three efficiency parameters for a specified level of estimation accuracy. And there are also tradeoffs between estimation

accuracy and algorithmic efficiency. Especially important is the need for schemes, such as extensive diverse initializations of the iterative algorithm, which avoid mistaking substantially suboptimum local maxima (of which there can be many) for the desired global maximum. These important topics on search algorithm research are outside the scope of this paper.

6 Error analysis

Substituting Eq. 1 into Eq. 10 and using an estimate $\hat{\alpha}$ of a cycle frequency yields

$$\hat{R}_{x_\varphi}^{\hat{\alpha}}(\tau) = \left\langle x(\psi[\varphi(t+\tau)])x^{(*)}(\psi[\varphi(t)])\exp(-j2\pi\hat{\alpha}t) \right\rangle_T \tag{13}$$

If φ is not exactly equal to the inverse of ψ , then there is some de-warping-function error, which is denoted by $e_\varphi \triangleq \varphi - \psi^{-1} \equiv \hat{\psi}^{-1} - \psi^{-1}$. This error may be due to error in estimating $\psi(t)$ and or error in inverting the estimated $\psi(t)$ or error in estimating $\psi^{-1}(t)$ directly.

In terms of this error, we have

$$\psi[\varphi(t)] = \psi[\psi^{-1}(t) + e_\varphi(t)] \triangleq t + e(t) \tag{14}$$

in which $e(t)$ denotes the time de-warping error created by the de-warping-function error:

$$\begin{aligned}
e(t) &\triangleq \psi[\psi^{-1}(t) + e_\varphi(t)] - \psi[\psi^{-1}(t)] \\
&= \psi[\psi^{-1}(t) + e_\varphi(t)] - t
\end{aligned} \tag{15}$$

Using this error definition, Eq. 13 can be re-expressed as

$$\begin{aligned}
\hat{R}_{x_\varphi}^{\hat{\alpha}}(\tau) &= \left\langle x(t+\tau+e(t+\tau))x^{(*)}(t+e(t)) \right. \\
&\quad \times \exp(-j2\pi\hat{\alpha}t) \left. \right\rangle_T.
\end{aligned} \tag{16}$$

Assuming that $\hat{\psi}^{-1}(t) = \varphi(t)$ and, therefore, $e(t)$ is not statistically dependent on $x(t)$ (this is, at best, an approximation when $\varphi(t)$ is determined from $x(t)$ in some data-adaptive manner, because then there is a deterministic relationship between $e(t)$ and $x(t)$), the probabilistic expected value (w.r.t the probability density function for $x(t)$) of Eq. 16 is given by

$$\begin{aligned}
E\left\{\hat{R}_{x_\varphi}^{\hat{\alpha}}(\tau)\right\} &= \left\langle \sum_\alpha R_x^\alpha(\tau+e(t+\tau)-e(t)) \right. \\
&\quad \times \exp(j2\pi[(\alpha-\hat{\alpha})t+ae(t)]) \left. \right\rangle_T \\
&= \left\langle \sum_\alpha R_x^\alpha(\tau[1+\langle\dot{e}(t)\rangle_\tau]) \right. \\
&\quad \times \exp(j2\pi[(\alpha-\hat{\alpha})t+ae(t)]) \left. \right\rangle_T
\end{aligned} \tag{17a}$$

where

$$\langle \dot{e}(t) \rangle_\tau \triangleq \frac{1}{T} \int_t^{t+\tau} \dot{e}(u) du.$$

This expression Eq. 17a also holds if $e(t)$ is statistically dependent on $x(t)$ provided that the expectation is conditional on the $x(t)$ -dependent $e(t)$ being any particular function of time. For an exact estimate of a cycle frequency, $\hat{\alpha} = \alpha_o$, Eq. 17a reduces to

$$\begin{aligned} E\left\{\hat{R}_{x_\varphi}^{\alpha_o}(\tau)\right\} &= \left\langle R_x^{\alpha_o}(\tau[1 + \langle \dot{e}(t) \rangle_\tau]) \right. \\ &\quad \times \exp(j2\pi\alpha_o e(t)) \left. \right\rangle_T \\ &+ \left\langle \sum_{\alpha \neq \alpha_o} \left\{ R_x^\alpha(\tau[1 + \langle \dot{e}(t) \rangle_\tau]) \right\} \right. \\ &\quad \times \exp[j2\pi\alpha e(t)] \\ &\quad \times \exp[j2\pi(\alpha - \alpha_o)t] \left. \right\rangle_T \end{aligned} \tag{17b}$$

For a signal model with a specified set of cyclic autocorrelation functions $\{R_x^\alpha(\tau)\}$, indexed by cycle frequency, Eqs. 17a and 17b can be used to study the sensitivity of the expected value of the objective function in Eq. 8 to the de-warping error and/or cycle frequency error.

The second term in Eq. 17b is called *cycle leakage* [2] (the term *cyclic leakage* used by some authors is conceptually misleading—the leakage is not cyclic; it represents the amount of the cyclic feature for each and every cycle frequency α and strength and phase $R_x^\alpha(\tau)$ that leaks into the measurement of the feature with cycle frequency $\hat{\alpha} = \alpha_o$ and strength and phase $R_x^{\alpha_o}(\tau)$). If the cyclicity, with cycle frequency $\alpha - \alpha_o$, in the product of the two factors in the sum in Eq. 17b that depends on t through the quantities $e(t)$ and $\langle \dot{e}(t) \rangle_\tau$ is negligible, this leakage term approaches zero as the averaging time T grows without bound. The first term in Eq. 17b is a time average of the actual cyclic autocorrelation prior to time warping in $x(t)$ and subjected to time-variant lag shift and complex-amplitude scaling.

If $\hat{\alpha} = \alpha$ is invalid for every cycle frequency α exhibited by $x(t)$, then the first term in the right member of Eq. 17b vanishes, and the value of the left member is due entirely to cycle leakage—the second term in the right member with α_o replaced by $\hat{\alpha}$. Also, it is noted that for sufficiently slowly varying $e(t)$, defined by $|\langle \dot{e}(t) \rangle_\tau| \ll 1$, Eq. 17b is closely approximated by

$$\begin{aligned} E\left\{\hat{R}_{x_\varphi}^{\alpha_o}(\tau)\right\} &= R_x^{\alpha_o}(\tau) \langle \exp(j2\pi\alpha_o e(t)) \rangle_T \\ &+ \sum_{\alpha \neq \alpha_o} R_x^\alpha(\tau) \langle \exp[j2\pi\alpha e(t)] \rangle_T \\ &\quad \times \exp[j2\pi(\alpha - \alpha_o)t] \left. \right\rangle_T. \end{aligned} \tag{18a}$$

in which the lag smoothing is negligible and the weighting function in the first term of the right member is a time-independent scalar β that can be re-expressed as

$\beta = \langle \cos(2\pi\alpha_o e(t)) \rangle_T + j \langle \sin(2\pi\alpha_o e(t)) \rangle_T$. It can be seen that, if $e(t)$ has an approximately even fraction-of-time amplitude density function, then the first term in β dominates the second term; the same result holds without this evenness assumption if the error is small enough, say $|e(t)| < 1/8|\alpha_o|$, and in this case the dominant term is close to 1. Therefore, the lower the cycle frequency α_o is, the larger the de-warping error that can be tolerated without significant attenuation of the actual cyclic autocorrelation, provided that $|\alpha_o e(t)| \ll 1$. In the event that $|\langle \dot{e}(t) \rangle_\tau| \ll 1$ is not satisfied, the first term in Eq. 18a is still a close approximation if $R_x^{\alpha_o}(\tau)$ varies very little over the range of $\langle \dot{e}(t) \rangle_\tau \tau$ for fixed τ ; this is satisfied, regardless of the size of the error $e(t)$, if $\tau = 0$ is selected for use.

For $\tau = 0$, Eq. 17a reduces to

$$E\left\{\hat{R}_{x_\varphi}^{\hat{\alpha}}(0)\right\} = R_x^\alpha(0) \left\langle \sum_\alpha \exp(j2\pi[(\alpha - \hat{\alpha})t + \alpha e(t)]) \right\rangle_T \tag{17c}$$

Generally speaking, it is to be expected that the more $\psi(t)$ or $\psi^{-1}(t)$ deviates from t (the stronger the warping or required de-warping), the larger $e_\varphi(t)$ and therefore $e(t)$ is likely to be and, as a consequence revealed by Eq. 17a, the weaker the cyclic autocorrelation of the de-warped data is likely to be. Interestingly, the scale for quantifying the size of the timing error $e(t)$ can be seen from the first term in Eq. 17b to be determined by the cycle frequency estimate $\hat{\alpha} = \alpha_o$.

For exact de-warping, $e(t) \equiv 0$, Eq. 17b reduces exactly to

$$\begin{aligned} E\left\{\hat{R}_{x_\varphi}^{\alpha_o}(\tau)\right\} &= R_x^{\alpha_o}(\tau) + \sum_{\alpha \neq \hat{\alpha}} R_{xx^{(*)}}^\alpha(\tau) \\ &\quad \times \langle \exp[j2\pi(\alpha - \alpha_o)t] \rangle_T \\ &= R_x^{\alpha_o}(\tau) + \sum_{\alpha \neq \hat{\alpha}} R_x^\alpha(\tau) \\ &\quad \times \frac{\sin(\pi(\alpha - \alpha_o)|T|)}{\pi(\alpha - \alpha_o)|T|} \\ &\quad \times \exp[j\pi(\alpha - \alpha_o)|T|] \end{aligned} \tag{18b}$$

and the expected value of the estimated cyclic autocorrelation is equal to the actual cyclic autocorrelation prior to time warping in $x(t)$ plus the *cycle leakage* term, which is inversely proportional to $\pi(\alpha - \alpha_o)|T|$. In the remaining sections of this paper, T will be used in place of $|T|$ to denote the length of the integration interval.

7 Basis-function expansion of de-warping function

To reduce the infinite-dimensional optimization problem in Eqs. 8 and 10 (searching over all functions $\varphi(t)$ defined on the time interval $[t_o, t_o + T]$ for any start time t_o), we can use the finite dimensional approximation

$$\varphi(t) = \sum_{k=1}^K a_k c_k(t) = \mathbf{a}^T \mathbf{c}(t) \triangleq \hat{\psi}^{-1}(t) \quad (19a)$$

where $(c_k(t); t \in [t_o, t_o + T])_{k=1}^K$ comprise a linearly independent set of functions chosen according to any available information about the time-warping function $\psi(t)$ or its inverse $\psi^{-1}(t)$, and where \mathbf{a}^T denotes the row vector obtained by matrix transposition of the column vector \mathbf{a} . For example, knowing nothing more than the spectral bandwidth of $\psi^{-1}(t)$, it is known that an optimum (minimum-dimension) set of basis functions that spans the space of all functions of duration no more than T and positive-frequency bandwidth of no more than $B_{\psi^{-1}}$ Hz consists of $K = 4B_{\psi^{-1}}T$ prolate spheroidal wave functions [15]. For all other sets of functions (that are not equivalent to this set in the sense of not being just K linearly independent linear combinations of the members of this set), larger values of K are required to span this same space.

If the functions used in Eq. 19a are chosen to be orthogonal, then the mean squared value of the error $e_\varphi(t) \triangleq \hat{\psi}^{-1}(t) - \psi^{-1}(t)$ is minimized by K independent minimizations of this same mean squared error w.r.t. to the K unknowns $\{a_k\}$ executed in any order. But this does not imply that the mean squared value of the error $e(t)$ defined by Eq. 14 behaves similarly. In general, a full joint minimization of this mean squared error w.r.t. all K unknowns $\{a_k\}$ must be executed. Yet, a perturbation analysis suggests that, for sufficiently small $e_\varphi(t)$, K independent minimizations of the mean squared value of the error $e(t)$ using orthogonal basis functions yields approximately the same result as a single joint minimization.

With no knowledge at all about $\psi^{-1}(t)$, except that it varies smoothly, harmonically related sinusoids or polynomials may be reasonable choices for $\{c_k(t)\}$.

