Cadzow's basic algorithm, alternating projections and singular spectrum analysis

JONATHAN GILLARD

After observing a noisy time series or signal, it is common practice to try to separate the noise from the observed measurements. Singular value decomposition based methods are often used to split the observed signal into a number of components. Components associated with noise may be removed from the signal, and subsequent analyses may be undertaken. This paper will describe two methods commonly used to remove noise from a signal; the so-called singular spectrum analysis, and Cadzow's basic algorithm. Connections between both methods will be drawn, and both will be related to the method of alternating projections, and structured low rank approximation (finding a lower rank approximation of a given matrix with specified structure). A simulation study and example based on real data will highlight and explain the differences between both methods.

1. INTRODUCTION

Let $Y_T = (y_1, \ldots, y_T)$ denote a noisy real-valued time series or signal. A common requirement is to apply some denoising methodology to Y_T in order to try to separate the noise from the signal. Common approaches map the one-dimensional Y_T to a multidimensional series of lagged vectors to be contained within some $L \times K$ matrix. This permits further analysis.

Such a mapping puts Y_T into a structured matrix (commonly Hankel) and singular value decomposition (SVD) methods are regularly used to deconstruct the matrix into a sum of rank-one components. Under certain conditions, components which are associated with noise, and components associated with the true signal may be separated. This procedure may be run iteratively. An example of a SVD-based method includes singular spectrum analysis (SSA) which, under certain conditions and constraints, may be viewed as one iteration of Cadzow's basic algorithm. Cadzow's iterations use SVD's within an alternating projections framework.

The aim of this paper is to draw connections between SSA and Cadzow's basic algorithm in relation to denoising a time series or signal. The corresponding implications for forecasting are discussed. Section 2 will describe the method of alternating projections in general. Section 3 will outline Cadzow's basic algorithm and SSA. Both methods will be related to the wider problem of finding a low-rank approximation of a structured matrix. Section 4 contains a simulation and an example comparing SSA and Cadzow's iterative algorithm. Section 5 will conclude the paper.

2. ALTERNATING PROJECTIONS

Let α and β denote two sets, and let P_α and P_β denote projections onto α and β respectively. The alternating projections algorithm (also known as the lift and project algorithm [4], pp. 254–257) begins with a point α_0 in α, and alternately projects onto α and β such that $\beta_i = P_\beta(\alpha_i)$, $\alpha_{i+1} = P_\alpha(\beta_i)$, for $i = 0, 1, \ldots$. If $\alpha \cap \beta \neq \emptyset$, then the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ converge to some value (not necessarily in a finite number of iterations). Further details are provided in [4], page 256. The method of alternating projections is the main idea behind Cadzow's iterative method described in the next section.

Generally, the method of alternating projections is slow and can be computationally intensive. The idea however is simple, and is readily extended to sequential or cyclic projections onto many sets. If the sets α and β are closed and convex, the method of alternating projections has a well described theory (see for example, [2, 4, 3, 6, 11, 8] and [14]) and has successful uses in many applications. In many applications, one or both of the sets may not be convex, or closed. The behaviour and reliability of the alternating projections algorithm also suffers as a result of having a non-convex set involved in the repeated projections [1].

3. CADZOW'S ALGORITHM AND SSA

This section will describe the two methods under discussion in this paper: Cadzow's basic algorithm [2], and SSA [9]. Both rely heavily on the implementation of SVD's.

3.1 Cadzow's basic algorithm

Let Y_T be mapped onto an $L \times K$ Hankel matrix X as follows:

$$(1) \quad X = [X_1, \ldots, X_K] = \begin{pmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_T \end{pmatrix}.$$

Using the terminology of SSA (described in the next section) L is a parameter known as the window length, an integer such that $2 \leq L \leq T$. The window length L is the sole parameter in this mapping. Selection of L depends on the problem in hand and on preliminarily information about Y_T. Namely, if we know that Y_T has a periodic component with an integer period, then for better separability of this component it is advisable to take L proportional to that period (see [13]). Theoretical results tell us that L should be large enough but not greater than $\frac{T}{2}$ (see for example [9]).

The SVD of X is given by $X = U\Sigma V^T$ where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_L)$, $\sigma_1 \geq \cdots \geq \sigma_L$. The singular values may be adjusted by applying some function f (in order to obtain a modified least squares estimate [21], a minimum variance estimate [5], a time-domain constraint estimate [7] or to achieve some other estimate [18]) and the SVD of X is truncated to rank r: $X_r = U_1 f(\Sigma_1) V_1^T$ where $U = [\underbrace{U_1}_{r} ; \underbrace{U_2}_{L-r}]$, $V = [\underbrace{V_1}_{r} ; \underbrace{V_2}_{K-r}]$, and $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r)$. The reconstructed signal component is estimated from X_r by averaging across its anti-diagonals. This is equivalent to finding a Hankel matrix approximation of X_r. These steps may be applied iteratively, where the algorithm can recommence with the estimated true signal. These steps aid in denoising the signal (if any noise is present), prior to estimation of any parameters. Typically, noise components are associated with small singular values.

Inherent within this method is the assumption that the pure signal lies in a low-dimensional subspace of $\mathbb{R}^T$. A typical aim of many signal processing methods is to approximate this subspace, and estimate the signal within it. This assumption may not hold exactly in all applications, but is nevertheless a good model for some signals [12] and has worked well in speech processing [16]. Denoising, in the context described above, should at least take us some way in approximating this low-dimensional subspace. It is difficult to quantify by how much we are able to denoise; this is likely to depend on the application, and the type of noise observed in the signal (if any). Further details are included in [10].

