○ Applied Informatics
a SpringerOpen Journal

CrossMark

Kernel fractional affine projection algorithm

Bilal Shoaib[1*†], Ijaz Mansoor Qureshi[2†], Shafqat Ullah Khan[3], Sharjeel Abid Butt[1] and Ihsan ul haq[1]

*Correspondence:
bilal.phdee42@iiu.edu.pk
†Bilal Shoaib and Ijaz
Mansoor Qureshi are equally
contributed to this article.
[1] Department of Electronic
Engineering, International
Islamic University, H-10,
Islamabad, Pakistan
Full list of author information
is available at the end of the
article

Abstract

This paper extends the kernel affine projection algorithm to a rich, flexible and cohesive taxonomy of fractional signal processing approach. The formulation of the algorithm is established on the inclusion of Riemann–Liouville fractional derivative to gradient-based stochastic Newton recursive method to minimize the cost function of the kernel affine projection algorithm. This approach extends the idea of fractional signal processing in reproducing kernel Hilbert space. The proposed algorithm is applied to the prediction of chaotic Lorenz time series and nonlinear channel equalization. Also the performance is validated in comparison with the least mean square algorithm, kernel least mean square algorithm, affine projection algorithm and kernel affine projection algorithm.

Keywords: Kernel affine projection algorithm, Riemann–Liouville derivative, Lorenz time series, Fractional signal processing approach

Background

Kernel-based learning algorithms gained interest since the last few years. Mercer's theorem is used in kernel-based learning algorithms to map the input data using some nonlinear kernel function to some higher dimensional feature space, known as reproducing kernel Hilbert space (RKHS), where the linear operations are easily performed on the input data. These kernel methods stem originally from support vector machines (Vapnik and Vapnik 1998; Hearst et al. 1998), a powerful tool in handling classification problems in the neural network architecture. Kernel principal component analysis (KPCA) and kernel regression (Scholkopf et al. 1997; Takeda et al. 2007; Hardle and Vieu 1992) also show desirable performance regarding classification in the complicated environment of statistical signal processing. However, these are batch mode methods and suffer the burden of high computational cost and memory usage. These issues are replaced by introducing the online kernel methods, such as kernel least mean square (KLMS) (Liu et al. 2008), kernel affine projection algorithm (KAPA) (Liu and Principe 2008), kernel recursive least squares (KRLS) (Engel et al. 2004; Liu et al. 2015) and extended kernel recursive least squares (Ex-KRLS) (Liu et al. 2009) algorithms. These online kernel algorithms are very much in common nowadays regarding system identification, weather forecasting, nonlinear channel equalization, prediction of stationary as well as nonstationary time

Shoaib *et al. Appl Inform* (2015) 2:12

Page 2 of 14

series. KLMS algorithm uses stochastic gradient method to minimize the mean square error-based cost function in its formulation on the transformed input data. In KAPA the gradient noise of the KLMS algorithm is removed by minimizing the cost function using the smoothed nonlinear Newton recursion method.

On the other hand, online learning algorithms based on fractional signal processing is introduced using the concept of fractional order calculus in the formulation of the algorithm. Ortigueira's work (Ortigueira et al. 2002; Ortigueira and Machado 2006; Ortigueira 2011) mainly considered as the pioneer in the field of fractional signal processing. Tseng et al. designed one- and two-dimensional finite impulse response filter using constraints regarding fractional derivative (Tseng and Lee 2012, 2013, 2014). Wang introduces fractional zero phase filtering based on Riemann–Liouville integrals (Wang et al. 2014). Raja and Qureshi introduced fractional least mean square algorithm (FLMS) (Zahoor and Qureshi 2009) for their work regarding system identification. In the recent past, FLMS algorithm has been applied to various multidimensional signal processing problems including parameter identification of nonlinear controlled autoregressive system, parameter estimation of CARMA systems (Zahoor and Chaudhary 2015), identification of Box-jenkins system, dual channel speech enhancement, Brownian motion modeling, performance analysis of the bessel beamformer, acoustic echo cancelation (Masoud and Osgouei 2011; Dubey and Rout 2012; Akhtar and Yasin 2012; Chaudhary et al. 2013), etc.

Recently a modified fractional least mean square algorithm (MFLMS) (Shoaib and Qureshi 2014a) is developed for stationary and nonstationary time series prediction, more specifically Mackey glass. Convergence of the MFLMS algorithm is also tested regarding the prediction of chaotic series along with different noise variances. To remove the guesswork existing in tuning the step size parameter of the MFLMS algorithm (Shoaib and Qureshi 2014b), a stochastic gradient-based method is introduced to adapt the step sizes of the MFLMS algorithm according to the mean square error and then its application towards the prediction of Mackey glass as well as Lorenz time series.

Kernel functions are widely used in obtaining the solution of fractional order nonlinear differential equations as discussed in Shoaib and Qureshi (2014a). They use different kernel function to model the fractional order nonlinear differential equation and then use heuristic computing techniques like genetic algorithm, particle swarm optimization (PSO), differential evolution (DE) to minimize the error function. Here in this paper we introduced a mechanism that combined the adaptive fractional learning algorithms and online kernel-based filtering algorithms. This idea greatly helps in improving the performance in solving nonlinear problems.

The main aim of this research work is the development of a kernel fractional affine projection algorithm (KFAPA). A method is introduced to adjust the Riemann–Liouville fractional derivative to formulate the KAPA algorithm to minimize the cost function based on mean square error using gradient-based smoothed nonlinear recursive method. The proposed algorithm is then applied on the prediction of only the X-component of the three-dimensional chaotic Lorenz time series and nonlinear channel equalization.

Organization of the paper is as follows: in "Affine projection and kernel affine projection algorithms" section, the brief introduction of affine projection and kernel affine projection algorithm is presented. "Fractional signal processing approach" section, presents the introduction of fractional signal processing and the proposed kernel fractional affine projection

Shoaib *et al. Appl Inform* (2015) 2:12

Page 3 of 14

algorithm. The experimental results are discussed in "Experimental results" and "Conclusion and future work" sections comprises of conclusion along with future directions.

