Fast Adaptive Identification of Stable Innovation Filters

A. P. Mullhaupt and K. S. Riedel
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012-1185

Abstract

The adaptive identification of the impulse response of an innovation filter is considered. The impulse response is a finite sum of known basis functions with unknown coefficients. These unknown coefficients are estimated using a pseudolinear regression. This estimate is implemented using a square root algorithm based on a displacement rank structure. When the initial conditions have low displacement rank, the filter update is $\mathcal{O}(n)$. If the filter architecture is chosen to be triangular input balanced, the estimation problem is well-conditioned and a simple, low rank initialization is available.

1 INTRODUCTION

We consider innovation models for state space systems with an unknown and possibly time dependent impulse response. Innovation models use the prediction fit errors as the stochastic input into the state space evolution and the feedback/gain matrix is estimated empirically. We use a pseudolinear regression (PLR) [LS, Section 3.7.3] to identify the unknown coefficients of the state space impulse response. PLR is a least squares estimate in which the unknown coefficients are recursively updated from past estimates of the residuals. We refer the reader to [LS] for an excellent exposition of the theoretical properties of innovation models and PLR.

We show that innovation filter systems possess a displacement structure [KS]. Using this displacement structure, we construct a fast $\mathcal{O}(n)$ square root filter. Our square root displacement filter (SRDF) effects a time-dependent change of coordinates that preserves the impulse response while transforming the empirical covariance matrix to the identity. A canonical representation of the state space system is used to simplify the SRDF initialization and to improve the condition of the estimation of these parameters.

In Section 2, the innovations filter representation of the impulse response is given and pseudo-linear regression is used to estimate the unknown coefficients. Sections 3 and 4 present a fast square root version of the PLR estimate that is based on low displacement rank. Section 5 describes a new matrix canonical representation, triangular input balanced (TIB) form and its applications to filter architecture. When the system advance matrix is in TIB form, the system is easily initialized and is always well-conditioned. Sections 6-7 discuss and summarize our results. Except where explicitly noted, our results apply for the multiple input multiple output (MIMO) case.

2 INNOVATIONS FILTER FOR SYSTEM IDENTIFICATION

Let $\mathbf{y}_t$ be a sequence of d-dimensional measurement vectors. We consider the system identification problem of associating to $\mathbf{y}_t$ an unknown multi-input/multi-output (MIMO) time invariant linear system of the form:

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{b}\,\boldsymbol{\epsilon}_t, \tag{2.1}$$

$$\mathbf{y}_t = \mathbf{c}^*\mathbf{z}_t + \boldsymbol{\epsilon}_t, \tag{2.2}$$

where $*$ denotes the Hermitian transpose. Here $\mathbf{z}_t$ is the n-dimensional state vector, $\mathbf{A}$ is the n-dimensional system matrix, and $\mathbf{b}$ and $\mathbf{c}$ are $n \times d$-dimensional matrices. Equations (2.1) and (2.2) are in *innovations model/ prediction error model* form, i.e. the measurement noise and the system noise are completely correlated via the d-dimensional innovations vector $\boldsymbol{\epsilon}_t$. Every stable linear time-invariant n-dimensional state space system may be represented in this form [HD].

Our goal is to identify the underlying impulse response of the system. It makes sense to assume that $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ is *minimal*, but instead we will simply assume that it is *controllable*. In addition, we will assume that $\mathbf{A}$ is stable and, to reduce the complexity of exposition, *nonsingular*. This implies that $(\mathbf{A}, \mathbf{b})$ is completely realizable.

The impulse response is preserved by the change of state space coordinates:

$$\mathbf{A} \rightarrow \mathbf{TAT}^{-1}, \quad \mathbf{b} \rightarrow \mathbf{Tb}, \quad \mathbf{c} \rightarrow \mathbf{T}^{-*}\mathbf{c}, \tag{2.3}$$

where $\mathbf{T}$ is an arbitrary nonsingular $n \times n$ matrix. Since the impulse response, $\mathbf{h}_j \equiv \mathbf{c}^*\mathbf{A}^j\mathbf{b}$, is preserved by this change of coordinates, we may choose an observationally equivalent state space system in which the system advance has desirable properties. In particular, by an appropriate choice of $\mathbf{T}$, we can represent the system in TIB form. (See Sec. 5.)

In practice, we prescribe the eigenvalues of $\mathbf{A}$ to correspond to the characteristic response times of the system. We then choose $(\mathbf{A}, \mathbf{b})$ to be in TIB form. This uniquely specifies $(\mathbf{A}, \mathbf{b})$ up to the transformation group: $(\mathbf{A}, \mathbf{b} \leftarrow (\Theta\mathbf{A}\Theta^*, \Theta\mathbf{b}\mathbf{U})$, where $\mathbf{U}$ is a $d \times d$ unitary transformation and Θ is a $n \times n$ *diagonal* unitary transformation. Thus, when the eigenvalues are prescribed, only $\mathbf{c}$ need be estimated to identify the impulse response.