Examples of the function f include $f(\Sigma) = (\Sigma^2 - L\sigma_{noise}^2 I)^{\frac{1}{2}}$, with $\sigma_r^2 > L\sigma_{noise}^2$ and σ_{noise}^2 is an estimate of the variance of the noise within the signal. The minimum variance method takes $f(\Sigma) = (\Sigma^2 - L\sigma_{noise}^2 I)\Sigma^{-1}$ [5].

Cadzow's basic algorithm takes $f(\Sigma) = \Sigma$ (see [2], [5]), with fixed L and K, and iterates by taking the outputted cleaned data vector, $\widehat{Y_T}$ say, and repeating the above method. This can be repeated for a number of iterations, or until the algorithm converges to some value within some specified tolerance.

3.2 Basic SSA

One iteration of Cadzow's basic algorithm under certain constraints corresponds to the basic SSA. The main purpose of SSA is to decompose the original series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic component, or noise (details of how to do this in practice are included in [9], and more recently in [13]). This is followed by a reconstruction of the original series. The SSA technique consists of two complementary stages: decomposition and reconstruction. Briefly, the basic SSA algorithm is as follows (for more details see [9]).

As in equation 1, let $X = [X_1, \ldots X_K]$ where $X_i = (y_i, \ldots, y_{i+L-1})^T \in \mathbb{R}^L$. Again we map Y_T onto a Hankel matrix consisting of L-lagged vectors $\{X_i, i = 1, \ldots, K\}$ such that $2 \leq L \leq N$. From the matrix XX^T the eigenvalues and eigenvectors are found. Let $XX^T = P\Lambda P^T$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_L)$ is a matrix of ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$ and P is the matrix containing corresponding eigenvectors $P = (P_1, \ldots, P_L)$.

To obtain a rank r approximation of X, the first r eigenvectors corresponding to $\lambda_1, \ldots, \lambda_r$ are chosen. The rank r approximation to X is then given by

$$\widehat{X} = \sum_{j=1}^{r} P_j P_j^T X.$$

Averaging over the anti-diagonals of $\widehat{X}$ yields the estimate $\widehat{Y_T}$ of Y_T.

3.3 Comparison of Cadzow's algorithm and SSA

One iteration of Cadzow's algorithm will yield identical results to SSA if the parameters L and r are chosen to be the same in both methods. Thus iterations of Cadzow's algorithm may be viewed as a form of iterated SSA. This is known as sequential SSA [9]. In [22] and [23] sequential SSA is described, and both papers offer reasons why it might be advantageous to alter L between iterations.

Cadzow [2] showed that by truncating the SVD to r terms, and iterating via his rank-reduction and averaging process does lead to a rank r Hankel matrix. De Moor [5] however demonstrated that this may not be the desired matrix. It is claimed however that averaging across the anti-diagonals works well in practice [12]. As SSA is just one iteration of Cadzow's algorithm, a reconstructed series obtained via SSA will not necessarily be of rank r.

Cadzow's algorithm is therefore a heuristic solution method, which is suboptimal in terms of the L_2 optimality criterion [5]. It is nevertheless a simple method that is readily implemented. The example used by De Moor [5] to demonstrate the sub-optimality of Cadzow's algorithm is simplistic however. Further details concerning the performance of Cadzow's algorithm are included in [19].

3.4 Structured low rank approximation

Both SSA, and Cadzow's iterative methods are closely associated with the structured low-rank approximation problem defined as follows (for further details see [15]). Given a structured matrix X and an integer $r < \min(L, K)$, the structured matrix low-rank approximation (SLRA) problem is as follows:

$$\min_{\hat{X}} \|X - \hat{X}\| \text{ subject to } \left\{ \begin{array}{l} \text{rank}(\hat{X}) \leq r; \\ \hat{X} \text{ has the same structure as } X. \end{array} \right.$$

In our context X is a Hankel matrix formed from some signal or time series. The idea is to perturb the matrix X by a small error ΔX such that $X + \Delta X$ becomes rank deficient (and at most r). The distance $\|\hat{X} - (X + \Delta X)\|$ should be 'small'. Examples of alternative structures for the matrix X include Toeplitz, block-Hankel, block-Toeplitz and circulant.

Finding the nearest matrix, of rank r, to X that possesses the same structure as X is linked to the problem of structured total least squares. Total least squares aims to find an approximate solution of the overdetermined linear system $Ax \approx b$ where both A and b are perturbed by error. There are obvious analogies with this problem, and that of finding a lower rank approximation of the matrix $X = [A\ b]$. There are many papers describing the methodology and applications of total least squares. A review of the problem is conducted in the book [20]. If A and b are also required to have structure, this is known as structured total least squares.

Structured total least squares aims to estimate the parameter x in the $Ax \approx b$. SSA is viewed as a non-parametric method, and so the estimation of the parameter x is not of primary importance. The SVD is the main tool used for total least squares and SSA, with both methods finding the closest subspace (in terms of Frobenius norm) to the observed data.

