

# Singular spectrum analysis of time series data from low-frequency radiometers, with an application to SITARA data

Jishnu N. Thekkepattu <sup>1,2★</sup>, Cathryn M. Trott <sup>1,3</sup> and Benjamin McKinley <sup>1,3</sup>

<sup>1</sup>International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia

<sup>2</sup>Raman Research Institute, C V Raman Avenue, Sadashivanagar, Bengaluru 560080, India

<sup>3</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Bentley, WA 6102, Australia

Accepted 2023 February 12. Received 2023 January 20; in original form 2022 May 10

## ABSTRACT

Understanding the temporal characteristics of data from low-frequency radio telescopes is of importance in devising suitable calibration strategies. Application of time-series analysis techniques to data from radio telescopes can reveal a wealth of information that can aid in calibration. In this paper, we investigate singular spectrum analysis (SSA) as an analysis tool for radio data. We show the intimate connection between SSA and Fourier techniques. We develop the relevant mathematics starting with an idealized periodic dataset and proceeding to include various non-ideal behaviours. We propose a novel technique to obtain long-term gain changes in data, leveraging the periodicity arising from sky drift through the antenna beams. We also simulate several plausible scenarios and apply the techniques to a 30-day time series data collected during 2021 June from SITARA – a short-spacing two element interferometer for global 21-cm detection. Applying the techniques to real data, we find that the first reconstructed component – the trend – has a strong anti-correlation with the local temperature suggesting temperature fluctuations as the most likely origin for the observed variations in the data. We also study the limitations of the calibration in the presence of diurnal gain variations and find that such variations are the likely impediment to calibrating SITARA data with SSA.

**Key words:** Dark ages, reionization, first stars – methods: data analysis.

## 1 INTRODUCTION

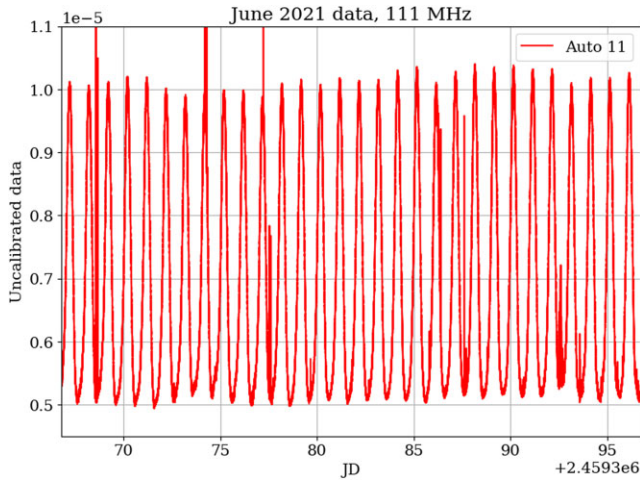
There is a renewed interest in low-frequency (<300 MHz) radio astronomy, due to the multitude of science cases that benefit from low-frequency observations. Several low-frequency radio telescopes such as LOFAR (van Haarlem et al. 2013), MWA (Tingay et al. 2013), HERA (DeBoer et al. 2017), and LWA (Hallinan et al. 2015) have been constructed and are currently observing, with the low-frequency Square Kilometer Array (SKA), SKA-low (Dewdney et al. 2009) in construction phase. Some of the key science goals for these metre-wavelength radio telescopes are cosmic dawn and epoch of reionization (CD/EoR) (Trott 2017), solar and heliospheric science (Nindos, Kontar & Oberoi 2019), and cosmic magnetism (Gaensler, Beck & Feretti 2004). Modern low-frequency radio telescopes differ from their higher-frequency counterparts in that they consist of aperture arrays constructed from a large number of antennas that are often beamformed in the analogue or digital domain and correlated against each other to observe the radio sky. The calibration requirements, calibration models, and the complexity are different to higher-frequency (cm wavelengths and above) radio telescopes.

For dish-based interferometers with a small number of antennas, single dish telescopes, and specialized low-frequency radiometers, calibration techniques such as Dicke switching or noise injection can be employed. Specifically, global 21-cm experiments constitute a group of low-frequency radiometers requiring precise and accurate

calibration of the systems to limit the systematics to less than one part in a million. Most of the global 21-cm experiments use single antennas as sensors and employ Dicke-switching ambient temperature matched loads and noise diodes (see Rogers & Bowman 2012; Nambissan T. et al. 2021) or noise injection (Singh et al. 2018) for bandpass calibration. The same calibrators are used to compensate for receiver gain drifts over extended periods of time. However, in order to maintain a stable excess noise ratio (ENR), the calibrators have to be maintained in temperature-controlled environments, and the noise diodes themselves require periodic re-calibration with laboratory standards to mitigate drift and ageing, especially if they are to be used in applications demanding high accuracy such as cosmology. Besides, with future telescopes such as the SKA-low potentially employing thousands of low-cost active antennas, calibration using dedicated noise diodes at each antenna becomes impossible. The aperture array nature of these instruments, i.e. a large number of stationary antennas of simple construction, often with integrated low-cost low-noise amplifiers (LNAs), necessitates calibration based on sky models. Therefore, development of novel mathematical tools to explore the long-term stability of these low-frequency telescopes and radiometers, and to determine the limits of calibration and data integration, become essential.

In this paper, we explore the potential of one such tool in analysing time series data from low-frequency radiometers and an application of it to real-valued time series data. The instrument under study is SITARA, a broadband two element interferometer targeting global 21-cm detection employing short spacing interferometry. SITARA consists of two MWA-style active antennas kept 1 m apart over a large

\* E-mail: [j.thekkepattu@curtin.edu.au](mailto:j.thekkepattu@curtin.edu.au)



**Figure 1.** SITARA raw data time series as a function of Julian Date (JD) for a single frequency channel of bandwidth 61 kHz, at a frequency of 111 MHz. The data have been extracted from a concatenated SITARA dataset for the month of 2021 June. Solar bursts contribute to most of the RFI seen in this time series.

ground-plane (35 m) and a correlator that records auto-correlations and cross-correlations with spectral resolution of about 61 kHz, across 0–250 MHz. The usable band is limited to 70–200 MHz, with reduced sensitivity in the 50–70 MHz band. SITARA is deployed within the radio quiet zone of the *Murchison Radioastronomy Observatory* (MRO) in Western Australia, which is also the site for future SKA-low. Data are collected round the clock and timestamped data are written out in *miriad* format (Sault, Teuben & Wright 1995). SITARA has been conceived as the first prototype to evolve tools and techniques for short-spacing interferometry; further details about SITARA can be found in Thekkeppattu et al. (2022). Single-frequency data from SITARA auto-correlations can essentially be treated as radiometric data, with SITARA behaving as an uncalibrated total power radiometer.

## 2 MOTIVATION

As an example to motivate this study, we consider real-valued auto-correlations from SITARA data. A plot of single-frequency channel time series data from SITARA auto-correlations, at a frequency of 111 MHz for the month of 2021 June, is shown in Fig. 1. The frequency of 111 MHz has relatively low radio frequency interference (RFI) and is also a frequency where mutual coupling between antennas does not cause large beam shifts (see Thekkeppattu et al. 2022) for details). The data have a periodic nature arising from the drifting of various regions of the radio sky through the antenna beams as the Earth rotates. If the radio telescope was perfectly calibrated, and there was no RFI, Fig. 1 would have shown a perfectly repeating pattern.

As can be seen, this is not the case. The data have some multiplicative gain variations as well as some additive RFI. Therefore, our aim is to decompose this time series into some components that can help us understand it better. A plausible decomposition is the Fourier transform, i.e. into sines and cosines. However, radiometric data can have aperiodic patterns such as a drift which gets distributed to various modes in a Fourier analysis, thus making an interpretation difficult. Therefore, it is desirable to adopt an analysis which selects a suitable basis from the data, in other words a data-driven technique.

Singular spectrum analysis (SSA) is a set of data-driven tools that can decompose a time series into elementary patterns such as trend and oscillatory components. The raw data from a low-frequency radio telescope with a fixed pointing are expected to have oscillatory components with a period corresponding to a sidereal day, while owing to environmental changes trend-like patterns (drift) are also expected. This makes SSA an ideal tool for analysis of such data.

SSA techniques appear in the analysis of dynamical systems (Broomhead & King 1986; Vautard & Ghil 1989) and SSA has been a popular tool for time series analysis in a variety of fields such as meteorology and climate science (Ghil et al. 2002) and geophysics (Dokht, Gu & Sacchi 2016). However the application of SSA in radio astronomy has been limited (Donskikh et al. 2016; Gürel, Hurley & Simeoni 2018). Indeed, to the best of our knowledge, SSA has not been applied for time series analysis of radiometric time series data. This paper aims to detail the necessary mathematical tools for SSA of time series data from a low-frequency radiometer and demonstrate them with SITARA data.

### 2.1 Notation and mathematical preliminaries

In this section, we describe the notation employed and define certain matrices that are useful for the subsequent analysis. Many of these definitions can be found in Davis (1979) as well as Olson et al. (2014). We use bold capital Roman letters such as  $\mathbf{X}$  to denote matrices, and small Roman letters such as  $n$  are used for indexing. We use zero-based numbering such that the indices start at 0. Small Roman letters with an arrow such as  $\vec{u}$  denote vectors. Integers are denoted by capital Roman letters such as  $L$ . We now proceed to define some basic matrices and associated linear algebra.

#### 2.1.1 Circulant and anti-circulant matrices

An  $N \times N$  square matrix  $\mathbf{C}$  is circulant if each row of the matrix is a *right*-shifted version of the previous row as shown in equation (1).

$$\mathbf{C} = \begin{bmatrix} c[0] & c[1] & \dots & c[N-1] \\ c[N-1] & c[0] & \dots & c[N-2] \\ c[N-2] & c[N-1] & \dots & c[N-3] \\ \vdots & \vdots & \dots & \vdots \\ c[1] & c[2] & \dots & c[0] \end{bmatrix}. \quad (1)$$

An  $N \times N$  square matrix  $\mathbf{C}$  is anti-circulant if each row of the matrix is a *left*-shifted version of the previous row as shown in equation (2).

$$\mathbf{C}_a = \begin{bmatrix} c[0] & c[1] & \dots & c[N-1] \\ c[1] & c[2] & \dots & c[0] \\ c[2] & c[3] & \dots & c[1] \\ \vdots & \vdots & \dots & \vdots \\ c[N-1] & c[0] & \dots & c[N-2] \end{bmatrix}. \quad (2)$$

Both the circulant and anti-circulant matrices can be obtained from a sequence  $c[n]$ ;  $n = 0, 1, \dots, N-1$  and are completely specified by that sequence. While useful in a wide variety of analyses, circulant matrices are not of much utility in this paper. We will be dealing with anti-circulant matrices instead. A matrix can also be interpreted to have been broken into blocks or submatrices that are themselves matrices.