However, since $\psi(t)$ will essentially always contain the additive term γt , where $\gamma = 1$ unless there exists a constant-velocity Doppler effect in the data $y(t)$, and since $\psi(t)$ will also contain the term ηt^2 if the data is affected by constant acceleration, then $\hat{\psi}^{-1}(t)$ may contain related terms like μt or $\nu t^{1/2}$. Consequently, even when sinusoids are used as basis functions, a more efficient approximate representation of $\psi^{-1}(t)$ may be obtained by adding a few such terms. For example, Eq. 19a can be replaced with

$$\begin{aligned} \varphi(t) &= \mu t + \nu t^{1/2} + \sum_{k=1}^K a_k c_k(t) \\ &= \mathbf{a}^T \mathbf{c}(t) \triangleq \hat{\psi}^{-1}(t) \end{aligned} \quad (19b)$$

where the vector \mathbf{a} has dimension $K + 2$ with the first two elements being $a_{-1} = \nu$ and $a_0 = \mu$. If $\nu = 0$, then

the dimension can be reduced to $K + 1$ and, if $\mu = 1$, then $a_1 = 1$ is fixed. In the case of Eq. 19b, $B_{\psi^{-1}}$ in the requirement $K \geq 4B_{\psi^{-1}}T$ might be taken to be the bandwidth of the component $\hat{\psi}^{-1}(t) - \mu t - \nu t^{1/2}$ of $\hat{\psi}^{-1}(t)$, which could be more well-defined than the bandwidth of $\hat{\psi}^{-1}(t)$.

Using Eq. 19, the set of equations Eq. 8 through Eq. 10 Eqs. 19a or 19b reduces to

$$\max_{\mathbf{a}} \left\{ \left| \hat{R}_{x_\varphi}^\alpha(\tau) \right|^2 \right\} \quad (20)$$

where

$$\hat{R}_{x_\varphi}^\alpha(\tau) = \left\langle y[\mathbf{a}^T \mathbf{c}(t + \tau)] y^{(*)}[\mathbf{a}^T \mathbf{c}(t)] \exp(-j2\pi \alpha t) \right\rangle_T \quad (21)$$

Once the optimum vector of coefficients $\mathbf{a} = \mathbf{a}_o$ is found from Eq. 8, the data transformation from Eq. 7,

$$\hat{x}(t) = y(\mathbf{a}_o^T \mathbf{c}(t)), \quad t \in [t_o, t_o + T], \quad (22)$$

approximately de-warps $y(t)$ to produce an approximation to the CS (or polyCS) data $x(t)$. Stated another way, the regularity of cyclicity of $y(\mathbf{a}_o^T \mathbf{c}(t))$ is higher than that of $y(t)$.

If the largest value of lag τ , at which the statistical dependence of $x(t)$ and $x^*(t - \tau)$ is not negligibly small, is denoted by τ_{\max} , then all pairs of lag products that are separated in time by at least $\tau_{\max} + \tau$ will be statistically independent. In this case, a generally applicable condition on the integration time T for obtaining a statistically reliable estimate of $R_x^\alpha(\tau)$ for all $|\tau| \leq \tau_{\max}$ is $\sqrt{T} \gg \sqrt{2\tau_{\max}}$. Here τ_{\max} upper bounds the coherence time of the data, but $\tau_{\max} + |\tau| \leq 2\tau_{\max}$ upper bounds the coherence time of the lag product of the data. In fact, the coefficient of variation of the estimate (the ratio of its standard deviation to the magnitude of its mean) is, under relatively broad conditions, roughly equal to $\sqrt{(\tau_{\max} + |\tau|)/T} \leq \sqrt{2\tau_{\max}/T}$. When possible, a value of $\sqrt{2\tau_{\max}/T}$ as small as 10% ($T = 200\tau_{\max}$) or even smaller is generally desirable; however, if the data with warped cyclicity (call it the signal) is corrupted by additive noise, with a *signal-to-noise ratio* (SNR) of average powers or mean squared values that is not sufficiently high, then T may need to be considerably larger.

The approximation $B_x \cong 1/\tau_{\max}$ is generally useful for the positive-frequency bandwidth of the power spectral density function of $x(t)$. (The exact relationship depends on the exact functional shape of the PSD and cyclic autocorrelation, and the particular definitions of width B_x, τ_{\max} .) Using this in the above

reliability condition, together with the accuracy condition $K \geq 4B_{\psi^{-1}}T$ (where, for Eq. 19b, $B_{\psi^{-1}}$ is the bandwidth of the component $\psi^{-1}(t) - \mu t - \nu t^{1/2}$), yields the alternative expression

$$K \geq 4B_{\psi^{-1}}T \gg 8B_{\psi^{-1}}/B_x \tag{23}$$

where for, high SNR, the symbol \gg as used here means at least 100 times greater as discussed above. The larger the ratio of bandwidths $B_{\psi^{-1}}/B_x$, the larger the number K of basis functions required to de-warp the data, unless the warping function is known except for the values of a “few” parameters, as illustrated below with two examples. To quantify \gg in relation Eq. 23 for SNR that is not high, one needs to know the value of SNR. This is addressed in Section 13.

For applications involving low-SNR data (e.g., SNR as low as 0 dB down to, say, -20 dB), which is one of the reasons CS processing is of interest [2, 3, 7, 8], one may need a time-bandwidth product $B_x T$ as large as, say, 10,000 to 1,000,000, instead of only 200 as in the case of high SNR.

As explained in Section 10, the iterative search algorithm proposed there is most practical when an analytical expression for the gradient vector of the objective function in Eq. 20 is available. Using standard differentiation methods for complex functions of a real variable, the following gradient expression can be derived from Eq. 21:

$$\nabla \left| \hat{R}_{x_\varphi}^\alpha(\tau) \right|^2 = 2 \operatorname{Re} \left[\hat{R}_{x_\varphi}^\alpha(\tau)^* \nabla \hat{R}_{x_\varphi}^\alpha(\tau) \right]. \tag{24}$$

For the case in which $\varphi(t)$ is given by Eqs. 19a or 19b, we have

$$\begin{aligned} \nabla \hat{R}_{x_\varphi}^\alpha(\tau) &= \langle \dot{y}[\mathbf{a}^T \mathbf{c}(t + \tau)] y^{(*)}[\mathbf{a}^T \mathbf{c}(t)] \\ &\quad \times \exp(-j2\pi a \tau) \mathbf{c}(t + \tau) \rangle_T \\ &+ \langle y[\mathbf{a}^T \mathbf{c}(t + \tau)] \dot{y}^{(*)}[\mathbf{a}^T \mathbf{c}(t)] \\ &\quad \times \exp(-j2\pi a \tau) \mathbf{c}(t) \rangle_T \end{aligned} \tag{25}$$

Equations 24 and 25 are valid as written as long as the vector \mathbf{a} is real valued, provided that one simply interprets the gradient symbol to mean the sum of the gradients of the real and imaginary parts of the function: $\nabla(\operatorname{re} + j\operatorname{im}) = \nabla \operatorname{re} + j\nabla \operatorname{im}$. However, when \mathbf{a} is complex valued (as it will be when $\mathbf{c}(t)$ is complex valued, e.g., complex sinusoids), the theory of complex gradient operators is needed to obtain the correct modification of these equations, cf. [16].

8 Inversion of warping functions

To obtain the best results, the choice for the basis functions $[c_k(t)]_{k=1}^K$ that is most efficient should be sought. For example, if the functional form of a time warp is known, it is sometime possible to deduce the functional form of its inverse. If the time-warp is a time-varying delay and/or advance, $\psi(t) = t + \delta(t)$ as it is when the warping is due to the Doppler effect resulting from possibly time-varying velocities and/or accelerations of data sensors and/or sources, then by using the definition $s \triangleq \psi(t)$, we obtain the equation $\psi^{-1}(s) = s - \delta(\psi^{-1}(s)) = t + \delta(t) - \delta(t) = t$ for the inverse. In some cases of practical interest, the equation $\psi^{-1}(s) = s - \delta(\psi^{-1}(s))$ can be solved for $\psi^{-1}(s)$. Some examples follow. But first, it is mentioned that the approximate inverse $\psi^{-1}(s) \cong s - \delta(s)$ can be quite accurate if the constraint $|\delta(t)/t| \ll 1$ is satisfied sufficiently strongly for the range of time values of interest.

For constant velocities of a source and sensor moving along a single straight line, $\delta(t) = a(t + b)$ and the solution to the above equation is $\psi^{-1}(s) = (s - ab)/(1 + a) = a'(s - b')$. But, more generally, constant velocities lead to a quadratic equation for $\delta(t)$ having 3 coefficients that are quadratic (or lower-order) functions of elapsed time, and quadratic (or lower-order) functionals of the velocity vectors and initial-position vectors. For constant accelerations $\delta(t)$ is the solution to a quartic equation, whose four coefficients are quartic (or lower-order) functions of elapsed time, and quadratic (or lower-order) functionals of the acceleration vectors and initial velocity and position vectors.

For the sake of simplicity in demonstrating exact inversion, a time advance that grows quadratically with time, $\delta(t) = at^2$ is considered here; then, we have

$$\begin{aligned} \psi^{-1}(s) &= s - \delta(\psi^{-1}(s)) \\ \psi^{-1}(s) &= s - a(\psi^{-1}(s))^2 \\ (\psi^{-1}(s))^2 &+ \frac{1}{a}\psi^{-1}(s) - \frac{s}{a} = 0 \end{aligned}$$

This quadratic equation has the solution

$$\begin{aligned} \psi^{-1}(s) &= -1/2a \pm \sqrt{1/4a^2 + s/a} \\ &= (1/2a)(\sqrt{1 + 4sa} - 1) \end{aligned}$$

which can be verified by substituting $s = t + at^2$. If we knew the time-warp had this functional form, we could simply use this form and maximize the objective function w.r.t. the single unknown parameter a . For example, the search algorithm presented in Section 10 could be used by calculating the gradient of the objective function w.r.t. a in $\hat{\psi}^{-1}$. Otherwise, we could choose a set of polynomials of order K or any other basis functions as in Eqs. 19–21.

As another example, if the time warping is believed to be a time-varying compression and/or expansion of time, $\psi(t) = t\varepsilon(t)$, say a linearly growing compression $\varepsilon(t) = at$, then we have $\psi(t) = at^2 = s$ and $\psi^{-1}(s) = \sqrt{s/a}$. Again, if this form were known, we could use it and maximize the objective function w.r.t. the unknown parameter a . Otherwise, we could choose a set of polynomials of order K or any other basis functions as in Eqs. 19–21.

9 Basis-function expansion of warping-compensation function

By analogy with the approach described in Section 7, we can use a finite dimensional approximation to reduce the infinite-dimensional optimization problem in Eqs. 8 and 12: searching over all functions $\varphi^{-1}(t)$ defined on the time interval $[\varphi^{-1}(t_o), \varphi^{-1}(t_o + T)]$ for any start time t_o . The analog of Eq. 19b is

$$\varphi^{-1}(t) = \gamma t + \eta t^2 + \sum_{k=1}^K b_k c_k(t) = \mathbf{b}^T \mathbf{c}(t) \triangleq \hat{\psi}(t). \tag{26}$$

Although the set of basis functions used in Eq. 19 for representing $\varphi(t)$ may very well be the same as those used in Eq. 26 for $\varphi^{-1}(t)$, they also can be different. If they are the same, the vectors of coefficients \mathbf{a} and \mathbf{b} will certainly differ from each other. It is expected that, when using the concepts in this paper, one would typically choose to use either Eq. 19 or Eq. 26, but not both except possibly for the objective of comparing these two approaches and selecting the one that seems to perform best according to any criteria the user may select.

Substituting Eq. 26 and its consequence $\dot{\varphi}^{-1}(t) = \mathbf{b}^T \dot{\mathbf{c}}(t)$ into Eq. 12 yields

$$\begin{aligned} \hat{R}_{x_\varphi}^\alpha(\tau) &= \frac{1}{T} \int_{\varphi(t_o)}^{\varphi(t_o+T)} y\left(\mathbf{u} + \Delta_\varphi^\tau[\varphi^{-1}(u)]\right) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\varphi^{-1}(u)] \dot{\varphi}^{-1}(u) du \\ &\cong \frac{1}{T} \int_{\varphi(t_o)}^{\varphi(t_o+T)} y(\tau/\dot{\varphi}^{-1}(u) + u) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\varphi^{-1}(u)] \dot{\varphi}^{-1}(u) du \\ &= \frac{1}{T} \int_{\varphi(t_o)}^{\varphi(t_o+T)} y(\tau/\mathbf{b}^T \dot{\mathbf{c}}(u) + u) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \mathbf{b}^T \dot{\mathbf{c}}(u) du \end{aligned} \tag{27a}$$

Under the condition that the interval of integration in Eq. 27a is not substantially different from $[t_o, t_o + T]$ on a percentage of overlap basis, Eq. 27a is usefully approximated by

$$\begin{aligned} \hat{R}_{x_\varphi}^\alpha(\tau) &\cong \frac{1}{T} \int_{t_o}^{t_o+T} y(\tau/\mathbf{b}^T \dot{\mathbf{c}}(u) + u) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \mathbf{b}^T \dot{\mathbf{c}}(u) du \end{aligned} \tag{27b}$$

which has the advantage of not requiring inversion of the function $\hat{\varphi}^{-1}(t) = \mathbf{b}^T \mathbf{c}(t)$ to obtain $\hat{\varphi}(t)$ at every search point \mathbf{b} . In fact, no function inversions are required by Eq. 27b. An example for which this condition leading to Eq. 27b might not be met is a substantial expansion or contraction of time over the entire interval $[t_o, t_o + T]$. But, even then, a substantial change in the length of the interval of integration need not have a significant impact on the gradient.