3.5 Conditions for separability

Given that the method of alternating projections used in the algorithms discussed in this paper depend heavily on the use of SVD, it is useful to identify the conditions in which the SVD can separate the noise from the signal. For the signal Y_T, it often follows that $Y_T = Y_{true} + Y_{noise}$, and so $X = X_{true} + X_{noise}$. This is an assumption, and might not always hold true [12]. Derivations of the conditions which allow for X_{true} to be reconstructed via an SVD are included

in [21]. Assume that the SVD of X_{true} is given by

$$X_{true} = \underbrace{[U_1}_{r} ; \underbrace{U_2]}_{L-r} \left[\begin{array}{cc} \Sigma_1 & 0 \\ 0 & 0 \end{array} \right] \underbrace{[V_1}_{r} ; \underbrace{V_2]}_{K-r}^T$$

where $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r)$.

The conditions which allow for the separability of X_{true} from X are:

1. $X_{true}^T X_{noise} = 0$;
2. V_1 and V_2 must be orthogonal in the inner product $X_{noise}^T X_{noise}$;
3. The singular value σ_r must be bigger than σ_{r+1} in the SVD of X.

The success of SVD-type algorithms depend on these assumptions, as then the rank of X_{true} and the subspaces generated by V_1 and V_2 may be estimated consistently. The SVD is however robust to mild violations of these assumptions. These assumptions can be summarised by the following orthogonalities: let S_l (S_r) be the linear spaces spanned by the left (right) singular vectors of X_{true}. Analogously let N_l and N_r be the linear spaces spanned by the left (right) singular vectors of X_{noise}. Let the sets of singular values of X_{true} and X_{noise} be Σ_S and Σ_N respectively. The conditions needed to reconstruct X_{true} from X are: $S_l \perp N_l$, $S_r \perp N_r$ and $\Sigma_S \cap \Sigma_N = \emptyset$. The condition $\Sigma_S \cap \Sigma_N = \emptyset$ can be made simpler by making the sets Σ_S and Σ_N distinct. Smaller singular values are usually associated with noise, and so in general all singular values of Σ_S are bigger than those in Σ_N. However, as stated in [20], the conditions described above are never satisfied exactly. The final condition is the one most likely to hold. The SVD is however robust to mild violations of these conditions. If $\|X_{true}^T X_{noise}\|$ is small (for any norm $\|\cdot\|$) the SVD will deliver satisfactory approximations to S_l and S_r (see [20], for example).

Cadzow's composite property algorithm and SSA decompose the SLRA into two smaller problems; namely that of

1. finding a low-rank matrix approximation of X where rank$(X) \leq r$,
2. finding a structured Hankel matrix.

In terms of the alternating projections described in Section 2, let α be the set of all matrices of size $L \times K$ with rank $\leq r$, and let β be the set of all Hankel structured matrices. For SSA and Cadzow's basic algorithm, P_α corresponds to the truncation of the SVD of X to r terms by setting the smallest $K - r$ singular values to 0. P_β corresponds to the averaging of the anti-diagonals of this new rank-reduced matrix, in order to obtain a Hankel approximation.

4. EXAMPLE AND DISCUSSIONS

4.1 Comparison of Cadzow's algorithm with SSA

This simulation example will use the signal originally used by Cadzow (in [2]) to illustrate his alternating projec-
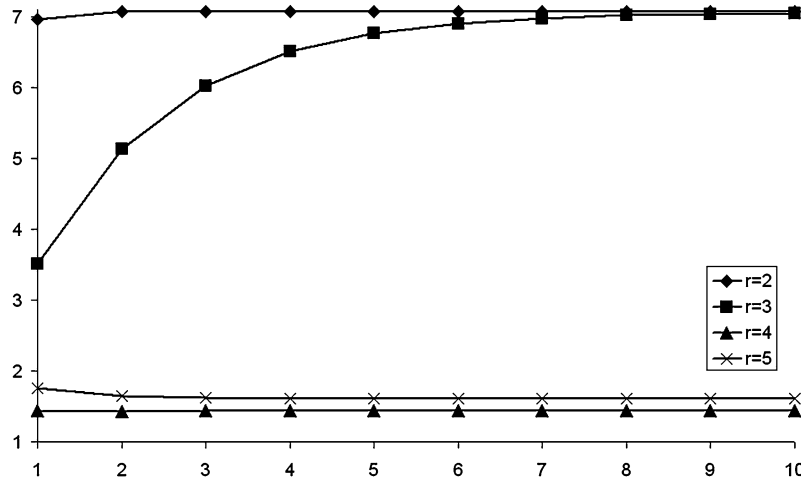
Figure 1. RMSE against number of Cadzow iterations for varying rank.

tions algorithm. The signal used by Cadzow had the functional form

$$f_t = 15\cos(0.25\pi t) + 10\cos(0.71\pi t), \quad t = 1, \ldots, 50.$$

In this section, the performance of Cadzow's basic algorithm in denoising a signal is described via root mean square error (RMSE). This is defined as

$$\sqrt{\sum_{t=1}^{T} \frac{(y_t - \widehat{y}_t)^2}{T}}$$

where T is the number of observed points. Figure 2 contains plots of RMSE (averaged over 1,000 simulations) against the number of Cadzow iterations, with different truncations of the SVD ($r = 2, 3, 4, 5$), applied to the signal $y_t = f_t + \varepsilon_t$ where each ε_t is a i.i.d. normally distributed random variable with zero mean, and variance 10. The rank of the noise-free signal is 4, and so the optimal truncation of the SVD is at the first 4 singular values ($r = 4$). For an observed signal in practise however, the rank of the true signal is unlikely to be known *a priori*. One iteration of Cadzow's basic algorithm corresponds to the basic SSA, and so this allows comparison of SSA with each additional Cadzow iteration. The parameter L was chosen to be half of the number of observations taken from the signal, and this is common practise (see [9]).