#### 2.1.2 Matrix products

The notations employed in this paper for the various matrix products are listed below.

- (i) Regular matrix-matrix multiplication is denoted with no specific operator.
- (ii)  $\otimes$  denotes the outer product of two vectors that results in a matrix.
- (iii)  $\odot$  represents an element-by-element multiplication of two matrices, known as the Hadamard product.

### 3 BASICS OF SINGULAR SPECTRUM ANALYSIS

Though the theory of SSA is covered in detail in references such as Golyandina, Nekrutkin & Zhigljavsky (2001), we provide a basic description of the steps involved for completeness. In this paper, we follow the SSA approach known as the Broomhead–King (BK) version, with the alternate being the Vautard–Ghil (VG) version. The VG version is only suitable for the analysis of a stationary time series and therefore is not discussed here. Following Golyandina & Korobeynikov (2014), the four major steps in BK SSA are given below.

- (i) Convert the 1-D time series into a 2-D matrix called a *trajectory matrix*. This step is called *embedding* in time series analysis.
- (ii) Decompose the trajectory matrix with singular value decomposition (SVD). The result consists of a set of left and right singular vectors and associated singular values.
- (iii) Reconstruct the constituent components of the trajectory matrix with selected singular vectors and singular values.
- (iv) Reconstruct the time series with these components by performing diagonal averaging.

In the embedding step, the trajectory matrix is constructed with columns consisting of elements from sliding a window of length  $L$  across the original time series. The window length is chosen in such a manner that each column of the resulting matrix consists of one cycle of the data. As an example, for drift sky data from a ground-based radiometer, as given in Section 4.1.1, this corresponds to one sidereal day. For each sliding, the elements inside the window are made into one column of the trajectory matrix, yielding a matrix with  $L$  rows and  $K = N - L + 1$  columns. Consider a time series  $x[n]$ ,  $n = 0, 1, \dots, N - 1$ . The embedding step converts this time series of length  $N$  into an  $L \times K$  matrix that has the elements of the time series as given in equation (3).

$$\mathbf{X} = \begin{bmatrix} x[0] & x[1] & x[2] & \dots & x[K-1] \\ x[1] & x[2] & x[3] & \dots & x[K] \\ x[2] & x[3] & x[4] & \dots & x[K+1] \\ \dots & \dots & \dots & \dots & \dots \\ x[L-1] & x[L] & x[L+1] & \dots & x[N-1] \end{bmatrix}. \quad (3)$$

From equation (3), it can be seen that the anti-diagonals of a trajectory matrix contain similar terms, and therefore it is like a Hankel matrix, although it is not square in general. It may be noted that some implementations pad the original time series with zeros to obtain  $K = N$ , though we do not employ this.

In the second step, the trajectory matrix is decomposed via SVD to yield left and right singular vectors as well as the corresponding singular values. This can be written as

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T. \quad (4)$$

The SVD operation decomposes the  $L \times K$  matrix  $\mathbf{X}$  into three matrices  $\mathbf{U}$ ,  $\mathbf{\Sigma}$ , and  $\mathbf{V}$ ; where  $\mathbf{U}$  is an  $L \times L$  unitary matrix,  $\mathbf{V}$  is a  $K \times K$  unitary matrix and  $\mathbf{\Sigma}$  is an  $L \times K$  diagonal matrix consisting of the singular values. Since we consider only real-valued matrices  $\mathbf{X}$  in this paper, the matrices  $\mathbf{U}$  and  $\mathbf{V}$  are real orthogonal matrices.

Each singular value  $\sigma_i$  and the corresponding singular vectors  $\vec{u}_i$  and  $\vec{v}_i$  form an eigentriple  $(\sigma_i, \vec{u}_i, \vec{v}_i)$ . The decomposition step can also be written as

$$\mathbf{X} = \sum_i \mathbf{X}_i; \quad \mathbf{X}_i = \sigma_i (\vec{u}_i \otimes \vec{v}_i), \quad (5)$$

where  $\otimes$  is the vector outer product of two vectors yielding a matrix.

These eigentriples can be grouped and used to reconstruct the various components of the time series. An inspection of the singular values can give an insight into the complexity of the data and the number of eigentriples to consider, and this is also related to the rank of the trajectory matrix. It may be noted that calculation of SVD of a data matrix and subsequent reconstruction is identical to principal component analysis or PCA (Jolliffe 1986). Seen in this light, SSA can be treated as PCA of the trajectory matrix formed out of a time series. However, PCA does not involve embedding or trajectory matrices, and therefore the interpretation of the PCA results is different to that of SSA.

The next step is to reconstruct the trajectory matrix with the selected eigentriples. Once this is accomplished, a reconstructed time series is obtained by performing an anti-diagonal averaging over the reconstructed matrix — this step is called ‘diagonal averaging’ (Golyandina & Korobeynikov 2014).<sup>1</sup> The diagonal averaging step applied to a trajectory matrix  $\mathbf{Y}$  yields a series  $y_s$  as shown in equation (6).

$$y_s[n] = \frac{\sum_{(l,k) \in \mathbb{A}_n} \mathbf{Y}[l][k]}{|\mathbb{A}_n|}, \quad (6)$$

where  $\mathbb{A}_n = \{(l, k); l + k = n, 0 \leq l \leq L - 1, 0 \leq k \leq K - 1\}$  and  $|\mathbb{A}_n|$  is the number of elements in the set  $\mathbb{A}_n$ . This step can be treated as the reverse of the embedding step, conversion of an anti-diagonal, Hankel-like matrix back into a time series. To gain an appreciation of this step, it is useful to take a closer look at the structure of a trajectory matrix as given in equation (3). Recalling that this matrix is formed by sliding a window and converting the data into columns of a matrix, it can be seen that the anti-diagonals consist of the same value from the time series. Therefore to convert a trajectory matrix back into a time series, all that one has to do is a computation of the mean of the redundant information in each anti-diagonal and form a time series out of them. This operation is what equation (6) does. It may also be noted that one can even select the first row and the last column of a trajectory matrix to reconstruct a time series; however, it is preferable to obtain a mean of the redundant information in the anti-diagonal elements. When applied to each  $\mathbf{X}_i$  in equation (5), diagonal averaging results in *reconstructed series*. However, one can also choose to form  $\mathbf{X}_i$  from groups of eigentriples, and then obtain reconstructed series. For example, one may pair the eigentriples, then each eigentriple used to form a matrix (by forming outer products and scaling with the corresponding singular value) and a pair of matrices added together to form one  $\mathbf{X}_i$ . Indeed, this grouping approach is what we employ in this paper. Therefore, the result is the decomposition of the original time series into a sum of reconstructed series as shown in equation (7).

$$x[n] = \sum_{i=0}^h \tilde{x}_i[n], \quad n = 0, 1, \dots, N - 1, \quad (7)$$

where  $\tilde{x}_i$  is the reconstructed series obtained from  $\mathbf{X}_i$  and  $h$  depends on the grouping of eigentriples. For elementary grouping,  $h = L - 1$

<sup>1</sup>The procedure is technically anti-diagonal averaging; however, we use the term ‘diagonal averaging’ to be consistent with the SSA literature.

as there can only be a maximum of  $L$  singular values for a matrix of dimensions  $L \times K$ . Depending on the selection of eigentriples, the reconstructed time series reveals the corresponding aspect of the data. For example, the first eigentriple contains information on the trend, while the subsequent ones are associated with oscillatory patterns. Higher eigentriples are typically associated with noise.

## 4 SSA OF PERIODIC TIME SERIES

In this section, we obtain the mathematical form of SSA when applied to a time series containing periodic data. We first consider the case where we have an ideal periodic time series. We find that in this case, the singular vectors obtained are sinusoidal in nature. The case where a periodic series is corrupted by multiplicative element is then considered. We also perform simulations to validate the algebra.

### 4.1 SSA applied to an ideal periodic time series

Consider a time series  $x[n]$ ,  $n = 0, 1, \dots, N - 1$  of length  $N$  that is strictly periodic with a period given by  $L$ .

$$x[k + pL] = x[k]. \quad (8)$$

We also assume that the length  $N$  of the time series obeys the relation  $N = zL - 1$ , where  $z$  is an integer. Forming a trajectory matrix out of this sequence gives the following matrix.

$$X = \begin{bmatrix} x[0] & x[1] & \dots & x[L-1] \\ x[1] & x[2] & \dots & x[0] \\ \dots & \dots & \dots & \dots \\ x[L-1] & x[0] & \dots & x[L-2] \end{bmatrix}_{L \times K}. \quad (9)$$

As  $K$  is an integer multiple of  $L$ , the trajectory matrix can be interpreted as a block matrix, and partitioned into  $M = \frac{K}{L}$  square submatrices as given in equation (10), where  $X_{ac}$  is an anti-circulant matrix of dimensions  $L \times L$ .

$$X = [X_{ac} \ X_{ac} \dots]_{L \times M}. \quad (10)$$

$$X_{ac} = \begin{bmatrix} x[0] & x[1] & \dots & x[L-1] \\ x[1] & x[2] & \dots & x[0] \\ x[2] & x[3] & \dots & x[1] \\ \dots & \dots & \dots & \dots \\ x[L-1] & x[0] & \dots & x[L-2] \end{bmatrix}. \quad (11)$$

Before proceeding further, some observations can be made.

(i) We find that  $\text{rank}(X) \leq L$ , as there can only be a maximum of  $L$  linearly independent columns in  $X$ . Consequently, there can be a maximum of  $L$  non-zero singular values for  $X$ .

