EFFICIENT MULTISCALE STOCHASTIC REALIZATION

Austin B. Frakt, Alan S. Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA 02139
http://vougeot.mit.edu/ssg.cgi/

ABSTRACT

Few fast statistical signal processing algorithms exist for large problems involving non-stationary processes and irregular measurements. A recently introduced class of multiscale autoregressive models indexed by trees admits signal processing algorithms which can efficiently deal with problems of this type. In this paper we provide a novel and efficient algorithm for translating any second-order prior model to a multiscale autoregressive prior model so that these efficient signal processing algorithms may be applied.

1. INTRODUCTION

Large statistical signal processing problems require fast algorithms. Yet the most widely used fast algorithms—for instance, those based on the fast Fourier transform or the wavelet transform—are applicable to a very restricted range of problems. For large problems involving non-stationary processes with correlations at many time-scales and for which sparse, non-local data corrupted by non-stationary measurement noise are available, few fast algorithms exist. Further complicating the signal processing challenge, there is a need in many applications to fuse data which represent measurements at different scales.

This paper builds on a recently introduced framework for multiscale signal processing which addresses all of the aforementioned challenges. This *multiscale autoregressive* (MAR) framework has already seen successful application in a variety of signal processing contexts [4–6] and is a generalization of the standard discrete-time autoregressive (AR) framework. The MAR framework admits fast sample path generation, fast likelihood calculation [8], and fast linear least squares estimation [3]. The estimator has computational complexity linear in problem size and is a generalization of the Kalman filter and Rauch-Tung-Striebel smoother. As such, it provides estimation error statistics with no additional computation beyond what is needed to compute the estimates themselves.

As in the AR setting, the challenge in applying the MAR framework is building an appropriate model. This is the *stochastic realization problem* one variant of which will be explored in this paper. More specifically, the contribution of this paper is a computationally efficient and completely general algorithm for mapping any second-order prior model to a MAR prior model. Once such a MAR model is in hand, efficient estimation, sample path generation, or likelihood calculation may be immediately performed.

This work was supported by the Office of Naval Research (grant N00014-91-J-1004) and the Boston University Multidisciplinary University Research Initiative (grant GC123919NGD). A.B. Frakt is supported by a National Defense Science and Engineering Graduate Fellowship awarded by the Defense Advanced Research Projects Agency.

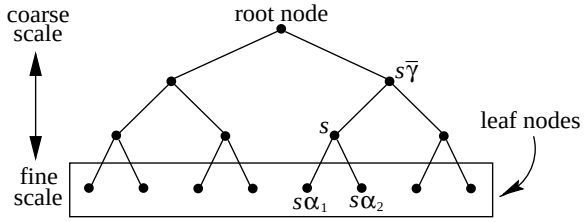

Figure 1: Node $s\bar{\gamma}$ is the parent of node s while $s\alpha_1, s\alpha_2$ are its children. Leaf nodes (in the box) represent the finest scale; the root node represents the coarsest scale.

This multiscale stochastic realization problem has been explored before [4, 7]. The work in [7] is based on a novel application of canonical correlations analysis which is a standard tool employed in the AR context [1, 2]. One drawback is that canonical correlations analysis requires the singular value decomposition of large covariance matrices. Consequently, the realization algorithm developed in [7] is quartic in problem size.

The approach in [4] is based on that in [7] but applied to the special case of self-similar processes with independent increments (e.g., fractional Brownian motions). What is shown in [4] is that the self-similarity and independent increments properties can be exploited to circumvent much of the computational effort required by the method of [7]. The resulting algorithm is cubic in problem size and is only applicable to this restricted class of processes. In contrast to the methods of [4, 7] our approach is only quadratic in problem size and, unlike the approach of [4], is completely general.

In Section 2 we review the MAR framework and state the stochastic realization problem. In Section 3 we define the concept of an internal process which, in turn, provides a convenient parameterization for the realization problem. In Section 4 we describe the computational engine of our realization procedure and contrast it with canonical correlations analysis. Our realization algorithm is described in Section 5. Examples are provided in Section 6 and conclusions and extensions are discussed in Section 7.

2. THE MAR FRAMEWORK AND STOCHASTIC REALIZATION PROBLEM

MAR processes are indexed by trees. While they may be defined on any tree, for our purposes dyadic trees will suffice (Figure 1). Let $\mathcal{S}$ be the set of nodes of a dyadic tree where $0 \in \mathcal{S}$ is the root node. We denote by $s\bar{\gamma}$ the parent node of s and by $s\alpha_1$, $s\alpha_2$ the left and right child of s, respectively. There is a natural notion of

scale associated with $\mathcal{S}$. The root node corresponds to the coarsest scale (scale zero) while the leaf nodes comprise the finest scale (scale M). More generally, scale n consists of the nodes in the set $\{s \in \mathcal{S} \mid s\bar{\gamma}^n = 0\}$.

