Comp. Appl. Math.
DOI 10.1007/s40314-016-0309-9

Weighted moving averaging revisited: an algebraic approach

Mantas Landauskas[1] · **Zenonas Navickas**[2] · **Alfonsas Vainoras**[3] · **Minvydas Ragulskis**[1]

Received: 28 July 2015 / Revised: 3 January 2016 / Accepted: 11 January 2016

Abstract An algebraic approach for the selection of weight coefficients for weighted moving averaging is proposed in this paper. The algebraic complexity of the sequence transformed by weighted moving averaging is set as a target criterion for the optimization problem of weight coefficients. A special computational setup is constructed in order to tackle the inevitable additive noise for real-world time series. Computational experiments prove that the proposed approach can outperform time series predictors based on classical moving averaging.

Keywords Moving average · Time series prediction · Weight coefficients

Mathematics Subject Classification 37M10 · 11B37 · 37M99

Communicated by Antonio José Silva Neto.

✉ Mantas Landauskas
mantas.landauskas@ktu.lt

Zenonas Navickas
zenonas.navickas@ktu.lt

Alfonsas Vainoras
alfonsas.vainoras@lsmuni.lt

Minvydas Ragulskis
minvydas.ragulskis@ktu.lt
http://www.personalas.ktu.lt/~mragul/

1 Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-147, Kaunas 51368, Lithuania

2 Department of Applied Mathematics, Kaunas University of Technology, Studentu 50-325, Kaunas 51368, Lithuania

3 Sport Institute, Lithuanian University of Health Sciences, Kaunas, Lithuania

Published online: 29 January 2016

1 Introduction

Predicting short-time series is a challenging problem in many areas of science and engineering. In general, the objective of all time series prediction techniques is to build a model of the process and then use this model to extrapolate past behavior into future. Time series prediction becomes even a more complicated task if the amount of available data for building the model is small.

It is agreeable that no single predictor will outperform all other predictors in all possible situations; however, it is admitted that a forecasting model is worth applying if it is better than the persistence model (also known as the naïve, or MA(1) predictor). For example, the root mean square error (RMSE) of the ANN model for predicting wind speed proposed in Ma et al. (2013) is 10.77 % less than RMSE of the persistence model.

Short-term wind speed prediction has attracted lots of attention recently creating a separate niche for short-time series prediction in sustainable engineering (Nan et al. 2013). Applications as the hourly scheduling of the generators is important in balancing power production and demand (Dong-xiao et al. 2012).

The development of different techniques for short-time series forecasting continues; some typical examples of such developments are presented in Ragulskis et al. (2011), Palivonaite and Ragulskis (2014) or Easton et al. (2014) to mention a few. An interesting combination of the naïve and other prediction models is utilized in Chen (2011) for forecasting the tourism demand. It is common in short-time series prediction that the naïve method might be the optimal method (depending on the type of the error estimator considered) although the overall performance of the naïve method can be lower compared to that of the combination models.

The demand of improved or specialized techniques for time series forecasting led to various other approaches such as the use of Hybrid Monte Carlo algorithm (Dong-xiao et al. 2012), decomposing time series using wavelets (Jia et al. 2014), using a linear combination of forecasters (Firmino et al. 2014), or reconstruction of near-optimal algebraic skeletons (Palivonaite and Ragulskis 2014). It was noted in Zhang et al. (2013) that stochastic characteristics of a real-world series need more complex approaches than statistical techniques for capturing the nonlinear relations.

Often data needs to be preprocessed in order for some forecasting method to be efficiently applied; such preprocessing might improve the results as demonstrated in Wang et al. (2014). Decomposition of the original time series into a finite number of separate primitives so that each primitive could be predicted separately can also improve the efficiency of short-time series predictors (Wang et al. 2014). The overview of preprocessing importance for track irregularity data could be found in Jia et al. (2014). Weighted moving average plays a central role in the preprocessing of original time series in Jia et al. (2014); the goal there is to find such weight coefficients which lead to the optimal decomposition of the series.

The main objective of this paper is to propose an alternative algebraic approach to weighted moving average. This paper is organized as follows. Averaging of a linear recurrence sequence is discussed in Sects. 3 and 4; a technique for time series forecasting based on algebraic weighted moving averaging is proposed in Sect. 5; numerical experiments are presented in Sect. 5.2 and concluding remarks are given in Sect. 6.

2 Preliminaries

An order n linear homogeneous recurrence relation with constant coefficients is an equation of the form:

$$x_j = \alpha_{n-1} x_{j-1} + \alpha_{n-2} x_{j-2} + \cdots + \alpha_0 x_{j-n}, \tag{1}$$

where coefficients $\alpha_k, k = 0, 1, \ldots, n-1$ are constants. A sequence which satisfies a relation of this form is a linear recurrence sequence (LRS). The initial values $x_k, k = 0, 1, \ldots, n-1$ uniquely determines the evolution of this LRS (Kurakin and Nechavev 1995; Kurakin 2001; Park and Elden 2003). The auxiliary polynomial to (1) reads:

$$P(\rho) = \rho^n - \alpha_{n-1}\rho^{n-1} - \alpha_{n-2}\rho^{n-2} - \cdots - \alpha_0 \tag{2}$$

whose n roots describes the sequence satisfying the recurrence. If the roots $\rho_1, \rho_2, \ldots, \rho_n$ are all distinct then the recurrence takes the form:

$$x_j = \mu_1 \rho_1^j + \mu_2 \rho_2^j + \cdots + \mu_n \rho_n^j, \tag{3}$$

where the coefficients $\mu_1, \mu_2, \ldots, \mu_n$ are determined in order to fit the initial conditions of the recurrence. Note that all roots are real or complex conjugate if only LRS is real. If though some roots coincide, then the recurrence reads:

$$x_j = \sum_{k=1}^{r} \sum_{l=0}^{n_k - 1} \mu_{kl} \binom{j}{l} \rho_k^{j-l}, \tag{4}$$

where r is the number of distinct roots; n_k is the multiplicity index of the k-th root; $n_1 + n_2 + \cdots + n_r = n$.