(ii) The anti-circulant matrix form given in equation (10) is real symmetric. Therefore, it has an eigendecomposition given as  $X_{ac} = Q\Lambda Q^T$ , where  $Q$  is an  $L \times L$  orthonormal matrix of eigenvectors and  $\Lambda$  is an  $L \times L$  matrix of the corresponding eigenvalues.

This enables us to write the trajectory matrix as a block matrix  $X = [Q\Lambda Q^T \ Q\Lambda Q^T \dots]$ . Without going through the pedagogical details, we state that  $X$  can be decomposed into three matrices as given below.

$$X = Q_{L \times L} [\sqrt{M}\Lambda \ \mathbf{0} \ \dots]_{L \times K} \frac{1}{\sqrt{M}} \begin{bmatrix} Q^T & Q^T & \dots \\ Q^T & Q^T & \dots \\ \dots & \dots & \dots \end{bmatrix}_{K \times K}, \quad (12)$$

$$= U_{L \times L} \Sigma_{L \times K} V^T_{K \times K}, \quad (13)$$

where, in the last step we recall the SVD of  $X$  to facilitate a direct comparison.

We now attempt to find the nature of this decomposition – specifically what one can expect from the singular vectors if the time series is periodic in nature. From equation (12), it can be noted that the singular values are a scaled version of the eigenvalues of anti-circulant matrix  $X_{ac}$ . Inspecting the block matrix representation in equation (12), it can also be seen that the matrix that forms to the left and right singular vectors is the matrix of eigenvectors  $Q$ . Therefore, all that we need to know is the nature of  $Q$ . A calculation of this is provided in Appendix A, where it is shown that  $Q$  consists of sinusoidal eigenvectors if the data are periodic. Therefore, the left singular vectors – which are exactly the eigenvectors contained in the matrix  $Q$  – are all sinusoidal. Besides, the vectors in  $Q$  are all periodic in  $L$  and thus when arranged as blocks in the right singular vector matrix, they form continuous sinusoids. Thus, the equivalence of decompositions in equation (12) shows that *when SSA is applied to a strictly periodic sequence, the singular vectors obtained are sinusoidal*.

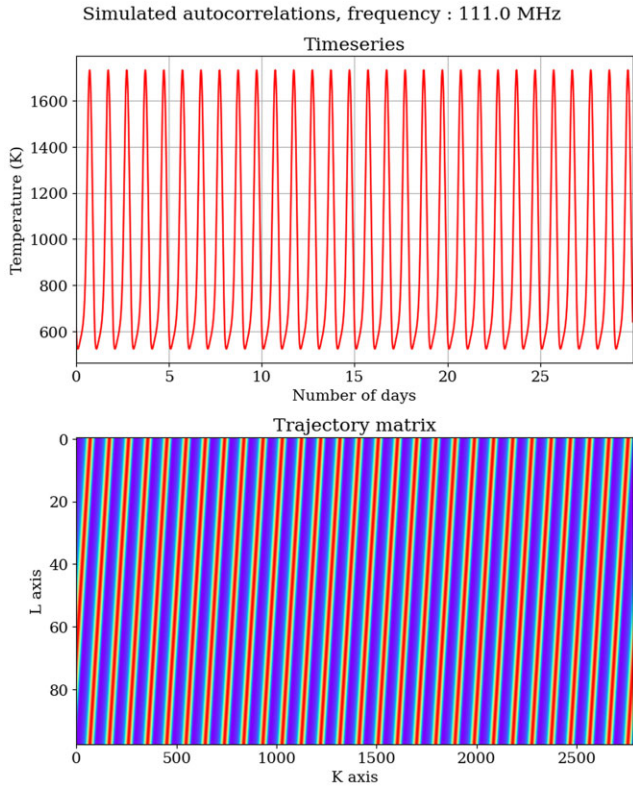
In Appendix A, it is also shown that the eigendecomposition of anti-circulant matrices such as  $X_{ac}$  has a deep relation to discrete Fourier transforms (DFT). Indeed, when applied to an ideal periodic series, SSA results are directly comparable to DFT of one cycle of periodic data; this is also demonstrated with simulations in Section 4.1.1. However, if the data have aperiodic structures, DFT of the entire span of data is not a suitable choice as aperiodic structures are not well represented in a sine/cosine Fourier basis. In such a scenario, SSA is better suited to reveal aperiodic structures in the data. Therefore, regardless of the nature of data, SSA can be applied to them and if the data are periodic, SSA reduces to a form of Fourier analysis.

Even though the decomposition given may be treated as an SVD of the trajectory matrix, some caution has to be exercised. Since the singular vector matrices have to be unitary, the right singular matrix gets divided by a scaling factor of  $\sqrt{\frac{K}{L}}$ , while the singular values get multiplied the same factor. Also, the left singular matrix in SVD is unitary, while the eigenvector form given in Appendix A has unit amplitude. The corresponding scaling applies to the singular value; however, it is inconsequential when the eigendecomposition and SVD are computed with numerical packages. Moreover, eigenvalues can be positive and negative, and as given in Appendix A, they occur in positive and negative pairs in this context. However, singular values are always non-negative and therefore the signs of the eigenvalues get applied to the singular vectors when equated with the SVD given in equation (12). Also, we assume that both the singular values and eigenvalues (and corresponding vectors) have been ordered in the same fashion – typically in descending magnitude.

#### 4.1.1 Simulations

To validate the above calculations, we analyse a simulated periodic time series. Since our aim is to gain a better understanding of SSA as applied to radiometric time series data, the simulations are beam multiplied sky temperature as would be seen by a radiometer. The simulation methodology used to obtain the time series is described in Thekkepattu et al. (2022), which we have extended to yield a time series spanning 30 days. For this, we make an important assumption that the radio sky is static and therefore, the true sky temperature is exactly the same across all days for each local sidereal time. The simulations are for auto-correlations at a frequency of 111 MHz with a time cadence of 15 min of sidereal time. Since the sky has a periodicity of one sidereal day, it is important that the cadence is chosen in sidereal time units.



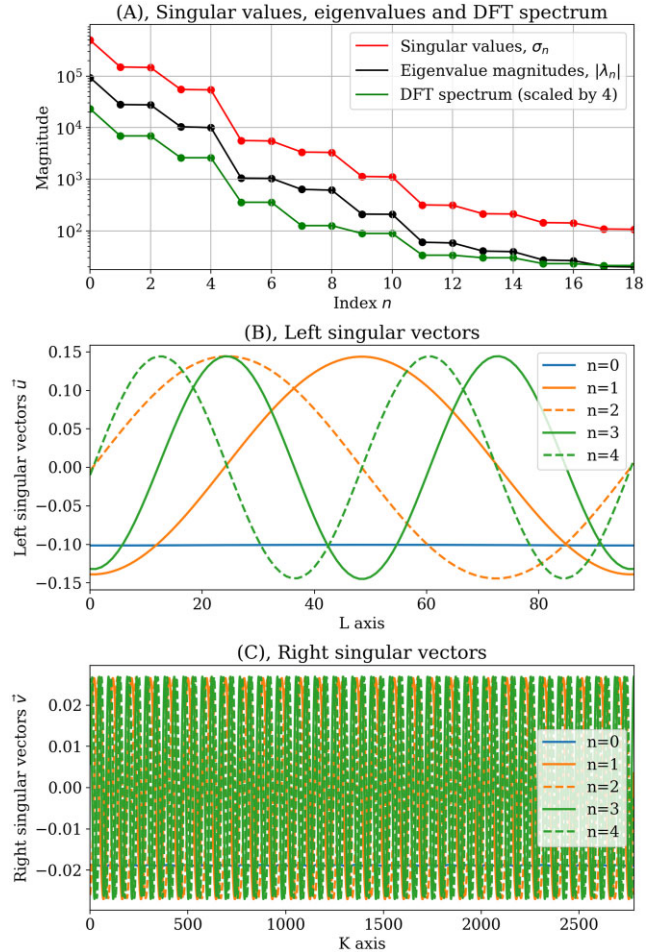


**Figure 2.** Simulated time series at a frequency of 111 MHz and the corresponding trajectory matrix formed by choosing an embedding dimension  $L = 96$ .

The simulation yields  $N = 2879$  data points that are then converted into a trajectory matrix. Key to embedding data is selection of an appropriate window length. Typically, a window length is chosen such that it is divisible by the fundamental of the known periodicity. As we have prior information that data have a periodicity corresponding to a sidereal day, a choice of the window length is the number of samples that correspond to one sidereal day. Therefore a suitable embedding dimension is  $L = 96$ ; as it corresponds to one sidereal day. This gives a matrix of dimensions  $96 \times 2784$  which can be partitioned into  $M = \frac{K}{L} = 29$  block matrices. The simulated time series and a representation of the trajectory matrix are shown in Fig. 2.

The trajectory matrix is then decomposed with SVD. The results from the SVD are shown in Fig. 3. Independently, an  $L \times L$  submatrix of the trajectory matrix (with  $L = 96$ ) is eigendecomposed. The eigenspectrum of this anti-circulant matrix is also given in Fig. 3. Moreover, the DFT of the underlying periodic sequence of length  $L = 96$  is also computed and plotted in the same figure with the DFT values sorted in descending order. As the sequence is real, the DFT spectrum is Hermitian and the sorted values appear twice in the spectrum.