A zero-mean MAR process[1] $x(\cdot)$ is autoregressive in scale:

$$x(s) = A(s)x(s\bar{\gamma}) + w(s) \tag{1}$$

where $w(\cdot)$ is a zero-mean white processes with auto-covariance $Q(s)$ and is uncorrelated with $x(0)$. The object $x(s)$ is referred to as the *state* at node s because, in analogy to the concept of state for AR time-series, conditioned on $x(s)$ the sub-processes which are indexed by nodes which are separated by s are mutually conditionally uncorrelated. For instance, in the case depicted in Figure 1, conditioned on $x(s)$, the three sub-processes $x(s\alpha_1)$, $x(s\alpha_2)$, and $\{x(t) \mid t \neq s\alpha_1, s\alpha_2, s\}$ are mutually conditionally uncorrelated. Hereafter we call this property the *Markov property*.

The second-order statistics of the root node state are[2] $x(0) \sim (0, P_{x(0)})$. The measurement equation associated with the MAR framework is

$$y(s) = C(s)x(s) + v(s) \tag{2}$$

where $v(s)$ is zero-mean white noise with auto-covariance $R(s)$ and is uncorrelated with $x(\cdot)$ and $w(\cdot)$.

The preceding parameterization implicitly provides a complete joint second-order characterization of $x(\cdot)$ and $y(\cdot)$. In particular, the second-order statistics of $x(\cdot)$ restricted to the finest scale (M) are implicitly determined by $A(\cdot)$, $Q(\cdot)$ and $P_{x(0)}$. We will denote the sub-process of $x(\cdot)$ restricted to the finest scale by x_M. The covariance of x_M will be denoted as P_M.

The stochastic realization problem addressed in this paper is the following. Consider a fine-scale signal $f \sim (0, P_f)$. How can we choose MAR parameters $A(\cdot)$, $Q(\cdot)$ and $P_{x(0)}$ so that the finest scale of the MAR process, x_M, has covariance P_M which matches or approximates (in some sense to be made precise) P_f?

3. INTERNAL MAR PROCESSES

There is a convenient parameterization for the realization problem based on the concept of an *internal process*. An internal process is one for which the state at each node s is a linear function of the finest scale process x_M:

$$x(s) = W_s x_M \,, \forall s \in \mathcal{S} \,. \tag{3}$$

The matrices W_s are called *internal matrices*. If all the internal matrices are specified, the realization problem is solved exactly (i.e., with $P_M \equiv P_f$) as follows. First, explicitly enforcing $P_M \equiv P_f$ and using (3) we get that

$$P_{x(s)} \stackrel{\triangle}{=} E[x(s)x(s)^T] = W_s P_M W_s^T = W_s P_f W_s^T \,, \tag{4}$$

$$P_{x(s)x(s\bar{\gamma})} \stackrel{\triangle}{=} E[x(s)x(s\bar{\gamma})^T] = W_s P_M W_{s\bar{\gamma}}^T = W_s P_f W_{s\bar{\gamma}}^T. \tag{5}$$

Next, recognizing that the first term of (1) is the linear least squares estimate of $x(s)$ based on $x(s\bar{\gamma})$ and the second term is the estimation error it follows that

$$A(s) = P_{x(s)x(s\bar{\gamma})} P_{x(s\bar{\gamma})}^{-1} \,, \tag{6}$$

$$Q(s) = P_{x(s)} - P_{x(s)x(s\bar{\gamma})} P_{x(s\bar{\gamma})}^{-1} P_{x(s)x(s\bar{\gamma})}^T \,. \tag{7}$$

Therefore, the remaining step is to find the internal matrices. This is a challenging problem for two reasons. First, there is often a desire for a reduced order realization, i.e., a realization for which the state dimensions are no larger than some integer d. The reason for imposing such a constraint is that the MAR estimator and likelihood calculator have computational complexity $O(Nd^3)$—linear in N, the length of x_M and cubic in the maximum state dimension, d. And sample path generation has complexity $O(Nd^2)$. Of course, if the state dimensions are constrained, one cannot expect $P_M \equiv P_f$. There will be some degree of approximation. Therefore, one would like to find internal matrices which prioritize the information of the state so that it is clear what information is least important and can be thrown out with minimal corruption in P_M (i.e., keeping the degree to which P_M approximates P_f small).

