EFFICIENT ALGORITHM TO IMPLEMENT SLIDING SINGULAR SPECTRUM ANALYSIS WITH APPLICATION TO BIOMEDICAL SIGNAL DENOISING

Muzammil Saeed   *Clive Cheong Took*   *Stephen R. Alty*⋆

Department of Electronic Engineering, Royal Holloway, University of London, TW20 0EX, U.K.

ABSTRACT

Previous work [1] has shown that Singular Spectrum Analysis (SSA) can be particularly effective at noise removal or signal separation in the case of single channel mixtures. The work presented here shows how the sliding or updating algorithm which performs best at signal separation can be implemented in a computationally efficient manner. The main computational burden involved in SSA is the evaluation of a full rank matrix Singular Value Decomposition (SVD). This process is well understood to be of $\mathcal{O}(n^3)$ where n is the rank of the matrix. Therefore, operation of the SSA algorithm in a sliding manner (once per new data sample) involves a very heavy computational cost. In this paper, we show it is possible to evaluate the rank-1 SVD update efficiently in $\mathcal{O}(n^2)$, thus dramatically increasing the speed of the sliding version of the SSA algorithm. Further, we demonstrate that our proposed sliding SSA can be particularly effective at removing ECG from EMG signals in an under-determined setting.

Index Terms— Singular Value Decomposition, Singular Spectrum Analysis, Sliding Window

1. INTRODUCTION

Singular Spectrum Analysis has received a great deal of attention recently [2–4] for its performance and simplicity in reliably separating signals, especially those of a biomedical nature, from single channel mixtures. It is often favoured over Independent Component Analysis (ICA) [5], Empirical Mode Decomposition (EMD) and Wavelet decomposition based methods due to its relatively straight-forward set up with little requirement for *a priori* knowledge of the type of signals under study. Previous work [1] showed how a sliding window based SSA technique performed significantly better than a block based method. By sliding window we mean one that moves by only one sample at a time with a given fixed sample window size. Whilst the performance in terms of signal separation improves, this is at the cost of significant extra computational burden in that each sample update requires a full-rank matrix singular value decomposition (a process requiring $\mathcal{O}(n^3)$ mathematical operations).

In this paper we present a method to dramatically speed-up the sliding window method by employing a novel updating algorithm based on [6] which can sequentially update the orthonormal singular vector matrices $\mathbf{u}_t$ and $\mathbf{v}_t$ and the diagonal singular value matrix $\mathbf{S}_t$ using only the new data sample and the previous iterations of the matrices, $\mathbf{u}_{t-1}$, $\mathbf{v}_{t-1}$ and $\mathbf{S}_{t-1}$.

This paper is organised as follows; the next section briefly introduces the theoretical basis of the standard Singular Spectrum Analysis technique. The Low-Rank Updating of SVD section describes the mathematical derivation of the low-rank update of the SVD matrices. The Updating sliding SSA (USSA) section details the mathematical development of the rank-1 updating 'USSA' algorithm including its computational benefits. Thereafter, there is a section of Results based on the application of the proposed USSA method to the effective removal of ECG interference from sampled diaphragmatic EMG data from an esophageal catheter. Finally, a Conclusions section provides a summary of the work at the end of the paper.

2. SINGULAR SPECTRUM ANALYSIS

Singular Spectrum Analysis has received renewed interest recently [3] being a powerful time-series analysis technique with relevance to signal source separation, financial modeling and biomedical signal denoising. SSA works by *embedding* time-series data into a Hankel matrix form, applying Singular Value Decomposition (SVD) and then extracting the so-called "*eigentriples*". Each *eigentriple* represent different components of the original signal including slowly varying trends, periodic components and unstructured noise. As such, SSA can be used as an efficacious denoising or source separation tool. SSA comprises two stages, namely *decomposition* and *reconstruction* and each stage further consists of two steps.