Several observations can be made from Fig. 3. The spectrum of singular values matches the spectrum of eigenvalue magnitudes exactly, except for a scaling (which is  $\sqrt{29}$ ). The spectrum of eigenvalue magnitudes matches exactly the DFT magnitude spectrum, thereby validating the results from Appendix A. The same scaling is also evident in the amplitudes of the right singular vectors. We also find that the singular vectors, except the  $n = 0$  component, are purely sinusoidal and occur in sine-cosine pairs as the calculations showed. We have also verified that the vector pairs are indeed orthogonal by calculating the inner product between such pairs. The singular vectors

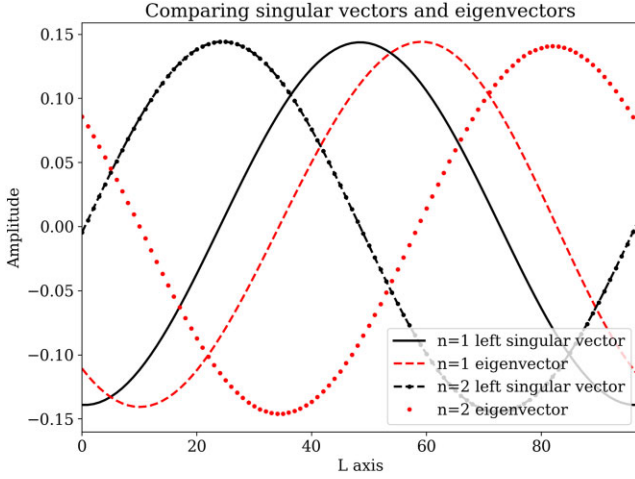


**Figure 3.** SVD of the trajectory matrix. Panel (A) shows singular values in a semilogarithmic scale. The same panel shows the eigenvalues of an  $L \times L$  submatrix as well as the DFT spectrum of the periodic sequence. The DFT spectrum has been sorted according to descending magnitude and artificially scaled by a value of 4 to make it distinguishable from the eigenspectrum. For clarity, only the first 19 values are plotted. Panel (B) shows the first five left singular vectors while Panel (C) shows the first five right singular vectors. In both plots, the orthogonal (sine-cosine) vector pairs are plotted with the same colour but with different line-styles.

are also periodic in  $L$  while the  $n = 0$  component is essentially the DC component of the data, similar to the zeroth component in Fourier transforms.

In Fig. 4, we compare the first pair of orthogonal left singular vectors with the corresponding pair of the eigenvectors. We find that the eigenvectors are exactly the same as the singular vectors, except for a sign reversal in one of the vectors in the pair. It has been verified that the corresponding eigenvalue carries a negative sign; thus, it is inconsequential to the overall analysis, and we can safely consider the equivalence between SVD and eigendecomposition in equation (12) to be valid.

Thus, we find that the singular vectors obtained from SSA of a periodic sequence are purely sinusoidal with them occurring in orthogonal pairs. It may also be noted that while such orthogonal features have been noticed in the SSA literature (see for e.g. Ghil et al. (2002)), we mathematically demonstrated using simple arguments why such orthogonal sinusoidal pairs are to be expected in the SSA of an ideal time series. We now inspect the reconstructed series given in Fig. 5. As we know that the singular vectors and



**Figure 4.** A comparison between the left singular vectors of the trajectory matrix and eigenvectors of a square submatrix. Only the vectors from the first orthogonal pair in each case are plotted.

corresponding values occur in pairs, we group the eigentriples into pairs (except the 0<sup>th</sup> component) and apply the diagonal averaging. As can be seen, the reconstructed series are also purely sinusoidal since the vectors are sinusoidal. Another important observation with Fig. 5 is the 0<sup>th</sup> component, which is often called ‘trend’ in SSA literature. For radiometric data this term corresponds to the sky-averaged component of the radio data and is the relevant observable for global 21-cm research. Since the origin of the 0<sup>th</sup> reconstructed component can be traced back to 0<sup>th</sup> eigenvector of the matrix  $\mathbf{Q}$ , we can see that the corresponding Fourier component has a frequency of zero. In other words for an ideal periodic time series, the 0<sup>th</sup> component in SSA is related to the zero frequency or DC term in the Fourier spectrum, such that its singular value is the same as the Fourier DC term.

In the above exercise, it is essential that the embedding dimension (window length) corresponds to exactly one sidereal day. If not, the decomposition will not lead to sinusoidal patterns of appropriate periodicity, and the various relations that we obtained between SVD, eigendecomposition, and Fourier analysis become invalid. With this important result, we now proceed to introduce some non-idealities into the time series.

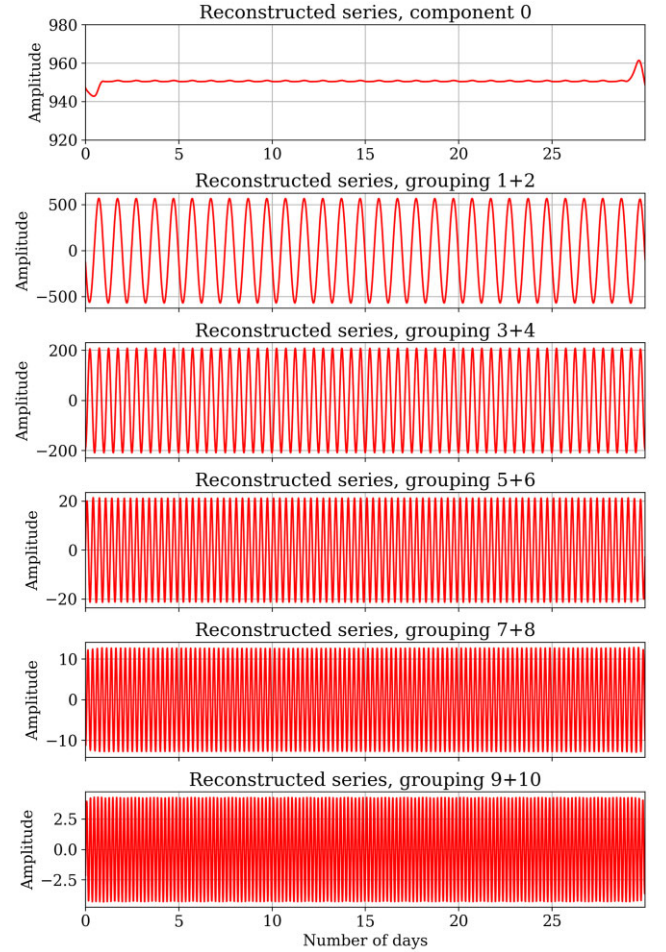
#### 4.2 SSA applied to periodic time series with time-varying gains

The calculations and simulations so far assume a case where the radiometer system has been fully calibrated to yield data calibrated to a reference plane outside the Earth’s ionosphere. Observational radiometric data have multiple non-idealities, the dominant ones are additive components such as receiver noise temperatures and multiplicative factors that are often called gains. Both of them can be time-varying and therefore are required to be known to correct the data. We now incorporate them into our model.

Assuming a simple system model for SITARA data (see Section 5.2 of Thekkepattu et al. (2022)), the trajectory matrix of the measured data can be expressed as a Hadamard product, given in equation (14).

$$\mathbf{X}' = (\mathbf{X}_s + \mathbf{R}) \odot \mathbf{G}, \quad (14)$$

where  $\mathbf{X}_s$  is the sky temperature trajectory matrix,  $\mathbf{R}$  is the receiver noise temperature trajectory matrix, and  $\mathbf{G}$  is the trajectory matrix of time-varying system gains. Since the low-frequency radio sky is



**Figure 5.** Reconstructed series from SSA of an ideal periodic series. The 0<sup>th</sup> component and the succeeding five grouped orthogonal pairs are given in this figure. Since the reconstructed series add up to the original time series, their x-axis is the number of days.

bright, the dominant time-varying component in the data is due to sky temperature multiplied with time-varying system gains. Therefore, we can apply a simplifying assumption that the receiver noise temperature is constant with time and absorb it into the ‘ideal’ sky matrix, and the sum can be written as  $\mathbf{X}$ . Thus we write

$$\mathbf{X}' = \mathbf{X} \odot \mathbf{G}. \quad (15)$$

While equations (14) and 15 may be written without resorting to a matrix formulation, expressing them with trajectory matrices enables SSA.

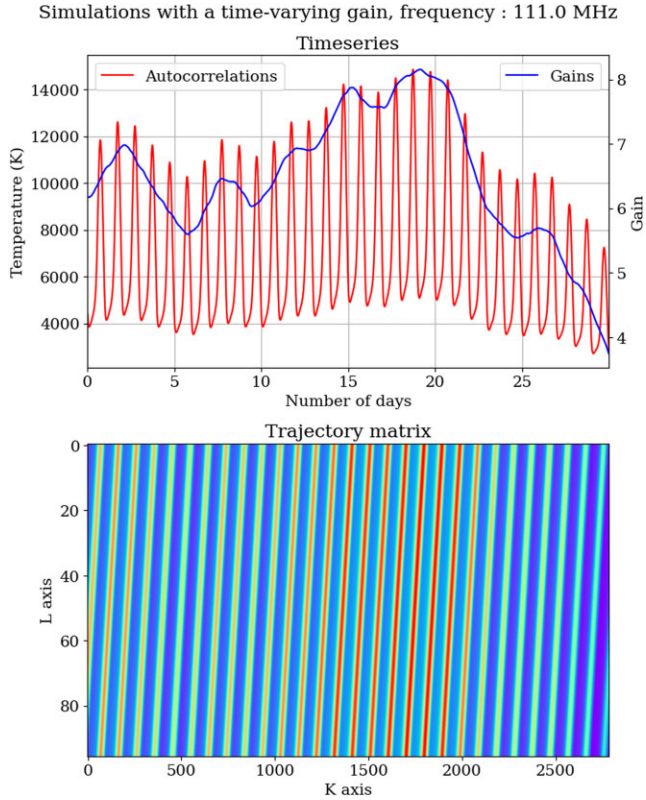
Based on the summation of  $\mathbf{X}$  from equations (5) and (15), the following relations can be derived.

