Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 86 (2006) 760–775

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Kalman filtering for self-similar processes

Birsen Yazıcı[a,*], Meltem Izzetoğlu[b], Banu Onaral[b], Nihat Bilgutay[c]

[a]*Electrical, Computer and Systems Engineering Department, Rensselaer Polytechnique Institute, Jonsson Engineering Center, Rm: 7008, 110 8th Street, Troy, NY 12180, USA*
[b]*School of Biomedical Engineering, Science and Health System, Drexel University*
[c]*Electrical and Computer Engineering Department, Drexel University*

Received 20 June 2003; received in revised form 10 June 2005; accepted 14 June 2005
Available online 18 August 2005

Abstract

In this paper, we develop a state space representation and Kalman filtering method for self-similar processes. Key components of our development are the concept of multivariate self-similarity and the mathematical framework of scale stationarity. We define multivariate self-similarity as joint self-similarity, in which the self-similarity is governed by a matrix valued parameter $\mathbf{H}$. Such a generalization suits the nature of Multi-Input Multi-Output (MIMO) systems, since each channel is likely to be governed by a different self-similarity parameter. The system and measurement models for the proposed Kalman filter are defined as $t\dot{\mathbf{x}}(t) = t^{\mathbf{H}}\mathbf{A}t^{-\mathbf{H}}\mathbf{x}(t) + t^{\mathbf{H}}\mathbf{B}\mathbf{u}(t)$ and $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{v}(t)$, respectively. Here, the derivative operator $t\dot{\mathbf{x}}(t)$ indicates that the memory of the process is stored in time scales, unlike the memory stored in time shifts for stationary processes. We exploit this fact in developing an insightful interpretation of the Riccati equation and the Kalman gain matrix, which lead to an efficient numerical implementation of the proposed Kalman filter via exponential sampling. Additionally, we include a discussion of network traffic modeling and communications applications of the proposed Kalman filter. This study demonstrates that the scale stationarity framework leads to mathematically tractable and physically intuitive formulation of Kalman filtering for self-similar processes.

Keywords: Self-similar processes; Kalman filtering; Exponential sampling

1. Introduction

$1/f$ or self-similar processes occur in a wide range of engineering and science applications such as network traffic, economics, noise in electronic devices, natural terrain formations, just to mention a few [1–5]. Typically, these processes are characterized with their empirical Fourier spectrum as being proportional to $1/f^{\gamma}$ for $\gamma > 0$. Such a power spectra is a manifestation of long term correlations and statistical self-similarity in these processes.

Estimation and prediction of $1/f$ processes is an important and necessary task in many engineering applications. For example, in self-similar network

*Correspondence author. Tel.: +1 518 276 2905; fax: +1 518 276 6261.
E-mail address: yazici@ecse.rpi.edu (B. Yazıcı).

0165-1684/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2005.06.009

traffic applications, the estimation of the buffer size or queue length is an important problem, which requires prediction of traffic flow rate in real time. Another communication application involves estimation of transmitted signals embedded in $1/f$ type environmental noise in wireless channels. Kalman filtering framework provides optimal recursive estimation and prediction methods for such problems involving multiple inputs and multiple outputs (MIMO).

In this paper, we propose a concept of "*multivariate self-similarity*" and develop a state space representation and Kalman filtering method for self-similar MIMO processes. To the best of our knowledge, there is only one study in the literature for the formulation and analysis of multivariate self-similarity [6]. In that work, self-similarity is modeled by using the long range dependent ordinary stationary processes and the correlation structure and power spectrum are formulated in Fourier domain. However, this does not provide a state space representation which can further be used in Kalman filtering formulation. Key components of our developments are the concept of multivariate self-similarity that depends on the mathematical framework of scale stationarity [1]. In multivariate self-similarity, we require each univariate random variable to be individually and jointly self-similar, but not necessarily with the same self-similarity parameter. Thus unlike in univariate case, multivariate self-similarity is governed by a matrix valued parameter $\mathbf{H}$. Such a generalization suits the very nature of multi-input-multi-output systems, since each channel is likely to be governed by a different self-similarity parameter in engineering applications. Note that multivariate self-similarity reduces to ordinary self-similarity when self-similarity matrix $\mathbf{H}$ is a multiple of the identity matrix.

There are a number of approaches for mathematical modeling and analysis of self-similar processes. These include fractional Brownian motion [2], wavelet transform based models [3], physics based models [4] and mathematical framework of scale stationarity [1]. In this work, we utilize the latter approach due to its powerful and intuitive mathematical framework, which leads to engineering oriented signal processing tools. This framework builds upon the concept of *scale stationarity* described as $E[x(t)x(\lambda t)] = R(\lambda)$, $t, \lambda > 0$. Similar to ordinary stationarity, this concept leads to spectral decomposition and Auto Regressive Moving Average (ARMA) type models for self-similar processes [1]. For recent applications of this approach, see [7].

In this paper, we utilize scale stationary ARMA models to develop state space representations for multivariate self-similar processes. The system and measurement models of the proposed representation are given by

$$t\dot{\mathbf{x}}(t) = t^{\mathbf{H}}(\mathbf{A}+\mathbf{H})t^{-\mathbf{H}}\mathbf{x}(t) + t^{\mathbf{H}}\mathbf{B}\mathbf{u}(t),$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{v}(t), \tag{1}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{H}$ are matrices for N states and M outputs. $\mathbf{u}(t)$ and $\mathbf{w}(t)$ are multivariate "scale stationary white noise" processes, uncorrelated with the current states and measurements, respectively. $\mathbf{H}$ is a diagonal self-similarity matrix with entries $H_1, H_2, \ldots, H_N$. The matrix $t^{\mathbf{H}} = e^{\mathbf{H}\ln t}$ is also a diagonal matrix with entries $e^{H_1 \ln t}, e^{H_2 \ln t}, \ldots, e^{H_N \ln t}$. Note that here, the derivative operator $t\dot{\mathbf{x}}(t) = \lim_{\Delta\to 1}(\mathbf{x}(t\Delta) - \mathbf{x}(t))/\ln\Delta$ can be interpreted as a rate of change in the process with respect to infinitesimal changes in time scales. This implies that the memory of the system evolves in scales, rather than in time shifts as indicated by the ordinary derivative operator.

Naturally, such a representation lends itself to Kalman filtering for recursive estimation and prediction. While, the proposed state space representation can be viewed as a special case of Kalman filter formulation with time varying parameters, our approach exploits the underlying self-similar structure to develop a constant vector–matrix interpretation of the state space representation. The key underlying idea that leads to such an interpretation is the fact that $t\dot{\mathbf{x}}(t)$ is a shift varying, but scale invariant operator. We exploit this fact in developing an insightful interpretation of the Riccati equation and the Kalman gain matrix, which leads to an efficient numerical implementation of the Kalman filter.

To the best of our knowledge, the work of Chou and Willsky is the only study that is directly related to the Kalman filtering of self-similar

processes [8,9]. This work utilizes wavelet based multiresolution techniques to derive ARMA type models on dyadic trees. Kalman filtering from fine to coarse scales corresponds to prediction and alternately from coarse to fine scales corresponds to smoothing operations. Kalman filtering for the prediction operation consists of the recursive application of three steps: a measurement update step, a fine to coarse prediction step and a fusion step, which has no counterpart in the time recursive Kalman filtering. In our proposed Kalman filter algorithm, we show that there is no need for extra steps as in the multiresolutional Kalman filter since evolution of the states occur on the time axis.

The content of the paper is summarized next. A general background of scale stationary processes is presented in Section 2. In Section 3, we introduce multivariate self-similarity and formulate a state space representation for self-similar processes based on first order self-similar AR processes. In Section 4, we introduce the Kalman prediction and backward smoothing algorithms and provide an optimal implementation of the proposed Kalman filter via exponential sampling. In Section 5, we provide two methods for the implementation of the proposed Kalman filter and demonstrate its performance in simulation examples. In Section 6, we discuss application of our work in communication systems. Finally, in Section 7, we conclude our discussion and briefly highlight areas of future research.

2. Background on scale stationary processes

In this section, we shall briefly review the properties of self-similar processes. For a more complete discussion of the topic, we refer the reader to [1].

We call a stochastic process $x(t)$, $-\infty<t<\infty$ as statistically self-similar with parameter H [1], if it satisfies

$$x(t)\equiv a^{-H}x(at),\quad -\infty<t<\infty\quad \text{for any } a>0, \tag{2}$$

where $\equiv$ denotes equality in the sense of probability distributions [2].