Affine projection and kernel affine projection algorithms

Affine projection algorithm

Affine projection algorithm (Haykin 2013) uses smoothed Newton's recursion to formulate the algorithm. By minimizing the cost function, we use the following input and desired sequence $[\mathbf{x}(i), d(i)]$ as

$$J(w) = \frac{1}{2} \sum_{i=0}^{n} (d(i) - w^{\mathrm{T}} \mathbf{x}(i))^2 \tag{1}$$

The gradient with respect to w is

$$\nabla_w J = -\sum_{i=1}^{n} x(i)(d(i) - w^{\mathrm{T}} \mathbf{x}(i)) \tag{2}$$

The weights of the APA is adjusted using stochastic Newton method as

$$\mathbf{w}(i) = \mathbf{w}(i-1) - \eta_{\mathrm{t}}(\nabla_w^2 J)^{-1} \nabla_w J \tag{3}$$

Initialize the weight vector to zero, η_{t} is the small positive step size. Here we perform a line search along the gradient descent direction to compute the weight vector. The corresponding steepest descent and Newton's recursion is

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_{\mathrm{t}} \mathbf{x}(i)[\mathbf{x}^{\mathrm{T}}(i)\mathbf{x}(i) + \epsilon \mathbf{I}]^{-1}[d(i) - \mathbf{x}(i)^{\mathrm{T}} \mathbf{w}(i-1)] \tag{4}$$

ϵ is the small positive constant, which prevents division by zero errors. To smoothen the Newton's recursion and to increase the convergence speed we proceed as

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_{\mathrm{t}}[\mathbf{x}(i)\mathbf{x}^{\mathrm{T}}(i) + \epsilon \mathbf{I}]^{-1}\mathbf{x}(i)[d(i) - \mathbf{x}(i)^{\mathrm{T}} \mathbf{w}(i-1)] \tag{5}$$

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_{\mathrm{t}}[\mathbf{x}(i)\mathbf{x}^{\mathrm{T}}(i) + \epsilon \mathbf{I}]^{-1}[\mathbf{x}(i)d(i) - \mathbf{x}(i)\mathbf{x}(i)^{\mathrm{T}} \mathbf{w}(i-1)] \tag{6}$$

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_{\mathrm{t}}[\mathbf{R}_{\mathrm{u}} + \epsilon \mathbf{I}]^{-1}[\mathbf{r}_{\mathrm{d}\mathbf{u}} - [\mathbf{R}_{\mathrm{u}} + \epsilon \mathbf{I}]\mathbf{w}(i-1)] \tag{7}$$

where the matrix $\mathbf{R}$ is invertible and strictly positive definite.

Kernel affine projection algorithm

Poor performance is examined, where the mapping between $\mathbf{x}$ and $\mathbf{d}$ is somehow highly nonlinear. Nonlinear mapping is introduced in Weifeng et al. (2011) as $\varphi(\mathbf{x}(i))$, which is a powerful model $\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}(i))$ than $\mathbf{w}^{\mathrm{T}}\mathbf{x}$. So using this model and finding $\mathbf{w}$ through smoothed stochastic Newton method may prove an efficient method towards nonlinear filtering as APA ensures for linear problems. Using the sequence $[\varphi(i), d(i)]$ to parameter weight vector $\mathbf{w}$ as

$$J(w) = \frac{1}{2} \sum_{i=0}^{n} (d(i) - w^{\mathrm{T}}\varphi(i))^2 \tag{8}$$

Shoaib *et al. Appl Inform* (2015) 2:12

Page 4 of 14

Minimizing the cost function, using the stochastic gradient descent, the weight adaptation equation becomes

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_t \Phi(i)[\mathbf{d}(i) - \Phi^T \mathbf{w}(i-1)] \tag{9}$$

and stochastic Newton method becomes

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_t [\Phi(i)\Phi(i)^T + \epsilon \mathbf{I}]^{-1} \Phi(i)[\mathbf{d}(i) - \Phi^T \mathbf{w}(i-1)] \tag{10}$$

Using Searle's matrix identity

$$[\Phi(i)\Phi(i)^T + \epsilon \mathbf{I}]^{-1}\Phi(i) = \Phi(i)[\Phi(i)^T\Phi(i) + \epsilon \mathbf{I}]^{-1} \tag{11}$$

The corresponding weight update equation becomes

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta_t \Phi(i)[\Phi(i)^T\Phi(i) + \lambda \mathbf{I}]^{-1}[\mathbf{d}(i) - \Phi^T \mathbf{w}(i-1)] \tag{12}$$

Therefore, KAPA only needs a $K \times K$ matrix inversion which can be computed easily by sliding window trick.

Fractional signal processing approach

Introduction to fractional derivative

Fractional calculus is a widely used signal processing algorithm autoregressive (AR) systems identification. It has been utilized in various fields of signal processing effectively including echo cancelation, dual channel speech enhancement and performance analysis of the bessel beamformer. Before going towards the proposed algorithm, we here present some basic concepts about the fractional calculus that focus on fractional integral and derivative.

$$I^\nu f(t) = \frac{1}{\Gamma(\alpha)} \int_0^T (t-\nu)^{\nu-1} f(\tau) \, d\tau \tag{13}$$

I^ν is the fractional integral of order ν. The fractional derivative is given as

$$\begin{aligned}
(D^\nu f)(t) &= \left(\frac{d}{dt}\right)^n (I^{n-\nu}f)(t) \\
(D^\nu f)(t) &= \frac{1}{\Gamma(\alpha)} \left(\frac{d}{dt}\right)^n \int_0^T (t-\tau)^{n-\nu-1} f(\tau) \, d\tau
\end{aligned} \tag{14}$$

D^ν is the fractional derivative and ν is the integer. To add little more detail we present Riemann–Liouville fractional derivative as follows. The fractional derivative of $f(t) = (t-a)^\alpha$ is,

$$D^\nu (t-a)^\alpha = \frac{\Gamma(1+\alpha)}{\Gamma(1+\alpha+\nu)} (t-a)^{\alpha-\nu} \tag{15}$$

where a and α are real constants.