The second challenge in finding internal matrices is that they are coupled. While this assertion can be made precise it is not hard to see intuitively why it is so. A state defined as $x(s) = W_s x_M$ carries some information about x_M given by the row-space of W_s. This information is passed down to x_M as described by (1). If the information retained at nodes which descend from s do not include that retained at node s then $x(s) \neq W_s x_M$. When this happens we say that there is a loss of *internal consistency*.

A necessary and sufficient condition to maintain internal consistency is that each state $x(s)$ is a linear function of its children states $x(s\alpha_1)$, $x(s\alpha_2)$. This fact follows from properties of linear least squares estimation and the Markov property of MAR processes. In our realization algorithm specified in Section 5 we will maintain internal consistency by enforcing this condition as we build internal matrices. We note that the realization approach developed in [7] is also based on the concept of an internal process but does *not* maintain internal consistency. This is one of two significant ways in which our approach differs from that of [7]. The second significant way in which our approach differs from that of [7] will be made clear in the next section.

4. DECORRELATING RANDOM VECTORS

The heart of the realization problem is finding for each $s \in \mathcal{S}$ the internal matrix W_s so that (1) the state $x(s) = W_s x_M$ contains the appropriate decorrelating information as dictated by the Markov property, (2) the number of rows of W_s is minimal, and (3) the information is prioritized so that it is clear which rows of W_s are least important and can be discarded if a reduced order realization is required. We will address these three challenges by considering a pair-wise decorrelation problem which is then easily generalized for the problem of decorrelating three (or more) random vectors.

Let $z = [z_1^T \; z_2^T]^T$ have zero-mean where z_i has length n_i and $n_1 \ll n_2$. Also,

$$P_i \stackrel{\triangle}{=} E[z_i z_i^T] \;\text{ and }\; P_{12} \stackrel{\triangle}{=} E[z_1 z_2^T] \,. \tag{8}$$

Let $\mathcal{M}_r$ be the set of $r \times n_1$ matrices. Canonical correlations analysis provides the matrix $V \in \mathcal{M}_r$ such that conditioned on Vz_1 the vectors z_1 and z_2 are maximally conditionally decorrelated. Since V is restricted to have r rows, we cannot expect that z_1 and z_2 are exactly conditionally decorrelated by Vz_1. What canonical correlations analysis provides is the best possible decorrelating

[1] To handle a non-zero mean we simply consider the deviation of the process from its mean.

[2] The notation $z \sim (m_z, P_z)$ means that the random vector z has mean vector m_z and covariance matrix P_z.

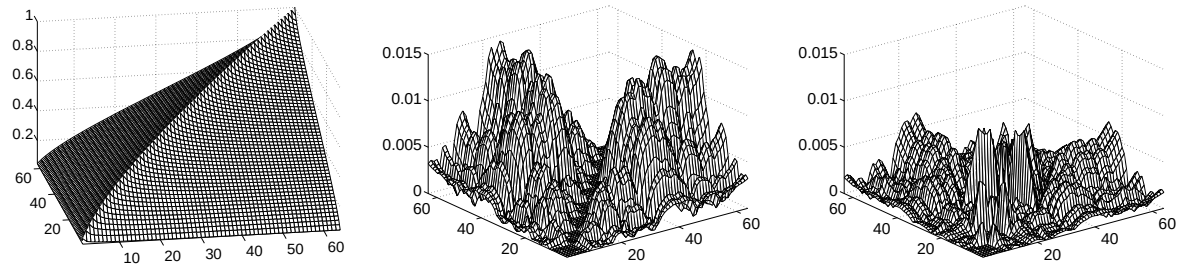

Figure 2: Figure (a) is P_f, the covariance matrix for 64 samples of fBm(0.3) on $(0, 1]$. The other figures illustrate $|P_f - P_M|$ where P_M is obtained from: (b) the method of [7]; (c) our approach. In both cases the maximum state dimension is four.

information subject to this row constraint. The drawback to this approach, which forms the engine of the method of [7], is that it is computationally intensive, requiring $O(n_2^3)$ operations.

Our approach[3] instead considers the mean square estimation error (MSEE) in estimating z_2 based on a r linear functions of z_1. The element of $\mathcal{M}_r$ which minimizes the MSEE is given by the first r rows of the matrix $U^T P_1^{-1/2}$ where $U\Lambda U^T$ is the eigen-decomposition of $P_1^{-1/2} P_{12} P_{12}^T P_1^{-T/2}$ with the eigenvalues forming the diagonal of Λ in decreasing order from upper-left to lower-right. The computational complexity of computing $U^T P_1^{-1/2}$ is $O(n_2)$. We will show in Section 5 that, in the context of the MAR realization problem, our MSEE approach provides decorrelating information comparable in quality to that provided by canonical correlations and with substantial savings in computational complexity.