Thus, we describe the self-similarity in the second order sense by the following conditions [1]:

$$\text{(i)}\quad E[x(t)]=\lambda^{-H}E[x(\lambda t)]\ \textit{for all}\ t>0,$$

$$\text{(ii)}\quad E[x(t)^2]<\infty\ \textit{for each}\ t>0,$$

$$\text{(iii)}\quad E[x(t_1)x(t_2)]=\lambda^{-2H}E[x(\lambda t_1)x(\lambda t_2)]\quad \textit{for all}\ t_1,t_2,\lambda>0. \tag{3}$$

For $H=0$, we refer to self-similar processes as wide sense scale stationary processes.

Before pursuing further, we want to point out that there is an isometry relationship between shift stationary and scale stationary processes. Given any shift stationary process $y(s)$ for $-\infty<s<\infty$, the process, $\tilde{x}(t)$ for $t>0$, obtained via the following exponential distortion:

$$\tilde{x}(t)=y(\ln(t)),\quad t>0 \tag{4}$$

is scale stationary. In fact, the isometry relationship stated above can be extended between self-similar and shift stationary processes because, self-similar processes are trended scale stationary processes, i.e., for any self-similar process, $x(t)$, with parameter $H\neq 0$, there is a scale stationary process, $\tilde{x}(t)$, such that

$$x(t)=t^H\tilde{x}(t),\quad t>0. \tag{5}$$

Thus, we appropriately referred $\tilde{x}(t)$ as the generating scale stationary process, and t^H as the trend term of a self-similar process, $x(t)$ [1].

Almost all the analytic properties of the self-similar processes can be derived from the theory of shift stationary processes via the isometry relationships given in (4) and (5). Here, we shall summarize some of the properties relevant to our subsequent developments.

It follows immediately from condition (iii) of wide sense self-similarity definition that for $H=0$

$$E[x(t)x(\lambda t)]=R(\lambda),\quad t,\lambda>0, \tag{6}$$

where $R(\lambda)$ is referred to as the autocorrelation function of a scale stationary process [1]. For

$H \neq 0$, (6) becomes

$$E[x(t)x(\lambda t)] = t^{2H}\lambda^{H}R(\lambda) = t^{2H}\Gamma(\lambda), \quad \forall t, \lambda > 0, \tag{7}$$

where $\Gamma(\lambda)$ is the autocorrelation of the underlying generating process. In this case, we refer $\Gamma(\lambda) = \lambda^{H}R(\lambda)$ as the basic autocorrelation function of a self-similar process [1].

It was shown in [1,7] that the generalized Mellin transform [10] decomposes self-similar processes into statistically independent components. As a result, under some regularity conditions, we define a spectrum for self-similar processes as follows

$$S(w) = \frac{1}{2\pi}\int_0^\infty \Gamma(\lambda)\lambda^{-H-jw}\,\mathrm{d}\,\ln\,\lambda, \tag{8}$$

where $\Gamma(\lambda)$ is the basic autocorrelation function, and we referred the function $S(w)$ as the scale spectral density function of a self-similar process [1].

An important class of self-similar processes is defined by the generalized Euler–Cauchy systems [1]. These processes are called self-similar autoregressive moving average (ARMA) processes. Symbolically, an Nth order self-similar ARMA process $y(t)$ with parameter H can be represented by the following time varying, ordinary differential equation:

$$\begin{aligned}\alpha_N t^N \frac{\mathrm{d}^N}{\mathrm{d}t_N}\,y(t) + \cdots + \alpha_1 t\,\frac{\mathrm{d}}{\mathrm{d}t}\,y(t) + \alpha_0 y(t)\\ = \beta_M t^{M+H}\,\frac{\mathrm{d}^M}{\mathrm{d}t_M}u(t) + \cdots + \beta_1 t^{1+H}\\ \times \frac{\mathrm{d}}{\mathrm{d}t}\,u(t) + \beta_0 t^H u(t),\end{aligned} \tag{9}$$

where $u(t)$ can be interpreted as a unit driving white noise process for a linear self-similar system. In [1], we have shown that under some regularity conditions, any self-similar process can be approximated by a finite order self-similar ARMA process in some sense. The basic autocorrelation function of an Nth order self-similar ARMA process is given by

$$\Gamma(\lambda) = \begin{cases} \sum_{j=0}^{n}\sum_{i=0}^{m_j} a_{ij}^2(\ln\,\lambda)^{-j}\lambda^{H+b_{ij}}, & 0 < \lambda \leqslant 1,\\ \sum_{j=0}^{n}\sum_{i=0}^{m_j} a_{ij}^2(\ln\,\lambda)^{j}\lambda^{H-b_{ij}}, & \lambda > 1. \end{cases} \tag{10}$$

The scale stationary white noise process $u(t)$ is defined in [1] as having autocorrelation function satisfying:

$$R_u(\lambda) = E[u(t)u(\lambda t)] = \sigma^2\tilde{\delta}(\lambda) \tag{11}$$

where $\tilde{\delta}(t)$ is referred to as the *unit driving force* and has properties

$$\text{(i)} \quad \tilde{\delta}(t) = 0, \quad t \neq 1, \quad t > 0, \tag{12}$$

$$\text{(ii)} \quad \int_0^\infty \tilde{\delta}(t/\lambda)\,\mathrm{d}\,\ln\,\lambda = 1, \ t > 0, \tag{13}$$

$$\text{(iii)} \quad x(t) = \int_0^\infty x(\lambda)\tilde{\delta}\left(\frac{t}{\lambda}\right)\mathrm{d}\,\ln\,\lambda. \tag{14}$$

In practical applications, often times the continuous time data is not available. Hence, we developed sampling methods for scale stationary processes by which the statistical structure of the continuous data can be recovered from its discrete samples in [11]. It is well known that for ordinary band limited stationary processes, the process can be recovered from its discrete samples recorded at equally spaced intervals. Since the characteristics of the signal remains the same with time scalings for statistically scale stationary signals, we developed a Shannon type optimal sampling procedure for such processes in terms of exponential sampling in [11,12]. It has also been shown in [13] that exponential sampling is the optimum sampling procedure for the discretization of the Mellin transform.

The exponential sampling scheme is summarized with the sampling theorem as [11].

Theorem 1. *Let $x(t)$ for $t > 0$ be a scale stationary process having autocorrelation function $R(\lambda)$ satisfying*

$$\int_0^\infty \mathrm{e}^{-\mathrm{j}w\,\ln\,\lambda}R(\lambda)\,\mathrm{d}\,\ln\,\lambda = 0 \quad \textit{for some } |w| > \Omega \tag{15}$$

then

$$R(\lambda) = \sum_{n=-\infty}^{\infty} R(T^n)\mathrm{sinc}(\Omega\ln(\lambda/T^n)), \quad \lambda > 0, \tag{16}$$

$$x(s_o\lambda) = x(s_o)\mathrm{sinc}(\Omega \ln(\lambda)) + \sum_{n=1}^{\infty} x(s_o T^n)\{\mathrm{sinc}(\Omega \ln(\lambda/T^n)) + \mathrm{sinc}(\Omega \ln(\lambda/T^{-n}))\}. \quad (17)$$

The condition of the theorem in (15) can be interpreted as band limitedness for scale stationary processes. The formula in (17) states that any band limited scale stationary process is completely determined by its sampled values, $x(s_o T^n)$, at exponentially spaced sampling intervals $s_o T^n$ for $n = 0, -+1, \ldots$.

3. Multivariate self-similarity and state space representation for self-similar processes

In this section, we will introduce a state space representation for the self-similar processes based on the scale stationary ARMA processes [1]. Before we introduce state space representation for self-similar processes, we want to introduce the concept of multivariate self-similarity.

Definition 1. We call the multivariate process variables $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \ldots\ x_N(t)]^{\mathrm{T}}$ strictly self-similar with self-similarity matrix $\mathbf{H}$ if it satisfies the following condition:

$$\mathbf{x}(t) \equiv \lambda^{-\mathbf{H}}\mathbf{x}(\lambda t), \quad -\infty < t < \infty \quad \text{for any } \lambda > 0, \quad (18)$$

where $\equiv$ denotes equality in finite dimensional probability distributions and $\mathbf{H}$ is a diagonal matrix as $\mathbf{H} = \mathrm{diag}(H_1\ H_2\ \ldots\ H_N)$.

A special case of self-similar state variables occur when the self-similarity matrix is null matrix $\mathbf{H} = \mathbf{0}$. For this case, each state becomes scale stationary, which means their statistics become absolutely independent of time scale. The definition of strictly scale stationary multivariate processes is given next.