Proposed kernel fractional affine projection algorithm

Here we introduced a mechanism to update the weights of kernel affine projection algorithm with the inclusion of fractional derivative term.

$$\mathbf{w}(i) = \mathbf{w}(i-1) - \eta_t (\nabla_w^2 J)^{-1} \nabla_w J - \nabla_w^\nu J \tag{16}$$

Shoaib *et al. Appl Inform* (2015) 2:12

Page 5 of 14

and the fractional derivative of the cost function is written as

$$
\begin{aligned}
\left(\frac{\partial}{\partial w(i)}\right)^{\nu} J &= -e(i)\mathbf{x}(i)D^{\nu}\mathbf{w}(i) \\
\left(\frac{\partial}{\partial w(i)}\right)^{\nu} J &= -e(i)\mathbf{x}(i)\left[\frac{1}{\Gamma(2-\nu)}\mathbf{w}^{1-\nu}(i)\right]
\end{aligned}
\tag{17}
$$

The weight update equation of the kernel fractional affine projection algorithm is

$$
\begin{aligned}
\mathbf{w}(i) &= \mathbf{w}(i-1) + \eta_t \Phi(i)[\Phi(i)^{\mathrm{T}}\Phi(i) + \lambda\mathbf{I}]^{-1}[\mathbf{d}(i) - \Phi^{\mathrm{T}}\mathbf{w}(i-1)] \\
&\quad + \mu_t(\mathbf{e(i)}^{\mathrm{T}}\varphi(i))\frac{\mathbf{w}^{1-\nu}(i-1)}{\Gamma(2-\nu)}
\end{aligned}
\tag{18}
$$

where η_t and μ_t are the small positive step sizes, typically lie between 0 and 1. In practice, we do not have access to the transformed weights $\mathbf{w}$ in the feature space, so the updated weights have to be evaluated through expansion coefficients as

$$
\mathbf{w}(i) = \sum_{j=1}^{i} \mathbf{a}_j(i)\varphi(j), \quad \forall i > 0.
\tag{19}
$$

Now to evaluate the $\mathbf{a}_j(i)$, setting the initial guess $\mathbf{w}(0) = 0$, and adopted procedure as

$$
\begin{aligned}
\mathbf{w}(0) &= 0 \\
\mathbf{w}(0) &= \eta d(1)\varphi(1) = \mathbf{a}_1(1)\varphi(1) \\
\vdots &= \vdots \\
\mathbf{w}(i-1) &= \sum_{j=1}^{i-1} \mathbf{a_j}(i-1)\varphi(j)
\end{aligned}
$$

similarly the $K \times K$ inversion in 17 is evaluated by the following Searle's identity

$$
\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}
\tag{20}
$$

where $\mathbf{A} = \Phi(i-1)^{\mathrm{T}}\Phi(i-1) + \epsilon\mathbf{I}$, $\mathbf{B} = \Phi(i)^{\mathrm{T}}\varphi(i)$, $\mathbf{C} = \varphi(i)^{\mathrm{T}}\Phi(i)$ and $\mathbf{D} = \varphi(i)^{\mathrm{T}}\varphi(i)$ $+\epsilon$ and the fractional part of (17) is efficiently evaluated as

$$
(\mathbf{e(i)}^{\mathrm{T}}\varphi(i))\mathbf{w}^{1-\nu}(i) = (\mathbf{e(i)}^{\mathrm{T}}\varphi(i))\left(\sum_{j=1}^{i} \mathbf{a}_j(i)\varphi(j)\right)^{1-\nu}
$$

ν is the order of fractional derivative and the term $(\mathbf{e(i)}^{\mathrm{T}}\varphi(i))\mathbf{w}^{1-\nu}(i)$ in Eq. 17 is evaluated using Eq. 20.

Experimental results

This section presents experimental results to reveal the performance of the proposed algorithm. The performance of KFAPA is validated by the prediction of X-component of Lorenz time series and equalization of nonlinear channel.

Shoaib *et al. Appl Inform* (2015) 2:12

Page 6 of 14

Time series prediction

One of the useful and demanding problems in signal processing is predicting a future value in the nonlinear time series. The mechanism used for one step prediction is by taking time series history to estimate a future value. A nonlinear time series is defined as a sequence of scalers or vectors that depends on time. Let us consider a time series T. S. $= \{u(n_0), u(n_1), u(n_2), \ldots, u(n_{k-1}), u(n_k), u(n_{k+1}), u(n_{k+2}), \ldots\}$. To predict a future value at a certain time, we use a process known as time embedding. The output contains a tapped delay line and then form a matrix shifted by one time sample in each column and is written as

$$
\begin{pmatrix}
u(n_k) & u(n_{k+1}) & \ldots & u(n_{k+N-1}) \\
u(n_{k-1}) & u(n_k) & \ldots & u(n_{k+N-2}) \\
u(n_{k-2}) & u(n_{k-1}) & \ldots & u(n_{k+N-3}) \\
\ldots & \ldots & \ldots \ldots \\
u(n_{k-M-1}) & u(n_{k-M}) & \ldots & u(n_{k+N-M})
\end{pmatrix}
\tag{21}
$$

Columns of the above-mentioned matrix represents the input used for training or testing. M is the order of the filter. The first input pattern is delivered to predictor for estimating the future value. Then the weight vector is updated by a law based on the function of mean square error as given below.
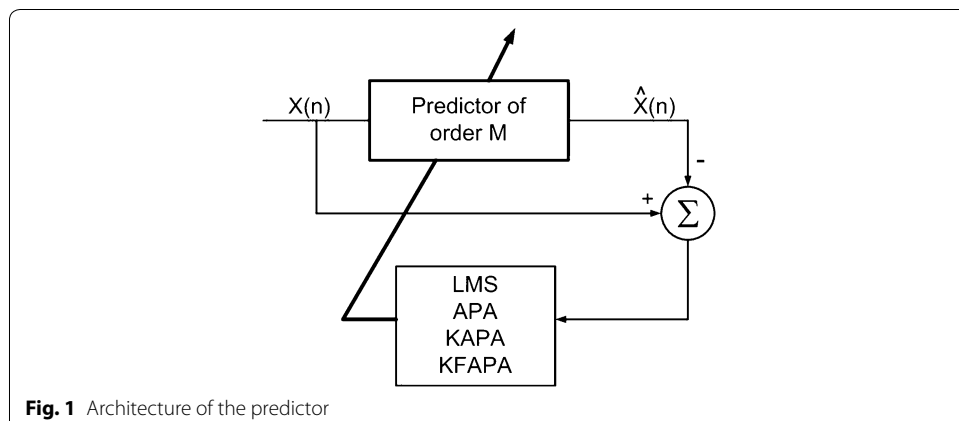
$$
e = |u(n_{k+1} - \widehat{u}(n_{k+1}))|^2
\tag{22}
$$

The cost function in Eq. 22 is used to adapt the parameters and Fig. 1 shows the architecure of the predictor.