The algorithm for the reconstruction of the model of LRS from a sequence $\{x_j\}_{j=0}^{+\infty}$ is more complex if the order of LRS is not known beforehand. Hankel transform of $\{x_j\}_{j=0}^{+\infty}$ yields the sequence $\{h_j\}_{j=0}^{+\infty}$ where $h_j = \det H_j$ and $H_j = (x_{k+l-2})_{1 \le k,l \le (j+1)}$ is a Hankel catalecticant matrix (matrix dimensions are $(j+1) \times (j+1)$). If there exists such $n \ge 1$ that $h_{n-1} \neq 0$ but $h_k = 0$ for all $k \ge n$, then $\{x_j\}_{j=0}^{+\infty}$ is LRS, its order is n and the auxiliary Equation 2 now reads:

$$\begin{vmatrix} x_0 & x_1 & \cdots & x_n \\ x_1 & x_2 & \cdots & x_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n-1} & x_n & \cdots & x_{2n-1} \\ 1 & \rho & \cdots & \rho^n \end{vmatrix} = 0 \tag{5}$$

As mentioned previously, we will consider only the case when roots of (5) are all distinct. Then the recurrence is uniquely described by (3).

Singular value decomposition (SVD) of a real matrix H results in the product $H = USV^T$, where U comprises orthonormal eigenvectors of HH^T; V comprises orthonormal eigenvectors of $H^T H$ and S is a diagonal matrix which elements are sorted square roots of eigenvalues of $H^T H$ (singular values of H).

Example 1 Let us consider a period-4 sequence $\{1, 0, -1, 0, 1, \ldots\}$ (note that periodic sequences comprise only a small fraction of the set of all possible LRS). It is clear that the order of this sequence is 2 because $\begin{vmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 1 \end{vmatrix}$ and all higher determinants are equal to 0. Then, the auxiliary equation reads:

$$\begin{vmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ 1 & \rho & \rho^2 \end{vmatrix} = -\left(1 + \rho^2\right) = 0. \tag{6}$$

Two distinct roots read: $\rho_1 = i; \rho_2 = -i$. Coefficients μ_k are determined from:

$$\begin{bmatrix} 1 & 1 \\ \rho_1 & \rho_2 \end{bmatrix} \cdot \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \mu_1 = \frac{1}{2}; \mu_2 = \frac{1}{2}. \tag{7}$$

Thus, finally,

$$x_j = \frac{1}{2} i^j + \frac{1}{2} (-i)^j; \quad j = 0, 1, 2, \ldots. \tag{8}$$

Now let us use this period-3 sequence and construct the following Hankel matrix:

$$H_4 = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 \end{bmatrix} \tag{9}$$

SVD of H_4 results into identical matrices U and V because $H_4^T H_4 = H_4 H_4^T$. The equality $\left| H_4^T H_4 - \lambda I \right| = 0$ yields the eigenvalues of $H_4^T H_4$:

$$\left| H_4^T H_4 - \lambda I \right| = \begin{vmatrix} 3-\lambda & 0 & -3 & 0 & 3 \\ 0 & 2-\lambda & 0 & -2 & 0 \\ -3 & 0 & 3-\lambda & 0 & -3 \\ 0 & -2 & 0 & 2-\lambda & 0 \\ 3 & 0 & -3 & 0 & 3-\lambda \end{vmatrix} = \lambda^3 \left(13\lambda - \lambda^2 - 36\right) = 0; \tag{10}$$

$\lambda_1 = 9$, $\lambda_2 = 4$ and $\lambda_{3,4,5} = 0$.

Lemma 1 *Let $\left\{x_j\right\}_{j=0}^{+\infty}$ is an LRS and its order is n. Let H_j is a $(j+1)$ order Hankel matrix of $\left\{x_j\right\}_{j=0}^{+\infty}$. Then the number of singular values of H_j not equal to zero is not higher than n.*

Proof Without loss of generality we assume a non-increasing order on the absolute values of the eigenvalues of H_j: $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq \left|\lambda_{j+1}\right|$.

Note that $H_j = H_j^T$, thus the singular values σ_k are the square roots of eigenvalues of H_j^2 which are equal to λ_k^2, $k = 1, \ldots, l$, where l is the number of distinct eigenvalues (Starzak 1989). Then $\sigma_k = |\lambda_k|$, $k = 1, \ldots, j+1$, since the singular values are computed without taking multiplicity into account.

Let $j \leq n-1$. The order of $\left\{x_j\right\}_{j=0}^{+\infty}$ is n. Thus, $\det(H_j) \neq 0$; $\lambda_1 \lambda_2 \ldots \lambda_{j+1} \neq 0$ and the number of nonzero singular values is 0.

Now, suppose that $j > n-1$. The nullity of H_j reads: $nul\left(H_j\right) = j + 1 - rank\left(H_j\right) = j + 1 - n$. The multiplicity of the zero eigenvalue of a matrix is equal to its nullity, thus $\sigma_1, \ldots, \sigma_n \neq 0$; $\sigma_{n+1}, \ldots, \sigma_{j+1} = 0$. □

Lemma 1 implies that SVD of the Hankel matrix of a sequence can be used as an effective computational tool for the determination of its order—if only that sequence is an LRS. However, every real-world time series is contaminated with inevitable noise. Thus, the order of a real-world time series is infinite even the evolution of the underlying model is governed by an LRS.