Definition 2. We shall call a multivariate process $\tilde{\mathbf{x}}(t) = [\tilde{x}_1(t)\ \tilde{x}_2(t)\ \ldots\ \tilde{x}_N(t)]^{\mathrm{T}}$ strictly scale stationary if it satisfies the following condition:

$$\tilde{\mathbf{x}}(t) \equiv \tilde{\mathbf{x}}(\lambda t), \quad -\infty < t < \infty \quad \text{for any } \lambda > 0. \quad (19)$$

It is straightforward to show that the isometry defined for self-similar and scale stationary processes given in (5) is also valid for self-similar and scale stationary multivariate processes, i.e.,

$$\mathbf{x}(t) = t^{\mathbf{H}}\tilde{\mathbf{x}}(t). \quad (20)$$

We shall define second order multivariate self-similarity as follows:

Definition 3. A multivariate process $\mathbf{x}(t)$ is wide sense self-similar with self-similarity matrix $\mathbf{H}$ if it satisfies the following conditions:

(i) $E[\mathbf{x}(t)] = \lambda^{-\mathbf{H}}E[\mathbf{x}(\lambda t)]$ for all $t, \lambda > 0$ i.e. $\lambda^{\mathbf{H}}$ is a constant matrix equal to $\lambda^{\mathbf{H}} = \mathrm{diag}(\lambda^{H_1}\ \lambda^{H_2}\ \ldots\ \lambda^{H_N})$
(ii) $\mathrm{trace}(E[\mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(t)]) < \infty$ for each $t > 0$
(iii) $E[\mathbf{x}(t_1)\mathbf{x}^{\mathrm{T}}(t_2)] = \lambda^{-\mathbf{H}}E[\mathbf{x}(\lambda t_1)\mathbf{x}^{\mathrm{T}}(\lambda t_2)]\lambda^{-\mathbf{H}}$ for all $t_1, t_2, \lambda > 0$.

where $\mathbf{H} = \mathrm{diag}(H_1\ H_2\ \cdots\ H_N)$.

Note that here finite variance of the state variables is needed to assure that multivariate process is physically realizable. From the last condition of Definition 3, it is clear that the scale stationary state variables satisfy the following relation:

$$E[\tilde{\mathbf{x}}(\lambda t)\tilde{\mathbf{x}}^{\mathrm{T}}(t)] = \mathbf{R}(\lambda). \quad (21)$$

We shall call $\mathbf{R}(\lambda)$ the *scale correlation matrix*. Using the isometry between the wide sense scale stationary states and the wide sense self-similar states we can show:

$$E[\mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(\lambda t)] = t^{\mathbf{H}}E[\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(\lambda t)]\lambda^{\mathbf{H}}t^{\mathbf{H}} = t^{\mathbf{H}}\mathbf{R}(\lambda)\lambda^{\mathbf{H}}t^{\mathbf{H}}. \quad (22)$$

Then the correlation matrix for the wide sense self-similar states can be introduced as:

$$E[\mathbf{x}(1)\mathbf{x}^{\mathrm{T}}(\lambda)] = \mathbf{R}(\lambda)\lambda^{\mathbf{H}} = \mathbf{\Gamma}(\lambda), \quad \lambda > 0. \quad (23)$$

We will refer to $\mathbf{\Gamma}(\lambda)$ as the *basic correlation matrix* of the multivariate process.

Note that, state self-similarity implies that not only each variable is self-similar (univariate), but it is also jointly self-similar. It is also interesting to note that $E[x_i(\lambda t)x_j(\lambda t)] = \lambda^{(H_i+H_j)}E[x_i(t)x_j(t)]$ for $i, j = 1, \ldots, N$. Hence, when $i \neq j$, $H_i + H_j$ can be

viewed as the self-similarity parameter of the joint process. Similarly, one can define self-similarity for higher order cross-covariance terms.

It is straightforward to extend the definition of univariate scale power spectral density [1], to the case of multivariate processes and obtain a spectral characterization of the multivariate self-similar processes. Here, we shall only state the results. Under some conditions on the entries of $\mathbf{\Gamma}(\lambda)$, one can show that, multivariate self-similar processes yield the following spectral decomposition:

$$\mathbf{S}(w) = \int_0^\infty \mathbf{\Gamma}(\lambda)\lambda^{-\mathbf{H}}\lambda^{-jw}\,\mathrm{d}\,\ln\,\lambda. \tag{24}$$

Or in matrix entries:

$$S_{ij}(w) = \int_0^\infty \Gamma_{ij}(\lambda)\lambda^{(H_i-H_j)-jw}\,\mathrm{d}\,\ln\,\lambda. \tag{25}$$

Before we proceed into state space representation for multi-input multi-output processes, we illustrate multivariate self-similarity with two simple examples.

Example 1. Assume that $x_1(t)$ is self-similar with parameter H and $\varepsilon(t)$ is a white noise process with variance λ^H, i.e., $E[\varepsilon(t)\varepsilon(\lambda t)] = \lambda^H\tilde{\delta}(\lambda)$, uncorrelated with $x_1(t)$, i.e., $E[x_1(t)\varepsilon(\lambda t)] = 0$.

Define $x_2(t) = ax_1(t) + \varepsilon(t)$. Then, $E[x_2(t)\,x_2(\lambda t)] = a^2\lambda^H R_{x_1}(\lambda) + \lambda^H\tilde{\delta}(\lambda)$ for $\lambda \geqslant 1$. Hence, $x_2(t)$ is also self-similar with parameter H. It can also be easily shown that $E[x_1(t)x_2(\lambda t)] = a\lambda^H R_{x_1}(\lambda)$. Hence, $[x_1(t)\ x_2(t)]^{\mathrm{T}}$ is multivariate self-similar with parameter matrix $\mathbf{H} = \mathrm{diag}(H, H)$.

Example 2. Let $x_1(t)$ and $\varepsilon(t)$ be defined as in Example 1. Define $x_2(t) = ax_1(\lambda_o t) + \varepsilon(t)$ for $0 < \lambda_o < 1$. Then, $E[x_2(t)x_2(\lambda t)] = a^2\lambda^H R_{x_1}(\lambda) + \lambda^H\tilde{\delta}(\lambda)$. Hence $x_2(t)$ is also self-similar with parameter H. Furthermore, $x_1(t)$ and $x_2(t)$ are jointly self-similar with cross-correlation function as $E[x_1(t)x_2(\lambda t)] = a(\lambda\lambda_o)^H R_{x_1}(\lambda\lambda_o)$. Thus $[x_1(t)\ x_2(t)]^{\mathrm{T}}$ is multivariate self-similar with parameter matrix $\mathbf{H} = \mathrm{diag}(H, H)$.

In many systems or measurement models, the unknowns are related to measurements within a linear regression framework. For example, in the case of univariate measurements:

$$y(t) = c_1x_1(t) + c_2x_2(t) + \cdots + c_Nx_N(t) + w(t), \tag{26}$$

where $w(t)$ is measurement noise. Such a linear model is particularly suitable for prediction and filtering problems. Motivated by these practical considerations, we propose the following system and measurement models for multivariate self-similar processes:

$$t\,\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{x}(t) = t^{\mathbf{H}}(\mathbf{A} + \mathbf{H})t^{-\mathbf{H}}\mathbf{x}(t) + t^{\mathbf{H}}\mathbf{B}\mathbf{u}(t), \tag{27}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{w}(t), \tag{28}$$

where $\mathbf{x}(t) = [x_1(t)\ \ x_2(t)\ \ \cdots\ \ x_N(t)]^{\mathrm{T}}$ is the $N \times 1$ state vector, $\mathbf{u}(t)$ is the $R \times 1$ input vector, $\mathbf{y}(t)$ is the $M \times 1$ output vector, $\mathbf{A}$ is an $N \times N$ matrix, $\mathbf{B}$ is an $N \times R$ matrix, $\mathbf{C}$ is an $M \times N$ matrix, $\mathbf{D}$ is an $M \times R$ and $\mathbf{H}$ is an $N \times N$ diagonal matrix having values $H_1, H_2, \ldots, H_N$ in its diagonal entries.

Note that, $\mathbf{y}(t)$ the measurement processes in the state space representation in (28) have some type of "cumulative self-similarity" i.e., measurements are a linear combination of self-similar univariate processes and self-similar input processes $\mathbf{u}(t)$, which are not necessarily characterized by the same self-similarity parameters. However, if the self-similarity parameter of one of the states is dominant over the others, the output can be approximated as a $1/f$ process with the self-similarity parameter of the dominant state.

Remarks.