The gradient w.r.t. \mathbf{b} of the squared magnitude of the statistic Eq. 27 is given by

$$\nabla \left| \hat{R}_{x_\varphi}^\alpha(\tau) \right|^2 = 2 \operatorname{Re} \left[\hat{R}_{x_\varphi}^\alpha(\tau)^* \nabla \hat{R}_{x_\varphi}^\alpha(\tau) \right] \tag{28}$$

where

$$\begin{aligned} \nabla \hat{R}_{x_\varphi}^\alpha(\tau) &\cong \frac{1}{T} \int_{t_o}^{t_o+T} \nabla [y(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \mathbf{b}^T \dot{\mathbf{c}}(u)] du \end{aligned} \tag{29}$$

In Eq. 29, the integrand is given by

$$\begin{aligned} \nabla[\dots] &= \dot{y}(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) \tau \nabla(1/\mathbf{b}^T \dot{\mathbf{c}}(u)) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \mathbf{b}^T \dot{\mathbf{c}}(u) \\ &\quad - y(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) y^{(*)}(u) \mathbf{b}^T \dot{\mathbf{c}}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] j2\pi\alpha \mathbf{c}(u) \\ &\quad + y(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) y^{(*)}(u) \\ &\quad \times \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \dot{\mathbf{c}}(u), \end{aligned} \tag{30}$$

within which we can use the equation

$$\nabla(1/\mathbf{b}^T \dot{\mathbf{c}}(u)) = -\dot{\mathbf{c}}(u) [\mathbf{b}^T \dot{\mathbf{c}}(u)]^{-2}$$

to obtain the re-expression

$$\begin{aligned} \nabla \hat{R}_{x_\varphi}^\alpha(\tau) &\cong \frac{1}{T} \int_{t_o}^{t_o+T} \dot{y}(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) \\ &\quad \times y^{(*)}(u) \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \\ &\quad \times [-\dot{\mathbf{c}}(u) \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)] du \\ &\quad + \frac{1}{T} \int_{t_o}^{t_o+T} y(u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)) \\ &\quad \times y^{(*)}(u) \exp[-j2\pi\alpha\mathbf{b}^T \mathbf{c}(u)] \\ &\quad \times \{\dot{\mathbf{c}}(u) - j2\pi\alpha\mathbf{b}^T \dot{\mathbf{c}}(u) \mathbf{c}(u)\} du \end{aligned} \tag{31a}$$

For both approaches described in this Section 9 and Section 7, the continuous time data represented by the

mathematics here must be time-warped or de-warped to achieve the indicated time-warping compensation in Eq. 31a or de-warping in Eq. 25 in order to evaluate the gradient at every iteration of the iterative gradient-ascent search method described in Section 10. Assuming that time-sampled data is used in the search algorithm for optimization described in the Section 10, the re-warping is implemented by re-interpolating the data. The computation and storage cost here depend significantly on the amount of data being processed. Using the minimum time-sampling rate $4B_x$, that avoids overlap of both aliased cycles and aliased spectral components, and a time-bandwidth product of $B_x T = 200$, we generally need about 800 time samples. But for low SNR (e.g., 0 dB or less), the number needed can be considerably larger (see Eq. 42).

Because $K \geq 4B_\psi T$, where B_ψ is the bandwidth of the component $\psi(t) - \gamma t - \eta t^2$ of $\psi(t)$, the stronger the inequality $T \gg 2/B_x$ that is required due to low SNR, the larger K , the number of basis functions, must be: $K \geq 4B_\psi T \gg 8B_\psi/B_x$. This is the counterpart, for warping compensation, to the requirement Eq. 23 for de-warping.

If the computational cost of the data interpolation is excessive, the following approach to circumventing data interpolation altogether should be considered. The term $u + \tau/\mathbf{b}^T \dot{\mathbf{c}}(u)$ in Eqs. 27–31a represents the only time-interpolation of the data that is required when using the warping-compensation method. When the cyclic autocorrelation of $x(t)$ is non-negligible at zero lag $\tau = 0$ (which is where the maximum commonly occurs), one can choose to use only this value of lag, in which case $\tau/\mathbf{b}^T \dot{\mathbf{c}}(u) = 0$ and the need for data interpolation vanishes. As one example, applications where $\tau = 0$ actually produces a maximum cyclic autocorrelation magnitude include those for which the signal consists of a periodic pulse train with random amplitudes and (i) flat pulses having width $\leq T_o/2$ (e.g., Eq. 43 with all $h_n = 1$ and $p(t)$ equal to a rectangle of width $W \leq T_o/2$), or (ii) arbitrary pulses having non-negative Fourier transforms. In contrast, for flat pulses having width $W > T_o/2$ (case (iii)), the value of the cyclic autocorrelation of $x(t)$ at $\tau = 0$ approaches zero as $W \rightarrow T_o$.

For $\tau = 0$, Eq. 31a reduces to

$$\begin{aligned} \nabla \hat{R}_{x_p}^\alpha(\tau) &\cong \frac{1}{T} \int_{t_o}^{t_o+T} y(u)y^*(u) \\ &\times \exp\{-j2\pi\mathbf{a}\mathbf{b}^T \mathbf{c}(u)\} \\ &\times [\dot{\mathbf{c}}(u) - j2\pi\mathbf{a}\mathbf{b}^T \dot{\mathbf{c}}(u)\mathbf{c}(u)] du. \end{aligned} \tag{31b}$$

In applications where the cyclic autocorrelation at zero lag is too weak to be used for warping compensation (e.g., case (iii) in the example above, with W close to T_o), but is sufficiently strong at one or more non-zero values of lag

(e.g., $\tau = T_o/2$ in the same example with $W = T_o$) for which the time-varying lag $\tau/\mathbf{b}^T \dot{\mathbf{c}}(u)$ has dynamic range over $u \in [t_o, t_o + T]$ that is small compared with some representative value, say τ/g , one can consider replacing $\tau/\mathbf{b}^T \dot{\mathbf{c}}(u)$ with the constant τ/g , rounded off to the nearest integer multiple of the data-sampling time-increment, which eliminates the need for data interpolation. For example, if $\mathbf{b}^T \mathbf{c}(u) = gu + \varepsilon(u)$ and $|\dot{\varepsilon}(t)| \ll g$, then $\mathbf{b}^T \dot{\mathbf{c}}(u) \cong g$. In this case, Eq. 31a reduces to the close approximation:

$$\begin{aligned} \nabla \hat{R}_{x_p}^\alpha(\tau) &\cong \frac{-\tau}{gT} \int_{t_o}^{t_o+T} \dot{y}(u + \tau/g)y^*(u) \\ &\times \exp[-j2\pi\mathbf{a}\mathbf{b}^T \mathbf{c}(u)] \dot{\mathbf{c}}(u) du \\ &+ \frac{1}{T} \int_{t_o}^{t_o+T} y(u + \tau/g)y^*(u) \\ &\times \exp[-j2\pi\mathbf{a}\mathbf{b}^T \mathbf{c}(u)] \\ &\times [\dot{\mathbf{c}}(u) - j2\pi\mathbf{a}\mathbf{g}\mathbf{c}(u)] du. \end{aligned} \tag{31c}$$

As another simplifying approximation, if the time-warping function varies slowly enough, relative to the cycle frequency, $|\dot{\mathbf{c}}(u)| \ll 2\pi|\alpha\mathbf{g}\mathbf{c}(u)|$, then the $\dot{\mathbf{c}}(u)$ term in Eq. 31b can possibly be deleted without having much impact on the gradient.

10 Iterative gradient-ascent search algorithm

Let $J(\mathbf{a})$ be an objective function—such as that in Eqs. 8 and 10, or Eqs. 20 and 21, or Eqs. 20 and 27—that is *maximized* when some vector of parameters such as \mathbf{a} in Eq. 21 or \mathbf{b} in Eq. 27, generically denoted by \mathbf{a} in this section, take on optimum values. There are many varied options for optimization algorithms. Solely for purposes of illustration, the algorithm selected here to optimize \mathbf{a} is Cauchy’s iterative gradient-ascent search algorithm, modified by use of an alternative step-size sequence: the candidate solution $\hat{\mathbf{a}}^{(i)}$ at iteration i is updated to obtain the next candidate solution $\hat{\mathbf{a}}^{(i+1)}$ at iteration $i + 1$ by moving from the i th candidate location in the direction of steepest ascent on the hypersurface $J(\mathbf{a})$ (when the step-size parameter is positive), which is the direction of the gradient vector evaluated at the i th location:

$$\hat{\mathbf{a}}^{(i+1)} = \hat{\mathbf{a}}^{(i)} + \mu(i) \nabla J(\hat{\mathbf{a}}^{(i)}) \tag{32}$$

where $\nabla J(\mathbf{a}) = [\partial J(\mathbf{a})/\partial a_1, \partial J(\mathbf{a})/\partial a_2, \dots, \partial J(\mathbf{a})/\partial a_K]^T$ is the gradient vector and $\mu(i)$ is the step size at iteration i . The following formula for the step-size sequence is one of many that have been proposed for iterative gradient-ascent algorithms:

$$\mu(i) = - \frac{[\hat{\mathbf{a}}^{(i)} - \hat{\mathbf{a}}^{(i-1)}]^T [\nabla J(\hat{\mathbf{a}}^{(i)}) - \nabla J(\hat{\mathbf{a}}^{(i-1)})]}{\|\nabla J(\hat{\mathbf{a}}^{(i)}) - \nabla J(\hat{\mathbf{a}}^{(i-1)})\|^2}. \quad (33a)$$

It has been reported to produce rapid convergence in a computationally efficient manner [17]. This step size can be re-expressed as

$$\mu(i) = \mu(i-1) \frac{1}{\|\nabla J(\hat{\mathbf{a}}^{(i)}) - \nabla J(\hat{\mathbf{a}}^{(i-1)})\|} \times \frac{\nabla J(\hat{\mathbf{a}}^{(i-1)})^T [\nabla J(\hat{\mathbf{a}}^{(i-1)}) - \nabla J(\hat{\mathbf{a}}^{(i)})]}{\|\nabla J(\hat{\mathbf{a}}^{(i-1)}) - \nabla J(\hat{\mathbf{a}}^{(i)})\|} \quad (33b)$$

from which it can be seen that the current step size is a scaled version of the previous step size, and the scalar consists of two factors: (i) the reciprocal of the size of the change in the gradient vector from previous to current step and (ii) the inner product between the previous gradient vector and the unit vector with direction equal to that of the difference between the previous and current gradient vectors. Consequently, if there is a large change in the gradient vector, then the first factor scales down the current step size from the previous one. The second factor can both scale the step size up or down and change its sign, which happens when the current gradient is larger and oppositely directed relative to the previous gradient. This possibility of occasional gradient descent, instead of consistent gradient ascent, means that Eq. 32, with step-size sequence Eq. 33, is generally a gradient-ascent algorithm but can occasionally produce a gradient descent to find a better point from which to resume ascending.

An alternative to the single-formula approach in Eq. 33 to specifying a step-size sequence, which has been said to better accommodate troublesome surfaces, is to alternate between two or more formulas for the step-size sequence [18].

The step-size algorithm Eq. 33 was used in the simulations reported in Section 14 simply as an example. The search algorithm Eqs. 32–33, like many search algorithms, may be challenged by either multimodal objective functions containing local peaks above hyperplanes of any dimension M from 1 to N and/or otherwise troublesome surfaces, such as those containing long narrow ridges of any dimension M from 2 to N . In some such cases, another search technique may be used in concert with Eqs. 32 and 33. There are many varied options among which are those that simply select multiple starting locations $\{\hat{\mathbf{a}}_n^{(0)} : n = 1, 2, \dots, N\}$ on a possibly

uniform grid of points covering what is considered to be the smallest admissible region of the search space and run the iteration Eqs. 32 and 33 for each and every starting location, and then simply compare all the local maxima found and select the largest one. This is called “brute-force” or “exhaustive” initialization.

11 Synchronization-sequence methods

Two methods are described here, one for initializing the iterative gradient-ascent algorithm described in Section 10 and another for replacing that algorithm with an elegant two-step method.

Another approach to initializing the iterative gradient-ascent method of Section 10, which can be used instead of the brute-force method, is as follows. A *synchronization sequence* is defined to be a point process for which the time locations of the points are the times at which some particular repeating feature in the data exhibiting irregular cyclicity is detected. The feature must be one that occurs, for the most part, once every cycle of the data exhibiting irregular cyclicity, and the phase at which it occurs in the cycle (the fraction of the cycle period that has elapsed when the feature occurs) must be, for the most part, approximately constant for best results. The assumed identifiability of such features limits applications of this method, but its simplicity merits its mention here.