$$\begin{aligned} \mathbf{X}' &= \sum_i \mathbf{X}'_i, \text{ and} \\ \mathbf{X}' &= \left( \sum_i \mathbf{X}_i \right) \odot \mathbf{G}, \\ &= \sum_i (\mathbf{X}_i \odot \mathbf{G}), \end{aligned} \quad (16)$$

where we have used the distributive property of Hadamard products. Thus we have,

$$\sum_i \mathbf{X}'_i = \sum_i (\mathbf{X}_i \odot \mathbf{G}). \quad (17)$$



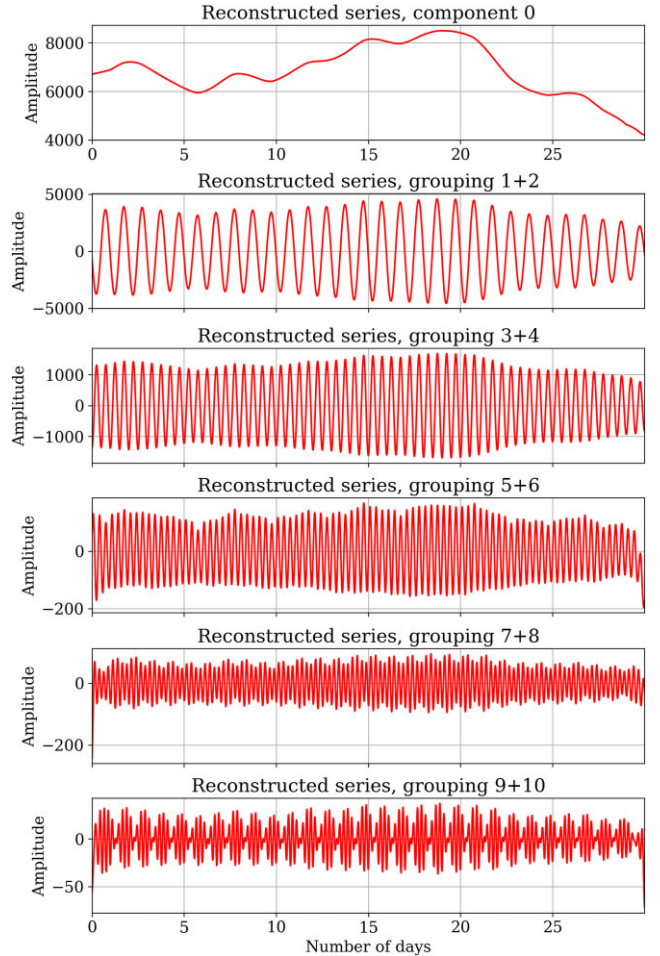


**Figure 6.** Simulated time series with gains at a frequency of 111 MHz and the corresponding trajectory matrix.

Though equation (17) appears trivial, it reveals a useful aspect of the decomposition. Applying diagonal averaging to each  $X'_i$  leads to gain multiplied reconstructed series, as the Hadamard product is an element-by-element multiplication. However, we already know the ‘true’ reconstructed series to be purely sinusoidal from Section 4. The multiplication of a pure sinusoid waveform by a signal of much lower frequency results in amplitude modulation of the sinusoid, a process that is called amplitude modulation (AM) in communication systems (Haykin & Moher 2008). Therefore, we may treat the gain multiplied reconstructed series as a form of AM signals with the sinusoids acting analogous to ‘carriers’ and using AM demodulation techniques to extract the temporal variations in gains.

Before proceeding further, it is instructive to verify the above calculations with simulations. For this, we begin with the basic time series from Section 4. To this series, we add a constant of 100 K as the receiver noise temperature and simulated radiometric noise (integration time of 15 min and bandwidth of 61 kHz) for the overall system temperature. The gain profile is modelled as a random-walk, which is generated as a cumulative sum of Gaussian white noise. The resulting profile is subsequently smoothed with box-car averaging to suppress variations of time cadence less than one sidereal day. The gains are also normalized to avoid negative excursions. The time series is then multiplied with the simulated gains and the resulting series and the associated trajectory matrix are shown in Fig. 6. It may be noted that the dimensions of the matrices and time series have all been kept the same.

The trajectory matrix is then decomposed with SVD and the eigen-triples grouped in the same manner as in Section 4 and diagonally averaged. It is easy to notice that the resulting reconstructed series shown in Fig. 7 are indeed amplitude-modulated sinusoids.



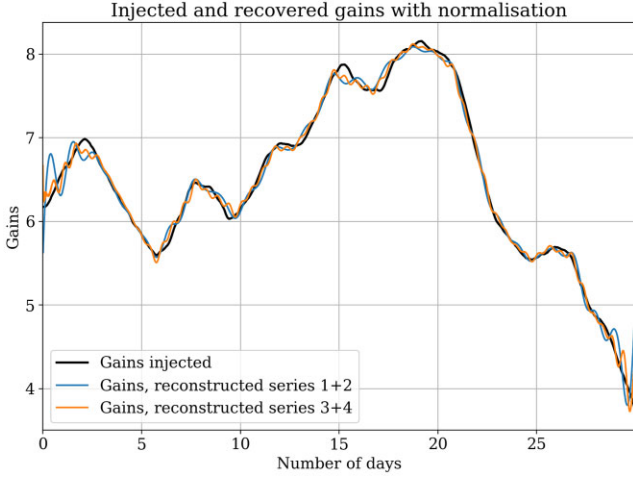
**Figure 7.** Reconstructed series from SSA of a simulated series with time-varying gains. The 0<sup>th</sup> series and the succeeding five components from grouping orthogonal pairs are given in this figure.

Several algorithms exist to demodulate such AM signals to obtain their envelopes. Here we use an algorithm that is mathematically simple to interpret. The first step in this algorithm is converting a real-valued signal into an analytic signal, which can be accomplished with Hilbert transforms. Taking the magnitude of this analytic signal yields the modulation envelope. We apply this procedure to the series given in Fig. 7. The resulting gain profiles are plotted in Fig. 8, in which the curves are also normalized according to the method described in Section 4.3.

Fig. 8 validates the interpretation of equation (17) that the reconstructed series have identical gain templates that can be recovered with amplitude demodulation. Indeed, a major outcome of this exercise is that the gain template for the DC component is identical to those of the periodic components. Therefore the time-variations in the DC term may be corrected for by using the gain templates from periodic components. This leads us to a novel calibration strategy for global 21-cm radiometers that can correct long-term drift and related systematics.

### 4.3 On the use of SSA to aid calibration

The recovered gains require normalization to enable a comparison with the injected gains and make them useful for calibration. We begin with equation (17) and incorporate equation (5) as shown in



**Figure 8.** Injected and recovered gains. The recovered gains have been normalized according to equation (22). The recovered gains from periodic components resemble the injected gains.

equation (18), where we also include the eigentriple grouping. The indices have also been modified for convenience.

$$X'_i = \mathbf{G} \odot \sum_{j=0}^1 \sigma_{i+j} (\vec{u}_{i+j} \otimes \vec{v}_{i+j}), i = 1, 3, \dots,$$

$$X'_i = \sigma_i \mathbf{G} \odot \sum_{j=0}^1 (\vec{u}_{i+j} \otimes \vec{v}_{i+j}), i = 1, 3, \dots, \quad (18)$$

where we used the property that singular values are identical for the eigentriples in a grouped pair. The vectors  $\vec{u}$  and  $\vec{v}$  are from unitary matrices and therefore their individual inner products equal to unity. However, for subsequent analysis, these vectors need to be normalized to have unity *amplitude* when diagonally averaged after grouping. The normalization can be achieved with the following operations, where the sinusoidal terms in the brackets have been normalized to have unity amplitude.

$$\vec{u}_i = \frac{1}{\sqrt{L/2}} (\vec{u}_i \sqrt{L/2}),$$

$$\vec{v}_i = \frac{1}{\sqrt{K/2}} (\vec{v}_i \sqrt{K/2}), \quad (19)$$

as the lengths of the vectors  $\vec{u}$  and  $\vec{v}$  are  $L$  and  $K$ , respectively. The grouped and diagonally averaged reconstructed series is given as

$$\tilde{x}_i[n] = \frac{\sigma_i}{\sqrt{L/2}\sqrt{K/2}} g[n] c_i[n], \quad (20)$$

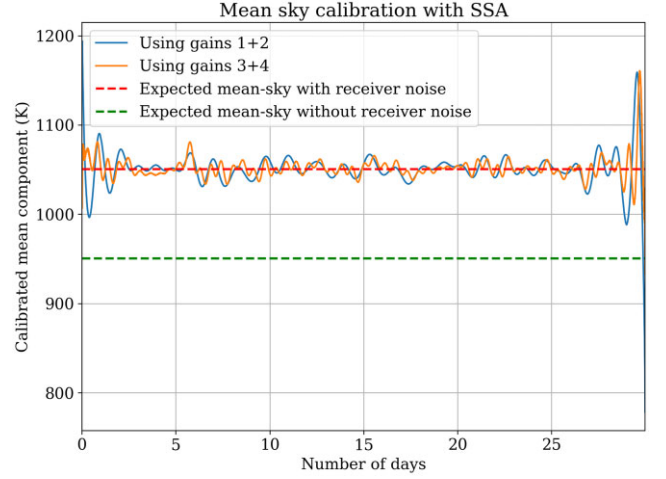
where  $g[n]$  is the gain series and  $c_i[n]$  is the ‘carrier’ sinusoid. The normalization in equation (19) ensures unity amplitude for  $c_i[n]$ . Upon demodulation of a series  $\tilde{x}_i[n]$ , the carrier is removed and a modulation envelope  $g_{i,t}[n]$  is obtained as

$$g_{i,t}[n] = \frac{\sigma_i}{\sqrt{L/2}\sqrt{K/2}} g'_i[n]. \quad (21)$$

From these envelopes, gain templates  $g'_i[n]$  can be recovered as

$$g'_i[n] = g_{i,t}[n] \frac{\sqrt{L/2}\sqrt{K/2}}{\sigma_i}. \quad (22)$$

For equation (22) to be useful for calibration, the only auxiliary information required is the singular value  $\sigma_i$ . This is the singular value of the component with the same periodicity of that we expect from the true sky and can be obtained from the SSA of the simulated



**Figure 9.** Calibrating the mean sky component using recovered gains from periodic components. For comparison, the expected mean levels with and without the 100 K receiver noise temperatures are also plotted.

ideal sky time series from Section 4. Using those singular values, we obtain the recovered and appropriately normalized gains that are plotted in Fig. 8.

We are now in a position to apply those gains to the 0<sup>th</sup> reconstructed series to achieve calibration and compare the calibration with the expected levels. The results are given in Fig. 9.

We find that the mean-sky component has been calibrated solely using the periodic component(s) of the sky with the aid of SSA. It has to be noted that the prescription provided here to calibrate the mean-sky relies only on the singular values of the simulated periodic components. Even without using the singular values, and only using the data, it is possible to achieve relative gain calibration, while applying gains with appropriate singular values to the mean-sky component establishes its proper brightness temperature scale.

Let us now try to obtain an intuitive appreciation of this calibration. The trend component obtained from SSA is essentially a box-car averaging of time series data. If the window used for averaging encompasses exactly one sidereal day, the periodic variations caused by the Galaxy transiting through the beam are averaged out. If the system were perfect the result would be a constant temperature, which is the sum of mean-sky and receiver noise temperatures. Thus the trend that we observe in the non-ideal system may be modelled by a constant value multiplied with the time-varying gains.