Figure 1 demonstrates that if the rank selected is smaller than the rank of the true signal (namely if the SVD is truncated to 1, 2 or 3 terms respectively) additional iterations of Cadzow's basic algorithm increase the RMSE, and then remain relatively stable after 10 iterations. If the SVD is truncated to too few terms, then the signal is over-smoothed and the approximation does not reflect the intricacies of the signal. The smallest RMSE is achieved after one Cadzow iteration, which corresponds to basic SSA. The smallest RMSE is achieved when the signals' true rank is selected, and the

RMSE is stable regardless of how many Cadzow iterations are performed. If too many terms are included in the truncation of the SVD ($r > 4$), the approximation is absorbing some of the error ε_t, $t = 1, \ldots, 50$. Analogously, the RMSE is stable regardless of how many Cadzow iterations are performed.

It is well documented however (see [17]) that performing an SVD on a given large matrix is computationally expensive. Given that each iteration of Cadzow's basic algorithm involves such an operation, it is computationally beneficial to avoid computing many SVD's. Figure 1 demonstrates that for the signal used in this example, there is little to be gained by repeating Cadzow iterations. For our example, it is optimal both computationally, and in terms of RMSE, to only apply one iteration of Cadzow's basic algorithm. That is, in this instance, the use of basic SSA is more beneficial than the use of Cadzow's basic algorithm. This is of particular importance if the rank of the true signal is underestimated.

4.2 Tracking projections – example

Figure 2 contains a plot of the yearly average temperature of Central England, measured between 1659 and 2009. The series is used regularly by scientists as the trends in temperatures since the mid-17th century can be followed. Temperatures fell during the period between 1659–1700 (approximately) and then rose in the early 1700s. During the 18th and 19th centuries, a cool period which coincided with snowy winters and generally cool summers, the temperatures fluctuated widely but with little trend. From 1910, temperatures increased slightly until about 1950 when they flattened before a sharp rising trend began in about 1975. Temperatures so far in the current decade appear different from the long-term average.

There is noise inherent in the data, and it is important to separate this noise from the trend. Figure 3 contains approximations found by running Cadzow's basic algorithm
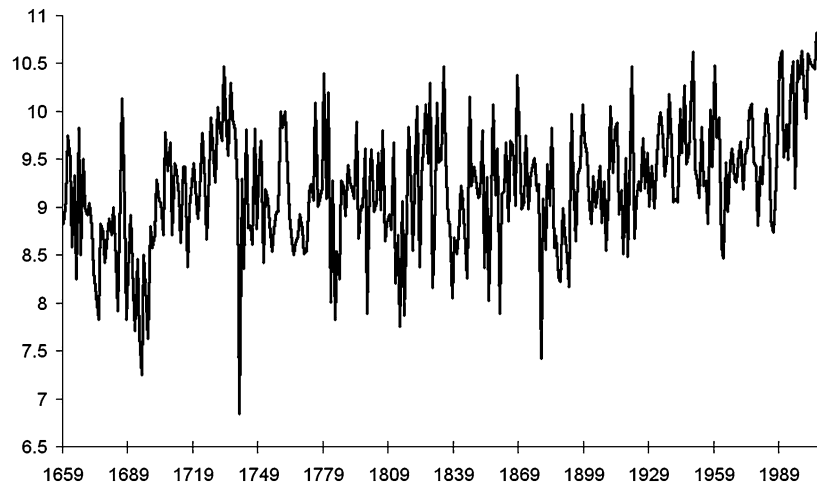
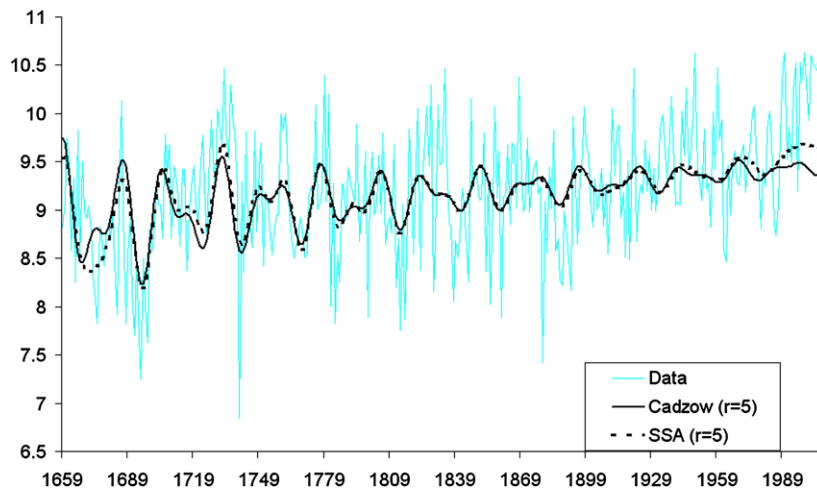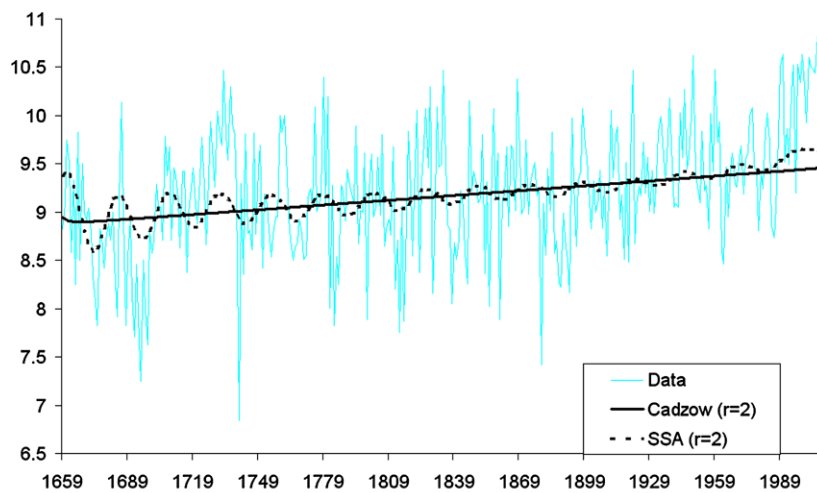Figure 2. Yearly average temperature, Central England: 1659–2009.