Lorenz time series

Lorenz series exhibits chaotic flow. Series is three dimensional, nonlinear, and deterministic, expressed as the following nonlinear partial differential equations.

$$
\begin{aligned}
\frac{dx}{dt} &= \sigma(y(t) - x(t)) \\
\frac{dy}{dt} &= -x(t)z(t) + \gamma x(t) - y(t) \\
\frac{dz}{dt} &= x(t)y(t) - Bz(t)
\end{aligned}
\tag{23}
$$

Fig. 1 Architecture of the predictor

Shoaib *et al. Appl Inform* (2015) 2:12

Page 7 of 14

Parameters of the Lorenz system on which its behavior becomes chaotic are $\sigma = 10$, $\gamma = 28$ and $B = \frac{8}{3}$. Set the initial values as $x(0) = 1$, $y(0) = 1$, $z(0) = 1$, a sampling period is taken 0.01 second, also it is fixed to obtain the sample data using first-order approximation method. The state trajectory of the Lorenz system is shown in Fig. 2.The next experiment is performed with training sample points 500–1000 of the X-component of the Lorenz series, and test sample points 1000–1200 to evaluate the performance of the proposed algorithm. The time embedding length or the order of the filter M is 5 for this experiment. To validate the performance of the proposed algorithm, learning curves in terms of mean square error (MSE) as a figure of merit are plotted in Fig. 3.

The learning curves clearly demonstrate the performance of the proposed algorithm is better in terms of mean square error in comparison with is counterparts. The X-component of the Lorenz time series is corrupted with white noise having different variances, the algorithms including the proposed one is tested and the results are displayed in the tabular form. Mean square error is observed after 200 Monte Carlo simulations and is listed in Table 1. Throughout this experiment, Gaussian kernel is used and the kernel
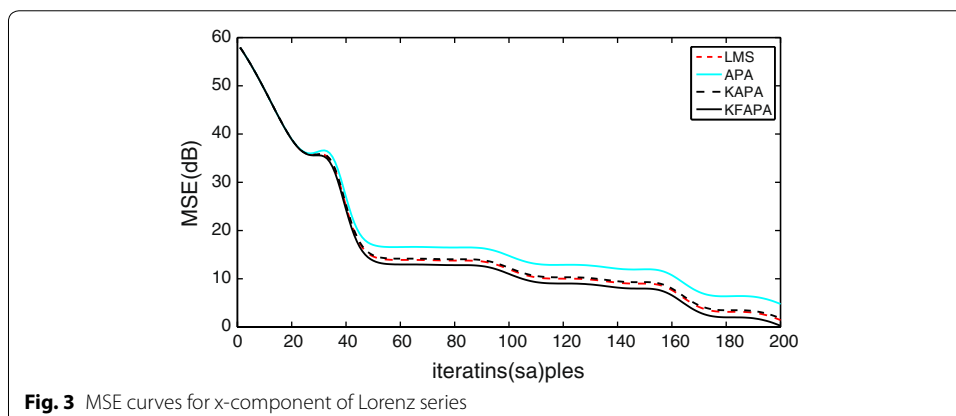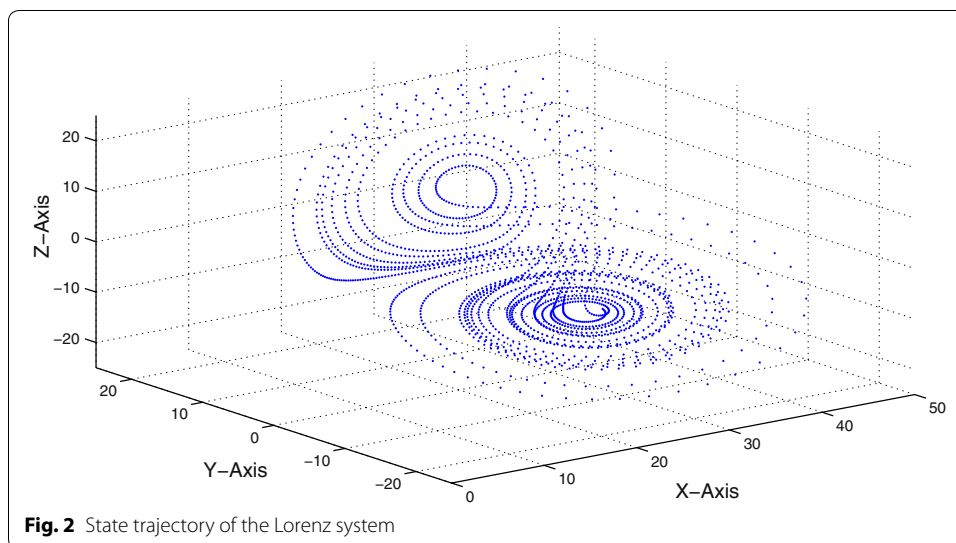

Fig. 2 State trajectory of the Lorenz system

Fig. 3 MSE curves for x-component of Lorenz series

Shoaib *et al. Appl Inform* (2015) 2:12

Page 8 of 14

Table 1 Performance comparison of LMS, APA, KAPA and KFAPA for X-component of Lorenz series prediction with different noise levels

Algorithm	LMS	APA	KAPA	KFAPA
Training MSE ($\sigma = 0.05$)	$0.02250 \pm 1.75e{-}005$	$0.01454 \pm 5.26e{-}004$	$0.041827 \pm 1.93e{-}065$	$0.02156 \pm 1.39e{-}005$
Training MSE ($\sigma = 0.05$)	$0.01583 \pm 0.22e{-}005$	0.07820 ± 0.00306	$0.017738 \pm 2.28e{-}003$	$0.091538 \pm 1.367e{-}004$
Training MSE ($\sigma = 0.02$)	0.01903 ± 0.003149	0.025175 ± 0.000198	0.001399 ± 0.000189	$0.0020052 \pm 6.50e{-}005$
Training MSE ($\sigma = 0.02$)	0.002970 ± 0.00170	0.005556 ± 0.0004324	0.001356 ± 0.000131	$0.0027892 \pm 9.49e{-}004$
Training MSE ($\sigma = 0.04$)	0.004349 ± 0.0003927	0.00680 ± 0.0007169	0.004117 ± 0.000251	0.0048219 ± 0.0001680
Training MSE ($\sigma = 0.04$)	0.0049979 ± 0.0004128	0.007162 ± 0.0009767	0.0051117 ± 0.000382	0.005822 ± 0.0003391
Training MSE ($\sigma = 0.1$)	0.015863 ± 0.0009678	0.010932 ± 0.0025963	0.026628 ± 0.00094904	0.045301 ± 0.00086739
Training MSE ($\sigma = 0.1$)	0.016166 ± 0.007296	0.019066 ± 0.0044241	0.035729 ± 0.00128194	0.0555606 ± 0.0023143
Training MSE ($\sigma = 0.5$)	0.42356 ± 0.10011	0.50001 ± 0.010242	0.82530 ± 0.2115	0.4209 ± 0.030332
Training MSE ($\sigma = 0.5$)	0.51752 ± 0.22074	0.69218 ± 0.028178	0.91918 ± 0.32231	0.52013 ± 0.047689

Shoaib *et al. Appl Inform* (2015) 2:12

Page 9 of 14

width is set as 0.1. The results show that in low noise levels, the proposed algorithm achieve better results in comparison to high noise levels.