2.1. Decomposition

The time-series data is decomposed by first embedding it in a Hankel type matrix. Consider the real-valued non-zero time series **s** where

$$\mathbf{s} = [s_1, s_2, ..., s_N] \qquad (1)$$

where $N \geq l + k$ is the minimum number of available samples, l is the window length and k is the embedding 'depth'.

⋆ Please address all correspondence to steve.alty@rhul.ac.uk

Here the embedding process relates to the mapping of $\mathbf{x}$ into k multidimensional lagged vectors of length l such that,

$$\mathbf{x}_i = [s_{i-1}, s_{i-2}, ..., s_{i+l-2}]^T \qquad (2)$$

where the embedding depth $1 \leq k \leq l$, and the superscript ^{T} denotes vector transpose operation. To complete the embedding process a so-called *trajectory* matrix is formed from vectors of $\mathbf{x}_i$. The trajectory matrix is thus,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k] \qquad (3)$$

or

$$\mathbf{X} = \begin{bmatrix} s_0 & s_1 & s_2 & \cdots & s_{k-1} \\ s_1 & s_2 & s_3 & \cdots & s_k \\ s_2 & s_3 & s_4 & \cdots & s_{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{l-1} & s_l & s_{l+1} & \cdots & s_{k+l-2} \end{bmatrix} \qquad (4)$$

$\mathbf{X}$ is clearly a Hankel matrix made up from the sample data whereby all its cross-diagonal elements are equal.

Now we define $\mathbf{M} = \mathbf{X}\mathbf{X}^T$ and since $\mathbf{M}$ must be positive definite, we know that its eigenvalues $\lambda_1, \lambda_2, ..., \lambda_l$ must be positive valued also, further we place the eigenvalues of $\mathbf{S}$ in descending order so that they are monotonically decreasing in value from elements 1 to l, such that, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_l \geq 0$ and $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_l$ are the corresponding orthonormal singular vectors such that $\|\mathbf{u}_i\| = 1$.

If we now define $\mathbf{v}_i = \mathbf{X}^T \mathbf{u}_i / \sqrt{\lambda_i}$, then the trajectory matrix can be broken down into individual components such that,

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_k \qquad (5)$$

where $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T$. The set of data elements $\{\sqrt{\lambda_i}, \mathbf{u}_i$ and $\mathbf{v}_i^T\}$ is often referred to as the ith *eigentriple* of the trajectory matrix $\mathbf{X}$ [8]. Each eigentriple contains progressively different statistical natured sub-components of the original time-series data and careful selection or *grouping* of the components can yield useful results in many source separation problems.

2.2. Reconstruction

The reconstruction stage comprises two parts, namely *grouping* and *diagonal averaging*. There is no general rule for grouping, it all depends on the application, the grouping rule will be defined by the special requirements of the problem and the statistical nature of the signal of interest or even its noise. In our case, we are primarily interested in separating ECG signals from EMG signals and therefore we can use statistical means to select the eigentriples which correspond to the relevant signal distribution function. Experimentation has shown that the fourth standardized moment, *kurtosis*, (see Eq. 6) is an effective statistical measure to differentiate ECG and EMG type signals; typically ECG has a significantly higher kurtosis figure than that of EMG being more noise-like in nature its kurtosis is typically less than 4. A simple threshold

can be set to group all eigentriples with low kurtosis as EMG and the rest to be ECG.

$$\kappa(\mathbf{x}) = \mathcal{E}\left[\left(\frac{\mathbf{x} - \mu}{\sigma} \right)^4 \right] \qquad (6)$$

The final stage of the reconstruction process involves the reconversion from matrix back to time-series data format and this is achieved by averaging the cross-diagonal elements of the reformed Hankel matrix which is simply a sum of the subset of desired eigentriples (see Eq. 5) according to the grouping rules stated in the previous sub-section. Once the new Hankel matrix is calculated it is then simple to *de-embed* the new time-series data. Normally, one Hankel matrix is reconstructed of entirely ECG selected eigentriples and then the resultant time series is simply subtracted from the original mixed times series to yield the de-noised EMG signal trace.