Let us assume that $\left\{x_j\right\}_{j=0}^{+\infty}$ is an order n LRS and $\left\{\varepsilon_j\right\}_{j=0}^{+\infty}$ is a discrete random variable. Then the sequence $\left\{x_j + \varepsilon_j\right\}_{j=0}^{+\infty}$ is not an LRS—however small is the additive noise (otherwise $\left\{\varepsilon_j\right\}_{j=0}^{+\infty}$ would be an LRS, which is contradictory to the definition of a random variable).

In other words, the computational procedure described in Lemma 1 is not applicable for the detection of the order of LRS—if only the investigated sequence is a real-world time series. Another approach is required for real-world time series.

Properties and computational aspects of the pseudospectrum of a square matrix is discussed in details in Wright and Trefethen (2002). Similar reasoning in respect of the pseudo-order of LRS could help to understand the effects introduced by the additive noise.

The spectrum of a square matrix A, denoted as $\Lambda(A)$ is the set of $z \in \mathbb{C}$ where the resolvent $(zI - A)^{-1}$ does not exist or is unbounded (Trefethen 1999), where I is the identity matrix. The ε-pseudospectrum of A is the set

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : z \in \Lambda(A + E) \text{ for some } E \text{ with } \|E\| < \varepsilon\} \tag{11}$$

for each $\varepsilon > 0$. In analogy to the classical definition of the spectrum of a square matrix we define the H-spectrum of the LRS as the set of roots of the auxiliary equation (5):

$$P(x_0, x_1, \ldots, x_{2n-1}) = \{\rho_k\}_{k=1}^{n}. \tag{12}$$

Then, the ε-H-spectrum of a square Hankel matrix is defined as:

$$P_\varepsilon(x_0, x_1, \ldots, x_{2n-1}) = P(x_0 + \varepsilon_0, x_1 + \varepsilon_1, \ldots, x_{2n-1} + \varepsilon_{2n-1}) \tag{13}$$

for some $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n-1} \in \mathbb{R}$ such that $\|\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n-1}\|_2 \leq \varepsilon$.

Such a definition of the ε-H-spectrum does enable a computational investigation of the effects introduced by the additive noise. In other words, such an approach helps to investigate such real-world time series where an underlying LRS is contaminated by noise.

Note that the coefficients of the auxiliary Equation (5) are computed as adjuncts with respect to the last line of the determinant (5). But Palivonaite and Ragulskis (2014),

$$\det(A + \varepsilon E) = \det(A) + \det(A)\, tr\left(A^{-1}E\right) \cdot \varepsilon + O\left(\varepsilon^2\right). \tag{14}$$

In other words, the ε-H-spectrum converges continuously to the H-spectrum as $\varepsilon \to 0$ if only A is not singular (what is true if the size of the Hankel matrix does correspond to the order of the LRS). Also, all roots of the perturbed sequence are either real numbers or complex conjugate numbers because the perturbation does not damage the symmetricity of the matrix.

Example 2 Let us consider the same sequence from Example 1—but perturb it by adding the external noise. Let us assume that $\varepsilon = 0.1$. Note that 4 elements of the sequence are required for the auxiliary Equation 5. Let us assume that $\tilde{\varepsilon}_0 = 0.1$; $\tilde{\varepsilon}_1 = -0.5$; $\tilde{\varepsilon}_2 = 0.3$ and $\tilde{\varepsilon}_3 = -0.2$ (it would be convenient to generate $\tilde{\varepsilon}_k$ as random numbers distributed uniformly in interval $(-1, 1)$). But $\tilde{n} = \|\tilde{\varepsilon}_0, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \tilde{\varepsilon}_3\|_2 \approx 0.6245 \neq 0.1$. The perturbation takes the form:

$$\varepsilon_k = \varepsilon \cdot \frac{\tilde{\varepsilon}_k}{\tilde{n}}; \quad k = \overline{0, 3}. \tag{15}$$

Note that $\|\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3\|_2 = \varepsilon$. The perturbed auxiliary equation reads:

$$\begin{vmatrix} 1.0160 & -0.0801 & -0.9520 \\ -0.0801 & -0.9520 & -0.0320 \\ 1 & \rho & \rho^2 \end{vmatrix} = -0.9736\rho^2 + 0.1088\rho - 0.9037 = 0. \tag{16}$$

Finally, $P_{0.1}(x_0, x_1, \ldots, x_{2n-1}) = \{0.0559 \pm 0.9618i\}$. Repeating this computational experiment many times helps to plot "clouds" of roots of perturbed auxiliary equations (Fig.