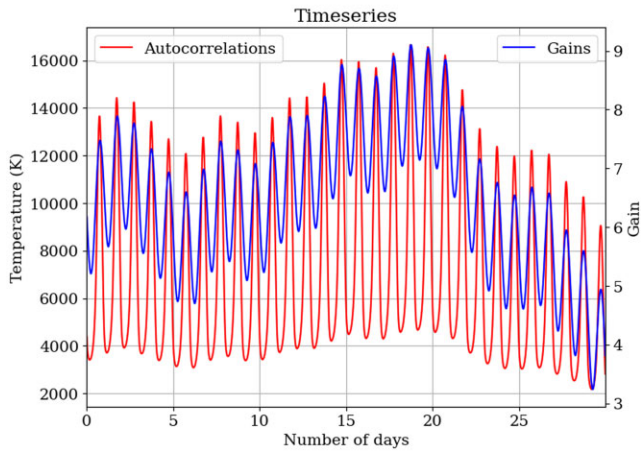
The gain variations that we have studied so far have smooth evolution over time with no periodicity. However, for real radio telescope systems that do not have temperature regulation, gain variations with local temperature are expected—details given in Section 5. Indeed, the rising and setting of the Sun can heat and cool the components in the analogue signal chain and induce diurnal variations in the gain that may be correlated with the sky drift. Yet another potential origin for diurnal patterns in gains is the ionosphere. Therefore, it is imperative to study such a scenario, which we perform next.

#### 4.4 SSA with diurnal gain variations

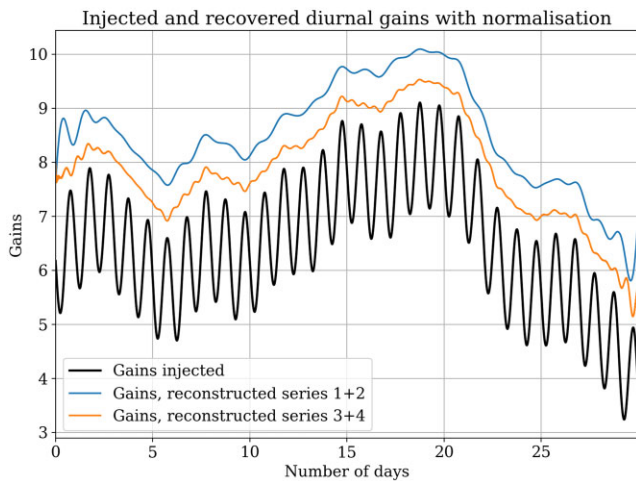
To the gains simulated in the previous section, we add a sinusoidal diurnal gain component. For a time series that spans a month, the



Simulations with diurnal gain variations, frequency : 111.0 MHz



**Figure 10.** Simulated radiometer data with gains that have a smoothly varying component and a diurnal component.



**Figure 11.** Injected and SSA-recovered gains when the gains have a diurnal component. Appropriate normalization has been applied.

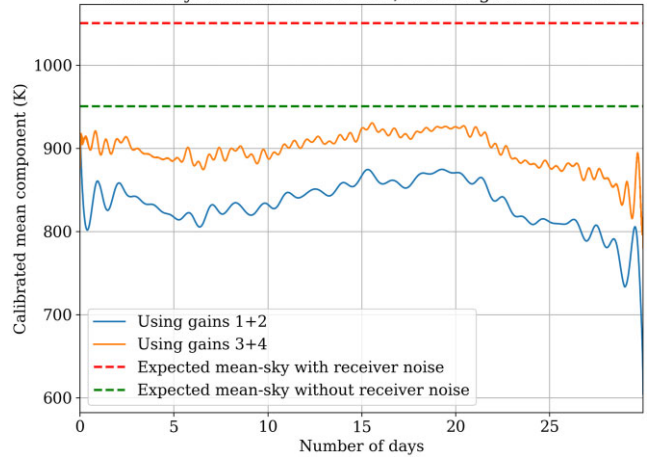
diurnal gain variations are expected to show a strong correlation with Galaxy transit. For longer time series, the difference between sidereal time and civil time reduces this correlation. The modified gains and the resulting time series are shown in Fig. 10.

We perform SSA in the exact same manner as in the previous case, and recover normalized gains, as shown in Fig. 11.

It is interesting to note that the correlation between gains and sky patterns reduces the separability of the two, and the recovered gains from the reconstructed series differ from one another. We attempt a calibration of the mean component with these gains and the results are given in Fig. 12.

The differences in the recovered gains in Fig. 11 as well as the calibrated temperatures in Fig. 12 demonstrate a fundamental limitation of the SSA technique in calibrating data when the gain variations are correlated with sky patterns. Lacking auxiliary information related to the cause of such gain variations, for example temperature measurements, the calibration technique fails and yields erroneous results as demonstrated. On the other hand, a disagreement between the recovered gains can be used as a practical indicator of gain variations that have frequencies coinciding with the periodicity induced by sky drift, thus pointing to a necessity for subsidiary

Mean sky calibration with SSA, diurnal gain variations



**Figure 12.** Application of recovered gains from periodic components to calibrate the mean-sky term, when the injected gains have a diurnal component.

information. In other words, differences in recovered gains between the groupings indicate presence of correlations between the gains and sky drift. It may also be noted that no *a priori* information is required to infer such a correlation. Having demonstrated the various cases of SSA, we now apply these techniques to data from SITARA observations.

## 5 SSA APPLIED TO SITARA DATA

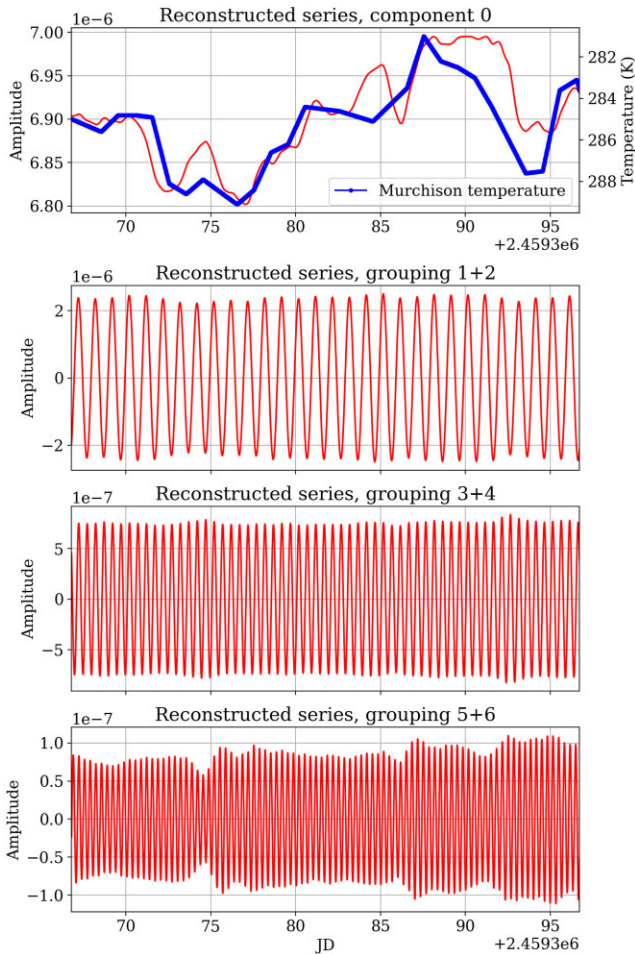
### 5.1 Data preparation

We use a time series of SITARA spectral data from the concatenated 30 days of 2021 June data. The individual data files in *miriad* format are converted to *hdf5* format and concatenated using custom tools written in *python*. From the concatenated dataset, auto-correlation data for a frequency of 111 MHz are extracted, as it is a frequency with relatively low RFI occupancy. The major source of RFI observed during this period was Solar radio bursts that are transient in nature. This raw data time series as a function of Julian Date (JD) is shown in Fig. 1.

The raw SITARA data spanning a month have about 749 000 points, which makes calculations difficult. Therefore, the data are binned into the same number of bins (2879) as done in the simulations, giving one sample per 15 min or 96 samples per sidereal day. Once again, in performing binning it is important to use sidereal time—and not Julian time—to ensure an equal number of samples for each cycle of sky drift. The data are then embedded into a trajectory matrix and SSA is applied in the same manner as before. The eigentriples are grouped and reconstructed series formed.

Fig. 13 shows first four of the reconstructed series. The trend plot also has a low-time cadence record of the temperature recorded within the Murchison Shire, which is available from NCEI-NOAA Integrated Surface Dataset (Lott et.al. 2001).<sup>2</sup> The temperature scale has been inverted, as the trend shows an anti-correlation with temperature. Further, the radio data and temperature data fall into two different families with different sampling, and therefore the temperature data have been interpolated to the same Julian day bins as the radio data. The Pearson correlation coefficient computed between

<sup>2</sup> <https://www.ncei.noaa.gov/access/search/index>



**Figure 13.** Reconstructed series from SITARA data SSA. A plot of the physical temperature recorded within the Murchison Shire is included in the trend plot to show the anti-correlation between them.

the trend and the temperature data is  $-0.83$ , which denotes a strong anti-correlation between the trend and temperature.

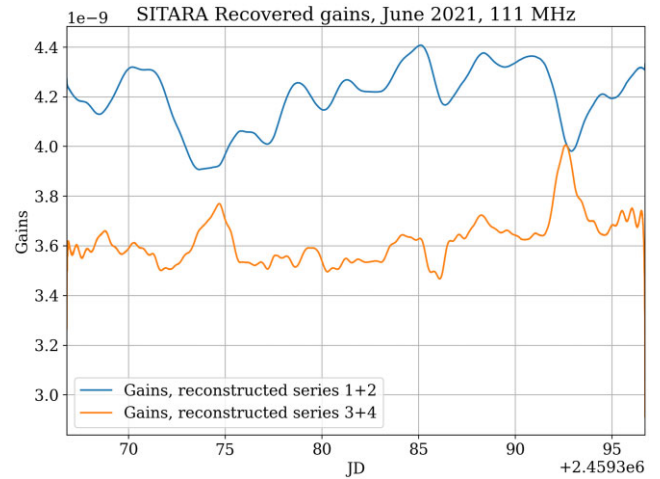
Apart from component ageing effects, there are two major reasons for such temperature-induced variations in radiometric data.