- An integral representation of the self-similar states and output processes are given by

$$\mathbf{x}(t) = t^{\mathbf{H}}\int_1^t \left(\frac{t}{\tau}\right)^{\mathbf{A}}\mathbf{B}\mathbf{u}(\tau)\,\mathrm{d}\,\ln\,\tau,$$

$$\mathbf{y}(t) = \mathbf{C}t^{\mathbf{H}}\int_1^t \left(\frac{t}{\tau}\right)^{\mathbf{A}}\mathbf{B}\mathbf{u}(\tau)\,\mathrm{d}\,\ln\,\tau + D\mathbf{w}(t), \tag{29}$$

  where the integrals are defined in the second order sense. Note that, when $\mathbf{u}(\tau)$ is a scale stationary "white" noise process, some care needs to be taken in the integral representation above. Here, we shall omit this technicality.

However, rigorous treatment can be found in [1].

- The dynamical model for the kth state variable in (28) is given by

$$t\frac{\mathrm{d}}{\mathrm{d}t}x_k(t) = (a_{k,k} + H_k)x_k(t) + t^{H_k}\sum_{\substack{l=1\\ l\neq k}}^{N} \times a_{k,l}t^{-H_l}x_l(t) + t^{H_k}\mathbf{Bu}(t). \quad (30)$$

As can be seen from this equation, the dynamics of a state is defined by the state itself, other states and the input depending on the **A** and **B** matrices. The dependency on the state itself can be seen as an intrinsic self-similarity, whereas the coupling to the other states may be treated as an additional input to the state under consideration. As such, the coupled states $x_l(t)$ with parameters H_l do not have any affect on the self-similarity of the state $x_k(t)$ which has a self-similarity parameter H_k.

- The derivative operator $t(\mathrm{d}/\mathrm{d}t)x(t)$ can be interpreted as follows:

$$t\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \lim_{\Delta\to 1}\frac{x(t\Delta) - x(t)}{\ln\Delta}. \quad (31)$$

This interpretation reveals that the memory of the system is stored in time scales rather than time shifts. Hence, the dynamical equations in the state space representation express the evolution of the states in infinitesimal time scalings. This interpretation highlights the difference between our state space representation and the multiresolutional state space representation in [9], where the states of the system is defined on a dyadic tree with dynamics evolving on the tree.

- The state space representation for self-similar processes can also be expressed with first order time varying ordinary differential equations $(\mathrm{d}/\mathrm{d}t)\mathbf{x}(t) = (t^{\mathbf{H}}\mathbf{A}t^{-\mathbf{H}}/t)\mathbf{x}(t) + (t^{\mathbf{H}}\mathbf{B}/t)\mathbf{u}(t)$, and time varying state space techniques can be used in their analysis. However, this representation does not provide any insight for the inner dynamics and time evolution of the self-similar states. Furthermore, the proposed representation along with the interpretation of the derivative operator $t(\mathrm{d}/\mathrm{d}t)$, leads to alternative and more efficient implementation techniques as we will discuss in Sections 4 and 5.
- The conditions on the stability, controllability and observability of the states can be easily derived from the system and measurement models using the techniques available for the classical state space representations [14].

4. Kalman filtering

In this section, we will address the problem of recursive estimation and prediction of the self-similar signals within the framework of Kalman filtering [15]. Naturally, the state space representation of self-similar signals introduced in the previous section provides the system and observation models necessary for Kalman machinery. In the following subsections we will derive the prediction and backward smoothing algorithms of the Kalman filter for self-similar processes where the optimal Kalman gain and the Riccati equations are derived.

4.1. Prediction

Consider the self-similar ARMA processes in state space representation:

$$t\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \underbrace{t^{\mathbf{H}}(\mathbf{A}+\mathbf{H})t^{-\mathbf{H}}}_{\mathbf{A}(t)}\mathbf{x}(t) + \underbrace{t^{\mathbf{H}}\mathbf{B}}_{\mathbf{B}(t)}\mathbf{w}(t), \quad (32)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{v}(t), \quad (33)$$

where, $w(t)$ is the system noise and $v(t)$ is the measurement noise where both are zero mean, Gaussian, white noise with covariances: $E[\mathbf{w}(t)\mathbf{w}^{\mathrm{T}}(\tau)] = \mathbf{Q}\tilde{\delta}(t/\tau)$, $E[\mathbf{v}(t)\mathbf{v}^{\mathrm{T}}(\tau)] = \mathbf{R}\tilde{\delta}(t/\tau)$. They are also uncorrelated with each other and the states.

If all the state space representation matrices are known, the same system can be easily established and the states and the outputs can be estimated if the initial conditions are known. However, since this will be an open loop system, the estimates will not be robust. Thus, in the Kalman filter, the estimated states $\hat{\mathbf{x}}(t)$ are obtained by feeding the error term obtained from the original measurements back to the original system model, whose

effect is determined by the Kalman gain matrix. The error state $\tilde{\mathbf{x}}(t)$ is obtained by the difference between the estimated states and the true ones $\tilde{\mathbf{x}}(t) = \hat{\mathbf{x}}(t) - \mathbf{x}(t)$. The optimum state estimates and the Kalman gain matrix coefficients are obtained by minimizing the covariance matrix of the error states in the minimum mean square sense. By estimating the states, a one-step ahead predictor for the output processes is also obtained. The proposed Kalman filter algorithm is summarized in Table 1.

As can be seen from the above equations, our algorithm does not need any extra steps as compared to the multiresolution Kalman filtering proposed in [9]. This is an expected result since our state dynamics evolve on the time scalings as opposed to a dyadic tree described in [9].

The Kalman filter equations derived by using the proposed state space representation for multivariate self-similar processes can also be obtained by using the classical time-varying Kalman filter equations. The major difference between our representation and the ordinary case lies in the state update and error covariance propagation equations. Here, the memory is captured in infinitesimal time scales as opposed to the ordinary case, which captures dynamics in time shifts.

The difference in the representation of the memory content of the state update and error covariance propagation in terms of time scales or time shifts provide us with two different approximation of the continuous time Kalman filter in terms of exponential or uniform discrete time samples, respectively. Hence, the major difference in these representations occur in the performance of their discrete time implementations. Since the exponential sampling is the Shannon type optimal sampling procedure for self-similar processes as explained in Section 2, Theorem 1 (interested readers are referred to [11]), we expect that the proposed Kalman filter via exponential samples approximates the continuous time Kalman filter better than the classical Kalman filter representation via uniform samples. In the following section, where we present simulation examples for the proposed Kalman filtering, we will also show numerically that exponential sampling provides better approximation than the uniform sampling.

4.2. Backward smoothing

As is well known, although Kalman filter allows real time application via recursive prediction, it suffers from the build up of errors in time. If offline processing is available and noise elimination is essential, then backward smoothing leads to the optimal Wiener filtering.

The backward smoothing algorithm for the proposed Kalman filtering can be easily obtained by adapting the Rauch–Tung–Striebel (RTS) algorithm [16,17] for the proposed system and measurement models. The backward dynamics of

Table 1
Kalman filtering algorithm

System Model	$t\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{w}(t)$
	$E[\mathbf{w}(t)] = 0\ E[\mathbf{w}(t)\mathbf{w}^{\mathrm{T}}(\tau)] = \mathbf{Q}\tilde{\delta}(t/\tau)$
	$\mathbf{A}(t) = t^{\mathbf{H}}(\mathbf{A}+\mathbf{H})t^{-\mathbf{H}}$, $\mathbf{B}(t) = t^{\mathbf{H}}\mathbf{B}$
Measurement Model	$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{v}(t)$
	$E[\mathbf{v}(t)] = 0\ E[\mathbf{v}(t)\mathbf{v}^{\mathrm{T}}(\tau)] = \mathbf{R}\tilde{\delta}(t/\tau)$
Initial Conditions	$E[\mathbf{x}(t_1)] = \hat{\mathbf{x}}(t_1)$,
	$E[(\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1))(\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1))^{\mathrm{T}}] = \mathbf{P}(t_1)$
Other Assumptions	$\mathbf{R}^{-1}$ exists
State Estimate	$t\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \underbrace{t^{\mathbf{H}}\mathbf{K}'(t)}_{\mathbf{K}(t)}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t))$
Error Covariance Propagation	$t\dot{\mathbf{P}}(t) = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^{\mathrm{T}}(t) + \mathbf{B}(t)\mathbf{Q}\mathbf{B}^{\mathrm{T}}(t) - \mathbf{K}(t)\mathbf{R}\mathbf{K}^{\mathrm{T}}(t)$
Kalman Gain Matrix	$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}$ when $E[\mathbf{w}(t)\mathbf{v}^{\mathrm{T}}(\tau)] = 0$

the smoothed signal estimate $\hat{\mathbf{x}}_s(t)$ and the covariance matrix $\mathbf{P}_s(t)$ are derived as

$$t\dot{\hat{\mathbf{x}}}_s(t) = \mathbf{A}(t)\hat{\mathbf{x}}_s(t) + \mathbf{B}(t)\mathbf{Q}\mathbf{B}^{\mathrm{T}}(t)\mathbf{P}^{-1}(t)[\hat{\mathbf{x}}_s(t) - \hat{\mathbf{x}}(t)],$$

$$\begin{aligned} t\dot{\mathbf{P}}_s(t) = {} & [\mathbf{A}(t) + \mathbf{B}(t)\mathbf{Q}\mathbf{B}^{\mathrm{T}}(t)\mathbf{P}^{-1}(t)]\mathbf{P}_s(t) \\ & + \mathbf{P}_s(t)[\mathbf{A}(t) + \mathbf{B}(t)\mathbf{Q}\mathbf{B}^{\mathrm{T}}(t)\mathbf{P}^{-1}(t)]^{\mathrm{T}} \\ & - \mathbf{B}(t)\mathbf{Q}\mathbf{B}^{\mathrm{T}}(t). \end{aligned} \tag{34}$$