Examples of such features include peaks, doublets (a positive (or negative) peak followed by a negative (or positive) peak), an end point of a “quiet subinterval” or “dead zone,” and the start time of a decaying oscillatory burst. For example, such a feature might be detectable in an EKG from one pair of sensors with a particular placement on the body while the data from another pair of sensors at some other particular placement is that which is to be analyzed. Or, the synchronization sequence could simply be the time of sunrise and/or sunset or other observable cyclic event like the time of full eclipse of Sun by Earth's Moon as observed from some specified location on Earth; or the time of onset of some specific easily identifiable phase of a cyclic chemical process or any other process that is cyclic. As discussed below, the more features per cycle that can be identified, the better.

One can synthesize a smooth de-warping function $\hat{\psi}^{-1}(t)$ from this point process. One approach, for the case in which the cycle frequency is unknown and there is one feature per cycle being detected is to first solve for the best-fitting set of equally spaced time points $\{\delta + nT_o\}_{n=1}^N$ to the measured set of unequally spaced points $\{t_n\}_{n=1}^N$. If the sum of squared differences between the values of these two sets of points is minimized w.r.t. the fixed time-separation value (period) T_o and the timing offset δ ,

$$\min_{\delta, T_o} \sum_{n=1}^N [\delta + nT_o - t_n]^2, \tag{34a}$$

then the optimum values are given by

$$\begin{aligned} \hat{T}_o &= \frac{1}{N} \sum_{n=1}^N n \left(t_n - \hat{\delta} \right) \left(\frac{1}{N} \sum_{n=1}^N n^2 \right)^{-1} \\ \hat{\delta} &= \frac{1}{N} \sum_{n=1}^N (t_n - n\hat{T}_o) \end{aligned} \tag{34b}$$

The implicit form of Eq. 34b is interesting because the solution is fully described in terms of temporal averages and correlations of the time-series $\{t_n\}$ and $\{n\}$, and the mean square value of $\{n\}$; but, because these two simultaneous equations are linear, they can easily be explicitly solved for each of $\hat{\delta}$ and \hat{T}_o .

As an aside, this is a possible method for initializing the unknown cycle frequency, $\hat{\alpha} = 1/\hat{T}_o$ or, possibly $\hat{\alpha} = 2/\hat{T}_o$ even if some other approach to initializing the search over the coefficient vector \mathbf{a} is to be used.

As a second step, one can find a smooth de-warping function $\hat{\psi}^{-1}$ that at least approximately satisfies $\hat{\psi}^{-1}(\hat{\delta} + n\hat{T}_o) = t_n, n = 1, 2, \dots, N$. This could be done by simply doing a least squares fit of the linear combination of the basis functions in Eq. 19 to the point process at the N measured time points:

$$\min_{\mathbf{a}} \sum_{n=1}^N \left[\hat{\psi}^{-1}(\hat{\delta} + n\hat{T}_o) - t_n \right]^2 \tag{34c}$$

Being a linear least squares problem, the solution is obtained by simply solving a set of N linear equations in K unknowns, \mathbf{a} , obtained by equating to zero the gradient of the quadratic objective function in Eq. 34c, with $\hat{\psi}^{-1}(t) = \mathbf{a}^T \mathbf{c}(t)$ substituted in. This approach cannot be expected to perform well if N is not large enough compared with K . However, the required value of K typically increases with increasing length T of data record, so the ratio N/K cannot be increased simply by increasing T . Instead the average number of features identified per interval of length \hat{T}_o must be increased, if possible.

If there are M identifiable features per cycle, then we can simply replace the single objective function in Eq. 34a with a sum of M such functions, each with its own timing offset $\delta^{(m)}$ and feature-occurrence times $t_n^{(m)}$ and all sharing the same average period T_o :

$$\min_{\delta, T_o} \sum_{m=1}^M \sum_{n=1}^N \left[\delta^{(m)} + nT_o - t_n^{(m)} \right]^2. \tag{35a}$$

In this case, the solution Eq. 34b is replaced with

$$\begin{aligned} \hat{T}_o &= \frac{1}{M} \sum_{m=1}^M \left\{ \sum_{n=1}^N n \left(t_n^{(m)} - \hat{\delta}^{(m)} \right) \right\} \left(\sum_{n=1}^N n^2 \right)^{-1} \\ \hat{\delta}^{(m)} &= \frac{1}{N} \sum_{n=1}^N \left(t_n^{(m)} - n\hat{T}_o \right), \quad m = 1, 2, 3, \dots, M. \end{aligned} \tag{35b}$$

Similarly, Eq. 34c becomes

$$\min_{\mathbf{a}} \sum_{m=1}^M \sum_{n=1}^N \left[\hat{\psi}^{-1}(\hat{\delta}^{(m)} + n\hat{T}_o) - t_n^{(m)} \right]^2 \tag{35c}$$

The performance can be expected to improve as the inequality $NM/K \gg 1$ is strengthened. An example of $M = 3$ is the well-known QRS complex in an EKG signal, which is repeated with every heartbeat. As in the case of the implicit solution Eq. 34b, Eq. 35b is a set of $M + 1$ simultaneous linear equations that are fully described in terms of temporal averages and correlations of $\{t_n^{(m)}\}$ and $\{n\}$, and the mean square value of $\{n\}$, and that are easily solved explicitly.

Besides using synchronization sequences $\{t_n^{(m)}\}_{n,m=1}^{N,M}$ for initialization of the iterative algorithm of Section 10, the initial time de-warping candidate $\hat{\psi}^{-1}(t)$ specified by Eq. 34c or Eq. 35c also can possibly be used to assess which of the conditions on $\psi(t)$, specified in this paper for validating approximations that are to be used in designing the search algorithm and predicting performance, are satisfied—although the solution to Eq. 36 below might be better suited.

The second method based upon synchronization sequences is revealed by the suggestion that the function $\hat{\psi}^{-1}(t)$ found in this elegant manner could conceivably be accurate enough to serve as the final de-warping function. Similarly, a warping function can be found by replacing Eq. 34c with Eq. 36

$$\min_{\mathbf{b}} \sum_{n=1}^N \left[\hat{\psi}(t_n) - n\hat{T}_o - \hat{\delta} \right]^2 \tag{36}$$

or its generalization for $M > 1$. When this approach is used, performance might be improved by identifying and discarding outliers in the data set $\{t_n^{(m)}\}$.

Unfortunately, there are many applications in which repeating features are hidden in erratic background fluctuations, preventing identification of such features and ruling out the methods of this section.

12 Pace irregularity

As inferred in preceding sections, de-warping or warping compensation may not be possible for severe cases of

time warping. In addition, there also are types of irregular cyclicity that are not due to time warping or cannot be modeled as time-warped regular cyclicity. In some such cases, it is possible that none of the methods described in this and preceding sections would be able to adequately reduce the irregularity of cyclicity. To illustrate, the following baseline signal model is considered:

$$y(t) = x(\psi(t)) = \sum_n q_n(\psi(t) - nT_o)$$

where the pulses $\{q_n(t)\}$ are random, independent, and identically distributed, and the synchronization times $\{t_n\}$ for some repeating feature in $y(t)$ or in the probability density function for the random process $y(t)$, due to the behavior of $q_n(t)$ at say t_* , satisfy $\psi(t_n) = t_* + nT_o$. This is expected to be a useful model for electrocardiograms with time-varying heart rate and corresponding time-varying width of the pulse complex within each beat. In general, signals of this type can be de-warped using the methods discussed in preceding sections, since

$$y(\psi^{-1}(t)) = x(\psi[\psi^{-1}(t)]) = \sum_n q_n(t - nT_o)$$

which is cyclostationary. However, if the baseline signal model is changed to the following *pace-irregular model*

$$y(t) = \sum_n q_n(t - t_n); \quad t_n = \psi^{-1}(2\pi n + \theta_*), \quad (37a)$$

then attempting to de-warp (using the symbol θ , instead of s as in Section 8, for $\psi(t)$) yields

$$\begin{aligned} x(\theta) &\triangleq y(\psi^{-1}(\theta)) \\ &= \sum_n q_n[\psi^{-1}(\theta) - \psi^{-1}(2\pi n + \theta_*)] \end{aligned} \quad (37b)$$

which is not cyclostationary as a function of θ unless the time-warping function is linear, $\psi(t) = \omega t$ for some constant ω , in which case the original data $y(t)$ also is cyclostationary.

The pace-irregular model Eq. 37a, and associated *rotation-angle* model (with unwrapped angle θ), as Eq. 37b shall be called, is useful for random vibrations $y(t)$ from rotating machinery with a fault point in some rotating component. In this model, the pulses or bursts $\{q_n(t)\}$ represent the machine structure's vibration response (typically damped ringing), which is modeled as independent of the times of occurrence of the causative impulsive shocks from the rotating fault. The advancing rotation angle can be expressed in terms of instantaneous frequency as follows:

$$\theta(t) = \int_{t_o}^t \omega(u) du + \theta(t_o).$$

The shapes of the bursts $\{q_n(t)\}$ are not affected in the model Eq. 37a by the warping which is determined by $\omega(t)$. Only their occurrence times t_n are affected. In fact, these occurrence times can be interpreted as warped versions of the equally spaced (unwrapped) angles $\{2\pi n + \theta_*\}$ at which the fault excites the system: $t_n = \psi^{-1}(\theta_n) \triangleq \psi^{-1}(2\pi n + \theta_*)$ where ψ^{-1} (not ψ) is the warping function, and $\psi(t) \equiv \theta(t)$ is the de-warping function.

Given $\omega(t)$ and using $\psi(t) = \theta(t)$, the warping function $\psi^{-1}(\theta) = t$ can, in principle, be solved for. For example, for $\omega = at$, the equation $\psi(t) = \theta(t) = (1/2)at^2 + \theta(t_o)$ can be solved to obtain $\psi^{-1}(\theta) = \sqrt{(2/a)[\theta - \theta(t_o)]}$. Or, for $\omega = \exp(at)$, $\psi^{-1}(\theta) = (1/a) \ln(a[\theta - \theta(t_o)])$.

When the speed of rotation $\omega = d\theta(t)/dt$ is constant (pace is regular), both $y(t)$ and $x(\theta)$ are cyclostationary. But when the speed changes with time, neither of these signals are cyclostationary! Moreover, there is no de-warping function that will render either of $y(t)$ or $x(\theta)$ cyclostationary! Nevertheless, depending on the shapes of the bursts $\{q_n(t)\}$, if they do not overlap each other too much, it may be possible to measure $\{t_n\}$, fit an estimated period \hat{T}_o to these measurements $\{\hat{t}_n\}$ (see Section 11), and then time-shift the individual bursts from $\{\hat{t}_n\}$ to $\{n\hat{T}_o\}$. If there were no errors in $\{\hat{t}_n\}$, then the burst-shifted signal would be cyclostationary. Also, if the errors in $\{\hat{t}_n\}$ are independent and identically distributed random variables, the burst-shifted signal would still be cyclostationary, but with lower degree of cyclostationarity (see Section 13). If the errors in $\{\hat{t}_n\}$ are small, this procedure can substantially increase the CCT (decrease the ISC) of the data, whether or not the burst-shifted signal is exactly cyclostationary.

More generally, for an irregular-paced signal model such as Eq. 37a, time warping as in Eq. 37b affects the pacemaker's rate and, as desired, can convert the irregular pulse-times of occurrence to regular pulse-angles (in the case of rotating machinery) of occurrence $\{\theta_n\} = \{2\pi n + \theta_*\}$; but, Eq. 37b also reveals that this time warping also affects the time scale of the individual paced pulses or bursts $\{q_n(t)\} = \{q_n(\psi^{-1}(\theta))\} \triangleq \{\tilde{q}_n(\theta)\}$ and the warped pulses $\{\tilde{q}_n(\theta)\}$ are no longer identically distributed. Consequently, neither the irregular-paced signal nor the time-warped regular-paced signal are cyclostationary. This is particularly important to the study of rotating machinery vibrations when the RPM (meaning revolutions per minute or rotations per minute) varies with time too fast to be treated in the data analysis as locally constant (meaning all vibration transients—e.g., from machine faults—have died away before the RPM changes substantially, in which case *quasi-static approximations* can yield accurate results).

It follows that the recently proposed *angle-time (two-variable) model* for vibration signals from rotating machinery [19] cannot be cyclostationary in angle or time unless the RPM is constant or only slowly varying. But, if it is constant, then nothing is gained from the angle-time model because angle is proportional to time and, consequently, the proposed *order* is just the scaled classical cycle frequency and the proposed *order-frequency spectral correlation function* [19] is just the classical spectral correlation function, with a scaled cycle frequency.