(i) The noise temperatures of active devices used in amplifiers increase with physical temperature.<sup>3</sup> The same holds true for passive components such as attenuators. If this is the cause of drift, the pattern so obtained is expected to be *correlated* with physical temperature.

(ii) The gain of amplifiers reduces as temperature is increased. In this case, the trend pattern and physical temperature would be *anti-correlated*.

It is therefore evident from Fig. 13 that the major contributor to the trend is temperature-induced gain variations. The receiver noise would inevitably vary as a function of physical temperature; however, when the overall system temperature is sky-dominated the impact of this would be secondary to gain variations. This informs our choice of calibration model given in Section 4.3, where we assume a constant receiver noise temperature and a time-varying gain.

Subsequently, the gain patterns are recovered from the reconstructed series with the demodulation technique. These recovered



**Figure 14.** Recovered gains from SITARA data, 2021 June. The gains differ across reconstructed series, pointing to potential diurnal gain variations.

gains for the first two periodic components are shown in Fig. 14. As can be seen, the normalized gains are different between the series, thus pointing to potential diurnal gain variations. Therefore, we refrain from applying these gains to the mean component to establish its brightness temperature scale.

## 6 DISCUSSION

In this paper, we developed the mathematical framework for SSA of a radiometric time series and demonstrated its application in analysing radiometric time series data from a radio telescope at a frequency with relatively low RFI. A major outcome of this work is a novel method to calibrate the mean-sky component of radio data using the periodic component of sky-drift patterns. Calibration using differential measurements of the sky to obtain the system gain has been successfully utilized in different low-frequency experiments (Rogers et al. 2004; Singh et al. 2017; Thekkeppattu et al. 2022). However, those applications were limited, as the calibration was performed on a day-to-day basis, assuming a constant value for the system gain for a specific day. With SSA, the periodic nature of data is leveraged to obtain gains that can vary with time, provided they do not have a diurnal component. Seen in this light, SSA can be viewed as an extension of the aforesaid technique.

Let us examine the implications of the gain recovery in Fig. 8. The gain patterns that we recovered are purely from the periodic components of the sky and the singular values of simulated periodic components. The procedure does not need simulations that include the mean-sky or the *zero-point* of the sky maps. Indeed, zero-point errors have been noticed in sky maps, since these maps are often products of combination of sky surveys conducted with different telescopes. Disagreements exist between estimates of the zero-points of sky maps across experiments, for example, Monsalve et al. (2021) report a different zero-point offset for the 150 MHz sky map (Landecker & Wielebinski 1970) compared to the measurements of Patra et al. (2015). Therefore, the methods outlined in this paper can be applied to experiments where zero-point levels have to be accurately known to verify that the gains do not have variations that can confuse sky signals. However, to get accurate estimates of system gains the input sky maps need to get an accurate temperature scale even though zero-points errors are acceptable. An associated caveat is that while SSA can provide multiplicative system gains, the

<sup>3</sup> It may be noted that this is one of the reasons for cryogenic cooling of radio telescope front-ends to achieve a low overall system temperature.

additive receiver noise temperature is practically indistinguishable from the mean-sky component. Therefore, experiments requiring accurate measurements of zero-level still need additional calibration to estimate the receiver noise temperature. Nonetheless, as demonstrated in this paper, SSA can be utilized as a diagnostic tool for such experiments.

### 6.1 Caveats and future work

Since this is the first work exploring SSA for radiometric data analysis and calibration, we have kept the analysis simple. Specifically, low-frequency radio data can have RFI and we ignored the flagging that has been performed. Flagging considerations lead to gaps in a time series dataset and may necessitate data in-painting or interpolation. Such explorations will be taken up in future work.

We have developed the mathematical background and demonstrated an application of SSA for auto-correlations at a single frequency, while SITARA has nearly 2000 usable frequency channels in auto-correlations and cross-correlations each. To use all the frequency channels and to consider complex cross-correlations, the mathematical framework developed in this paper has to be expanded. For a proper treatment of multiple frequency channels, multivariate SSA techniques will be explored in future work.

While we demonstrated via simulations a calibration technique to remove gain variations that evolve smoothly with time, the application of it to SITARA data is hampered by diurnal gain variations. This shows the necessity to maintain temperature regulation or inclusion of thermometers at points in the signal chain to track such gain variations—this will be incorporated in future revisions of the SITARA system. Additionally, the data in frequency channels below 50 MHz, where receiver temperature dominates over sky temperature, show diurnal periodicity arising from such gain variations that may be exploited as a template to improve calibration.

### 6.2 A potential application of SSA for space-based 21-cm experiments

As we demonstrated, a major limitation with the SSA technique for calibration is the confusion between diurnal gain variations and sky drift, as both have approximately the same periodicity. If the periodicities can be made to differ, the calibration can be improved significantly. Specifically, if the sky drift can be made faster than the gain variations, substantially better calibration can be expected. While it is difficult for ground-based experiments to introduce such a separation, a radiometer payload on a spin-stabilized satellite can have beams that rapidly scan the sky, thereby increasing the sky-drift rate. For example, a broadband dipole antenna could be placed on a satellite spinning such that the nulls sweep the Galactic plane at a rate much faster than any gain variations, including the flicker noise of the electronic systems. Spinning a satellite at a rate faster than the  $1/f$  knee frequency of the radiometric system, to reduce the deleterious impact of gain fluctuations on the images, has been employed in CMB missions such as Planck (Bersanelli et al. 2010). As there are a few projects proposed, planned, or launched targeting the global 21-cm signal from a satellite platform such as DARE (Burns et al. 2012), Longjiang/Chang’e-4 (Jia et al. 2018), DSL (Chen et al. 2021), PRATUSH, etc., we opine that SSA would be an ideal tool for analysis and calibration of time series data from such radiometers in space.

## 7 CONCLUSIONS

In this paper, we introduced SSA as a powerful tool to study radiometric data. We showed the deep connections between the SSA techniques and Fourier transforms and leverage that to obtain long-term gain evolution from radio data. A novel technique to calibrate 21-cm experiments using periodicity in the sky-drift patterns has been proposed and simulated. The limitations of that technique in the presence of diurnal variations that can confuse with the sky drift is studied. Upon application of SSA to SITARA data, we find that the obtained decomposition has the features as expected, with the trend showing strong anti-correlation with temperature. However, the gains obtained point to diurnal variations as a limiting factor in using SSA for gain calibration of SITARA data.

## ACKNOWLEDGEMENTS

This work is funded through CT’s ARC Future Fellowship, FT180100321 – ‘Unveiling the first billion years: enabling Epoch of Reionisation science’. This work makes use of the *Murchison Radio-astronomy Observatory*, operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. JNT acknowledges Dr Adrian T. Sutinjo for his Linear Algebra over Lunch lectures. This work uses the following python packages and we would like to thank the authors and maintainers of these packages: `numpy` (Harris et al. 2020), `scipy` (Virtanen et al. 2020), `healpy` (Zonca et al. 2019), `astropy` (Astropy Collaboration et al. 2013, 2018), `aipy` (Parsons 2016), `matplotlib` (Hunter 2007), `ephem` (Rhodes 2011), `pygdsn` (Price 2016), and `h5py` (<https://www.h5py.org/>).

## DATA AVAILABILITY

The SITARA data used in this work will be made available based on reasonable requests. The temperature data used are publicly available Integrated Surface Dataset (Global) from NOAA National Centers for Environmental Information.

## REFERENCES

- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Bersanelli M. et al., 2010, *A&A*, 520, A4
- Broomhead D., King G. P., 1986, *Physica D*, 20, 217
- Burns J. O. et al., 2012, *Adv. Space Res.*, 49, 433
- Chen X., Yan J., Deng L., Wu F., Wu L., Xu Y., Zhou L., 2021, *Phil. Trans. R. Soc. A*, 379, 2019056633222649
- Davis P. J., 1979, *Circulant Matrices*. John Wiley, New York
- DeBoer D. R. et al., 2017, *Publ. Astron. Soc. Pac.*, 129, 045001
- Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, *IEEE Proc.*, 97, 1482
- Dokht R. M., Gu Y. J., Sacchi M. D., 2016, *Geophys. J. Int.*, 208, 1430
- Donskikh G. I., Ryabov M. I., Sukharev A. L., Aller M., 2016, *Astrophysics*, 59, 199
- Gaensler B., Beck R., Feretti L., 2004, *New Astron. Rev.*, 48, 1003
- Ghil M. et al., 2002, *Rev. Geophys.*, 40, 3–1
- Golyandina N., Korobeynikov A., 2014, *Comput. Stat. Data Anal.*, 71, 934
- Golyandina N., Nekrutkin V., Zhigljavsky A. A., 2001, *Analysis of Time Series Structure: SSA and Related Techniques*. CRC press, New York
- Gray R. M., 2006, *Found. Trends Commun. Inf. Theory*, 2, 155
- Gürel N. M., Hurlley P., Simeoni M., 2018, in 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Salt Lake City, Utah. p. 1983