Nonlinear channel equalization

Nonlinear channel model considered here in this experiment as a test bench, consists of serial connection of linear filter and a memoryless nonlinearity. This type is commonly used to model digital communication channels and digital magnetic recording channels. A binary signal $[b(1), b(2), b(3),...,b(k)]$ is fed into a nonlinear channel, while adding static nonlinearity and additive white Gaussian noise the signal will be observed as $[r(1), r(2), r(3),...,r(k)]$. The channel model is defined as $h(i) = b(i) + 0.5b(i-1)$ and output is $r(i) = h(i) - 0.9h(i)^2 + n(i)$, where $n(i)$ is the additive white Gaussian noise having variance of 0.01. We aim here in this experiment to reproduce the original signal with low error rate. The time embedding length or the order of the filter is 5. 5000 symbols are used to train the coefficients of the nonlinear channel and the mean square error during training is displayed in Fig. 4. Figure 5 shows that during training, the MSE curve of the proposed algorithm is slightly better than its counterparts and the results are also displayed in tabular form in Table 2. The performance of the proposed algorithm is also tested in Fig. 6 by inserting an abrupt change at iteration 500. It can be easily
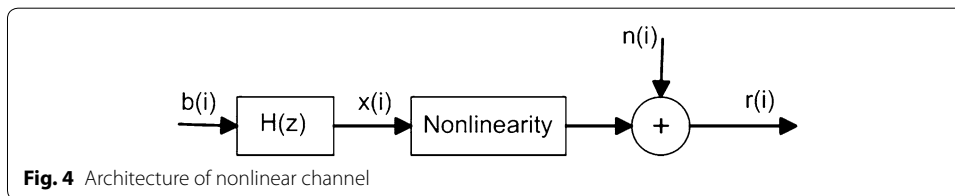

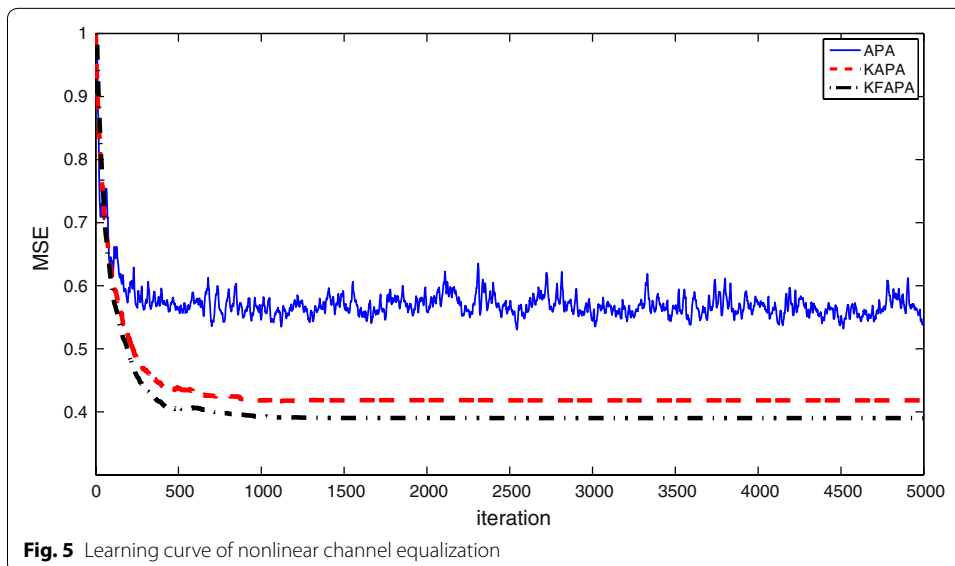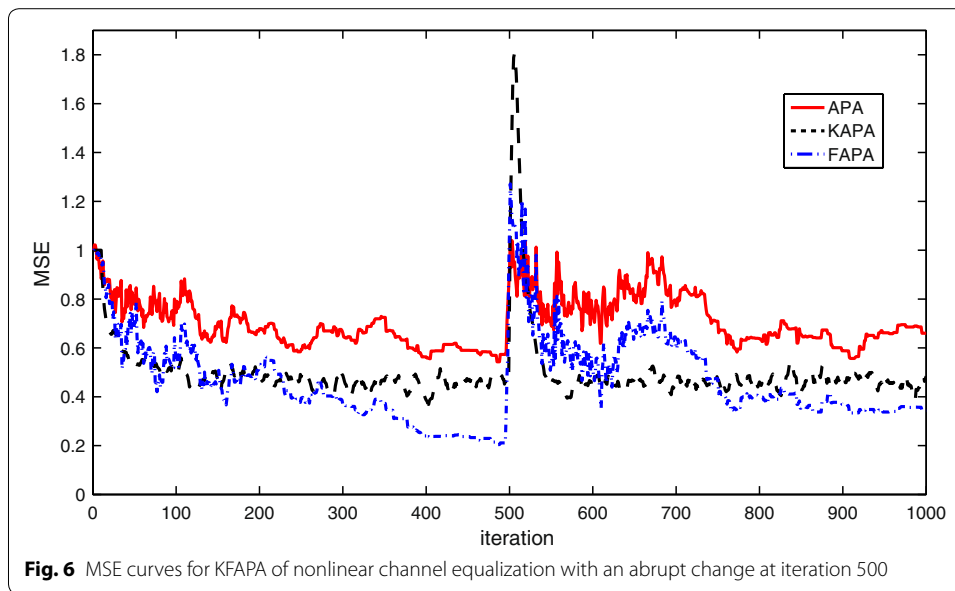**Fig. 4** Architecture of nonlinear channel

Fig. 5 Learning curve of nonlinear channel equalization

Shoaib *et al. Appl Inform* (2015) 2:12

Page 10 of 14

Fig. 6 MSE curves for KFAPA of nonlinear channel equalization with an abrupt change at iteration 500

observed that the proposed algorithm is able to recover efficiently in comparison with its counterparts and the improvement of 0.1 dB is achieved.