Figure 3. Yearly average temperature, Central England: 1659–2009 with $r = 2$ and $r = 5$ (Cadzow and SSA) approximations respectively.

Table 1. RMSE of retrospective forecast of various lengths (in brackets), obtained by truncating the time series to 300 terms

Rank	Cadzow (2)	SSA (2)	Cadzow (5)	SSA (5)	Cadzow (10)	SSA (10)
$r = 5$	0.6963	0.5857	0.6147	0.6061	0.6218	0.5904
$r = 10$	0.7770	0.5871	0.6894	0.6195	0.6386	0.5881
$r = 15$	1.2802	0.6674	1.0185	0.6958	0.8535	0.6190
$r = 20$	1.3422	0.8367	1.0530	0.7382	0.8565	0.6467

Figure 4. Yearly average temperature, Central England: 1659–2009. The difference between the Cadzow and SSA approximations for $r = 10$ is plotted for each year.

until convergence, for different truncations of the SVD of the original signal. As more components are taken in the truncated SVD, the reconstructed signal will tend towards the observed signal. The higher rank approximations ($r = 10$, $r = 15$ and $r = 20$) demonstrate the elevated temperature experienced over the past decade, but the $r = 5$ approximation remains fairly constant across the range of the signal. The higher rank approximations are more able to pick on the seasonality and periodicity usually present within such data, and are more able to reflect some of the sharp peaks and troughs in the series. It is worth noting however that the amount of Cadzow iterations needed for the algorithm to finally converge to a low-rank approximation were 36, 103, 112 and 127 for the rank 5, 10, 15 and 20 approximations respectively.

Figure 3 contains approximations to the data in figure 2 found by using SSA and Cadzow's basic algorithm, for $r = 2$ and $r = 5$. As more components are taken in the truncated SVD, the reconstructed signal will tend towards the observed signal. Figure 4 contains the difference between the Cadzow and SSA approximations for $r = 10$, for each year. As r is increased, the approximations become indistinguishable. The only noticeable differences are at the beginning and the end of the series.

It is worth noting that the computational time taken to run SSA is much smaller than repeating Cadzow iterations. The output trajectory matrix as a result of running SSA will not however possess the rank-deficiency property that

would have been obtained by running Cadzow's algorithm until convergence.

The $r = 5$ SSA approximation demonstrates a more accelerated upward trend towards the end of the series than its Cadzow equivalent. The higher rank Cadzow approximations of the series demonstrate an inflated predicted temperature for the first year, 1659 (see figure 5). The SSA equivalents do not demonstrate this inflation as severely. It is worth noting that the SSA approximations across the range of the series seem less turbulent, and so the Cadzow approximations appear a closer fit to the original data.

Once the signal has been de-noised, the reconstructed signal may be used for forecasting. Figures 5 and 6 use the rank 5, 10, 15 and 20 ($r = 5, 10, 15, 20$) reconstructions (of the initial trajectory matrix) to forecast 20 points into the future using Cadzow's basic algorithm (until convergence) and basic SSA respectively. The forecast based on SSA seems less erratic than the corresponding Cadzow forecasts. Both the rank 5 approximations obtained via Cadzow's algorithm and SSA are stable, and do not demonstrate any severe predicted increase in temperature for the next 20 years. The forecasts using Cadzow's algorithm are more volatile than the SSA forecasts. The higher rank approximations, for both methods, do imply an increase in temperature over the next 20 years.