Here, the starting conditions are $\hat{\mathbf{x}}_s(T) = \hat{\mathbf{x}}(T)$ and $\hat{\mathbf{P}}_s(T) = \hat{\mathbf{P}}(T)$ where T is the final time of the measured signal. Therefore, the smoothed signal is estimated by backward processing of the Kalman filtering algorithm given in the previous subsection by starting from $t = T$ ending at $t = t_1$ where t_1 was the initial time of recording.

5. Numerical and simulation experiments

5.1. Experiment I

Throughout this simulation experiment, we use the following first order AR process where $H = -0.3$, $A = -0.2$, $B = 1$, $C = 1$, $Q = 1$ to test the performance of the proposed Kalman filter. The system and the measurement models are:

$$t\dot{x}(t) = -0.5x(t) + t^{-0.3}w(t),$$

$$y(t) = x(t) + v(t), \tag{35}$$

We generate the simulation data via its covariance matrix and Karhunen–Loeve (KL) transform (please see [18] and the references therein). The auto-covariance of $x(t)$ for the first order system given above is

$$C_{xx}(t_1, t_2) = \beta(t_1 t_2)^{(A+H)}(\max(t_1, t_2)^{(-2A)} - 1), \tag{36}$$

where $\beta = B^2Q/(-2A)$. Using covariance matrix above and the KL transform, we generate $x(t)$ and $y(t)$ for $1 < t < 3.14$. We propose two methods to implement the continuous time Kalman filter for discrete data. These methods essentially differ in the way the derivative operator $t(\mathrm{d}/\mathrm{d}t)$ is discretized as described in the previous section.

5.2. Kalman predictor and smoother via exponential sampling

In the Kalman filtering algorithm, the estimated state in continuous time is given by

$$t\dot{\hat{x}}(t) = A(t)\hat{x}(t) + K(t)(y(t) - C\hat{x}(t)). \tag{37}$$

Here, we discretize $t(\mathrm{d}/\mathrm{d}t)x(t) \approx (x(t\Delta_e) - x(t)/\ln \Delta_e)$ and obtained the following approximation to Eq. (37):

$$\hat{x}(\Delta_e t) = \hat{x}(t) + \ln \Delta_e(A(t)\hat{x}(t) + K(t)(y(t) - C\hat{x}(t))). \tag{38}$$

Similarly, the Riccati equation solution for $P(\Delta_e t)$ for each $\Delta_e t$ time:

$$\begin{aligned} P(\Delta_e t) = {} & P(t) + \ln \Delta_e(A(t)P(t) + P(t)A^{\mathrm{T}}(t) \\ & + B(t)Q(t)B^{\mathrm{T}}(t) - K(t)R(t)K^{\mathrm{T}}(t)). \end{aligned} \tag{39}$$

Note that, in our application $A(t) = -0.5$, $B(t) = t^{-0.3}$ and $C = 1$. As expected, the continuous time approximation becomes more accurate when the scale step Δ_e is selected as close to 1 as possible. In our application we select the scale step as $\Delta_e = 1.0045$.

We test the performance of the Kalman filter for four different SNRs of 20, 10, 5 and 2 dB using 100 Monte Carlo Runs. The SNR of the signal is calculated as:

$$\mathrm{SNR} = \mathrm{var}(x)/\mathrm{var}(v). \tag{40}$$

A sample data $x(t)$ (solid line), $y(t)$ (dash–dot line) and the estimated data $\hat{x}(t)$ (dashed line) for SNR = 20, 10, 5 and 2 dBs out of 100 runs are given in Fig. 1(a)–(d), respectively. The estimation SNRe for each estimated signal $\hat{x}(t)$ is calculated as:

$$\mathrm{SNRe} = \mathrm{var}(x)/\mathrm{var}(x - \hat{x}). \tag{41}$$

Note that the proposed Kalman filter also suffers from the same problems as the conventional Kalman filter, that is from the build up of the "random walk" type errors as the prediction time increases. This problem can be overcome by using the backward smoother.

Again, the smoothed state estimate in continuous time is approximated using backwards scale derivative definition given in (31) for exponential

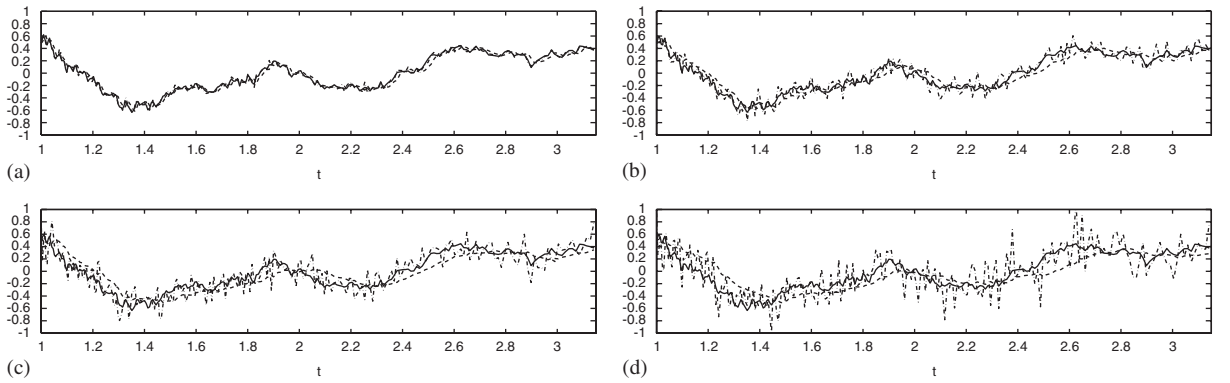

Fig. 1. The input signal, $x(t)$ (solid line), the observed signal, $y(t)$ (dash–dot line) and the predicted signal, $\hat{x}(t)$ (dashed line) for: (a) input SNR = 20 dB; (b) input SNR = 10 dB; (c) input SNR = 5 dB; (d) input SNR = 2 dB.

time intervals as

$$\hat{x}_s(t) = \hat{x}_s(\Delta_e t) - \ln \Delta_e (A(\Delta_e t)\hat{x}_s(\Delta_e t) + B(\Delta_e t) \times QB^T(\Delta_e t)P^{-1}(\Delta_e t)[\hat{x}_s(\Delta_e t) - \hat{x}(\Delta_e t)]). \quad (42)$$

We process 100 Monte Carlo Runs to test the performance of the smoother for SNRs of 20, 10, 5 and 2 dB. The sample data from the same seeds as in the prediction case, $x(t)$ (solid line), $y(t)$ (dash–dot line) and the estimated smoothed data $\hat{x}_s(t)$ (dashed line) for SNR = 20, 10, 5 and 2 dB out of 100 runs are given in Fig. 2(a)–(d), respectively.

The smoothing SNRse for each smoothed signal $\hat{x}_s(t)$ is calculated as

$$\text{SNRse} = \text{var}(x)/\text{var}(x - \hat{x}_s). \quad (43)$$

5.3. Classical Kalman predictor and smoother via uniform sampling

We will now compare the performance of the implementation of the proposed Kalman prediction and smoothing algorithms with the usual Kalman predictor and smoother. In this case, the system and measurement models are expressed as follows:

$$\dot{x}(t) = -0.5x(t)/t + t^{-0.3}w(t)/t,$$

$$y(t) = x(t) + v(t) \quad (44)$$

and the derivative operator is discretized as $(d/dt)x(t) \approx x(t + \Delta_u) - x(t)/\Delta_u$. We create $1/f$ data using the covariance matrix definition as given in (36) and KL transform for uniform time intervals between $1 < t < 3.14$.