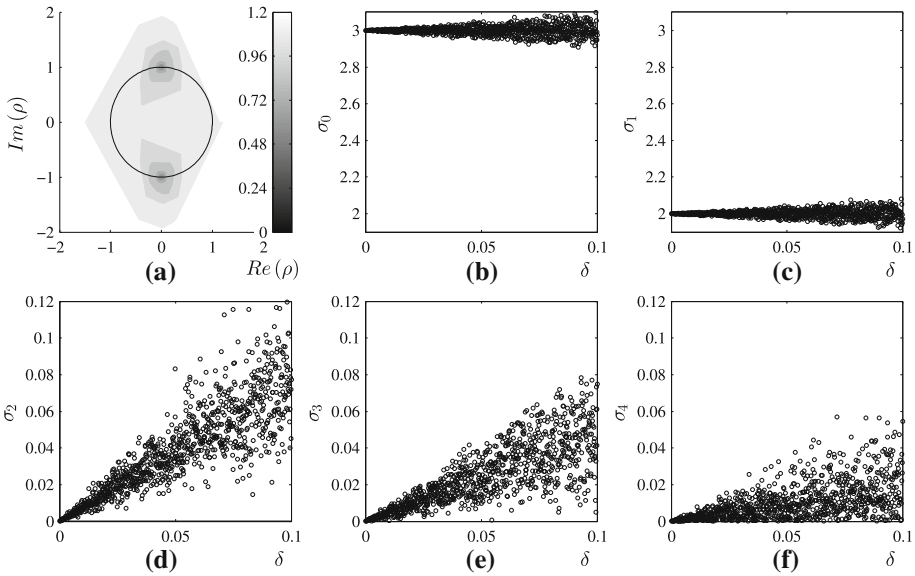

Fig. 1 The ε-H-spectrum converges continuously to the H-spectrum as $\varepsilon \to 0$. The distribution of roots of the auxiliary equation for the perturbed sequence in Example 2 are illustrated in **a**. The continuous convergence of SVD Eigenvalues σ_k; $k = \overline{0,4}$ of the perturbed matrix H_4 are illustrated in **b**–**f**

1a). Computational experiments illustrate continuous convergence of the ε-H-spectrum to the H-spectrum as $\varepsilon \to 0$.

Similar computational experiments can be performed for the Eigenvalues of $H_4^T H_4$. Note that the length of the perturbation vector is 9 (H_4 is a square Hankel matrix of order 5). The perturbation does not damage the symmetricity of the matrix; the matrix remains catalecticant. SVD of H_4 produces 5 Eigenvalues—all of them are shown in Fig. 1b–f accordingly. Contradictory to the previous experiment with the auxiliary equation, we do perform a single perturbation for every discrete value of ε. It can be seen that SVD Eigenvalues of the perturbed matrix do converge continuously to Eigenvalues of H_4 as $\varepsilon \to 0$. Moreover, SVD of the Hankel matrix of a perturbed sequence can be still used as an effective computational tool for the determination of the order of the non-perturbed sequence—one just needs to setup the error level and count how many Eigenvalues are higher than this error level. The techniques for the selection of this error level are discussed in next sections.

3 Weighted moving average and the order of LRS

A classical moving average (MA) method transforms the original sequence $\{x_j\}_{j=0}^{+\infty}$ into a sequence

$$\{y_j\}_{j=0}^{+\infty} = \left\{ \frac{1}{L} \sum_{s=0}^{L-1} x_{s+j} \right\}_{j=0}^{+\infty}$$

where L is the width of the observation window. MA techniques are widely used in signal smoothing (Manikandan and Soman 2012), time series forecasting applications (Holt 2004).

It is well known that MA(1) (MA at $L = 1$, or naive prediction techniques) are one of the best predictors for highly random, complex time series (Sauer et al. 1991).

Weighted moving average method WMA(L) differs from MA(L) in the sense that the weight coefficients w_s are not equal:

$$\{y_j\}_{j=0}^{+\infty} = \left\{ \sum_{s=0}^{L-1} w_{L-s} x_{s+j} \right\}_{j=0}^{+\infty}. \tag{17}$$

Let us consider that $\{x_j\}_{j=0}^{+\infty}$ is an LRS. It is clear that WMA(L) does transform an LRS into an LRS. Equations 4 and 17 yield:

$$y_j = \sum_{s=0}^{L-1} w_{L-s} \sum_{k=1}^{r} \sum_{l=0}^{n_k-1} \mu_{kl} \binom{s+j}{l} \rho_k^{s+j-l} = \sum_{k=1}^{m} \sum_{t=0}^{n_k-1} \hat{\mu}_{kt} j^t \rho_k^j.$$

In general, it may happen that a particular combination of weight coefficients w_k and roots ρ_k may result into such situations when some of the expressions $\hat{\mu}_{kt}$ become equal to zero. In such cases the order of the transformed LRS can become lower than the order of the original sequence. It can be noted that the expressions $\hat{\mu}_{kt}$ do not depend on j.

Example 3 Consider a sequence of period-6. Let this sequence be described by the roots of the characteristic equation $\rho_k = e^{\frac{\pi}{3}(k-1)i}$, $k = \overline{1,6}$.

We choose μ values in the way that makes y_j, $j = 1, 2, \ldots$, real numbers. Thus let:

$$\mu_1 = 0.5,\ \mu_2 = -0.2 + 0.1i,\ \mu_3 = 0.3 + 0.2i,$$

$$\mu_4 = -0.3,\ \mu_5 = 0.3 - 0.2i,\ \mu_6 = -0.2 - 0.1i.$$

The defined variables result in period-6 sequence with first members:

$$0.4,\ 0.3 - 0.3\sqrt{3},\ 0.1 + 0.1\sqrt{3},$$

$$1.8,\ 0.1 - 0.1\sqrt{3},\ 0.3 + 0.3\sqrt{3}, \ldots$$

Consider $L = 6$ and $w_i = \frac{1}{6}$, $i = \overline{1, L-1}$. It is natural that WMA yields in a new constant sequence $\frac{1}{2},\ \frac{1}{2},\ \frac{1}{2} \ldots$.

But suppose $L = 4$ and $\{w_i\}_{i=0}^{3} = \{0.1; 0.4; 0.4; 0.1\}$. Now the resulting new sequence is of period-6, but has 2 elements which are the same:

$$0.38 - 0.08\sqrt{3},\ 0.8,\ 0.8,$$

$$0.38 + 0.08\sqrt{3},\ 0.32 + 0.08\sqrt{3},\ 0.32 - 0.08\sqrt{3}, \ldots$$

The roots of the characteristic equation of the new sequence comprise with ρ_k, $k = \overline{1,6}$ except now there is no ρ_4.