- Hallinan G. et al., 2015, in American Astronomical Society Meeting Abstracts #225, p. 328.01
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Haykin S., Moher M., 2008, An Introduction to Analog and Digital Communications, 2nd Edition. Wiley, New York
- Hunter J. D., 2007, *Comp. Sci. Eng.*, 9, 90
- Jia Y., Zou Y., Ping J., Xue C., Yan J., Ning Y., 2018, *Planet. Space Sci.*, 162, 207
- Jolliffe I. T., 1986, Principal component analysis, Springer, Berlin
- Landecker T. L., Wielebinski R., 1970, *Aust. J. Phys. Astrophys. Suppl.*, 16, 1
- Lott N., Baldwin R., Jones P., 2001, NCDC Technical Report No. 2001-01
- Monsalve R. A. et al., 2021, *ApJ*, 908, 145
- Nambissan T. J. et al., 2021, *Exp. Astron.*, 51, 193
- Nindos A., Kontar E., Oberoi D., 2019, *Adv. Space Res.*, 63, 1404
- Olson B. J., Shaw S. W., Shi C., Pierre C., Parker R. G., 2014, *Appl. Mech. Rev.*, 66, 040803
- Parsons A., 2016, Astrophysics Source Code Library, record ascl:1609.012
- Patra N., Subrahmanyan R., Sethi S., Udaya Shankar N., Raghunathan A., 2015, *ApJ*, 801, 138
- Price D. C., 2016, Astrophysics Source Code Library, record ascl:1603.013
- Rhodes B. C., 2011, Astrophysics Source Code Library, record ascl:1112.014
- Rogers A. E. E., Bowman J. D., 2012, *Radio Sci.*, 47, RS0K06
- Rogers A. E. E., Pratap P., Kratzenberg E., Diaz M. A., 2004, *Radio Sci.*, 39, RS2023
- Sault R. J., Teuben P. J., Wright M. C. H., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E. eds, ASP Conf. Ser. Vol. 77, Astronomical Data Analysis Software and Systems IV. Astron. Soc. Pac., San Francisco. p. 433
- Singh S. et al., 2017, *ApJ*, 845, L12
- Singh S., Subrahmanyan R., Shankar N. U., Rao M. S., Girish B. S., Raghunathan A., Somashekar R., Srivani K. S., 2018, *Exp. Astron.*, 45, 269
- Thekkeppattu J. N., McKinley B., Trott C. M., Jones J., Ung D. C. X., 2022, *PASA*, 39, e018
- Tingay S. J. et al., 2013, *PASA*, 30, e007
- Trott C. M., 2017, *Proc. Int. Astron. Union*, 12, 92
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Vautard R., Ghil M., 1989, *Physica D*, 35, 395
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261 32015543
- Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *J. Open Source Softw.*, 4, 1298

## APPENDIX A: EIGENDECOMPOSITION OF ANTI-CIRCULANT MATRICES

The relation between circulant matrices and circular convolution, as well as the relation between the eigendecomposition of such matrices and DFT, are well known – see for e.g. Gray (2006) for an approachable introduction to this relation – and put to use in several signal processing applications. However, the matrices of relevance for this paper are anti-circulant and therefore we provide some useful relations between those matrices and DFT.

Consider a real-valued sequence  $x[n]$ ,  $n = 0, 1, \dots, N-1$ . This sequence can be used to construct an  $N \times N$  real-valued anti-circulant matrix as given in equation (A1).

$$\mathbf{X}_{ac} = \begin{bmatrix} x[0] & x[1] & \dots & x[N-1] \\ x[1] & x[2] & \dots & x[0] \\ x[2] & x[3] & \dots & x[1] \\ \dots & \dots & \dots & \dots \\ x[N-1] & x[0] & \dots & x[N-2] \end{bmatrix}. \quad (\text{A1})$$

The eigendecomposition of  $\mathbf{X}_{ac}$  can be written as

$$\mathbf{X}_{ac} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \quad (\text{A2})$$

,where

$$\mathbf{X}_{ac} \mathbf{y}_n = \lambda_n \mathbf{y}_n, \quad (\text{A3})$$

where  $\lambda_n$  are the eigenvalues and  $\mathbf{y}_n$  corresponding eigenvectors with  $n$  being the index. The following set of equations are obtained from eigendecomposition.

$$x[0]y_n[0] + \dots + x[N-1]y_n[N-1] = \lambda_n y_n[0],$$

$$x[1]y_n[0] + \dots + x[0]y_n[N-1] = \lambda_n y_n[1],$$

$$\dots$$

$$x[N-1]y_n[0] + \dots + x[N-2]y_n[N-1] = \lambda_n y_n[N-1], \quad (\text{A4})$$

which can be rearranged into,

$$x[0]y_n[0] + \dots + x[N-1]y_n[N-1] = \lambda_n y_n[0],$$

$$x[0]y_n[N-1] + \dots + x[N-1]y_n[N-2] = \lambda_n y_n[1],$$

$$\dots$$

$$x[0]y_n[1] + \dots + x[N-1]y_n[0] = \lambda_n y_n[N-1]. \quad (\text{A5})$$

The above set of equations can be recast into a summation equation as given in equation (A6).

$$\sum_{k=0}^{m-1} x[k]y_n[k-m] + \sum_{k=m}^{N-1} x[k]y_n[k-m+N] = \lambda_n y_n[m], \quad (\text{A6})$$

where  $m = 0, 1, \dots, N-1$ . The matrix is real symmetric and therefore the eigenvalues are real valued. The eigenvectors are fully real and we assume them to be  $N$  periodic as given in equation (A7).

$$y_n[k-m+N] = y_n[k-m]. \quad (\text{A7})$$

This simplifies equation (A6) into

$$\sum_{k=0}^{N-1} x[k]y_n[k-m] = \lambda_n y_n[m]; m = 0, 1, \dots, N-1. \quad (\text{A8})$$

It may be noted that the LHS of the above equation is a case of circular correlation between  $x$  and  $y_n$ . Therefore we can write

$$y_n \circledast x = \lambda_n y_n, \quad (\text{A9})$$

where  $\circledast$  denotes correlation. Applying Fourier transforms, we get

$$\mathcal{F}(y_n \circledast x) = \mathcal{F}(\lambda_n y_n). \quad (\text{A10})$$

Fourier transform of correlation can be written as a product, giving us

$$\overline{\mathcal{F}(y_n)}\mathcal{F}(x) = \lambda_n \mathcal{F}(y_n). \quad (\text{A11})$$

We know that  $y_n$  and  $\lambda_n$  are real. Since  $y_n$  is periodic, a suitable choice is sinusoidal functions.

### A1 Cosine case

We first assume a cosine form for  $y_n$  such that  $y_n[m] = \cos(\omega_n m - \phi_i)$ , where  $\omega_n = \frac{2\pi}{N}n$ . The Fourier transform of such a cosine function is a pair of delta functions as given in equation (A12).

$$\mathcal{F}\{\cos(\omega_n m - \phi_i)\} = e^{-j\phi_i} \delta(\omega - \omega_n) + e^{j\phi_i} \delta(\omega + \omega_n). \quad (\text{A12})$$

The real valued normalization factors have been ignored for the Fourier transform as they cancel out in the subsequent steps. The correlation equation can now be written as

$$[e^{j\phi_i} \delta(\omega - \omega_n) + e^{-j\phi_i} \delta(\omega + \omega_n)]\mathcal{F}(x) = \lambda_n [e^{-j\phi_i} \delta(\omega - \omega_n) + e^{j\phi_i} \delta(\omega + \omega_n)]. \quad (\text{A13})$$

Note that the above equation becomes non-zero only when  $\omega = \omega_n$  or  $\omega = -\omega_n$ . Let us take  $\omega = \omega_n$  case. We have

$$e^{j\phi_i} \mathcal{F}(x)[\omega_n] = \lambda_n e^{-j\phi_i}, \quad (\text{A14})$$

where the  $\omega_n$  component of  $\mathcal{F}(x)$  is

$$\mathcal{F}(x)[\omega_n] = |\mathcal{F}(x)[\omega_n]| e^{j\angle\mathcal{F}(x)[\omega_n]}. \quad (\text{A15})$$

Then

$$e^{j\phi_i} |\mathcal{F}(x)[\omega_n]| e^{j\angle\mathcal{F}(x)[\omega_n]} = \lambda_n e^{-j\phi_i}, \quad (\text{A16})$$

leading to

$$|\mathcal{F}(x)[\omega_n]| e^{j(\angle\mathcal{F}(x)[\omega_n] + 2\phi_i)} = \lambda_n. \quad (\text{A17})$$

Since  $\lambda_n$  is real, we have  $\angle\mathcal{F}(x)[\omega_n] + 2\phi_i = 0$  or  $\phi_i = -\frac{\angle\mathcal{F}(x)[\omega_n]}{2}$ . As  $x$  is real, the Fourier spectrum is Hermitian and therefore it can be shown that the above relation holds for  $\omega = -\omega_n$ . Thus, the eigenvalue and the corresponding cosine eigenvector for a given  $n$  are

$$\lambda_n = |\mathcal{F}(x)[\omega_n]| \quad \text{and} \\ y_n[m] = \cos\left(\omega_n m + \frac{\angle\mathcal{F}(x)[\omega_n]}{2}\right). \quad (\text{A18})$$

### A2 Sine case

Another plausible eigenvector is  $y_n[m] = \sin(\omega_n m - \phi_i)$ , giving

$$\mathcal{F}\{\sin(\omega_n m - \phi_i)\} = -j(e^{-j\phi_i} \delta(\omega - \omega_n) - e^{j\phi_i} \delta(\omega + \omega_n)). \quad (\text{A19})$$

Calculating the Fourier transform and simplifying, we obtain

$$[e^{j\phi_i} \delta(\omega - \omega_n) - e^{-j\phi_i} \delta(\omega + \omega_n)] \mathcal{F}(x) = -\lambda_n [e^{-j\phi_i} \delta(\omega - \omega_n) - e^{j\phi_i} \delta(\omega + \omega_n)]. \quad (\text{A20})$$

Once again, let us consider  $\omega = \omega_n$  case. We have

$$(e^{j\phi_i}) \mathcal{F}(x)[\omega_n] = -\lambda_n (e^{-j\phi_i}), \quad (\text{A21})$$

leading to

$$|\mathcal{F}(x)[\omega_n]| e^{j(\angle\mathcal{F}(x)[\omega_n] + 2\phi_i)} = -\lambda_n. \quad (\text{A22})$$

Once again, since  $\lambda_n$  is real we have  $\angle\mathcal{F}(x)[\omega_n] + 2\phi_i = 0$ . Thus, the eigenvalue and the corresponding sine eigenvector for a given  $n$  are

$$\lambda_n = -|\mathcal{F}(x)[\omega_n]| \quad \text{and} \\ y_n[m] = \sin\left(\omega_n m + \frac{\angle\mathcal{F}(x)[\omega_n]}{2}\right). \quad (\text{A23})$$

Summarizing, the spectrum of an anti-circulant matrix consists of the *absolute* value of the DFT spectrum of the underlying periodic pattern, with the values occurring in positive–negative pairs. The eigenvectors are sines and cosines with phases determined by the phase of the DFT.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.