Furthermore, it is well known that time averages cannot accurately approximate expected values for nonstationary signals that are not either slowly nonstationary or CS or polyCS or almost-CS. Therefore, the proposed angle-time model, which is a generally nonstationary stochastic process (generally not stationary or CS or polyCS or even time-warped CS), for rotating machinery with non-slow variation in RPM, does not provide a basis for a probabilistic theory of empirical signal processing because the probabilistic parameters cannot be estimated using empirical time averages and cannot provide a viable avenue for extending the theory of CS to vibration signals from rotating machinery with rapidly changing RPM. Similarly, the related concept of a *cyclo-nonstationary* signal model [19], besides being burdened with an unfortunate name, cannot do any better because, again, the idealized-model characteristics obtained from expected values cannot be accurately approximated with empirical time averages. Any attempt to develop a sound probabilistic theory for the time-average processing of empirical data for such nonstationary processes is destined to fail. This is not a new result [20], but it is a result that should be more broadly known and understood by signal processing researchers wishing to add a novel theoretical flavor to their work:

The parameters in generally nonstationary stochastic-process models cannot be accurately estimated using time-averages [20].

But there is a way to obtain a vibration signal with rapidly time-varying RPM that is polyCS: If the variation of RPM with time is periodic, then the nonstationarity can be cyclic—the vibration signal can be polyCS with a period of CS that is varied periodically as in a frequency-modulated sine wave with a periodically time-varying frequency.

This provides motivation for machine testing with RPM intentionally varied periodically; and it is hereby suggested that such *periodic-RPM test protocols* be investigated for purposes of exciting machine faults in a manner that enhances their detectability.

13 Design guidelines

Two key parameters in the basis-function approach to the optimization problem described in Sections 4, 5, 6, 7, 8, 9, and 10 are the integration time T and the model order (number of basis functions) K . The parameter T must be large enough to suppress additive noise that corrupts the signal and to adequately average the erratic (or random) fluctuations in the signal itself in order to reveal the cyclic autocorrelation present. At the same time, this parameter must be no larger than necessary to keep the number of basis functions K required in the model of the time-warping function down to a minimum for those applications where the warping function is not known to within the unknown values of a small fixed set of parameters. As explained in Sections 7 and 8, Eq. 23 ideally needs to be satisfied for the de-warping method described in Section 7, and the counterpart of Eq. 23, in which ψ^{-1} is replaced with ψ ideally needs to be satisfied for the warping-compensation method described in Section 9. But just how large does the ratio of the LHS to the far-RHS of Eq. 23 need to be in order to satisfy the “much greater than” requirement? One approach to answering this question is presented here.

Qualitatively, the lower the *degree of cyclostationarity (DCS)* of the measured data after ideal de-warping is, the longer the integration time must be. The following results can be straightforwardly obtained from the fundamentals of the theory of CS developed in [2]. The DCS of the ideally de-warped measured noisy data $\tilde{x}(t) \triangleq y(\psi^{-1}(t))$, where $y(t) = s(\psi(t)) + n(t)$ from which $\tilde{x}(t) = s(\psi[\psi^{-1}(t)]) + n(\psi^{-1}(t)) = s(t) + \tilde{n}(t)$, is defined to be the complex-valued *correlation coefficient* (this is also called the *cyclic correlation coefficient* and is one of several definitions of DCS—the utility of each definition is strongly application-dependent (see [21])—

$$DCS^\alpha(\tau) = \frac{|R_{\tilde{x}\tilde{x}}^\alpha(\tau)|}{R_{\tilde{x}\tilde{x}}^\alpha(0)} \tag{38}$$

(usage of the shorthand $R_x^\alpha(\cdot) \equiv R_{xx^{(*)}}^\alpha(\cdot)$ from above continues here and below). It can be shown that

$$DCS^\alpha(\tau) = |\rho_s^\alpha(\tau)| \frac{SNR}{1 + SNR} \tag{39}$$

where $SNR \triangleq R_s(0)/R_{\tilde{n}}(0)$ is the ratio of mean squared signal to mean squared noise, and $\rho_s^\alpha(\tau) \triangleq R_{ss^{(*)}}^\alpha(\tau)/R_{ss^{(*)}}^\alpha(0)$ is the cyclic correlation coefficient (assuming the signal has zero mean value or, equivalently, contains no finite-strength additive sine-wave components), which has magnitude less than or equal to unity.

A key objective to be met for the optimization to be effective is to ensure that the *coefficient of variation (CV)* for the estimated cyclic autocorrelation for the de-warped data is small compared with unity, say 1/10 as an example target:

$$CV \triangleq \frac{\sqrt{\text{var}\{\hat{R}_x^\alpha(\tau)\}}}{\left|\text{mean}\{\hat{R}_x^\alpha(\tau)\right|} < 1/10 \quad (40)$$

The CV can be shown [2, 5] to be approximated by

$$CV \cong \frac{1}{\sqrt{B_s T}} \left(\frac{1}{DCS^\alpha(\tau)} \right). \quad (41)$$

Substituting Eq. 39 into Eq. 41 transforms the requirement in Eq. 40 into

$$CV \cong \frac{1}{\sqrt{B_s T}} \left(\frac{1}{|\rho_s^\alpha(\tau)|} \right) \left(\frac{1 + \text{SNR}}{\text{SNR}} \right) < 1/10$$

which can re-expressed as

$$T > \frac{100}{B_s} \left(\frac{1}{|\rho_s^\alpha(\tau)|^2} \right) \left(\frac{1 + \text{SNR}}{\text{SNR}} \right)^2. \quad (42)$$

Because the order of the time-warping model, which is the number of basis functions K , required to obtain a good fit to the time de-warping function component $\psi^{-1}(t) - \mu t - \nu t^{1/2}$ can be no smaller than $4B_{\psi^{-1}}T$ (assuming the basis spans the set of signals having duration T and bandwidth $B_{\psi^{-1}}$), K is, from Eq. 42, lower bounded by

$$K \geq 4B_{\psi^{-1}}T > 400 \frac{B_{\psi^{-1}}}{B_s} \left(\frac{1}{|\rho_s^\alpha(\tau)|^2} \right) \left(\frac{1 + \text{SNR}}{\text{SNR}} \right)^2 \quad (43)$$

Equality in Eq. 43 might well suffice when the optimum (most efficient) set of basis functions, which is the smallest set that spans the space of all possible time-warping or warp-compensating functions, is used. For example, for the space defined by a maximum possible time-bandwidth product (e.g., not exceeding $B_{\psi^{-1}}T$), the optimum set of basis functions are the prolate spheroidal wave functions [15]. For any other basis, K will generally need to be larger than $4B_{\psi^{-1}}T$ for this particular space of possible warping functions. It follows from Eq. 43 that the smallest the model order can be is the minimum of 1 and $400 B_{\psi^{-1}}/B_s$. Thus, we have the following guideline:

The smaller the ratio of the bandwidth of the de-warping function to the bandwidth of the signal is, the less computationally costly the optimization can be.

One would hope for a bandwidth ratio no larger than about 5% to 10%, in which case K need not

exceed 20 to 40 (for the above selected example target value of CV).

The *averaging time requirement* in Eq. 42 applies generally to both the *cyclicity-restoral and cyclicity-irregularity-compensation* approaches introduced in Sections 4, 5, 6, 7, 8, and 9. However, the *model-order-requirement* in Eq. 43 applies only to the former (de-warping) method of Section 7. For the latter (warping compensation) method of Section 9, the condition $K \geq 4B_{\psi^{-1}}T$ used to derive Eq. 43 must be replaced with $K \geq 4B_{\psi}T$, in which case, we obtain

$$K \geq 4B_{\psi}T > 400 \frac{B_{\psi}}{B_s} \left(\frac{1}{|\rho_s^\alpha(\tau)|^2} \right) \left(\frac{1 + \text{SNR}}{\text{SNR}} \right)^2. \quad (44)$$

Generally speaking, there does not appear to be any specific relationship between the bandwidths of a function $\psi(t)$ and its inverse $\psi^{-1}(t)$. For this reason and others, there is no general guideline presented here for determining which of the methods of Sections 7 and 8 is more efficient (requires smaller model order K) or is otherwise superior for a particular application. Nevertheless, one can, in principle, apply both methods for multiple values of K and/or different sets of basis functions and compare results. Two of various optional criteria for selecting the “best” method are (i) that which produces the smaller CV for a specified model order and (ii) that which produces the smaller model order for a specified CV. (It is not generally true that if the specified CV is smaller than but close enough to the smallest [over these two methods] CV for a specified K , then the smallest K [over these two methods] for this specified CV will be achieved with the same method.)

It is emphasized that the condition on averaging time given in Eq. 42 is necessary and sufficient for obtaining a low value for the CV, and the conditions on model order given in Eqs. 43 or 44 are generally necessary for obtaining a close approximation to ψ^{-1} or ψ in cases for which there is no knowledge of the functional form of these functions. As explained in Section 6, at least partly, the more $\psi(t)$ or $\psi^{-1}(t)$ deviates from t or more generally from any trend such as $ct + dt^2$ or $gt + ht^{1/2}$ (i.e., the stronger the warping or required de-warping), the weaker the cyclic autocorrelation of the data de-warped with a candidate estimate of $\psi^{-1}(t)$ is likely to be. And the weaker it is, the poorer the quality of the optimization of \mathbf{a} based on maximizing it is likely to be.

The extent of deviation of $\psi(t)$ or $\psi^{-1}(t)$ from a trend is for the most part captured by the bandwidths of these deviations, denoted by B_{ψ} and $B_{\psi^{-1}}$. Consequently, Eqs. 43 and 44 capture for the most part the impact of these deviations on the required values for the parameters T and K . For example, scaling the derivative of the deviation $\Delta_{\psi}(t) \triangleq \psi(t) - \text{trend}$ (without changing the

range of $\Delta_\psi(t)$, which requires time compression or expansion, scales its bandwidth B_ψ by the same amount: $d\Delta_\psi(\beta t)/dt = \beta(d\Delta_\psi(t)/dt)$ and $B_{\psi(\beta t)} = \beta B_{\psi(t)}$.

Due in part to the fact that there is generally no amount of data (integration time T) and/or model order K that will produce an arbitrarily small specified CV , we have the following guideline:

Some amount of experimentation with parameter values, T and K , and basis functions will typically be required to obtain the best attainable results.

The performance of any particular method applied to any particular data can generally be expected to exhibit a minimum CV , for any specified set of basis functions and any available computational precision, at some particular pair of values for T and K (unless the warping function is of known form with a small fixed number of unknown parameters, regardless of the value of T —in which case, the larger T is, the smaller the CV will generally be); values larger than these optimum values for T and K can be expected to degrade performance.

Summarizing, the integration time T required is dictated largely by the condition given by Eq. 42, which is independent of the order K of the model $\hat{\psi}(t)$ for $\psi(t)$. On the other hand, the model order required, for a good fit of $\hat{\psi}(t)$ to $\psi(t)$, with no knowledge about $\psi(t)$ other than its bandwidth, is dictated by the requirement $K \geq 4B_\psi T$. Given a value of T large enough to produce a sufficiently small CV when the exact $\psi^{-1}(t)$ or $\psi(t)$ is used for de-warping or warping compensation (call this a target value of CV), the choice for a value of K affects primarily how well the model $\hat{\psi}(t)$ or $\hat{\psi}^{-1}(t)$ can fit $\psi^{-1}(t)$ or $\psi(t)$ and therefore, how close to the target value of CV the actual value is. The value of K can, in principle, be chosen as large as needed, which is generally dictated by T , to obtain any desired precision of fit. The impact of large values of K is the effect it has on the convergence of the iterative search algorithm described in Section 10 or the accuracy of the model fit using any method for optimizing the vector of parameters **a** or **b**. This is a numerical issue, unlike the statistical issue of reliability, characterized here in terms of the CV . While the performance characteristics described in this section are promising in terms of the suggested apparent breadth of applicability of the methods introduced here, there are almost certainly limitations on applicability depending on the bandwidths B_s and B_ψ or $B_{\psi^{-1}}$ which further research should seek to characterize.

14 Numerical example

Because the viability of cyclostationarity exploitation in data processing in many fields of science and

engineering has been amply demonstrated in the literature in recent decades using real-world data, the only purpose of this section is to demonstrate that the *theory of cyclicity restoration* and/or *irregular cyclicity compensation* by time de-warping or time-warp compensation, presented herein, is itself viable.

14.1 Experimental setup

The signal model to be used for the example is the following pulse-amplitude/pulse-width modulated pulse train with uniform spacing between pulse starting times:

$$s(t) = \sum_n g_n p(h_n [t - nT_o]) \tag{45}$$

where the two sequences of random variables $\{g_n = \pm 1, \text{ iid uniform}\}$ and $\{h_n = 1, 2, 3, \text{ iid uniform}\}$ are statistically independent of each other, and the nominal pulse shape $p(t)$ in this pulse stream is as shown in Fig. 1.