4 WMA and real-world time series

As discussed previously, the order of a real-world time series is infinite. But it has also been shown that SVD can be used for the identification of the order of an LRS contaminated by an additive noise—if only this LRS does exist.

Let us construct the following optimization problem.

4.1 Preprocessing

Given a time series $\{x_j\}_{j=0}^{n}$ (this time series is not necessarily an LRS), set the order of the Hankel matrix m and the error level $\varepsilon > 0$. Set the width of the observation window L and perform MA: $\{y_j\}_{j=0}^{+\infty} = \left\{\frac{1}{L}\sum_{s=0}^{L-1} x_{s+j}\right\}_{j=0}^{+\infty}$. Construct the catalecticant Hankel matrix from $\{y_j\}_{j=0}^{+\infty}$ (starting from y_0). Perform SVD of the Hankel matrix; SVD Eigenvalues are denoted as $\{\sigma_k^2\}_{k=1}^{m}$. Compute the number of SVD Eigenvalues greater than ε: $N_\varepsilon = \sum_{k=1}^{m} \delta_k$, where $\delta_k = \begin{cases} 1 \; if \;\; \sigma_k^2 \geq \varepsilon \\ 0 \; if \;\; \sigma_k^2 < \varepsilon \end{cases}$. Store MA weight coefficients $\{\frac{1}{L}\}_{j=1}^{L}$ and N_ε as the current best result.

4.2 Target function and constraints

The goal of the optimization procedure is to find such weight coefficients $\{w_j\}_{j=1}^{L}$ that the order of the WMA is the lowest possible: $\min_{\{w_1,\ldots,w_L\}} N_\varepsilon$. Such an optimization problem has a trivial solution: $w_1 = \cdots = w_L = 0$. Additional constraints for the weight coefficients are required to make this a well posed optimization problem. We set natural requirements:

$$\sum_{j=1}^{L} w_j = 1; \;\; w_j > 0 \quad \text{for all } j. \tag{18}$$

Note that MA weight coefficients belong to the feasible set of the constrained optimization problem.

4.3 Iterations

Find such weight coefficients of WMA that result into a minimum order of the LRS. Generate a vector of weight coefficients $\{w_j\}_{j=1}^{L}$ satisfying constraints (18) and perform WMA: $\{y_j\}_{j=0}^{+\infty} = \left\{\sum_{s=0}^{L-1} w_{L-s} x_{s+j}\right\}_{j=0}^{+\infty}$. Compute the number of SVD Eigenvalues of the Hankel matrix of the WMA sequence greater than ε : N_ε. If the value of N_ε is lower than the best stored value, save the current vector of weight coefficients $\{w_j\}_{j=1}^{L}$ and N_ε as the best current arguments and the best current value of the target function. Otherwise, execute the same iteration again.

4.4 The optimization strategy

The strategy for the generation of a sequence of weight coefficients $\{w_j\}_{j=1}^{L}$ is an intrinsic feature of the optimization algorithm—a wise selection of this strategy may guarantee an effective computational solution of the optimization problem (note that the defined target function maps $\mathbb{R}^n$ into $\mathbb{N}_0$). However, neither deterministic strategies, nor evolutionary approaches based on particle swarm optimization (PSO) or genetic algorithms (GA) do not produce satisfactory results. That can be explained by the topology of the target function (a simple two-dimensional computational experiment at $L = 2$ is used for clarity).

Let us consider the chaotic logistic map (Weisstein 2015) defined by:

$$x_{j+1} = r x_j (1 - x_j)$$

at $r = 4$ and starting from the initial condition $x_0 = 0.44$. Initially, we release constraints 18 and require only that $0 \leq w_j \leq 1$; $j = 1, 2$. N_ε is computed at the grid-points in the whole domain—the results are illustrated in Fig. 2a. It is clear that the global minimum is reached at $w_1 = w_2 = 0$.

Note that the constraint $w_1 + w_2 = 1$ yields a feasible set which is geometrically represented as a line interval in the domain of Fig. 2a; the problem becomes a single variable optimization problem. In general, constraints (18) reduce the dimension of the optimization problem by one. But the generation of a vector of the weight coefficients which does belong to the feasible set is not so straightforward if the dimension of the problem is higher than 2. For example, one could generate random values of w_1 and w_2 distributed uniformly in the interval [0,1]—but $w_3 = 1 - w_1 - w_2$ does not always yield a feasible point in case $L = 3$ (consider a situation when $w_1 = 0.9$ and $w_2 = 0.9$).

Therefore, the generation of numerical values of the weight coefficients $\{w_j\}_{j=1}^{L}$ is executed according to the following algorithm. First of all, a preselected search strategy algorithm is allowed to generate numbers $\{e_j\}_{j=1}^{L}$ with the only limitation $0 \leq e_j \leq 1$; $j = 1, \ldots, L$. Then, the weight coefficients are computed as follows:

$$w_j = \frac{e_j}{\sum_{k=1}^{L} e_k}; \quad j = 1, \ldots, L. \tag{19}$$

The vector $\{e_j\}_{j=1}^{2}$ is allowed to span over the whole domain $[0, 1] \times [0, 1]$ in Fig. 2b. The resulting distribution of N_ε (Fig. 2b) clearly shows that the discontinuous surface of the values of the target function comprises relatively wide flat zones (which is predetermined by the mapping $\mathbb{R}^n \rightarrow \mathbb{N}_0$). In other words, gradient decent methods and even PSO or GA methods would fail in generating a sequence of points converging to a local minimum. Therefore, random Monte Carlo (MC) search strategy is selected for all further computational experiments.