Regardless of which method is used—canonical correlations or our MSEE-based approach—it can be generalized to the case of three random vectors (a case we will encounter in the realization problem on dyadic trees). To mutually conditionally decorrelate the three sub-vectors of $z = [z_1^T \; z_2^T \; z_3^T]^T$ we simply consider the two pair-wise problems of finding the best[4] $V_1 z_1$ to conditionally decorrelate z_1 and $[z_2^T \; z_3^T]^T$ and the best $V_2 z_2$ to conditionally decorrelate z_2 and $[z_1^T \; z_3^T]^T$. The stacked vector

$$\begin{bmatrix} V_1 z_1 \\ V_2 z_2 \end{bmatrix} \tag{9}$$

will exactly conditionally decorrelate z_1, z_2, and z_3 if no restrictions on the number of rows of V_1 and V_2 are applied. Approximate conditional decorrelation is obtained under a row constraint.

5. AN EFFICIENT REALIZATION ALGORITHM

It turns out that for internally consistent processes the Markov property is equivalent to the following: $x(s)$ at scale $n \neq M$ conditionally decorrelates $x(s\alpha_1)$, $x(s\alpha_2)$, and $\{x(t) \mid t \text{ is at scale } n+1, t \neq s\alpha_1, s\alpha_2\}$. I.e., this *scale-recursive Markov property* is equivalent to the Markov property defined in Section 2 for internally consistent processes.

In our realization approach we ensure internal consistency by explicitly forcing $x(s)$ to be a linear function of its children states. Therefore, it is sufficient to appeal to the simpler scale-recursive Markov property rather than to the Markov property. The algorithm proceeds as follows. For each scale $n = M-1, M-2, \ldots, 0$ and for each node s at scale n do the following:

1. For each $i = 1, 2$, find the vector $V_{s\alpha_i} x(s\alpha_i)$ which minimizes the MSEE in estimating all the other states at scale $n+1$ from $V_{s\alpha_i} x(s\alpha_i)$. This is a pair-wise problem as described in Section 4.
2. Form the stacked vector

$$\begin{bmatrix} V_{s\alpha_1} x(s\alpha_1) \\ V_{s\alpha_2} x(s\alpha_2) \end{bmatrix}. \tag{10}$$

   This stacked vector solves the three-way problem involving $x(s\alpha_1)$, $x(s\alpha_2)$ and all the other states at scale $n+1$ as described previously.
3. If the resulting state dimension is too large, delete the appropriate number of rows of the $V_{s\alpha_i}$ matrices. Because the rows of $V_{s\alpha_i}$ are in priority order, it is a simple matter to decide which information is least important.

Together the $\{V_{s\alpha_i}\}$ define the internal matrices. Notice that in the above algorithm each node is visited once and the complexity of the calculation at each node is $O(N)$ where N is the size of f. Therefore, the overall complexity is $O(N^2)$—constant computational complexity per element of the covariance matrix we are trying to realize, P_f.

6. EXAMPLES

In our first example we compare our approach to that described in [7] for the case where f consists of samples of fractional Brownian motion (as defined in [9]) with Hurst parameter $H = 0.3$ (denoted by fBm(0.3)). Figure 2(a) illustrates P_f, the covariance matrix for 64 equally spaced samples of fBm(0.3) on the interval $(0, 1]$. Figure 2(b) illustrates $|P_f - P_M|$ where $|\cdot|$ is element-wise absolute value and P_M is obtained by the method described in [7] based on canonical correlations. Figure 2(c) illustrates $|P_f - P_M|$ where P_M is obtained by our MSEE-based method. For both cases the maximum state dimension is four.

The main message of Figure 2 is that the two methods yield approximations to P_f of comparable quality. However, our approach is $O(N^2)$ while that of [7] is $O(N^4)$. Moreover, our approach yields an internally consistent model. While space constraints prevent us from providing additional examples, the one depicted in

[3] A similar approach applied to the classical stationary AR stochastic realization problem is discussed in [2].

[4] "Best" is defined by the approach used. Canonical correlations minimizes correlation coefficients while our approach minimizes MSEE.