The time-sampling increment is T_s , the pulse repetition period is $T_o = 160T_s$, and the signal bandwidth is approximated by the reciprocal of the width of the auto-correlation function [2]: $B_s \cong 12/T_o$. The averaging time (data-record length) is $T = 32,768T_s \cong 205T_o$ includes 205 cycles. It can be shown that the cyclic correlation coefficient is $|\rho_s^\alpha(\tau)| \cong 1$.

The measured data $y(t)$ contains the signal, time-warped by $\psi(t)$, in additive white Gaussian noise $n(t)$: $y(t) = s(\psi(t)) + n(t)$ with $SNR = +10$ dB (ratio of mean squared values of s and n equals 10) in case A (strong signal), and with $SNR = -10$ dB in case B (weak signal). From Eq. 1, we have $x(t) = s(t) + n(\psi^{-1}(t))$.

The time-warping function is given by the sum of un-warped time t and a non-periodic function

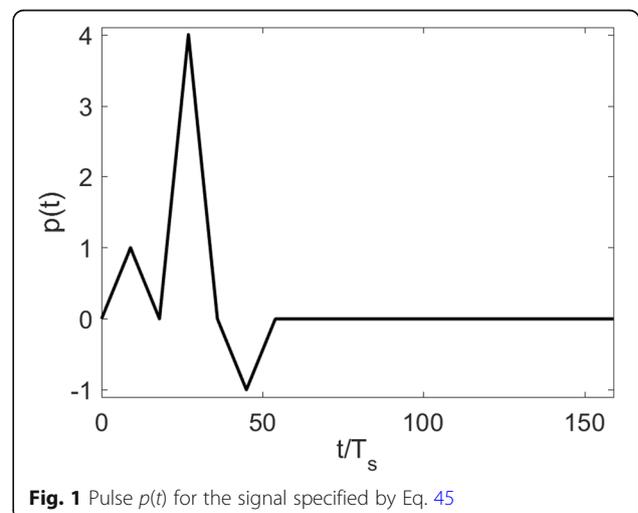


Fig. 1 Pulse $p(t)$ for the signal specified by Eq. 45

$$\begin{aligned} \psi(t) = t + \frac{3}{20\omega_o} \cos(\omega_o t + \pi/3) \\ + \frac{1}{20\omega_o} \cos((3 + \pi/1000)\omega_o t) \end{aligned} \tag{46}$$

where $\omega_o = 1/10T_o$, and the bandwidth of the component $\psi(t) - t$ is approximately $B_\psi \cong \omega_o/2\pi$, which yields $B_\psi/B_s \cong 0.002$. The warping function estimate is given by $\hat{\psi}(t) = \mathbf{b}^T \mathbf{c}(t)$ from Eq. 26 with $\gamma = 1$ and $\eta = 0$. From Eq. 44, we require $K > 4B_\psi T = 13 > 0.48/|\rho_s^\alpha(\tau)|^2 \cong 0.5$ provided that the basis functions span the space of functions with duration T and bandwidth $B_\psi \cong \omega_o/2\pi$. By using partial knowledge of $\psi(t)$ below, $K < 13$ becomes feasible.

Taking advantage of approximate knowledge of $\psi(t)$, we select the $K = 4$ basis functions, $c_1 = \cos(\omega_* t)$, $c_2 = \sin(\omega_* t)$, $c_3 = \cos(3\omega_* t)$, $c_4 = \sin(3\omega_* t)$, with fundamental frequency $\omega_* = 1.001\omega_o$ (whose frequencies do not exactly match the true frequencies in $\psi(t)$) for case A, and with ω_* and $3\omega_*$ replaced with the exact values ω_o and $(3 + \pi/1000)\omega_o$ for case B.

14.2 Discussion of results

The results of executing the iterative gradient-ascent optimization algorithm Eqs. 32 and 33 to estimate the vector of basis-function coefficients \mathbf{b} in Eq. 26 with $\gamma = 1$, $\eta = 0$, and $K = 4$, using the gradient expression Eq. 31a with the conjugation choice $(*) = *$, are shown in the eight figures in Section 14.2. Because the objective function Eq. 8 is highly multimodal, a substantial computational effort is needed to find the best initialization of the iteration. This costly task was circumvented by selecting a starting vector for \mathbf{b} known to be in the vicinity of the optimum vector. Also, because the warping function is of the form $\psi(t) = t + \varepsilon(t)$, with $|\varepsilon(t)| \ll t$ moderately well satisfied, the approximate formula $\hat{\psi}^{-1}(t) \cong t - \hat{\varepsilon}(t)$ for the inverse of the estimate $\hat{\psi}(t)$ was used as an expedient and, for $\hat{\varepsilon}$, the estimate $\hat{\varepsilon}(t) = \hat{\psi}(t) - t$ was used. Quantitatively, $|\varepsilon(t)| < 1/5\omega_o = 320T_s$; so a sufficient requirement for $|\varepsilon(t)| \ll t$ is $t/T_s \gg 320$, and the range of t/T_s used is $[0, 32768]$. Therefore, the requirement is met for only the latter 90% of the data, suggesting this may not be a highly accurate approximation and better results than those shown may be obtained using a more accurate approximation to the inverse of $\hat{\psi}(t)$, or restricting t to be greater than 320.

It can be seen that time warping with $\psi(t)$ (Fig. 2) significantly suppresses the cyclicity of the signal $s(t)$ (Figs. 3, 4, 5, 6), but that $\psi(t)$ can be estimated quite accurately from noisy data (Fig. 2), and inverted and used to de-warp the noisy warped signal well enough to substantially restore its cyclicity (Figs. 7, 8, 9). These results reveal the substantial noise tolerance exhibited by cyclostationary signals. Using the peak

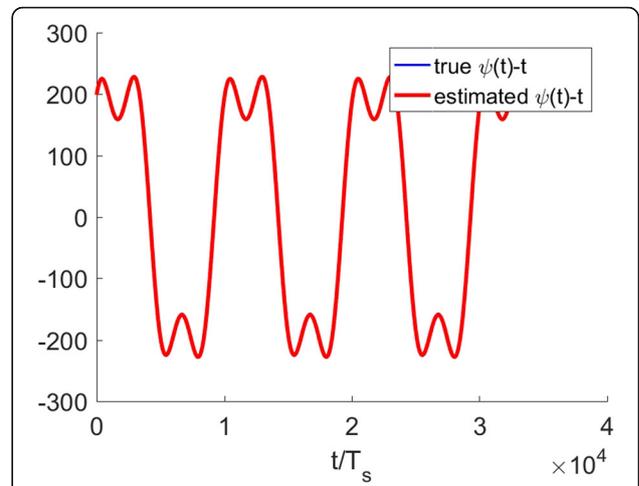


Fig. 2 Case A: time-warping function deviation $(\psi(t) - t)/T_s$ (blue), and estimate $(\hat{\psi}(t) - t)/T_s$ thereof (red), using inexact warping frequencies with $SNR = 10$ dB. Case B: for exact warping frequencies and $SNR = -10$ dB, results are graphically indistinguishable from those shown for case A

value of the cyclic correlogram (which is the quantity the optimization algorithm seeks to maximize) as a metric, it can be seen that the de-warping method for a signal with power level of only 1/10 that of the noise (Fig. 9) performs almost as well as it does for a signal with power level of 10 times that of the noise (Fig. 8).

14.3 Remaining challenges

Compared with the above example, there are, no doubt, substantially more challenging examples that arise in many fields of science involving data exhibiting irregular statistical cyclicity. Although the approach presented

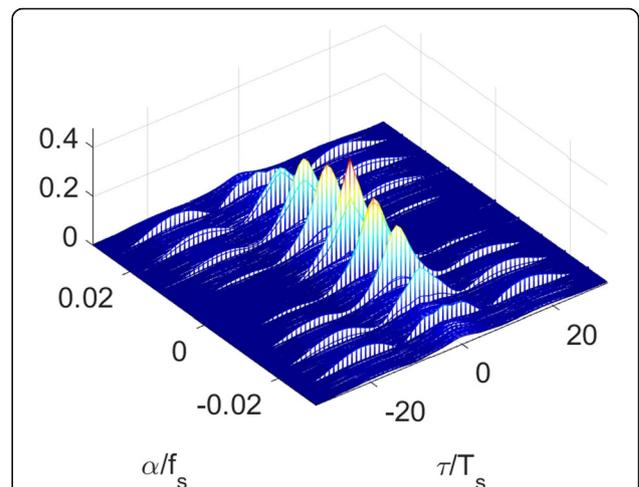
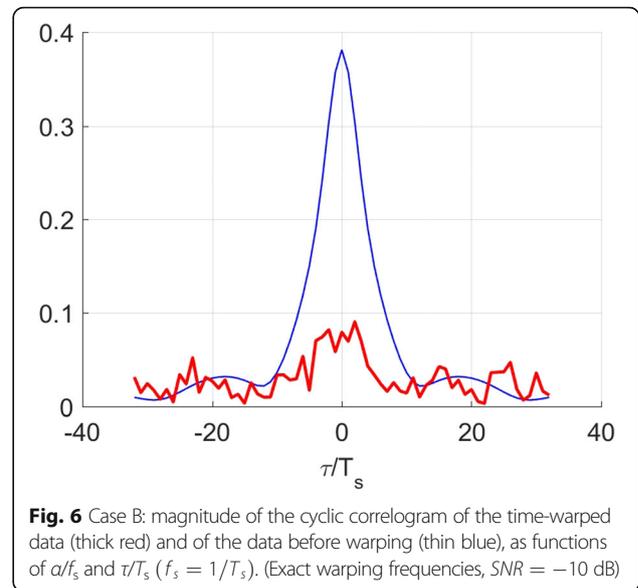
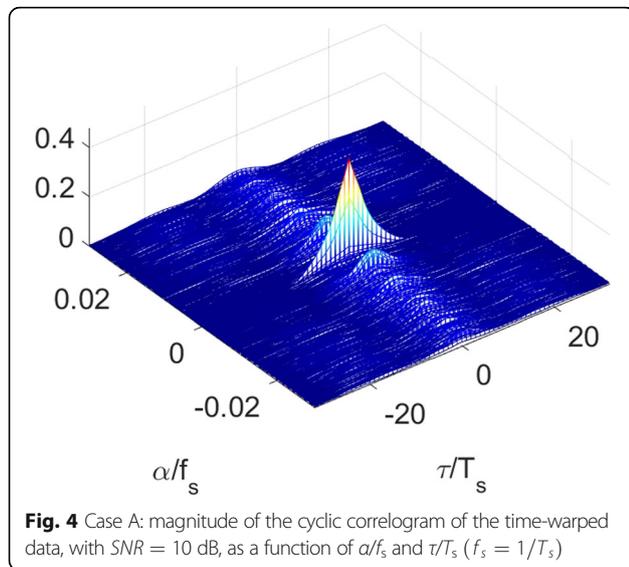


Fig. 3 Case A: magnitude of the cyclic correlogram of the un-warped data, with $SNR = 10$ dB, as a function of normalized cycle frequency and lag, α/f_s and τ/T_s ($f_s = 1/T_s$)

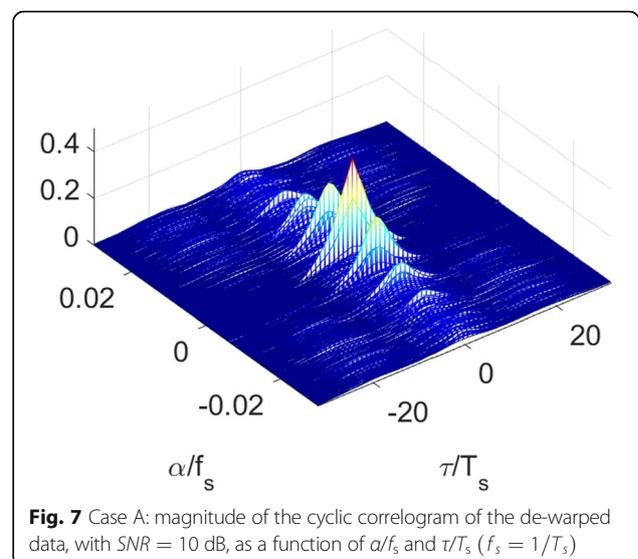
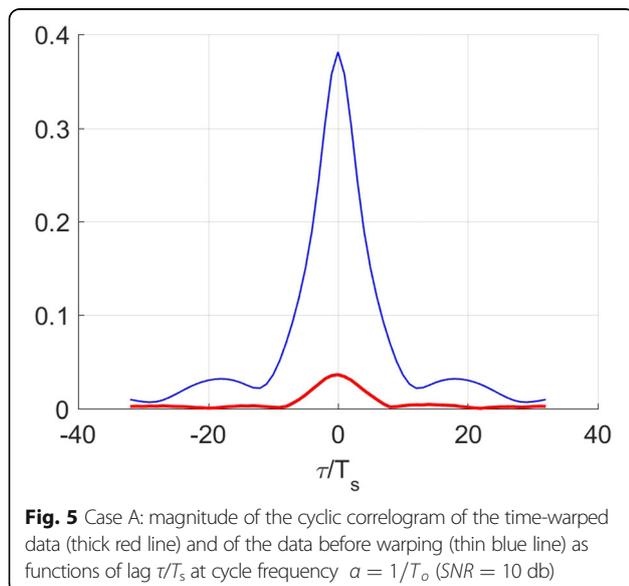


here is quite tolerant of noisy data, it is expected to be sensitive to the extent of the irregularity of the cyclicity. Time-warping functions with too large of a bandwidth can result in a required model order that may be too high (according to some as yet unidentified criterion). In addition, some important types of irregular cyclicity, such as pace irregularity, cannot be modeled in terms of time warping, as discussed in Section 12. Additional numerical examples are provided in [22, 23].