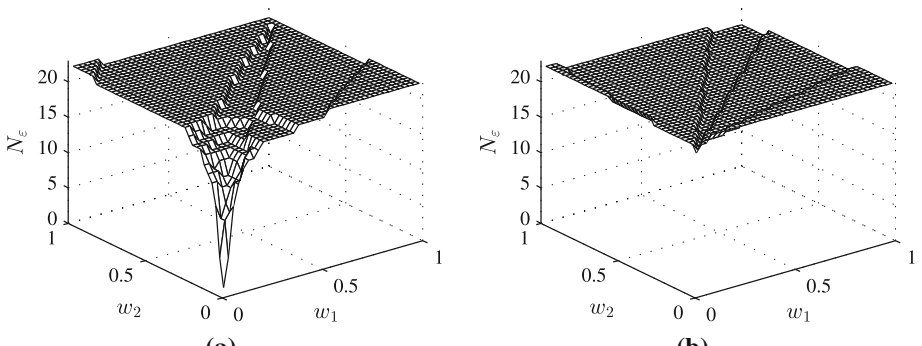

Fig. 2 The geometrical shape of the target function N_ε in respect of weight coefficients w_1 and w_2. The constraint $w_1 + w_2 = 1$ is released in **a** and is in force according to Eq. (18) in **b**

4.5 Computational experiments

Let us consider a random Gaussian noise time series with zero mean and the standard deviation equal to 1. Let $L = 10$ and $\varepsilon = 0.1$. It is well known that it is impossible to find an optimal embedding time delay vector for a random noise sequence (Sauer et al. 1991)—it is hard to expect that WMA would produce better results compared to MA. As mentioned previously, MC techniques will be used for the generation of a vector of random weight coefficients according to (19). But first of all, the preprocessing (MA) produces $N_\varepsilon = 89$ (this result is shown as a thin solid horizontal line in Fig. 3a). One thousand MC iterations are executed; the value of N_ε in each trial is visualized as an empty circle in Fig. 3a along the horizontal axis. It appears that 1000 MC trials cannot generate weight coefficients which would result in a lower value of N_ε than produced by MA. The weight coefficients resulting into the best N_ε are shown in Fig. 3b.

Next, we continue computational experiments with a synthetic sequence $x_j = \sin\left(\frac{j}{10}\right) + 0.01 \cdot rand_j$; $j = 0, 1, \ldots$, where $rand_j$ is a discrete normal Gaussian random variable $N(0, 1)$ generated at the j-th step. Now, MA produces $N_\varepsilon = 43$. But 1000 MC iterations manage to produce such weight coefficients which do result in a lower value of N_ε compared to MA (Fig. 4a). The best set of weight coefficients is illustrated in Fig. 4b.

5 Time series forecasting based on AWMA

5.1 The forecasting algorithm

As mentioned previously, MA techniques are often used as simple but effective time series forecasting tools. The ability to identify underlying algebraic relationships by varying the weight coefficients of WMA suggests an alternative approach for time series prediction. The schematic diagram for algebraic weighted moving average (AWMA) prediction technique is illustrated in Fig. 5.

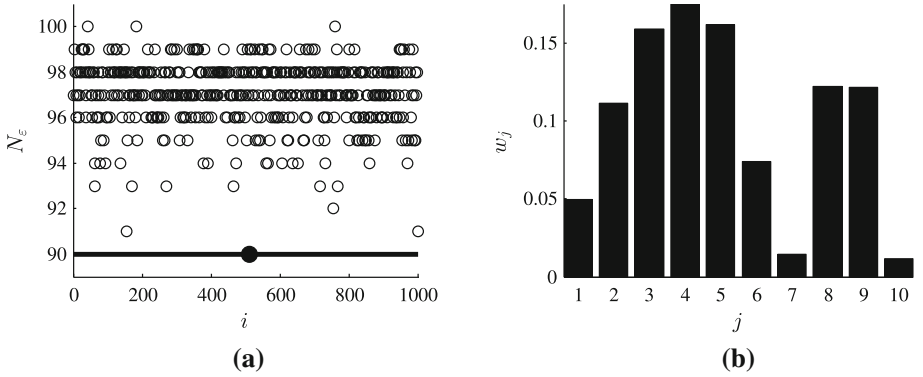

Fig. 3 Near-optimal set of weight coefficients for the Gaussian noise; the width of the observation window is 10. 1000 MC trials are illustrated in **a**; *thick solid horizontal line* denotes the result for the classical MA; the *black dot* shows the best result produced by random search. **b** The weight coefficients for the best result (the *black dot* in **a**). Index i denotes an MC trial number; j—is the weight coefficient index

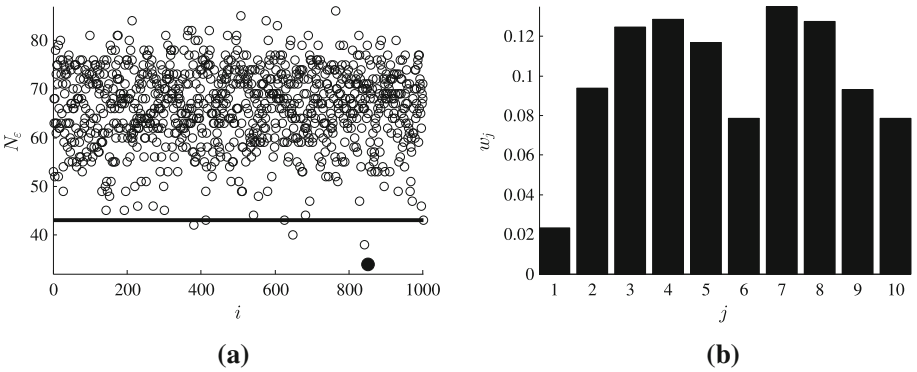

Fig. 4 Near-optimal set of weight coefficients for the synthetic time series contaminated with noise; the width of the observation window is 10. 1000 MC trials are illustrated in **a**; *thick solid horizontal line* denotes the result for the classical MA; the *black dot* shows the best result produced by random search. **b** The weight coefficients for the best result (the *black dot* in **a**)