3. LOW-RANK UPDATING OF SVD

Brand [6] showed that it is possible to update the SVD of a matrix efficiently if the change in the matrix, either by updating (adding new data rows) or downdating (removing data rows) is of low-rank (i.e. of a small number relative to the original dimensions of the matrix to be decomposed). This method can be readily adapted to develop an efficient algorithm to perform a sliding window update of the SSA algorithm as follows. If we apply SVD to the original 'trajectory' Hankel matrix Eq. 4 we get;

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{X} \qquad (7)$$

where $\mathbf{U}$ and $\mathbf{V}$ are the orthonormal matrices of singular vectors and $\mathbf{S}$ is the diagonal matrix of singular values. Such that $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$ the identity matrix of rank k (i.e. the same as the embedding depth). Then the update process is as follows, $(\mathbf{X} + \mathbf{A}\mathbf{B}^T)$ where $\mathbf{A}$ and $\mathbf{B}$ are the new data matrices. Next, we need to define $\mathbf{P}$ as an orthogonal basis of $(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{A} = \mathbf{A} - \mathbf{U}\mathbf{U}^T\mathbf{A}$, where the component of $\mathbf{A}$ is orthogonal to $\mathbf{U}$ as would be obtained from the QR-decomposition;

$$[\mathbf{U}, \mathbf{P}] \begin{bmatrix} \mathbf{I} & \mathbf{U}^T\mathbf{A} \\ \mathbf{0} & \mathbf{R}_A \end{bmatrix} \xleftarrow{\text{QR}} [\mathbf{U}, \mathbf{A}] \qquad (8)$$

which can be computed via the modified Gram-Schmit procedure (MGS). $\mathbf{R}_A$ is an upper-triangular matrix. Likewise, let $\mathbf{Q}$ be the orthogonal basis of $\mathbf{B} - \mathbf{V}\mathbf{V}^T\mathbf{B}$. Then,

$$[\mathbf{U}, \mathbf{P}]^T (\mathbf{X} + \mathbf{A}\mathbf{B}^T)[\mathbf{V}, \mathbf{Q}]$$

$$= \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + [\mathbf{U}, \mathbf{P}]^T \mathbf{A}\mathbf{B}^T [\mathbf{V}, \mathbf{Q}]$$

$$= \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{U}^T\mathbf{A} \\ \mathbf{R}_A \end{bmatrix} \begin{bmatrix} \mathbf{V}^T\mathbf{B} \\ \mathbf{R}_B \end{bmatrix}^T \qquad (9)$$

The goal then is to re-diagonalise Eq. 9. Let $\mathbf{U}'\mathbf{S}'\mathbf{V}'^T$ be the low-rank SVD of the right-hand side of Eq. 9. Then,

$$
\begin{aligned}
\mathbf{U}''\mathbf{S}''\mathbf{V}''^T &= ([\mathbf{U},\mathbf{P}]\mathbf{U}')\mathbf{S}'([\mathbf{V},\mathbf{Q}]\mathbf{V}') \quad (10) \\
&= (\mathbf{X}+\mathbf{A}\mathbf{B}^T) \quad (11)
\end{aligned}
$$

Note that we never need the original data matrix $\mathbf{X}$ to compute the update.