For usual Kalman filtering, the continuous time state estimate $\hat{x}(t)$ is approximated using uniform time intervals as:

$$\hat{x}(t + \Delta_u) = \hat{x}(t) + \Delta_u(A(t)\hat{x}(t) + K(t)(y(t) - C\hat{x}(t))). \quad (45)$$

For the Riccati equation solution,

$$P(t + \Delta_u) = P(t) + \Delta_u(A(t)P(t) + P(t)A^T(t) + B(t)Q(t)B^T(t) - K(t)R(t)K^T(t)). \quad (46)$$

Naturally, continuous time approximation improves as the uniform sampling intervals get closer to zero. Here, the time interval is selected as $\Delta_u = 0.0065$.

We test the performance of the usual Kalman filter for the SNR levels of 20, 10, 5 and 2 dB using 100 Monte Carlo runs. A sample data $x(t)$ (solid line), $y(t)$ (dash–dot line) and the estimated data $\hat{x}(t)$ (dashed line) for SNR = 20, 10, 5 and 2 dB out of 100 runs are given in Fig. 3(a), (b), (c) and (d), respectively.

We apply the usual smoothing algorithm of Rauch–Tung–Striebel [16,17]. The smoothed state

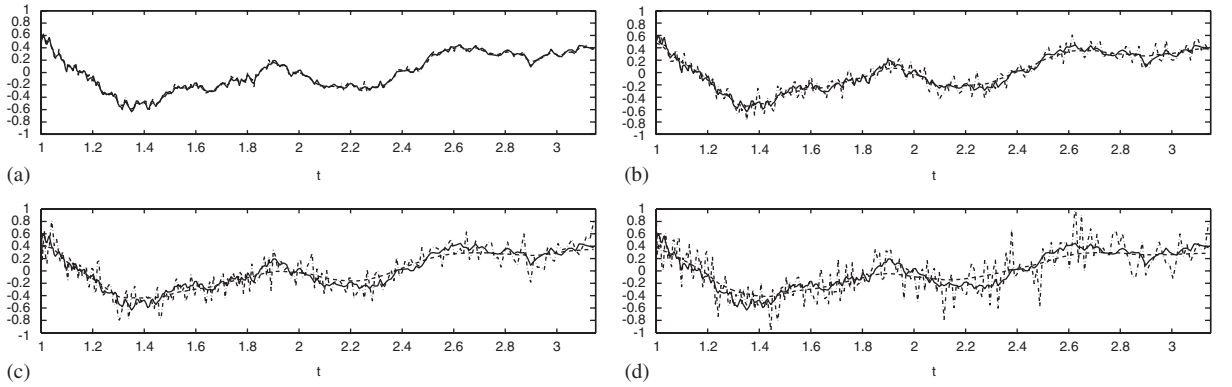

Fig. 2. The input signal, $x(t)$ (solid line), the observed signal, $y(t)$ (dash–dot line) and the smoothed signal, $\hat{x}_s(t)$ (dashed line) for: (a) input SNR = 20 dB; (b) input SNR = 10 dB; (c) input SNR = 5 dB; (d) input SNR = 2 dB.

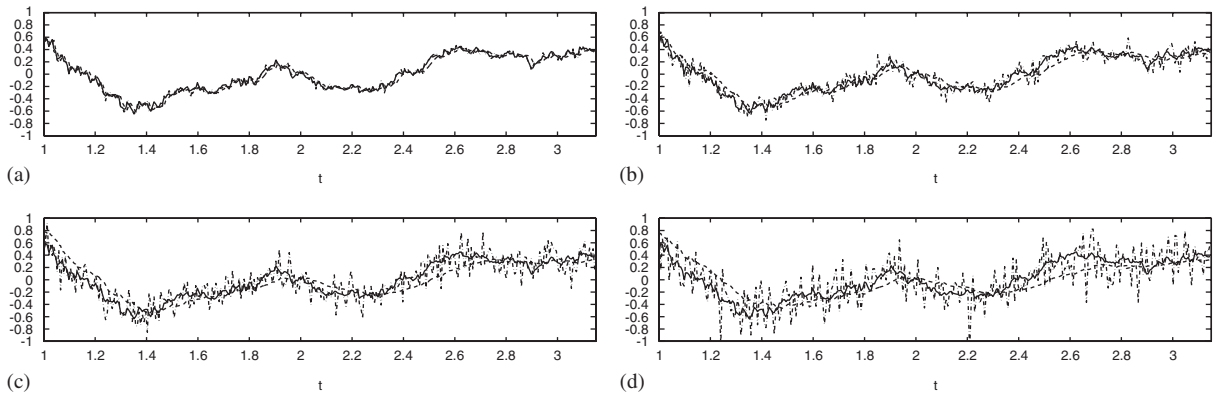

Fig. 3. The input signal, $x(t)$ (solid line), the observed signal, $y(t)$ (dash–dot line) and the predicted signal, $\hat{x}(t)$ (dashed line) for: (a) input SNR = 20 dB; (b) input SNR = 10 dB; (c) input SNR = 5 dB; (d) input SNR = 2 dB.

estimate in continuous time is approximated as follows:

$$\hat{x}_s(t) = \hat{x}_s(t+\Delta_u) - \Delta_u(A(t+\Delta_u)\hat{x}_s(t+\Delta_u) + B(t+\Delta_u)QB^T(t+\Delta_u)P^{-1}(t+\Delta_u) \times[\hat{x}_s(t+\Delta_u) - \hat{x}(t+\Delta_u)]). \quad (47)$$

We again process 100 Monte Carlo Runs to test the performance of the smoother for SNRs of 20, 10, 5 and 2 dBs. We use the same random number generator seeds to provide visual comparison between the performances of two implementation techniques. These results are presented in Fig. 4(a), (b), (c) and (d), respectively.

In order to compare the performances of the proposed and the conventional Kalman filter algorithms, using 100 Monte Carlo Runs, we present in errorbars the estimation SNRe versus the input SNR for the prediction in Fig. 5(a) and the smoothing SNRse versus the input SNR for the smoothing in Fig. 5(b). In these figures the solid line is used for the proposed algorithm with exponential sampling and dashed line is used for the classical algorithm with uniform sampling.

What has been shown analytically in [11,12], has also been verified numerically in this section, that the implementation of the continuous time

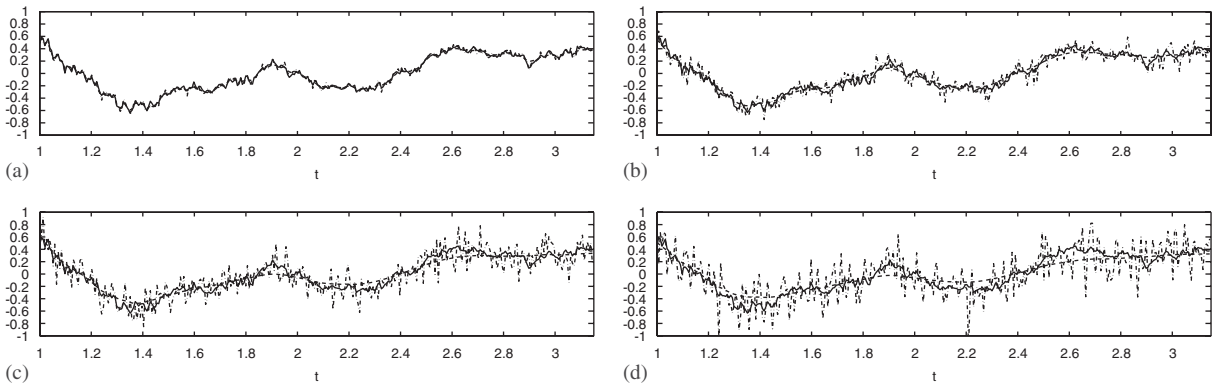

Fig. 4. The input signal, $x(t)$ (solid line), the observed signal, $y(t)$ (dash–dot line) and the smoothed signal, $\hat{x}_s(t)$ (dashed line) for: (a) input SNR = 20 dB; (b) input SNR = 10 dB; (c) input SNR = 5 dB; (d) input SNR = 2 dB.

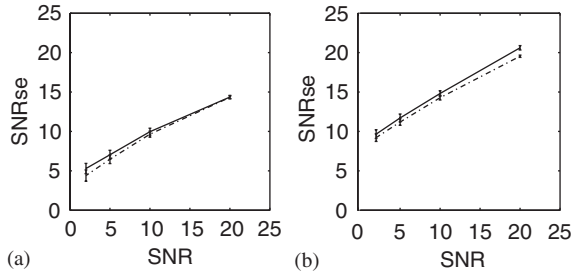

Fig. 5. (a) The prediction SNRe versus the input SNR of the proposed (solid line) and the usual (dashed line) Kalman predictor; (b) the smoothing SNRse versus the input SNR of the proposed (solid line) and the usual (dashed line) Kalman smoother.