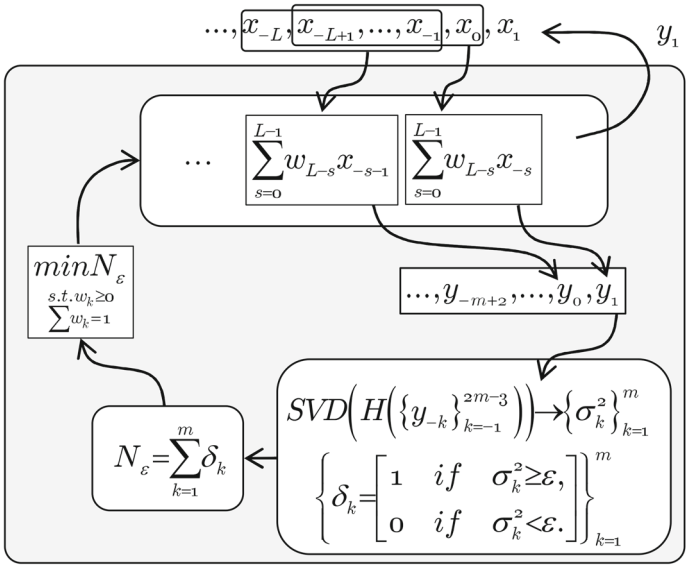

Fig. 5 The schematic diagram of AWMA prediction

Let x_0 be the value of the original sequence at the present time moment; thus x_s; $s = -1, -2, \ldots$ will denote the past values of the sequence.The present and past values of the sequence are used to forecast x_1. The algorithm described in Sect. 4 is used to identify a set of near-optimal weight coefficients $\{w_j\}_{j=1}^{L}$; the weighted average $y_1 = \sum_{s=0}^{L-1} w_{L-s} x_{-s}$ is used as a direct estimate of x_1. Note, that the WMA forecasting becomes MA forecasting if only it appears that it is impossible to find a better set of weight coefficients than the MA coefficients (the preprocessing step in Sect. 4.1): $y_1 = \frac{1}{L} \sum_{s=0}^{L-1} x_{-s}$.

The selection of the parameter ε is considered in Sect. 4—but the selection of the width of the observation window L has not been discussed yet. In fact, the strategy for the selection of L is straightforward. First of all, one needs to determine the optimal width of the observation

window for MA prediction; the optimality criterion is the RMSE of MA prediction. Then, at each time step one should try to find such set of weight coefficients which would results into a lower value of N_ε compared to the one produced by MA. One step forward WMA prediction is executed if the search was successful; ordinary MA prediction is performed otherwise.

5.2 Computational experiments

Quarterly percentage increase in estimated resident population of Australia Sep 71–Jun 93 (also named as *erppc*) was chosen for computational experiments (the dataset is available online at the Time Series Data Library Hyndman 2012).

Table 1 RMSE as a function of the observation window for MA prediction

L	2	3	4	5	6	7	8	9
RMSE	0.0749	0.0707	0.0689	0.0697	0.0709	0.0727	0.0739	0.0752

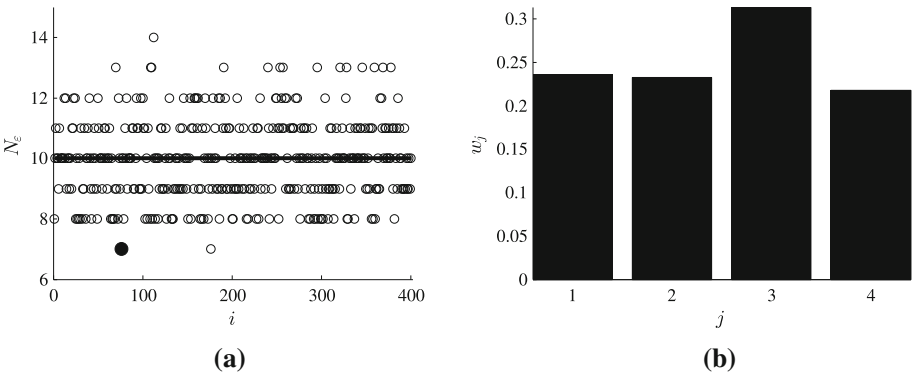

Fig. 6 Near-optimal set of weight coefficients for *erppc*; the width of the observation window is 4. 400 MC trials are illustrated in **a**; *thick solid horizontal line* denotes the result for the classical MA; the *black dot* shows the best result produced by random search. **b** The weight coefficients for the best result (the *black dot* in **a**)