4. UPDATING SLIDING SSA (USSA)

As our sliding algorithm only requires the addition of a single row of new data per update and a single down-date (which is achieved by shifting the original Hankel matrix up by one row) only a rank-1 update is required, this reduces the matrices $\mathbf{A}$ and $\mathbf{B}$ to row vectors $\mathbf{a}$ and $\mathbf{b}$ decreasing computational complexity yet further. We will now introduce the subscript of lower case '$_t$' to signify the latest time series data sequence (implying '$_{t-1}$' is the previous time iteration). Suppose our update can be written as follows;

$$
\mathbf{X}_t = \hat{\mathbf{X}}_{t-1} + \mathbf{a}\mathbf{b}_t^T \quad (12)
$$

where

$$
\hat{\mathbf{X}}_{t-1} =
\begin{bmatrix}
s_1 & s_2 & s_3 & \cdots & s_k \\
s_2 & s_3 & s_4 & \cdots & s_{k+1} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
s_{l-1} & s_l & s_{l+1} & \ddots & s_{k+l-2} \\
0 & 0 & 0 & \cdots & 0
\end{bmatrix}
\quad (13)
$$

is the previous iteration of Hankel matrix which has had the top row of data removed and a row of zeroes appended to the bottom row,

$$
\mathbf{a} = [0,0,0,\ldots,1] \quad (14)
$$

a row vector of zeroes with a one as its final element and

$$
\mathbf{b}_t = [s_l, s_{l+1}, s_{l+2}, \cdots, s_{k+l-1}] \quad (15)
$$

a row vector that contains the new time series data to be appended to the updated Hankel matrix, $\hat{\mathbf{X}}_{t-1}$, to create $\mathbf{X}_t$.

We can compute the updated orthonormal matrices of singular vectors and diagonal matrix of singular values directly and efficiently according to the modified Gram-Schmidt algorithm (MGS) if we expand Eq. 8 with the following definitions; $\mathbf{m} = \mathbf{U}^T\mathbf{a}$; $\mathbf{p} = \mathbf{a} - \mathbf{U}\mathbf{m}$; $p = \sqrt{\mathbf{p}^T\mathbf{p}} = \sqrt{\mathbf{a}^T\mathbf{p}}$; $\mathbf{P} = \mathbf{p}/p$. Similarly, $\mathbf{n} = \mathbf{V}^T\mathbf{b}_t$; $\mathbf{q} = \mathbf{b}_t - \mathbf{V}\mathbf{n}$; $q = \sqrt{\mathbf{q}^T\mathbf{q}} = \sqrt{\mathbf{b}_t^T\mathbf{q}}$; $\mathbf{Q} = \mathbf{q}/q$. Then it can be shown that the right most term of Eq. 9 becomes an outer vector product

$$
\begin{bmatrix} \mathbf{U}^T\mathbf{A} \\ \mathbf{R}_A \end{bmatrix}
\begin{bmatrix} \mathbf{V}^T\mathbf{B} \\ \mathbf{R}_B \end{bmatrix}^T =
\begin{bmatrix} \mathbf{m} \\ p \end{bmatrix}
\begin{bmatrix} \mathbf{n} \\ q \end{bmatrix}^T
\quad (16)
$$