Table 1 contains the RMSE for various retrospective forecast of the series. That is, the initial series was truncated to 300 values, and future values (2, 5, and 10) were pre-
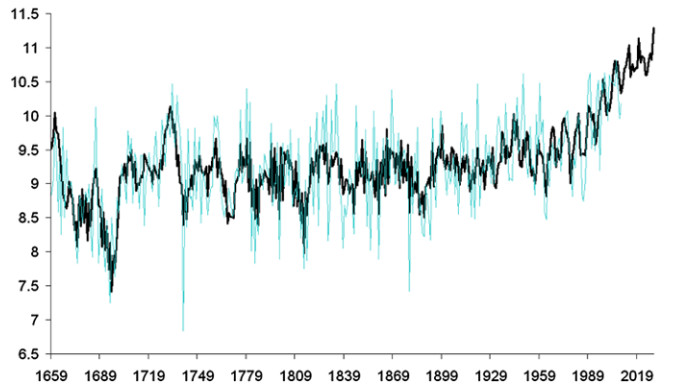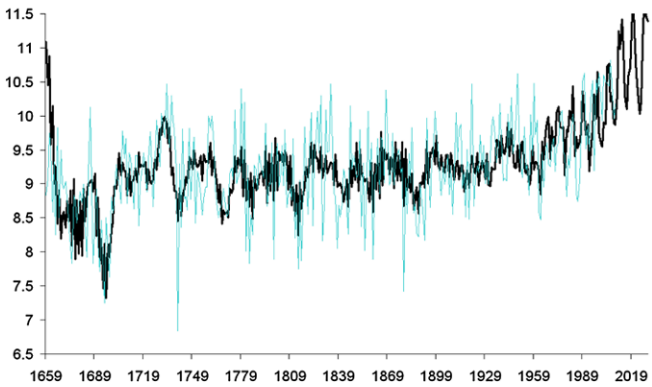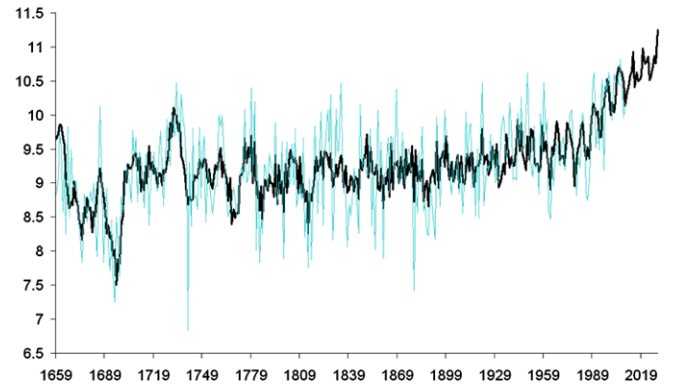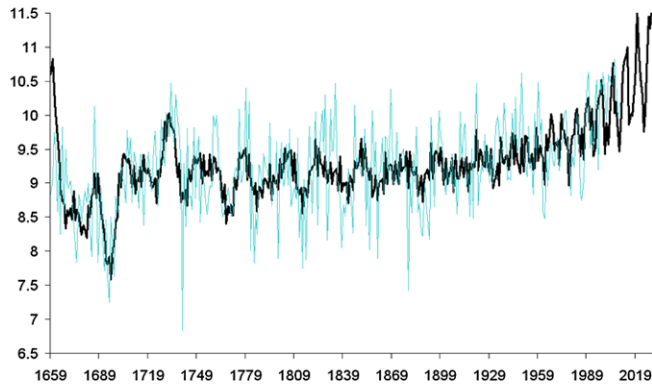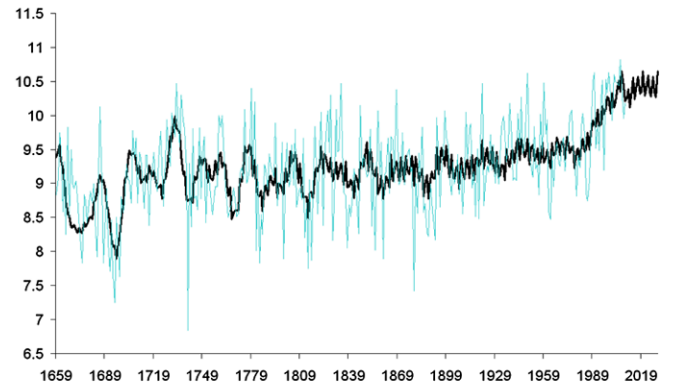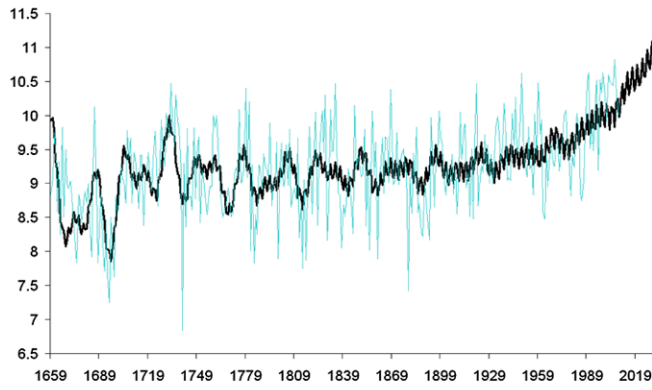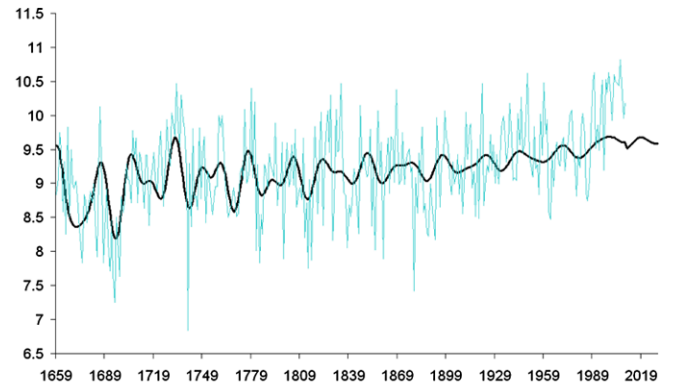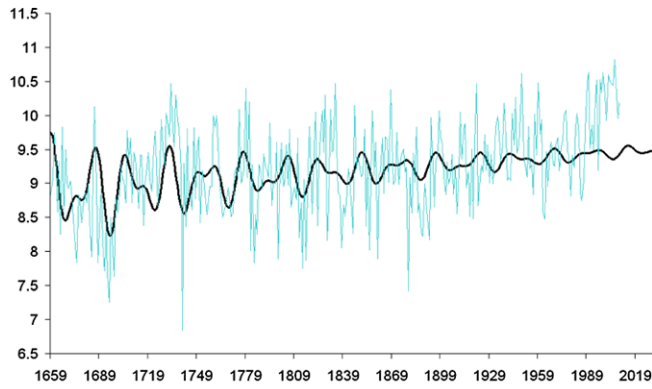
Figure 5. Yearly average temperature, Central England, with $r = 5$, $r = 10$, $r = 15$ and $r = 20$ (Cadzow) approximations. Forecast ahead for 20 years is included.