Kalman filter for self-similar processes depending on the proposed state space representation based on exponential sampling provides better performance than the classical time-varying Kalman filter based on uniform sampling intervals. Note that the numerical simulation results shown in Fig. 5 are obtained by using uniformly sampled data with 331 sample points for the classical Kalman filtering; and the same data exponentially sampled with 256 sample points for the proposed approach. While the improvements upon the classical approach appears limited, further improvements in performance can be achieved by increasing the number of data points in the proposed Kalman filtering approach. However, the simulation results shown in Fig. 5 indicate that proposed approach provides at least as good results as the standard approach with less number of data points and hence less computation.

5.4. Experiment II

In the first experiment, we used a univariate system and showed the performance of the Kalman filter to exponential and uniform samples. In this experiment, we will study a multivariate system. We implement the following second order AR process where $\mathbf{H} = [-0.2\ 0;\ 0\ \ -0.4]$, $\mathbf{A} = [-0.2\ 0;\ -0.1\ \ -0.3]$, $\mathbf{B} = [1\ 0;\ 0\ 1]$, $\mathbf{C} = [1\ 1;\ 0.2\ 0.6]$, $\mathbf{Q} = [1\ 0;\ 0\ 1]$ to test the performance of the proposed Kalman filter to exponential samples. In this experiment, the exponential sampling period is selected as $\Delta_e = 1.006$. The system and the measurement models are:

$$t\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = t^{\mathbf{H}}\underbrace{\left(\begin{bmatrix}-0.2 & 0\\ -0.1 & -0.3\end{bmatrix}+\begin{bmatrix}-0.2 & 0\\ 0 & -0.4\end{bmatrix}\right)}_{\mathbf{A}+\mathbf{H}} \times t^{-\mathbf{H}}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + t^{\mathbf{H}}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}w_1(t)\\ w_2(t)\end{bmatrix},$$

$$\begin{bmatrix}y_1(t)\\ y_2(t)\end{bmatrix} = \begin{bmatrix}1 & 1\\ 0.2 & 0.6\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \begin{bmatrix}v_1(t)\\ v_2(t)\end{bmatrix}. \quad (48)$$

The simulated data for the states $x_1(t)$, $x_2(t)$ and the outputs $y_1(t)$ and $y_2(t)$ obtained from the above system and observation equations are presented in Fig. 6 (a), (b), (c) and (d), respectively.

Next, we applied the proposed Kalman filter given in Table 1 to exponential samples and obtain the predicted and smoothed states. In Fig. 7 (a) and (b), we present the original (solid line) and predicted (dashed line) states $x_1(t)$ and $x_2(t)$, respectively. In Fig. 8 (a) and (b), we present the original (solid line) and smoothed (dashed line) states $x_1(t)$ and $x_2(t)$, respectively. As can be seen from this experiment, the proposed Kalman filter estimates multivariate self-similar states optimally from the observation data.

6. Application areas

Many estimation and prediction tasks in signal processing involving $1/f$ processes can be formulated within the proposed Kalman filtering framework. Some applications include network traffic prediction, which has potential implications in network management and quality service provisioning and matched filtering in communications applications to eliminate $1/f$ type interference or environmental noise.

In this section, we will explain how proposed methods can be employed to develop solutions for aforementioned problems.

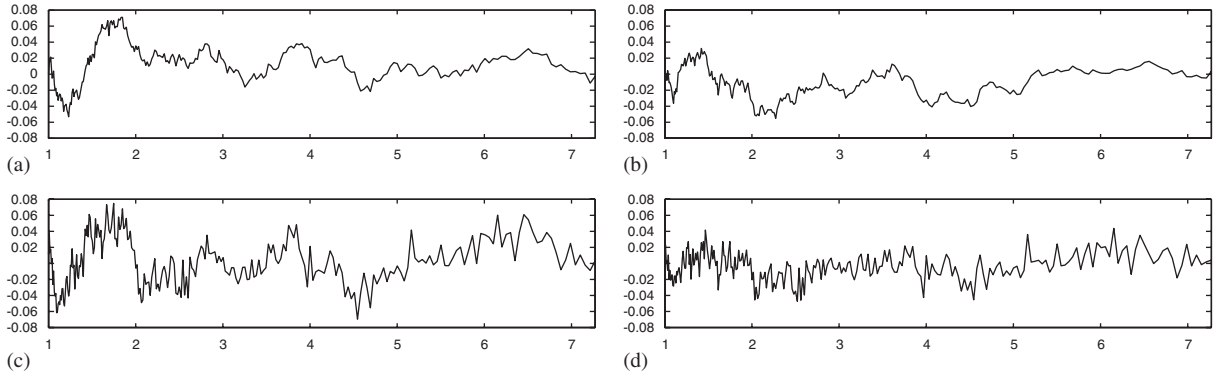

Fig. 6. (a) The input signal, $x_1(t)$; (b) the input signal, $x_2(t)$; (c) the observed signal, $y_1(t)$ and (d) the observed signal, $y_2(t)$.

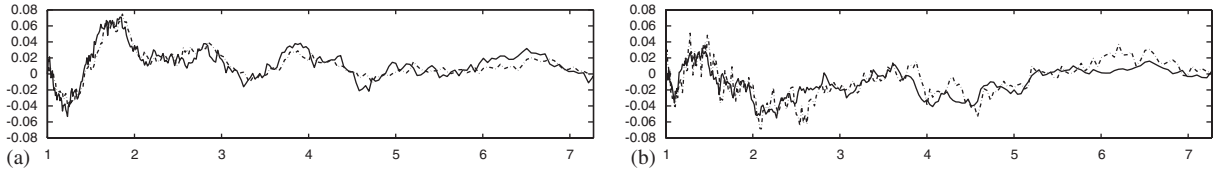

Fig. 7. The original (solid line) and the predicted (dashed line) input signals: (a) $x_1(t)$; (b) $x_2(t)$.

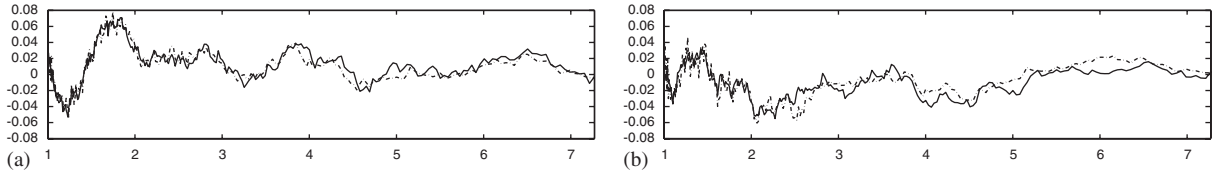

Fig. 8. The original (solid line) and the smoothed (dashed line) input signals: (a) $x_1(t)$; (b) $x_2(t)$.

6.1. Self-similar network traffic modeling and prediction

Recent measurements of local area (LAN) and wide area (WAN) networks such as in Ethernet, ISDN packet networks, signaling (CCSN/SS7) networks for public telephone networks [5,18–25], have shown that network traffic exhibits self-similar statistics as opposed to the early assumptions of Poisson distribution. Due to the attractive theoretical properties of Poisson processes, such as independence between arrivals and exponential distribution for interarrivals, the network traffic was at first modeled using such processes. However, subsequent experimental studies on the behavior of both Ethernet LAN and WAN traffic show that the distribution of packet interarrivals clearly differs from exponential distributions [5,19,22]. Especially in the seminal work of Leland et al. [5], the packet arrivals that are aggregated over non-overlapping blocks in different time scales showed the same statistics in each time scale and long range correlations which decays hyperbolically over different time scales.

Subsequent to this work, a string of studies has appeared dealing with various aspects of traffic self-similarity [18–25]. In most of these studies, the self-similarity parameter is considered as the major characteristic of self-similarity, and its reliable estimation from the network measurements is an important task for both characterization and simulation of the traffic [5,19,23,24,26]. In [20] and [25] it was shown that long-range dependent traffic is likely to degrade the performance and the queue length distribution under self-similar traffic decays, much more slowly than with short range dependent sources. In the studies mentioned above the magnitude of the traffic is shown to be self-similar. Additionally, recent study [21] shows that the traffic delay, which is encountered in the encoding, transmission, receiver buffering and decoding, is also self-similar. This fact is expected to have a fundamental impact on the quality of service (QoS) of real time applications.