Further computational efficiencies are outlined in [6]. It can be shown, however, that the rank-1 update process can be achieved in $\mathcal{O}(k^2)$ time [7], dramatically increasing the speed of the sliding window algorithm when compared to full re-computation of the SVD which requires $\mathcal{O}(k^3)$.
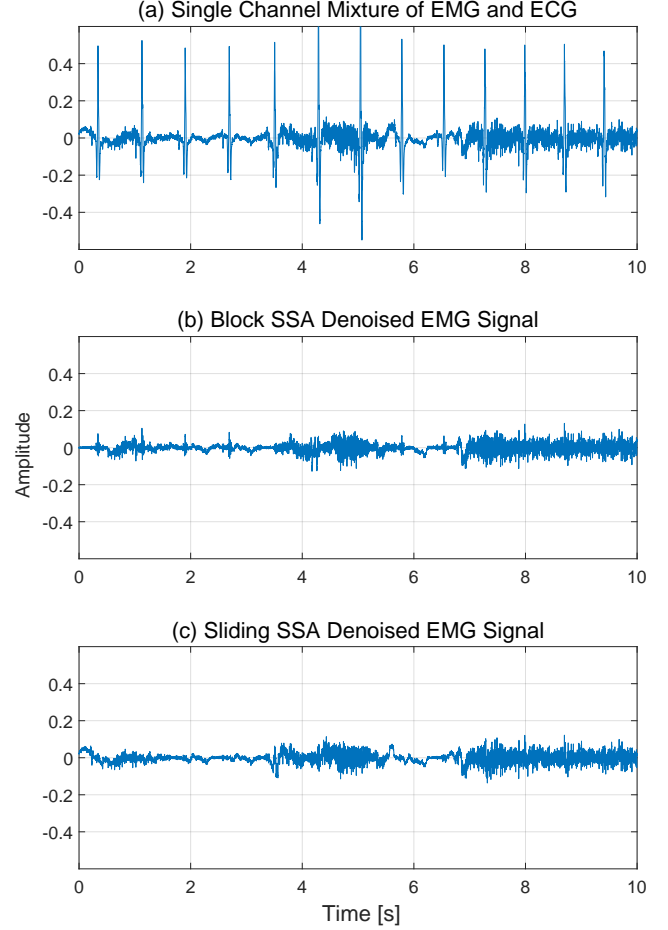

Fig. 1. (a) single channel mixed EMG and ECG data, denoised using (b) block based SSA and (c) sliding USSA.

5. RESULTS

The real EMG data used in this paper was acquired from patients at the Royal Brompton Hospital (UK) suffering from some form of respiratory disease. Figure 1(a) shows a 10s segment of the test signal (sample rate = 2000 Hz) that originates from one channel of data acquired from the esophageal catheter which shows significant unknown mixing of EMG and ECG signals. Figure 1(b) shows the results from the application of the block-based SSA algorithm to the test signal. Figure 1(c) shows the results from the sliding window based USSA algorithm when applied to the same signal. Whilst the results for the block based method are satisfactory, especially considering the demanding source of the data, some ECG artefacts remain. The sliding window method, however, shows a clear advantage with virtually no evidence of ECG interference remaining. Our work improves upon the block SSA result significantly, the initial settings of the SSA method remain unchanged; the window length used was, $l = 500$ samples, the depth of the decomposition was set to, $k = 50$ for both methods as was the kurtosis threshold, κ, where ex-
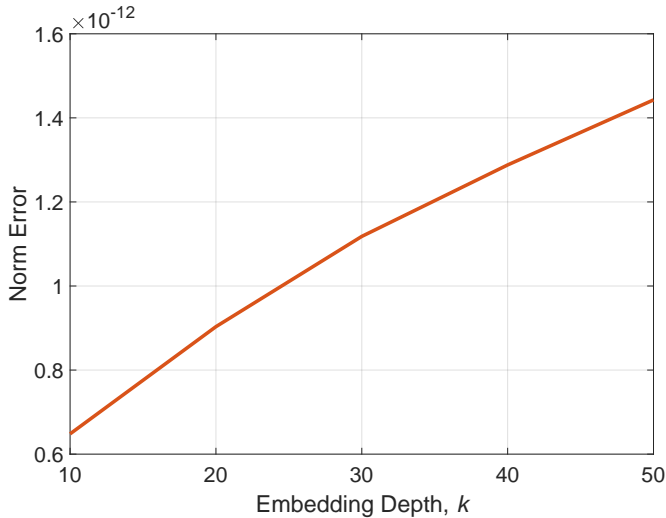
Fig. 2. Frobenius norm error for window length, $l = 500$, versus embedding depth, k.

Fig. 3. Relative computational gain of USSA over sliding SSA [1] for $l = 500$ and different embedding depths, k.

perimentation on synthetic mixed data yielded a threshold of around 5 to be optimal [1].