15 Conclusions

It is shown in this work that statistical inference from time-series data based on enhancement or restoral of the property of cyclostationarity can be performed to achieve two data processing preparatory tasks; this

achievement enables further processing of the time series data based on exploitation of cyclostationarity once it has been enhanced or restored and thereby identified by determining one of its cycle frequencies. The preparatory tasks are (1) determination of a time de-warping function or its equivalent time-warping compensation function and (2) use of that determined function to increase the degree of cyclostationarity thereby rendering it more amenable to cyclostationarity exploitation techniques that are well known to be effective for various types of statistical inference and decision objectives, particularly in situations where the signal of interest present in the data is masked by noise and/or interfering signals also present in the data.



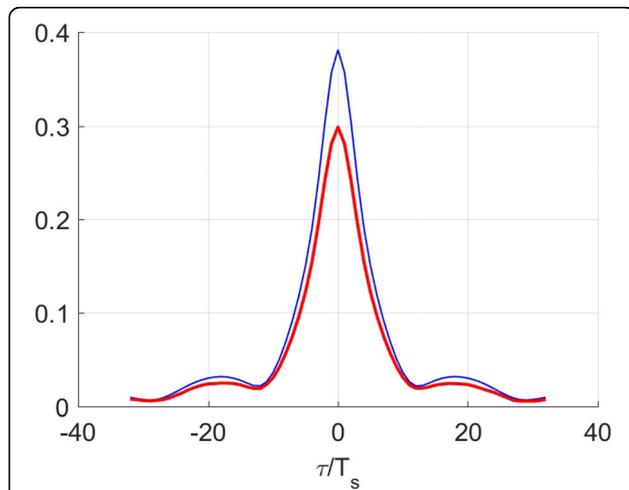


Fig. 8 Case A: magnitude of the cyclic correlogram of the de-warped data (thick red line) and of the data before warping (thin blue line) as functions of lag τ/T_s at cycle frequency $\alpha = 1/T_o$ (inexact warping frequencies, $SNR = 10$ dB)

Since its inception 30 to 40 years ago (cf. [1–5, 7–11]), cyclostationarity exploitation has proven to be an unusually versatile tool for extracting information that is, in some sense, hidden or buried in the available data. The achievement of the work reported here is the extension of this important and now widely used paradigm for signal processing from data exhibiting regular cyclicity to more challenging data exhibiting only irregular cyclicity. Strictly speaking, exactly regular cyclicity exists only in mathematics. Measurement or observation data obtained from the physical world can exhibit only irregular cyclicity. Depending on the phenomenon giving rise to the data, the degree of regularity in the cyclicity

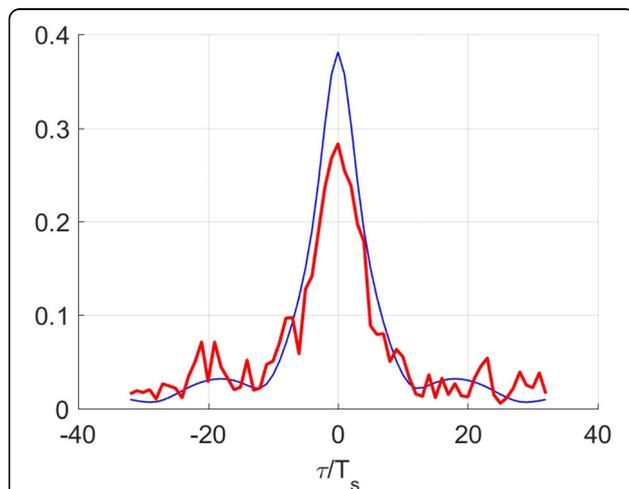


Fig. 9 Case B: magnitude of the cyclic correlogram of the de-warped data (thick red) and of the data before warping (thin blue), as functions of α/f_s and τ/T_s ($f_s = 1/T_s$). (Exact warping frequencies, $SNR = -10$ dB)

can be very high, moderate, very low, or absent altogether. Although cyclicity is ubiquitous in our world, as a consequence of rotation of Earth about its axis, its revolution about the Sun, the Moon's revolution about Earth, other planets' revolutions about the Sun, and, indeed, the astrodynamics of stars, galaxies, etc.—which has only recently been convincingly argued is a result of the central role of electromagnetism in the behavior of the Universe—the degree of irregularity in this naturally occurring cyclicity is more often than not too high to be ignored. This means the efficacy of cyclostationarity exploitation can often be substantially limited if the degree of irregularity is not decreased through time de-warping. And this typically requires statistical inference of an appropriate de-warping function.

As a consequence of the theoretical framework developed in this paper, the cyclostationarity paradigm can be expected to be extended, through statistically inferred time de-warping or time-warp compensation, from its present broad and diverse array of applications to a considerably more ubiquitous field of applications.

16 Highlights of converting irregular cyclicity to regular cyclicity

- Conversion of irregular cyclicity in time-series data to regular cyclicity is demonstrated.
- Data with regular cyclicity can be modeled as cyclostationary.
- The cyclostationarity paradigm for statistical inference has proven to be a rich resource.
- Cyclostationarity exploitation offers noise- and interference-tolerant signal processing.
- Cyclostationarity exploitation can now be extended to many more fields of science/engineering.

17 Appendix

17.1 The use of cyclostationary data models in science

As a simple means of assessing the utility of the concept of *cyclostationarity* in various fields of science and engineering, a web search using <https://scholar.google.com/>, performed in April 2018, was based on just under 50 nearly distinct applications areas in science and engineering, and the search terms were chosen with the intent of being minimally redundant: minimum number of hits, each of which results from more than one application area. The results are shown in Table 1, where it can be seen that the total number of hits was about 136,000. Analysis showed that the hits grew from a trickle of 1 to 2 figures per decade in the 1960s/1970s to a total of 6 figures over the ensuing

half century. The same is true, with 5 figures, for the search performed on the single general search term “cyclostationary OR cyclostationarity”. Also, as shown in Table 2, a search was performed using just over 20 search terms that represent partially-redundant general subjects in science and engineering; that is, there were substantial numbers of hits, each of which result from more than one subjected. The total number of hits was about 258,000. These hits also grew from a trickle per decade in the 1960s/1970s to a total of 6 figures over the ensuing half century. The number of hits for just the compound term comprised of the adjective and corresponding noun “cyclostationary OR cyclostationarity” was over 25,000 and has grown by a factor of approximately 4 every decade since the 1960s. To facilitate use of this data, it has been ordered alphabetically by search application (specific and general) in Table 1 or search subject (specific and general) in Table 2 and numerically by number of hits in Tables 3 and 4. Despite the concerted effort to use the search-term operators AND and OR judiciously and to select the search applications in a manner that minimizes the likelihood of more than one term producing the same hit (here called “search-term redundancy”), the search results obtained are suspicious. Given that there are only about 25,000 hits for the search subject “cyclostationary OR cyclostationarity” then, by “ANDing” this term with each of the approximately 50 approximately non-redundant applications, the total of all the hits for the “ANDed” terms should not exceed 25,000 by very much, yet it is about 5 times larger than this! This large of a total requires that either 1) the result of 25,000 is artificially limited by some search algorithm employed (not taken into account here out of ignorance) being used by the Google Scholar search engine, or 2) the AND and OR operators are not functioning correctly, or 3) the search application areas are much more redundant than expected—or possibly all three of these potential causes are in effect. The results of future analysis and possible refinement of this search study is planned to be made available in [6].

Table 5 highlights a particular scientific field in which cyclicity is central, and Table 6 highlights a few applications in which data measurement/analysis is enhanced by artificially instilling cyclicity into the data. These include use of optical spectral cloning for ultrafast/wideband photonic implementations of signal-processing algorithms (e.g., for radio frequency data) that were previously limited to slower and more narrowband electronic implementation and use of spectral redundancy insertion, such as spectral cloning, for introducing some level of immunity to noise and/or interference when used in conjunction with frequency-shift filtering.

Table 1 Nearly Distinct Application Areas^a

1	"aeronautics OR astronautics OR navigation" AND "CS/CS"	3,190
2	"astronomy OR astrophysics" AND "CS/CS"	864
3	"atmosphere OR weather OR meteorology OR cyclone OR hurricane OR tornado" AND "CS/CS"	2,230
4	"cognitive radio" AND "CS/CS"	8,540
5	"comets OR asteroids" AND "CS/CS"	155
6	"cyclic MUSIC"	512
7	"direction finding" AND "CS/CS"	1,170
8	"electroencephalography OR cardiography" AND "CS/CS"	742
9	"global warming" AND "CS/CS"	369
10	"oceanography OR ocean OR maritime OR sea" AND "CS/CS"	3,060
11	"physiology" AND "CS/CS"	673
12	"planets OR moons" AND "CS/CS"	274
13	"pulsars" AND "CS/CS"	115
14	"radar OR sonar OR lidar" AND "CS/CS"	5,440
15	"rheology OR hydrology" AND "CS/CS"	639
16	"seismology OR earthquakes OR geophysics OR geology" AND "CS/CS"	1,090
17	"SETI OR extraterrestrial" AND "CS/CS"	83
18	autoregression AND "CS/CS"	2,040
19	bearings AND "CS/CS"	3,980
20	biology AND "CS/CS"	2,030
21	biometrics AND "CS/CS"	309
22	chemistry AND "CS/CS"	2,020
23	classification AND "CS/CS"	10,900
24	climatology AND "CS/CS"	811
25	communications AND "CS/CS"	21,200
26	cosmology AND "CS/CS"	172
27	ecology AND "CS/CS"	356
28	economics AND "CS/CS"	2,050
29	galaxies OR stars AND "CS/CS"	313
30	gears AND "CS/CS"	2,000
31	geolocation AND "CS/CS"	676
32	interception AND "CS/CS"	2,270
33	mechanical AND "CS/CS"	4,770
34	medical imaging OR scanning AND "CS/CS"	1,370
35	medicine AND "CS/CS"	2,990
36	modulation AND "CS/CS"	17,000
37	physics AND "CS/CS"	4,539
38	plasma AND "CS/CS"	542
39	quasars AND "CS/CS"	47
40	Sun AND "CS/CS"	4,320
41	UAVs AND "CS/CS"	238

Table 1 Nearly Distinct Application Areas^a (Continued)

42	universe AND "CS/CS"	209
43	vibration OR rotating machines AND "CS/CS"	3,240
44	walking AND "CS/CS"	990
45	wireless AND "CS/CS"	15,100
	TOTAL	135,628

^a"CS/CS" is an abbreviation for "cyclostationary OR cyclostationarity"**Table 2** Partially-Redundant General Subjects^a

1	"almost "CS/CS""	8,840
2	"almost periodically correlated" AND "sequences OR processes"	352
3	"Cyclic Wiener Filtering" OR "FRESH filtering" OR "frequency-shift filtering"	676
4	"cyclostationary EOF" OR "cyclostationary empirical orthogonal functions" OR CSEOF	453
5	"exploiting "CS/CS""	11,900
6	"Gardner relation" OR "Cyclic Wiener Relation"	75
7	"periodically correlated" AND "sequences OR processes"	1,740
8	"signal analysis" AND "CS/CS"	3,210
9	"signal processing" AND "CS/CS"	19,000
10	"spatial filtering" AND "CS/CS"	571
11	"spectral redundancy" AND "CS/CS"	1,170
12	computers AND "CS/CS"	17,700
13	cyclic spectrum AND "CS/CS"	9,580
14	cyclostationary OR cyclostationarity	25,000
15	equalization AND "CS/CS"	6,360
16	estimation AND "CS/CS"	20,800
17	filtering AND "CS/CS"	18,400
18	filtering OR smoothing AND "CS/CS"	12,400
19	higher-order OR cumulant AND cyclostationarity	6,040
20	identification AND "CS/CS"	15,700
21	mathematics AND "CS/CS"	10,900
22	prediction AND "CS/CS"	9,540
23	sensing AND "CS/CS"	14,900
24	spectrum AND "CS/CS"	22,200
	TOTAL	237,507