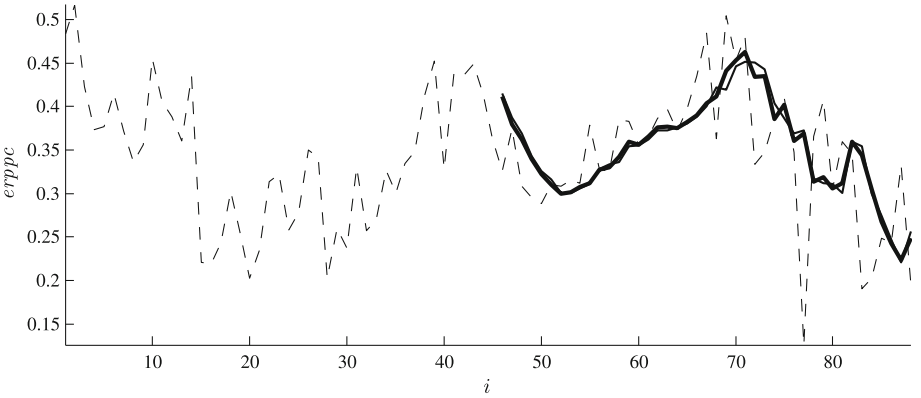

Fig. 7 The application of AWMA to *erppc* series in comparison to the standard moving average at $n = 4$; $\varepsilon = 0.04$. Original data are plotted by the dashed line; MA prediction—by *thin solid line*; AWMA—by *thick solid line*

Table 1 shows RMSE for MA prediction at different values of the length of the observation window. The prediction was most accurate with $L = 4$.

AWMA results for the *erppc* series are shown in Figs. 6 and 7. Each prediction was generated by simulating 10,000 sets of the weight coefficients for WMA and then minimizing the rank of underlying skeleton sequence. Prediction errors for MA are 0.0689; for AMWA are 0.0647. Note that AWMA also outperforms the naïve prediction which in this case results in prediction errors equal to 0.0797.

6 Concluding remarks

An effective time series forecasting technique based on algebraic weighted moving averaging is proposed in this paper. The weight coefficients are selected in such a way that the algebraic complexity of the resulting averaged time series is minimal. The complex optimization problem is solved by employing random MC search strategy. The functionality and feasibility of the proposed short-time series forecasting technique is demonstrated by computational experiments with real-world time series.

The size of the algebraic model is automatically selected by minimizing the prediction errors produced by the classical MA for different observation windows. An alternative (but much more time consuming) approach would be to vary the observation window of the algebraic model independently of the MA results. Such an approach would probably result into even better predictions, but these investigations remain as a tangible topic for future research.

Acknowledgments This work was supported by the Lithuanian Science Council under project No. MIP-078/2015.

References

Chen KY (2011) Combining linear and nonlinear model in forecasting tourism demand. Exp Syst Appl 38(8):10368–10376

Dong-xiao N, Hui-feng S, Desheng DW (2012) Short-term load forecasting using bayesian neural networks learned by hybrid monte carlo algorithm. Appl Soft Comput 12(6):1822–1827

Easton JF, Stephens CR, Angelova M (2014) Risk factors and prediction of very short term versus short/intermediate term post-stroke mortality: a data mining approach. Comput Biol Med 54:199–210

Firmino PRA, de Mattos Neto PS, Ferreira TA (2014) Correcting and combining time series forecasters. Neural Netw 50:1–11

Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast 20(1):5–10

Hyndman RJ (2012) Time series data library. http://data.is/TSDLdemo. Accessed 16 Jan 2012

Jia C, Wei L, Wang H, Yang J (2014) Study of track irregularity time series calibration and variation pattern at unit section. Comput Intell Neurosci 2014:14. http://dx.doi.org/10.1155/2014/727948

Kurakin V (2001) Linear complexity of polinear sequences. Disctrete Math Appl 11(1):1–51

Kurakin, VL, Kuzmin AVMAS, Nechavev AA (1995) Linear complexity of polinear sequences. J Math Sci 76:2793–2915

Ma R, Hu SJ, Xu HH (2013) Very short-term wind speed prediction of a wind farm based on artificial neural network. Adv Mater Res Trans Tech Publ 608:677–682

Manikandan MS, Soman K (2012) A novel method for detecting R-peaks in electrocardiogram (ECG) signal. Biomed Signal Process Control 7(2):118–128

Nan X, Li Q, Qiu D, Zhao Y, Guo X (2013) Short-term wind speed syntheses correcting forecasting model and its application. Int J Electric Power Energy Syst 49:264–268

Palivonaite R, Ragulskis M (2014) Short-term time series algebraic forecasting with internal smoothing. Neurocomputing 127:161–171

Park H, Elden L (2003) Matrix rank reduction for data analysis and feature extraction. Tech. rep., technical report

Ragulskis M, Lukoseviciute K, Navickas Z, Palivonaite R (2011) Short-term time series forecasting based on the identification of skeleton algebraic sequences. Neurocomputing 74(10):1735–1747

Sauer T, Yorke JA, Casdagli M (1991) J Stat Phys. Embedology 65(3–4):579–616

Starzak ME (1989) Mathematical methods in chemistry and physics. Springer Science & Business Media, Berlin

Trefethen LN (1999) Computation of pseudospectra. Acta Numerica 8:247–295

Wang J, Hu Mm, Ge P, Ren Py, Zhao R (2014) A forecasting model for short term tourist arrival based on the empirical mode decomposition and support vector regression. In: Proceedings of 2013 4th international Asia conference on industrial engineering and management innovation (IEMI2013). Springer, Berlin, pp 1009–1021

Weisstein EW (2015) Logistic map. From MathWorld–a wolfram web resource. http://mathworld.wolfram.com/LogisticMap.html

Wright TG, Trefethen LN (2002) Pseudospectra of rectangular matrices. IMA J Numer Anal 22(4):501–519

Zhang W, Wang J, Wang J, Zhao Z, Tian M (2013) Short-term wind speed forecasting based on a hybrid model. Appl Soft Comput 13(7):3225–3233