Atmospheric CO_2 concentration forecasting

The data consist of monthly average CO_2 concentrations (in parts per million by volume ppmv) in atmosphere collected at Mauna Loa observatory Hawaii, between 1958 and 2008, with 600 total observations. the first 400 points are used for training while the other 200 for testing. The kernel function for this specific problem handles long-term rising, seasonal effect, periodicity and some irregularities. The kernel function is

$$k(\mathbf{x}, \mathbf{x}^{'}) = k_1(\mathbf{x}, \mathbf{x}^{'}) + k_2(\mathbf{x}, \mathbf{x}^{'}) + k_3(\mathbf{x}, \mathbf{x}^{'}) + k_1(\mathbf{x}, \mathbf{x}^{'}) \tag{24}$$

$k_1(\mathbf{x}, \mathbf{x}^{'})$ is used to model the rising trends and is defined as

$$k_1(\mathbf{x}, \mathbf{x}^{'}) = a_1^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}^{'})^2}{2a_2^2}\right) \tag{25}$$

a_1 is the amplitude and a_2 is the kernel width. A seasonal effect is modeled through periodic kernel with a time period of 1 year. A Gaussian kernel is used to put decay away from the exact periodicity.

Table 2 Performance comparison of APA, KAPA and KFAPA in nonlinear channel equalization

Algorithm	MSE (dB)
APA	0.6 ± 0.2
KAPA	0.55 ± 0.05
KFAPA	0.4 ± 0.1

Shoaib *et al. Appl Inform* (2015) 2:12

Page 11 of 14

$$k_2(\mathbf{x}, \mathbf{x}') = a_3^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{2a_2^4} - \frac{\sin^2 \pi((\mathbf{x} - \mathbf{x}'))}{a_5^2}\right) \tag{26}$$

a_3, a_4 and a_5 are magnitude, smoothing factor and the periodic component. To handle the irregularities in the observed dataset $k_3(\mathbf{x}, \mathbf{x}')$ is defined as

$$k_3(\mathbf{x}, \mathbf{x}') = a_6^2 \left(1 + \frac{(\mathbf{x} - \mathbf{x}')^2}{2a_8^2 a_7^2}\right)^{-a_8^2} \tag{27}$$

a_6 is magnitude, a_7 and a_8 are the smoothing factor and shape parameter, respectively. To model the noise component $k(\mathbf{x}, \mathbf{x}')$ is defined as
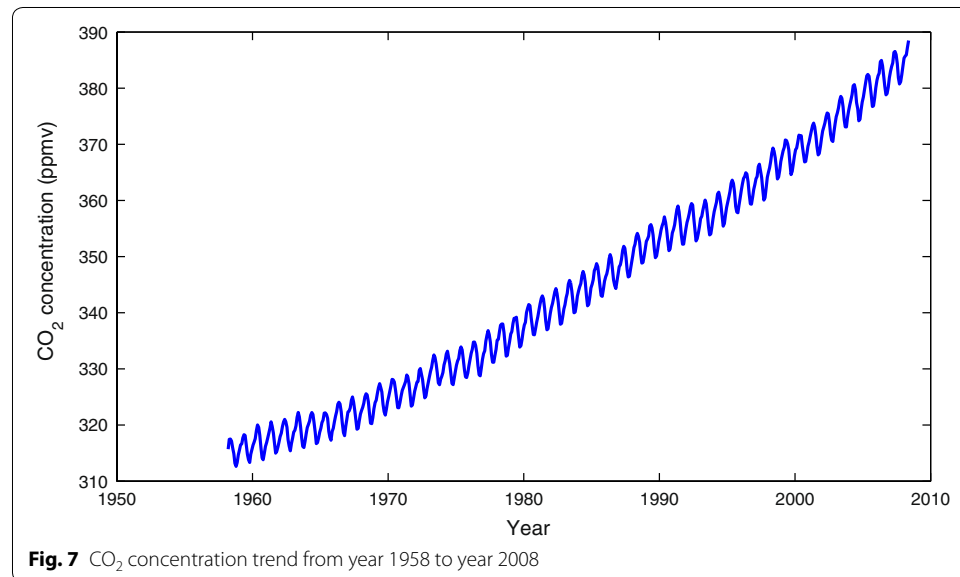
$$k_4(\mathbf{x}, \mathbf{x}') = a_9^2 \left(\frac{(\mathbf{x} - \mathbf{x}')^2}{2a_{10}^2} + a_{11}^2 \delta(\mathbf{x} - \mathbf{x}')\right) \tag{28}$$
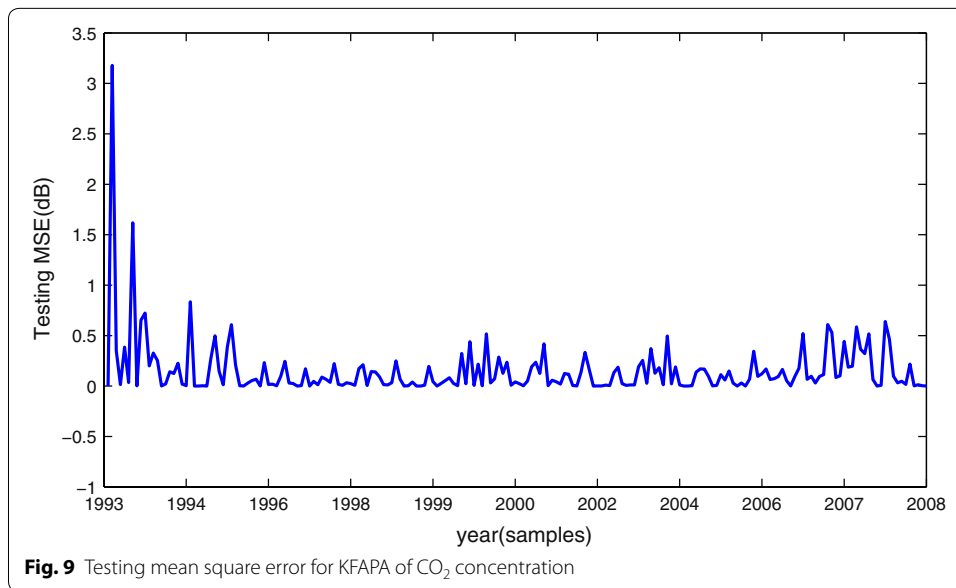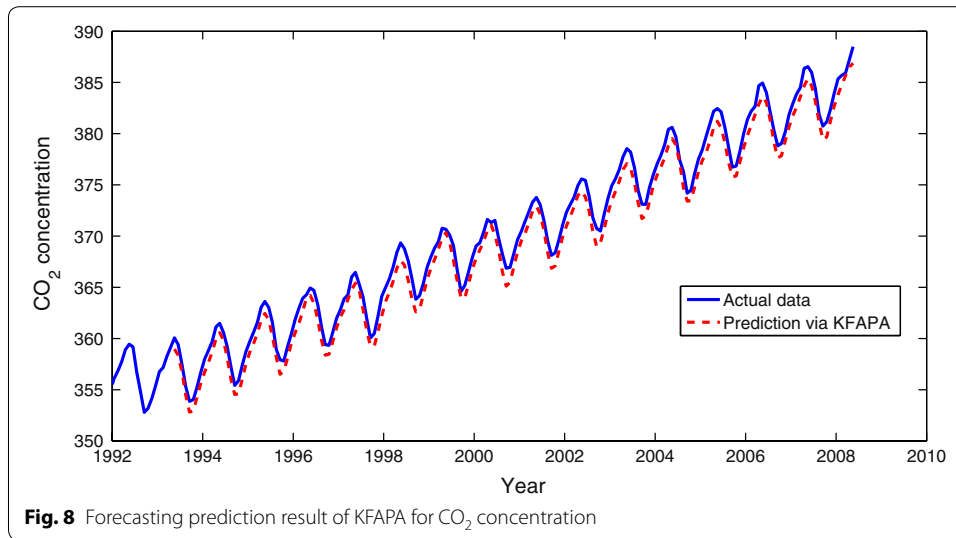
a_9 and a_{10} are magnitude and smoothing factor of the colored noise. a_{11} is the magnitude of white noise. The values of these kernel function parameters are listed in Table 3.