Figure 6. Yearly average temperature, Central England, with $r = 5$, $r = 10$, $r = 15$ and $r = 20$ (SSA) approximations. Forecast ahead for 20 years is included.
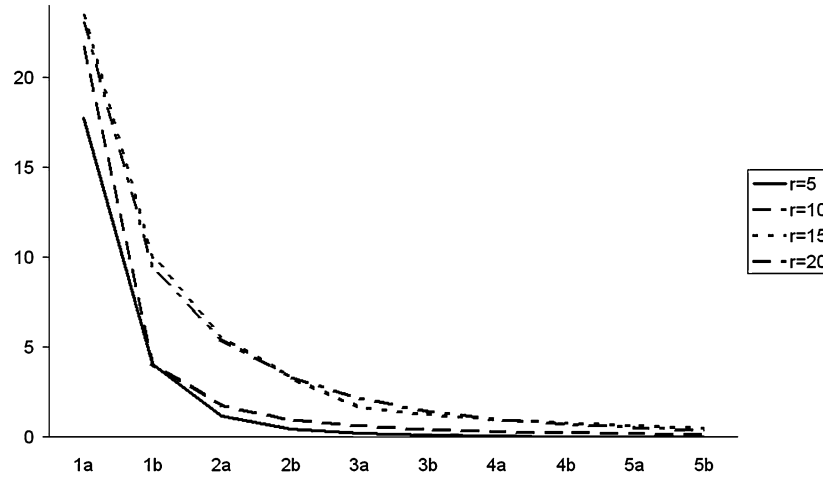
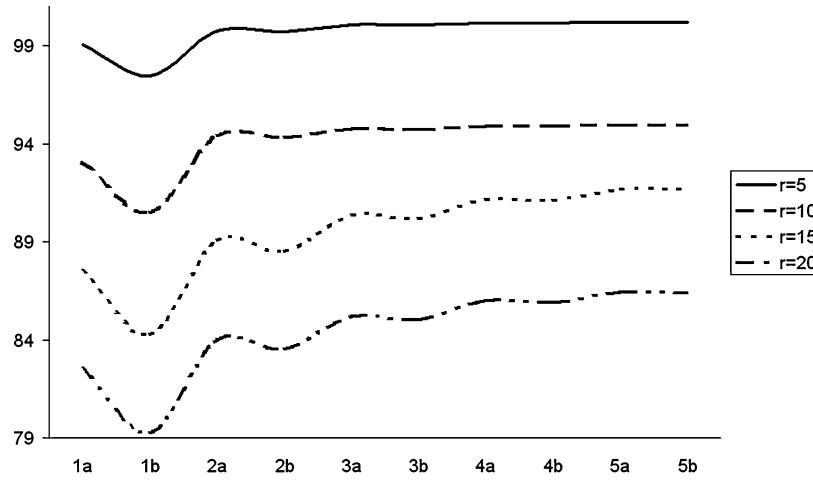Figure 7. Frobenius distances within and between Cadzow iterations, for $r = 5, 10, 15, 20$.

Figure 8. Frobenius distances from the original trajectory matrix, to the output matrices at each Cadzow iteration, for for $r = 5, 10, 15, 20$.

dicted and compared with the observed values. Both basic SSA and Cadzow's basic algorithm were used to forecast the remaining values at different truncations of the SVD ($r = 5, 10, 15, 20$). For more information on forecasting using Cadzow's algorithm, and SSA, see [21] and [9] respectively. The RMSE for both methods are similar, but are smaller (for each value of r) via basic SSA.

The main stages within Cadzow's basic algorithm are

a Finding a low-rank approximation of the trajectory matrix;

b The Hankelisation of the low-rank approximation by averaging over the anti-diagonals.

Figure 7 contains the Frobenius distances between the matrices at each of these stages within 5 iterations of Cadzow's basic algorithm. For example, '1a' and '1b' denote the low-rank approximation (of the original trajectory matrix), and Hankelisation of this low-rank approximation within the

first Cadzow iteration, respectively. The biggest difference between Frobenius distances is within the first iteration. Across the remaining iterations the distances get smaller, and convergence to a solution is slow. It seems then that the most benefit in taking the initially observed trajectory matrix onto a lower dimension subspace appears after the first iteration.

Figure 8 contains the Frobenius distances between the matrix obtained at each of these stages within 5 iterations of Cadzow's basic algorithm, and the original trajectory matrix formed by the initial data. Recall that one iteration of Cadzow's basic algorithm corresponds to basic SSA. The SSA approximations to the data remain closest to the original data. Each Cadzow iteration takes the approximation further away (in terms of Frobenius norm) from the original trajectory matrix. However, Cadzow's basic algorithm upon convergence gives a rank-r Hankel matrix as its output. SSA does not have this property.

5. CONCLUSION

This paper has discussed the connections between Cadzow's basic algorithm, SSA, and put both in the wider context of alternating projection methods, and structured low rank approximation. The algebraic connections are clear, with the SVD the main mathematical construct behind both Cadzow's basic algorithm and SSA.