Figure 2 shows the average Frobenius norm error accumulated per iteration across the entire Hankel matrix when comparing rank-1 updated SVD to full SVD computation. The plot shows the norm error in the Hankel (trajectory) matrix for $l = 500$ when embedding depth of matrix, $k = 10, 20, 30, 40$ & 50. It can be seen that even the largest matrix of 25,000 elements has a very low error in the order of 1.4×10^{-12} which implies that many thousands of iterations could be performed before accumulated errors became significant, at which point a 'key' full SVD could be undertaken to refresh the decomposition.

Finally, Figure 3 shows the relative comparison of computational complexity of the full SVD sliding SSA algorithm versus rank-1 update SVD algorithm USSA algorithm for window length, $l = 500$, versus embedding depths k. Significant gains, that is to say reduction of computational complexity are apparent for values of $k \geq 25$. In this application we found that an embedding depth of 50 was best suited to the task, thus yielding an approximate computational gain of 5 times over the earlier full rank sliding SVD approach.

6. CONCLUSIONS

The results of the updating sliding window USSA algorithm shows clear improvements over the block based methods whilst simultaneously reducing the computational burden over previously published work. Accumulated propagation errors grow linearly and can be reset every few thousand iterations with a fresh full SVD. This could enable the algorithm to be readily deployed in real-time applications.
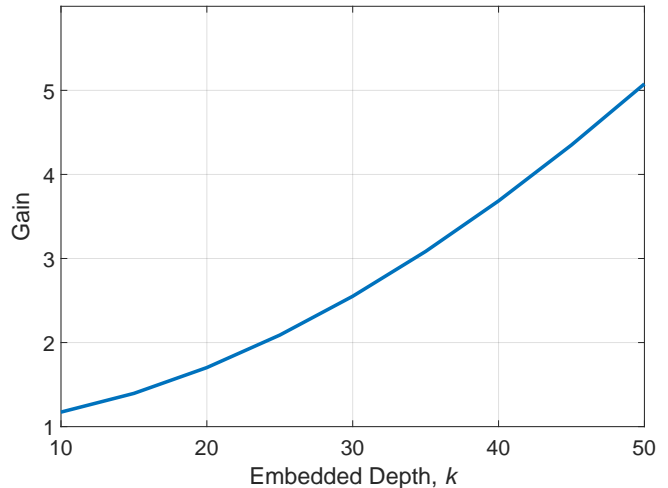
7. REFERENCES

[1] M. Saeed and S. R. Alty, "Removal of Electrocardiogram Interference from Diaphragmatic Electromyogram Signals using Sliding Singular Spectrum Analysis," *40th IEEE EMBC,* pp. 1238–1241, July 2018.

[2] L. Souza, B. Gatto and K. Fukui, "Classification of Bioacoustic Signals with Tangent Singular Spectrum Analysis," *44th IEEE ICASSP,* pp. 351–355, May 2019.

[3] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals*, CRC Press, 2015.

[4] J. Barrios-Muriel, F. Romero, F. J. Alonso, and K. Gianikellis, "A simple SSA-based de-noising technique to remove ECG interference in EMG signals," *Biomedical Signal Processing and Control*, pp. 117–126, 2016.

[5] Z. Zhang, H. Li, and D. Mandic, "Blind source separation and artefact cancellation for single channel bioelectrical signal," *13th IEEE Int. Conf. on Wearable and Implantable Body Sensor Networks*, pp. 2376–8894, 2016.

[6] M. Brand, "Fast online SVD revisions for lightweight recommender systems" *SIAM 3rd International Conference on Data Mining,* pages 37–46, May 2003.

[7] M. Gu and S. C. Eisenstat. "A stable and fast algorithm for updating the singular value decomposition" *Tech. Report YALEU/DCS/RR-966*, Department of Computer Science, Yale University, New Haven, CT, 1993.

[8] T. Alexandrov, "A method of trend extraction using singular spectrum analysis," *REVSTAT Statistical Journal*, vol. 7, pp. 1–22, April 2009.