The CO_2 concentration in atmosphere is modeled as a function of time and is shown in Fig. 7. Prediction performance is shown in Fig. 8, and the performance in mean square error is plotted in Fig. 9. The figure shows the MSE curve.

Table 3 Kernel function parameter values

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
66	0.075	0.40	0.0576	1.0878	0.6600	0.4167	0.78	0.18	3.7509	0.1900

Fig. 7 CO_2 concentration trend from year 1958 to year 2008

Shoaib *et al. Appl Inform* (2015) 2:12

Page 12 of 14

Fig. 8 Forecasting prediction result of KFAPA for CO_2 concentration

Fig. 9 Testing mean square error for KFAPA of CO_2 concentration

Static function approximation

This example is about static function approximation, where the desired output data are generated by

$$d(i) = cos(\omega x(i) - \tau) + v(i) \qquad (29)$$

where $\tau \geq 0$, input $\mathbf{x}(i)$ is uniformly distributed over $[\tau, \tau + 2]$, and $[v(i)]$ is a zero mean Gaussian noise with variance σ_v^2. In this experiment, $N = 2000$ samples are generated with $\sigma_v^2 = 0.01$, $\omega = 2$ and $\tau = 1.0$. 500 samples are used for training and another 200 are used for testing. The test pattern is shown in Fig. 10. Figure 11 illustrates the convergence curves for APA, KAPA and KFAPA. MSE denotes the mean square error. Simulation results clearly indicate that the performance of the proposed algorithm has been perfectly good as listed in Table 4.
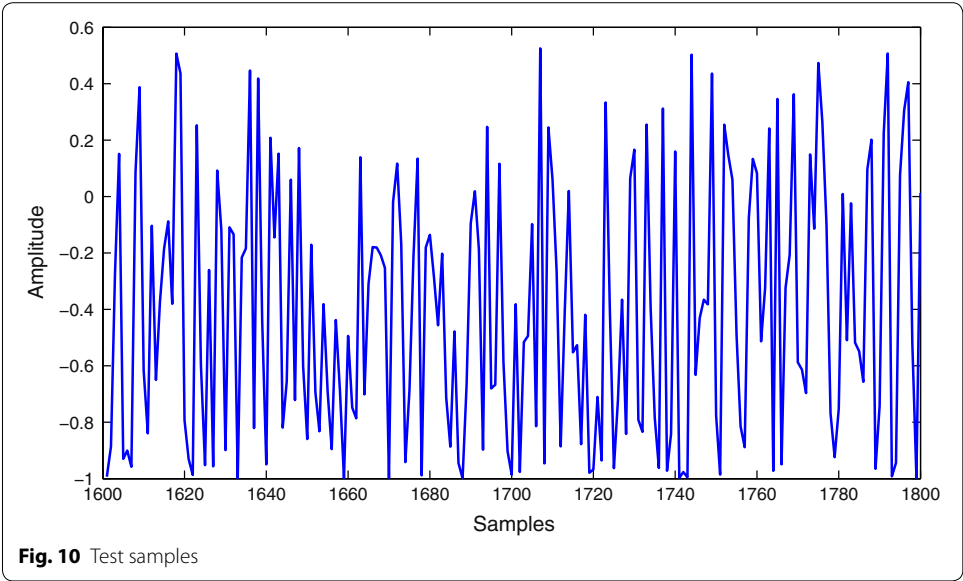
Shoaib *et al. Appl Inform* (2015) 2:12
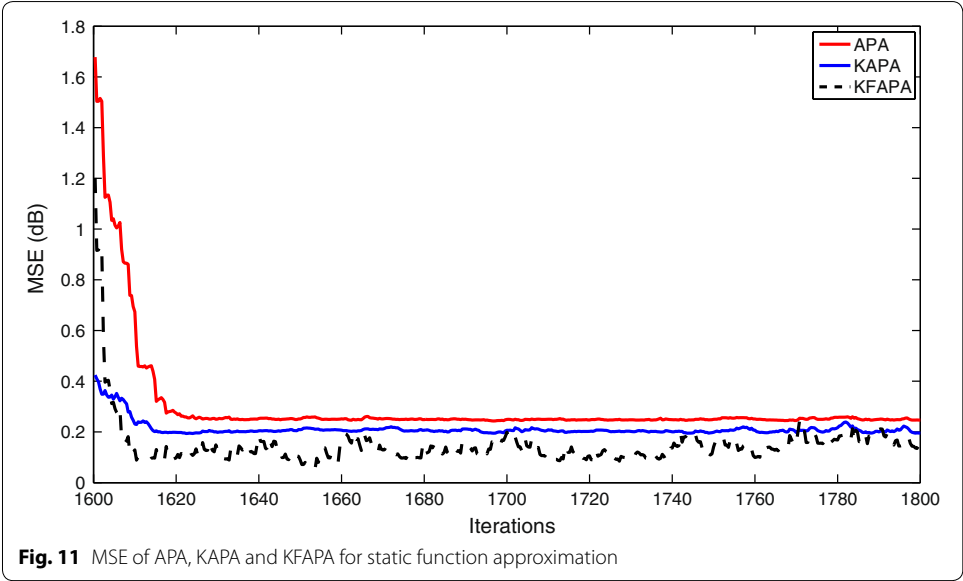
Page 13 of 14