^a"CS/CS" is an abbreviation for "cyclostationary OR cyclostationarity"**Table 3** Nearly Distinct Application Areas^a

1	communications AND "CS/CS"	21,200
2	modulation AND "CS/CS"	17,000
3	wireless AND "CS/CS"	15,100
4	classification AND "CS/CS"	10,900
5	"cognitive radio" AND "CS/CS"	8,540
6	"radar OR sonar OR lidar" AND "CS/CS"	5,440
7	mechanical AND "CS/CS"	4,770
8	physics AND "CS/CS"	4,539
9	Sun AND "CS/CS"	4,320
10	bearings AND "CS/CS"	3,980
11	vibration OR rotating machines AND "CS/CS"	3,240
12	"aeronautics OR astronautics OR navigation" AND "CS/CS"	3,190
13	"oceanography OR ocean OR maritime OR sea" AND "CS/CS"	3,060
14	medicine AND "CS/CS"	2,990
15	interception AND "CS/CS"	2,270
16	"atmosphere OR weather OR meteorology OR cyclone OR hurricane OR tornado" AND "CS/CS"	2,230
17	economics AND "CS/CS"	2,050
18	autoregression AND "CS/CS"	2,040
19	biology AND "CS/CS"	2,030
20	chemistry AND "CS/CS"	2,020
21	gears AND "CS/CS"	2,000
22	medical imaging OR scanning AND "CS/CS"	1,370
23	"direction finding" AND "CS/CS"	1,170
24	"seismology OR earthquakes OR geophysics OR geology" AND "CS/CS"	1,090
25	walking AND "CS/CS"	990
26	"astronomy OR astrophysics" AND "CS/CS"	864
27	climatology AND "CS/CS"	811
28	"electroencephalography OR cardiography" AND "CS/CS"	742
29	geolocation AND "CS/CS"	676
30	"physiology" AND "CS/CS"	673
31	"rheology OR hydrology" AND "CS/CS"	639
32	plasma AND "CS/CS"	542
33	"cyclic MUSIC"	512
34	"global warming" AND "CS/CS"	369
35	ecology AND "CS/CS"	356
36	galaxies OR stars AND "CS/CS"	313
37	biometrics AND "CS/CS"	309
38	"planets OR moons" AND "CS/CS"	274
39	UAVs AND "CS/CS"	238
40	universe AND "CS/CS"	209

Table 3 Nearly Distinct Application Areas^a (Continued)

41	cosmology AND "CS/CS"	172
42	"comets OR asteroids" AND "CS/CS"	155
43	"pulsars" AND "CS/CS"	115
44	"SETI OR extraterrestrial" AND "CS/CS"	83
45	quasars AND "CS/CS"	47
	TOTAL	135,628

^a"CS/CS" is an abbreviation for "cyclostationary OR cyclostationarity"

Table 4 Partially-Redundant General Subjects^a

1	cyclostationary OR cyclostationarity	25,000
2	spectrum AND "CS/CS"	22,200
3	estimation AND "CS/CS"	20,800
4	"signal processing" AND "CS/CS"	19,000
5	filtering AND "CS/CS"	18,400
6	computers AND "CS/CS"	17,700
7	identification AND "CS/CS"	15,700
8	sensing AND "CS/CS"	14,900
9	filtering OR smoothing AND "CS/CS"	12,400
10	"exploiting "CS/CS""	11,900
11	mathematics AND "CS/CS"	10,900
12	cyclic spectrum AND "CS/CS"	9,580
13	prediction AND "CS/CS"	9,540
14	"almost "CS/CS""	8,840
15	equalization AND "CS/CS"	6,360
16	higher-order OR cumulant AND cyclostationarity	6,040
17	"signal analysis" AND "CS/CS"	3,210
18	"periodically correlated" AND "sequences OR processes"	1,740
19	"spectral redundancy" AND "CS/CS"	1,170
20	"Cyclic Wiener Filtering" OR "FRESH filtering" OR "frequency-shift filtering"	676
21	"spatial filtering" AND "CS/CS"	571
22	"cyclostationary EOF" OR "cyclostationary empirical orthogonal functions" OR CSEOF	453
23	"almost periodically correlated" AND "sequences OR processes"	352
24	"Gardner relation" OR "Cyclic Wiener Relation"	75
	TOTAL	237,507

^a"CS/CS" is an abbreviation for "cyclostationary OR cyclostationarity"

Table 5 Chronobiology

Chronobiology (45,300 search hits on Google Scholar). The study of how solar- and lunar-related rhythms are governed by living pacemakers within organisms constitutes the scientific discipline of chronobiology (sometimes called chronomics). Parallel to the familiar spatial cellular structure of living cells, temporal (time) organization is a vital part of the survival and normal functioning of every species. Adaptations evolved by organisms to cope with regular geophysical cycles in their environment are evident in nearly every aspect of their lives. In fact, biological timekeeping is a core property of life on a rotating and revolving planet. Thus, cyclicity (or rhythmicity) is central to much of chronobiology. That is chronobiology is in large part the study of biological rhythms which represent irregular cyclicity and include physiological or behavioral, and internally- (endogenous) or externally- (exogenous) controlled rhythms; diurnal, nocturnal, and crepuscular rhythms; circadian, ultradian, and infradian rhythms; and tidal, lunar, and gene oscillation rhythms.

Chronobiological studies include but are not limited to comparative anatomy, physiology, genetics, molecular biology and behavior of organisms within biological rhythms mechanics [Chronobiology: Biological Timekeeping, by J. C. Dunlap; J. J. Loros, and P. J. DeCoursey, Sinauer Associates Inc., 2009.] Other aspects include development, reproduction, ecology and evolution. <https://qz.com/1091769/the-2017-nobel-prize-in-medicine-goes-to-hall-rosbash-and-young-for-their-work-explaining-circadian-rhythm/>.

Table 6 Spectral cloning in optics

Instilling cyclicity: The utility of the concept of cyclostationarity in signal processing is further broadened by the fact that some data processing systems can benefit from artificially introducing cyclostationarity into data that exhibits no cyclicity whatsoever. An example of this is systems for amplifying and detecting fast transients that occur, for example, in optical data processing, as briefly discussed below.

Minimum-noise optical amplification is "the ability to amplify optical signals with minimum excess noise [and] is of critical interest in a wide variety of photonic-systems applications that span optical communications, imaging, sensing, metrology, and quantum state processing, to name a few" [Stojan Radic, *Advances in Optics and Photonics*, Vol. 5, Issue 3, pp. 318–384 (2013) <https://doi.org/10.1364/AOP.5.000318>]. Spectral cloning is a relatively new approach to phase-sensitive "subnoise" or "noiseless" optical amplification, which—to quote Radic—is "opening diverse applications that rely on low-noise optical operation: wavelength conversion, phase regeneration, and signal spectral replication". Spectral cloning is achieved with nonlinear optical parametric amplification, but the underlying principle is that of replicating the spectrum of a signal in multiple distinct bands and then frequency-shifting all such bands to a common band where the replicas are added. Because the spectral replication is able to be performed prior to the processing stage that is the dominant source of receiver noise, the creation and addition of the replicas results in coherent processing gain against the low noise (idealistically called "noiseless" or "subnoise" amplification), coming out of the cloning stage, and produces a pre-amplified signal that can then be further amplified or processed by subsequent noisier stages. This instillation and exploitation of spectral redundancy is identical to instillation and exploitation of cyclostationarity. This approach to performing "subnoise" detection of fast random events is described in [V. Ataie, D. Esman, B. P.-P. Kuo, N. Alic, and S. Radic, "Subnoise detection of a fast random event", *Science*, [sciencemag.org](https://doi.org/10.1126/science.1256666), 11 Dec 2015, Vol. 350 Issue 6266].

Abbreviations

CCT: Cycle coherence time; CS: Cyclostationary/Cyclostationarity; CV: Coefficient of variation; DCS: Degree of cyclostationarity; ISC: Irregular statistical cyclicity; polyCS: Polycyclostationarity; RSC: Regular statistical cyclicity

Acknowledgements

The author expresses his gratitude to Professor Antonio Napolitano for performing the simulations reported in Section 14 and for his helpful technical comments on the manuscript.

Author's contributions

The author read and approved the final manuscript.

Competing interests

The author declares that he has no competing interest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 July 2017 Accepted: 12 June 2018

Published online: 18 September 2018

References

1. WA Gardner, in *Introduction to Random Processes with Applications to Signals and Systems*. Cyclostationary processes (Macmillan, New York, Ed. 1, 1985; McGraw-Hill, New York, Ed. 2, 1990)
2. WA Gardner, in *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Part II: periodic phenomena (Prentice-Hall, Englewood Cliffs, 1987)
3. WA Gardner, *Cyclostationarity in Communications and Signal Processing* (IEEE Press, Piscataway, 1994)
4. WA Gardner, *Stationarizable Random Processes*, IEEE Transactions on Information Theory, IT-24 (1978), pp. 8–22
5. A Napolitano, *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications* (Wiley/IEEE Press, West Sussex, 2012)
6. William A. Gardner. A comprehensive tutorial website under construction. (2017). <https://www.cyclostationarity.com>. Accessed 1 Aug 2018.
7. WA Gardner, A Napolitano, L. Paura, Cyclostationarity: half a century of research. *Signal Process.* **86**, 639–697 (2006)
8. WA Gardner, Exploitation of spectral redundancy in cyclostationary signals. *IEEE Signal Process. Mag.* **8**, 14–36 (1991)
9. William A. Gardner, University of California, Davis–Webpage, 2000; link: <http://faculty.engineering.ucdavis.edu/gardner/publications/>. Accessed 1 Apr 2018.
10. WA Gardner, Two alternative philosophies for estimation of the parameters of time-series. *IEEE Trans. Inf. Theory* **37**, 216–218 (1991)
11. WA Gardner, WA Brown, Fraction-of-time probability for time-series that exhibit cyclostationarity. *Signal Process.* **23**, 273–292 (1991)
12. J Leskow, A Napolitano, Foundations of the functional approach for signal analysis. *Signal Process.* **86**, 3796–3825 (2006)
13. BG Agee, SV Schell, WA Gardner, Spectral self-coherence restoral: a new approach to blind adaptive signal extraction using antenna arrays. *Proc. IEEE* **78**, 753–767 (1990)
14. WA Gardner, CK Chen, Signal-selective time-difference-of-arrival estimation for passive location of man-made signal sources in highly corruptive environments. *IEEE Trans. Signal Process.* **40**, 1168–1184 (1992) [Some of the quantitative performance evaluation results from simulations in Sec IV of Part II of this reference are misleading because the TDOA to be estimated is an integer-multiple of the time-sampling interval, which inadvertently reduces TDOA estimation error variance by quantizing to zero all individual errors less than half the sampling interval in magnitude. Particularly suspect are the mean squared errors curves that cross the CRB in Figs. 6(c), (d), 7, and 8.]
15. LE Franks, *Signal Theory* (Prentice-Hall, Englewood Cliffs, 1969)
16. K Kreutz-Delgado, *The Complex Gradient Operator and the CR-Calculus* (UCSD, San Diego, 2009) link: <https://arxiv.org/pdf/0906.4835.pdf>
17. J Barzilai, JM Borwein, Two-point step size gradient methods. *IMA J. Numer. Anal.* **8**, 141–148 (1988)
18. H Zhang, WW Hager, A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.* **14**, 1043–1056 (2004)
19. D Jerome Antoni, SB Abboud, in *Cyclostationarity: Theory and Methods, Lecture Notes in Mechanical Engineering*, ed. by F Chaari, J Leskow, A

- Napolitano, A Sanchez-Ramirez. Time-angle periodically correlated processes (Springer International Publishing, 2014), pp. 3–14
20. WA Gardner, Correlation estimation and time-series modeling for nonstationary processes. *Signal Process.* **15**, 31–41 (1988)
21. GD Zivanovic, WA Gardner, Degrees of cyclostationarity and their application to signal estimation and detection. *Signal Process.* **22**, 287–297 (1991)
22. A Napolitano, WA Gardner, Algorithms for Analysis of Signals with Time-Warped Cyclostationarity, Proceedings of 50th Asilomar Conference on Signals, Systems, and Computers (2016), pp. 539–543
23. A Napolitano, Time-warped almost-cyclostationary signals: characterization and statistical function measurements. *IEEE Trans. Signal Process.* **64**, 5526–5541 (2017)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com