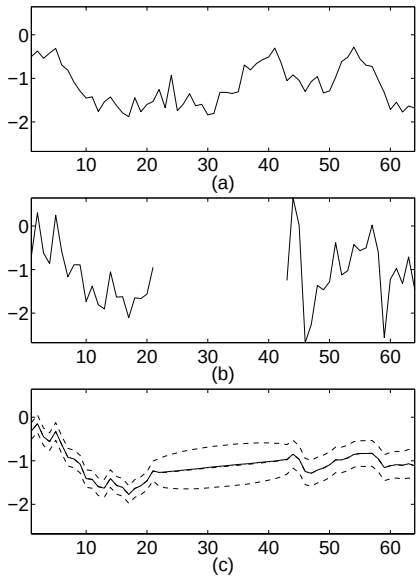

Figure 3: Figure (a) is a sample path of x_M generated by our MSEE-based model of fBm(0.3). Figure (b) are noisy measurements taken over the first and last third of the process in (a). The solid line of (c) are the MAR estimates which coincide with the optimal estimates based on the true fBm(0.3) statistics. The dotted lines indicate one standard deviation errors.

Figure 2 is typical of all the examples we have tried. In all cases our approach and that of [7] yield comparable results.[5]

In our second example we illustrate the problem of estimating an fBm(0.3) process from irregular measurements corrupted by non-stationary noise. We emphasize that this is a problem which *cannot* be handled with fast transform techniques due to the non-stationarity of the process to be estimated and the process noise and the irregularity of the measurements. Figure 3(a) is a sample path of x_M based on our MSEE-based realization of fBm(0.3) from the previous example. Figure 3(b) illustrates noisy irregular measurements of figure (a). Measurements are taken over the first and last third of the process. No measurements are available over the middle third. The white measurement noise has variance 0.1 over the first third sub-interval and 0.5 over the last third sub-interval. Figure 3(c) shows the output of the MAR estimator [3] (solid line) with one-standard-deviation error bars (dotted lines). The optimal estimate based on the exact fBm(0.3) statistics (rather than our approximate model of them) is also plotted in figure (c). However it is not distinguishable from the MAR estimate since the two nearly coincide. This demonstrates that the degree to which our MAR model deviates from the exact model is statistically irrelevant.

[5]We have tried fBm(H) for $0 < H < 1$, discrete fractional Gaussian noises, fractionally differenced Gaussian noises, reciprocal processes of several orders, a damped sinusoidal covariance function, and randomly chosen covariance matrices.

7. CONCLUSION

The contribution of this paper has been to provide a novel and computationally efficient realization algorithm for a recently introduced class of multiscale models. Our algorithm is quadratic in problem size while the only other known general purpose realization algorithm is quartic in problem size. In addition, our realized models are internally consistent.

It is worth emphasizing that the approach discussed here generalizes easily to a MAR process defined on an arbitrary tree (irregular or symmetric). In particular, we have also applied our approach to realization of quad-tree-based multiscale stochastic image models with great success. It is in this context that our approach, while fast, is still far too slow to handle large problems of interest (say, 1024×1024 images). Our future work will focus on approximate algorithms which are faster ($O(N)$ or $O(N \log(N))$, say).

8. REFERENCES

[1] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM Journal of Control*, 13(1):162–173, January 1975.

[2] K. S. Arun and S. Y. Kung. Balanced approximation of stochastic systems. *SIAM Journal of Matrix Analysis and Applications*, 11(1):42–68, January 1990.

[3] Kenneth C. Chou, Alan S. Willsky, and Albert Benveniste. Multiscale recursive estimation, data fusion, and regularization. *IEEE Transactions on Automatic Control*, 39(3):464–478, March 1994.

[4] Michael M. Daniel. *Multiresolution statistical modeling with application to modeling groundwater flow*. PhD thesis, Massachusetts Institute of Technology, February 1997.

[5] Paul W. Fieguth and Alan S. Willsky. Fractal estimation using models on multiscale trees. *IEEE Transactions on Signal Processing*, 44(5):1297–1300, May 1996.

[6] T. T. Ho, P. W. Fieguth, and A. S. Willsky. Multiresolution stochastic models for the efficient solution of large-scale space-time estimation problems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, volume 6, pages 3097–3100, Atlanta, Georgia, USA, May 1996.

[7] W. W. Irving, W. C. Karl, and A. S. Willsky. A theory for multiscale stochastic realization. In *Proceedings of the 33rd IEEE Conference on Decision and Control*, volume 1, pages 655–62, Lake Buena Vista, Florida, USA, December 1994.

[8] Mark R. Luettgen and Alan S. Willsky. Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination. *IEEE Transactions on Image Processing*, 4(2):194–207, February 1995.

[9] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.