Fig. 10 Test samples

Fig. 11 MSE of APA, KAPA and KFAPA for static function approximation

Table 4 Training and testing MSE

Algorithm	Training MSE	Testing MSE
APA	0.35 ± 0.05	0.3 ± 0.02
KAPA	0.2 ± 0.04	0.22 ± 0.05
KFAPA	0.11 ± 0.1	0.13 ± 0.3

Conclusion and future work

In this paper, a new kernel fractional affine projection algorithm is presented. Affine projection and kernel affine projection algorithms has also been discussed. One application of predicting a chaotic three-dimensional Lorenz system is presented that demonstrates

Shoaib *et al. Appl Inform* (2015) 2:12

Page 14 of 14

the performance of the proposed algorithm in comparison with LMS, APA, KAPA and in terms of mean square error as a figure of merit. Proposed algorithm is also tested on nonlinear channel equalization. This new formulation is another contribution in the field of nonlinear signal processing.

Authors' contributions
BS proposed and implemented the idea. Dr IMQ and Dr IU are supervisor and cosupervisor respectively. SAB and SUK did the drafting and paper writng process. All authors read and approved the final manuscript.

Author details
[1] Department of Electronic Engineering, International Islamic University, H-10, Islamabad, Pakistan. [2] Electrical Engineering Department, Air University, Sector: E-9, Islamabad, Pakistan. [3] Department of Electronic Engineering, School of Engineering and Applied Sciences, ISRA University, Sector I-10, Islamabad, Pakistan.

Competing interests
The authors declare that they have no competing interests.

References
Akhtar P, Yasin M (2012) Performance analysis of bessel beamformer and LMS algorithm for smart antenna array in mobile communication system. In: Emerging Trends and Applications in Information Communication Technologies, vol 281. Springer Berlin Heidelberg, pp 52–61
Chaudhary NI, Raja MAZ, Khan JA, Aslam MS (2013) Identification of input nonlinear control autoregressive systems using fractional signal processing approach. Sci World J 2013:1–13 **(ID 467276)**
Dubey SK, Rout NK (2012) FLMS algorithm for acoustic echo cancellation and its comparison with LMS. In: Proceedings of the 1st international conference on IEEE recent advances in information technology (RAIT)
Engel Y, Mannor S, Meir R (2004) The kernel recursive least-squares algorithm. IEEE Trans Signal Process 52(8):2275–2285
Hardle W, Vieu P (1992) Kernel regression smoothing of time series. J Time Ser Anal 13(3):209–232
Haykin S (2013) Adaptive filter theory, 5 edn. Pearson Education, Limited, India **(revised)**
Hearst MA et al (1998) Support vector machines. Intell Syst Appl IEEE 13(4):18–28
Liu W, Pokharel PP, Principe JC (2008) The kernel least-mean-square algorithm. IEEE Trans Signal Process 56(2):543–554
Liu W et al (2009) Extended kernel recursive least squares algorithm. IEEE Trans Signal Process 57(10):3801–3814
Liu W, Principe JC, Haykin S (2011) Kernel adaptive filtering: a comprehensive introduction, vol 57. John Wiley & Sons
Liu W, Principe JC (2008) Kernel affine projection algorithms. EURASIP J Adv Signal Process 1(2008):784292
Liu W, Principe JC, Haykin S (2010) Kernel recursive least-squares algorithm. In: Kernel Adaptive Filtering: A Comprehensive Introduction, pp 94–123
Masoud G, Osgouei SG (2011) Dual-channel speech enhancement using normalized fractional least-mean-squares algorithm. In: Proceedings of the 19th Iranian conference on electrical engineering (ICEE)
Ortigueira, MD, Machado JT, de Almeida R (2002) Special issue on fractional signal processing and applications. Signal Proc 82:1515
Ortigueira MD, Machado JAT (2006) Fractional calculus applications in signals and systems. Signal Proc 86(10):2503–2504
Ortigueira MD (2011) Fractional calculus for scientists and engineers, vol 84. Springer Science and Business Media
Raja MAZ, Chaudhary NI (2015) Two-stage fractional least mean square identification algorithm for parameter estimation of CARMA systems. Signal Process 107:327–339
Scholkopf B, Smola A, Muller KR (1997) Kernel principal component analysis. Artificial Neural Networks ICANN 97. Springer, Berlin Heidelberg, pp 583–588
Shoaib B, Qureshi IM (2014) A modified fractional least mean square algorithm for chaotic and nonstationary time series prediction. Chin Phys B 23(3):030502
Shoaib B, Qureshi IM (2014) Adaptive step-size modified fractional least mean square algorithm for chaotic time series prediction. Chin Phys B 23(5):050503
Takeda H, Farsiu S, Milanfar P (2007) Kernel regression for image processing and reconstruction. IEEE Trans Image Process 16(2):349–366
Tseng CC, Lee SL (2012) Design of linear phase FIR filters using fractional derivative constraints. Signal Process 92(5):1317–1327
Tseng CC, Lee SL (2013) Designs of two dimensional linear phase FIR filters using fractional derivative constraints. Signal Proc 93(5):1141–1151
Tseng CC, Lee SL (2014) Designs of fractional derivative constrained 1D and 2D FIR filters in the complex domain. Signal Proc 95:111–125
Vapnik VN, Vapnik V (1998) Statistical learning theory, vol 1. Wiley, New York
Wang J et al (2014) Fractional zero phase filtering based on the Riemann Liouville integral. Signal Proc. 98:150–157
Zahoor RMA, Qureshi IM (2009) A modified least mean square algorithm using fractional derivative and its application to system identification. Eur J Sci Res 35(1):14–21