In the simulation study within this paper, it has been demonstrated that repeated iterations of Cadzow's basic algorithm (in an attempt to separate the noise from the signal) may result in an increased RMSE from the true signal. This is particularly the case if the number of terms selected in the truncated SVD is smaller than the rank of the true signal. Indeed, if the number of terms in the truncated SVD is chosen to correspond (or is larger than) the rank of the true signal, the RMSE remains fairly constant. Thus there is no perceived benefit in running more than one iteration of Cadzow's basic algorithm. As stated earlier, one iteration of Cadzow's basic algorithm corresponds to the basic SSA.

In terms of forecasting, there are some subtle differences between Cadzow's basic algorithm and that of SSA. For the example considered within this paper, the forecast based upon SSA appears less volatile than the corresponding forecast made by using Cadzow's basic algorithm. By truncating the series, a retrospective forecast demonstrated that the RMSE obtained using Cadzow's basic algorithm was larger than the corresponding value obtained using basic SSA.

As a result the Cadzow's methodology seems to have no major superiority over SSA, but Cadzow's methodology involves the computation of many more SVD's. It is therefore suggested that SSA may be a viable alternative in approximating and forecasting a signal.

REFERENCES

[1] BAUSCHKE, H. H. and BORWEIN, J. M. (1993). On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Values Analysis.* **1** 185–212. MR1239403

[2] CADZOW, J. (1988). Signal enhancement: a composite property mapping algorithm. *IEEE Trans. on Acoustics, Speech, and Signal Processing.* **36**(2) 49–62.

[3] CHENEY, W. and GOLDSTEIN, A. A. (1959). Proximity maps for convex sets. *Proc. Amer. Math. Soc.* **10** 448–450. MR0105008

[4] CHU, M. T. and GOLUB, G. H. (2005). *Inverse Eigenvalue Problems: Theory, algorithms and applications*, OUP, Oxford. MR2263317

[5] DE MOOR, B. (1994). Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Trans. on Signal Processing.* **42**(11) 3104–3113.

[6] DEUTSCH, F. (2001). Accelerating the convergence of the method of alternating projections via a line search: a brief survey. *in Inherently parallel algorithms in feasibility and optimization and their applications* (Haifa, 2000), 203–217. MR1853223

[7] EPHRAIM, Y. and VAN TREES, H. L. (1995) A signal subspace approach for speech enhancement. *IEEE Trans. on Speech and Audio Processing.* **3**(4) 251–266.

[8] GLUNT, W., HAYDEN, T. L., HONG, S. and WELLS, J. (1990) An alternating projection algorithm for computing the nearest Euclidean distance matrix. *SIAM J. on Matrix Analysis and Applications.* **11**(4) 589–600. MR1066161

[9] GOLYANDINA, N., NEKTUTKIN, V. and ZHIGLJAVSKY, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, Boca Raton. MR1823012

[10] GOLYANDINA, N. (2010) On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and Its Interface.* **3** 259–279.

[11] HAN, S-P. (1988) A successive projection method. *Math. Programming.* **40**(1) 1–14. MR0923692

[12] HANSEN, P. C. and JENSEN, S. H. (2007). Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: survey and analysis. *EURASIP Journal on Advances in Signal Processing.* **Volume 2007** 1–24. MR2366752

[13] HASSANI, H. (2007) Singular spectrum analysis: methodology and comparison. *Journal of Data Science.* **5** 239–257.

[14] HIGHAM, N. J. (2002) Computing the nearest correlation matrix – a problem from finance. *IMA J. Numer. Anal.* **22**(3) 329–343. MR1918653

[15] MARKOVSKY, I. (2008). Structured low-rank approximation and its applications. *Automatica.* **44** 891–909. MR2530933

[16] MCAULAY, R. J. and QUATERI, T. F. (1986) Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. on Acoustics, Speech and Signal Processing.* **34**(4) 744–754.

[17] SCHMIDT, K. M. and MOSKVINA, V. (2003). Approximate projectors in singular spectrum analysis. *SIAM J. on Matrix Analysis and Applications.* **24** 932–942. MR2003313

[18] THORPE, A. J. and SCHARF, L. L. (1995) Data adaptive rank-shaping methods for solving least squares problems. *IEEE Trans. on Signal Processing.* **43**(7) 1591–1601.

[19] TUFTS, D. W. and SHAH, A. A. (1993) Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix. *IEEE Trans. on Signal Processing.* **41**(4) 1716–1721.

[20] VAN HUFFEL, S. and VANDEWALLE, J. (1991). *The Total Least Squares Problem: Computational aspects and analysis*, SIAM, Philadelphia. MR1118607

[21] VAN HUFFEL, S. (1993). Enhanced resolution based on minimum variance estimation and exponential data modelling. *Signal Processing.* **33**(3) 333–355.

[22] VARADI, F., PAP, J. M., ULRICH, R. K., BERTELLO, L. and HENNEY, C. J. (1999) Searching for signal in noise by random-lag singular spectrum analysis. *The Astrophysical Journal.* **526** 1052–1061.

[23] YIOU, P., SORNETTE, D. and GHIL, M. (2000) Data-adaptive wavelets and multi-scale singular spectrum analysis. *Physica D.* **142** 254–290. MR1771905

Jonathan Gillard
School of Mathematics
Cardiff University
E-mail address: GillardJW@Cardiff.ac.uk