For the modeling of the self-similar traffic, fractional Brownian motion [5,23,25] and ON/OFF modeling [20,22,24] have been proposed and studied. While these models have been successful in developing insights to the source of self-similarity in network traffic, they do not have predictive and recursive capability. However, prediction of the traffic is an important problem for the improvement of the QoS. For example, it will allow optimal buffer size estimation that will eliminate packet loss or it will provide optimum rate channel assign (RCA) for the optimal usage of the resources by the base station controller in IS-2000. Using the proposed Kalman filtering technique, the self-similar network traffic for each user can be predicted recursively in real time, which will allow estimation of optimal RCA or optimal buffer size leading to better QoS.

In order to use the proposed Kalman filtering framework, the system and measurement models parameters for each user, i.e., $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{H}$ have to be known or estimated. We believe that the estimation of the model parameters and prediction should be interleaved or performed simultaneously within the expectation-maximization framework [27,28] or by using the extended Kalman filtering algorithm of Tsatsanis et al. [29] that adaptively estimates the system parameters using a least squares technique. These topics are open research areas and will be addressed in our future work.

6.2. Matched filtering for communication systems

Another potential application area for the proposed Kalman filter is matched filtering for communication systems. The received signal $y(t)$ for matched filtering is conventionally modeled as the summation of the transmitted signal $x(t)$ and the environmental noise $w(t)$ as

$$y(t) = x(t) + w(t). \tag{49}$$

Then, the matched filtering problem becomes one of finding a linear filter that maximizes the output SNR.

It is shown that in many communication applications such as in radar, sonar, underwater acoustics, the noise that comes from the environment is non-stationary or time-varying [29–33]. Especially in underwater acoustics and for active sonar, the reverberation, which is caused by the reflection of the transmitted signal on

several interfaces such as air-sea or ground-sea, is considered as non-stationary noise [30]. Similarly, the transmitted signal can also be non-stationary [33]. In such cases, the received signal, which is a linear combination of self-similar process embedded in additive long term correlated noise, can be modeled using the proposed state space model with self-similar states of first order AR processes. Then, the proposed Kalman filtering technique can be used for matched filtering to eliminate the noise. Here again, the parameter and state estimation needs to be performed iteratively or simultaneously.

7. Conclusion

In this paper, we developed continuous time state space representation and an optimal Kalman filtering algorithm for self-similar processes. In the system model, we introduced multivariate self-similarity concept where self-similarity of the states is captured in the self-similarity matrix and the memory is captured in infinitesimal time scalings. The concept of multivariate self-similarity forms the essence of our development.

Using the system and measurement models, we formulate the continuous time Kalman filter to estimate, predict or smooth self-similar processes. Although the algorithm is in the same form as the usual Kalman filter, the major difference is in the memory content or the dynamics of the estimated state and the error covariance matrix, which is appropriate for capturing the self-similar nature of the processes. We show analytically and also numerically that approximation of this scale memory content of the self-similar state and error covariance estimates in the Kalman filtering implementation via exponential sampling provides better than the uniform sampling.

This work can be extended to several further research areas. Estimation of model parameters, including the self-similarity matrix, has to be addressed for those applications where model parameters are unknown. Another important direction is to extend these ideas to discrete time models for numerically efficient implementation.

References

[1] B. Yazici, R. Kashyap, A class of second-order stationary self-similar processes for $1/f$ phenomena, IEEE Trans. Signal Process. 45 (2) (1997) 396–410.
[2] B.B. Mandelbrot, J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, Siam Rev. (1968) 422–437.
[3] G.W. Wornell, Signal Processing with Fractals: A Wavelet-Based Approach, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[4] M.S. Keshner, $1/f$ Noise, Proc. IEEE 70 (1982) 212–218.
[5] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, On the self-similar nature of ethernet traffic (extended version), IEEE/ACM Trans. Network. 2 (1) (1994) 1–15.
[6] M. Bladt, Multivariate self-similar processes: second-order theory, J. Appl. Prob. 31 (1994) 139–147.
[7] C.J. Nuzman, H.V. Poor, Linear estimation of self-similar processes via Lamperti's transformation, J. Appl. Prob. 37 (2000).
[8] K.C. Chou, A.S. Willsky, A. Benveniste, Multiscale Recursive Estimation, Data Fusion and Regularization, IEEE Trans. Automat. Control 39 (3) (1994) 464–478.
[9] K.C. Chou, A.S. Willsky, R. Nikoukhah, Multiscale systems, Kalman filters, and Riccati equation, IEEE Trans. Automat. Control 39 (1994) 479–492.
[10] R. Bracewell, The Fourier Transform and its Applications, McGraw-Hill, New York, 1965.
[11] B. Yazici, R.L. Kashyap, Signal modeling and parameter estimation for $1/f$ processes using scale-stationary models, in: Proceedings of ICASSP'96, 1996, pp. 2841–2844.
[12] M. Izzetoglu, Wiener and Kalman filtering for self-similar processes, Ph.D. Thesis, Drexel University, 2002.
[13] J. Bertrand, P. Bertrand, J.P. Ovarlez, Discrete Mellin Transform for Signal Analysis, in: Proceedings of IEEE ICASSP-90, 1990, pp. 1603–1606.
[14] T. Kailath, Linear Systems, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[15] M.S. Grewal, A.P. Andrews, Kalman Filtering: Theory and Practice, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[16] H.E. Rauch, F. Tung, C.T. Striebel, Maximum likelihood estimates of linear dynamic systems, AIAA J. 3 (8) (1965) 1445–1450.
[17] A. Gelb, Applied Optimal Estimation, The MIT Press, Cambridge, MA, 1974.
[18] M.A. Thornton, The Karhunen–Loeve transform of discrete MVL functions, in: IEEE International Symposium on Multiple-Valued Logic (ISMVL), 2005, pp. 194–199.
[19] V. Paxson, S. Floyd, Wide area traffic: the failure of Poisson modeling, IEEE/ACM Trans. Network. 3 (3) (1995) 226–244.
[20] X. Yang, A.P. Petropulu, The extended alternating fractal renewal process for modeling traffic in high-speed communication networks, IEEE Trans. Signal Process. 49 (7) (2001) 1349–1363.

[21] H.S. Borella, S. Uludag, G.B. Brewster, I. Sidhu, Self-Similarity of Internet Packet Delay, Proceedings of International Conference on Communications 1 (1997) 513–517.

[22] K. Park, G. Kim, M. Crovella, On the relationship between file sizes, transport protocols, and self-similar network traffic, in: International Conference on Network Protocols, 1996, pp. 171–180.

[23] J. Beran, R. Sherman, M.S. Taqqu, W. Willinger, Long-range dependence in variable-bit-rate video traffic, IEEE Trans. Commun. 43 (2/3/4) (1995) 1566–1579.

[24] W. Willinger, M.S. Taqqu, R. Sherman, D.V. Wilson, Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level, IEEE/ACM Trans. Network. 5 (1) (1997) 71–86.

[25] I. Norros, On the use of fractional Brownian motion in the theory of connectionless networks, IEEE J. Selected Areas Commun. 13 (1995) 953–962.

[26] W.E. Leland, Self-similarity and traffic measurement, in: Proceedings of Eighth Workshop on Computer Communications, 1993, pp. 32–35.

[27] H. Zamiri-Jafarian, S. Pasupathy, EM-based recursive estimation of channel parameters, IEEE Trans. Commun. 47 (9) (1999) 1297–1302.

[28] L.A. Johnston, V. Krishnamurthy, Derivation of a sawtooth iterated extended Kalman smoother via the AECM algorithm, IEEE Trans. Signal Process. 49 (9) (1999) 1899–1909.

[29] M.K. Tsatsanis, G.B. Giannakis, G. Zhou, Estimation and equalization of fading channels with random coefficients, Proceedings of ICASSP'96 2 (1996) 1093–1096.

[30] G. Ginolhac, Detection in presence of reverberation, IEEE Conference and Exhibition OCEANS 2000 MTS 2 (2000) 1047–1051.

[31] N.S. Kontoyannis, J.R. Mitchell, A.A. Louis Beex, Robust Digital Communication in Time-Varying Noise, Proceedings of ICASSP'92 5 (1992) 145–148.

[32] A.A. Louis Beex, J.M. Wilson, Robust communication in a time-varying environment, Proceedings of ICASSP'88 3 (1988) 1546–1549.

[33] B.S. Chen, Y.C. Chung, D.F. Huang, A wavelet time-scale deconvolution filter design for non-stationary signal transmission systems through a multipath fading channel, IEEE Trans. Signal Process. 47 (5) (1999) 1441–1446.