At each time step, we observe $\mathbf{y}_t$, and estimate the coefficients using the PLR of [LS]. We denote the estimate of $\mathbf{z}_t$ and $\mathbf{c}$ using data up to and including $\mathbf{y}_t$ by $\hat{\mathbf{z}}_t$ and $\hat{\mathbf{c}}_t$. The predictive residual error, $\hat{\boldsymbol{\epsilon}}_t = y_t - \hat{\mathbf{c}}_{t-1}^*\hat{\mathbf{z}}_t$, is evaluated, where $\hat{\mathbf{z}}_t$ is the estimate of $\mathbf{z}_t$ given by using the time advance equation, (2.1) with the substitutions, $\mathbf{z}_t \leftarrow \hat{\mathbf{z}}_t$, $\boldsymbol{\epsilon}_t \leftarrow \hat{\boldsymbol{\epsilon}}_t$ and $\mathbf{c} \leftarrow \hat{\mathbf{c}}_{t-1}$. Let δ be a 'forgetting factor' with $0 < \delta \leq 1$ and $1 - \delta << 1$. We define the (weighted) empirical covariance $\hat{\mathbf{P}}_t^\delta$ and the empirical cross-covariance $\mathbf{d}_t^\delta$:

$$\hat{\mathbf{P}}_t^\delta \equiv \delta^t\hat{\mathbf{P}}_0^\delta \ + \ \sum_{k=1}^{t} \delta^{t-k}\hat{\mathbf{z}}_k\hat{\mathbf{z}}_k^*, \quad \mathbf{d}_t^\delta \equiv \delta^t\mathbf{d}_0^\delta \ + \sum_{k=1}^{t} \delta^{t-k}\hat{\mathbf{z}}_k\mathbf{y}_k^*, \tag{2.4}$$

where $\hat{\mathbf{P}}_0^\delta$ is the initial covariance, and the initial cross-covariance is $\mathbf{d}_0^\delta = \hat{\mathbf{P}}_0^\delta\hat{\mathbf{c}}_0$. The PLR estimate of $\mathbf{c}$ is $\hat{\mathbf{c}}_t = \left[\hat{\mathbf{P}}_t^\delta\right]^{-1}\mathbf{d}_t^\delta$. Note that the value of $\mathbf{y}_t$ is used in the estimate $\hat{\mathbf{c}}_t$. In summary, the PLR estimate of (2.1)-(2.2) given the measurements $\{\mathbf{y}_s | 1 \leq s \leq t\}$ is

$$\hat{\mathbf{z}}_{t+1} = \mathbf{A}\hat{\mathbf{z}}_t + \mathbf{b}\hat{\boldsymbol{\epsilon}}_t, \tag{2.5}$$

$$\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_t - \hat{\mathbf{c}}_t^*\hat{\mathbf{z}}_t, \tag{2.6}$$

$$\hat{\mathbf{c}}_t = \left[\hat{\mathbf{P}}_t^\delta\right]^{-1}\mathbf{d}_t^\delta \ . \tag{2.7}$$

In the "pre-windowed" case, the filter is initialized as $\hat{\mathbf{z}}_1 \equiv \hat{\mathbf{z}}_{t=1} = \hat{\mathbf{c}}_{t=0} = 0$. In order to *predict* $\mathbf{y}_{t+1}$, we may use $\hat{\mathbf{y}}_{t+1|t} = \hat{\mathbf{c}}_t^*\hat{\mathbf{z}}_{t+1}$ and $\hat{\mathbf{z}}_{t+1} = \mathbf{A}\hat{\mathbf{z}}_t + \mathbf{b}\left(\mathbf{y}_t - \hat{\mathbf{c}}_{t-1}^*\hat{\mathbf{z}}_t\right)$. The empirical covariances satisfy

$$\hat{\mathbf{P}}_t^\delta = \delta\hat{\mathbf{P}}_{t-1}^\delta \ + \ \hat{\mathbf{z}}_t\hat{\mathbf{z}}_t^*, \quad \mathbf{d}_t^\delta = \delta\mathbf{d}_{t-1}^\delta + \hat{\mathbf{z}}_t\mathbf{y}_t^*. \tag{2.8}$$

Let $\hat{\mathbf{P}}_0^\delta$ be invertible and define $\mathbf{\Phi}_t = \left[\hat{\mathbf{P}}_t^\delta\right]^{-1}$. By the matrix inversion identity, $\mathbf{\Phi}_t = \delta^{-1}\left[\mathbf{\Phi}_{t-1} - \frac{\mathbf{\Phi}_{t-1}\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t^*\mathbf{\Phi}_{t-1}^*}{\delta+\hat{\mathbf{z}}_t^*\mathbf{\Phi}_{t-1}\hat{\mathbf{z}}_t}\right]$, and the PLR update for $\hat{\mathbf{c}}_t$ is

$$\hat{\mathbf{c}}_t = \hat{\mathbf{c}}_{t-1} - \frac{\mathbf{\Phi}_{t-1}\hat{\mathbf{z}}_t\left(\mathbf{y}_t - \hat{\mathbf{c}}_{t-1}^*\hat{\mathbf{z}}_t\right)^*}{\delta + \hat{\mathbf{z}}_t^*\mathbf{\Phi}_{t-1}\hat{\mathbf{z}}_t}. \tag{2.9}$$

The PLR estimate determines $\hat{\mathbf{c}}_t$ in terms of the previous estimates of the prediction errors $\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_t - \hat{\mathbf{c}}_{t-1}^*\hat{\mathbf{z}}_t$. For large times, $t > 1/(1-\delta)$, the expectation of empirical covariance tends to $\mathrm{E}[\hat{\mathbf{P}}_t^\delta] = \mathbf{P}_\infty/(1-\delta)$, where $\mathbf{P}_\infty$ satisfies Stein's equation [LT]:

$$\mathbf{P}_\infty - \mathbf{A}\mathbf{P}_\infty\mathbf{A}^* = \sigma^2\mathbf{b}\mathbf{b}^* \ . \tag{2.10}$$

3 DISPLACEMENT STRUCTURE of the COVARIANCE

In Section 4, we describe a fast square root algorithm for computing $\hat{\mathbf{c}}_t$. The fast algorithm is based on the *displacement structure* of $\hat{\mathbf{P}}_t^\delta$. This is similar to, but not the same as, the fast update of Sayed and Kailath [SK1]. Here, we compute a generator and its signature for the displacement of the empirical covariance, $\hat{\mathbf{P}}_t^\delta - \delta^{-1}\mathbf{A}\hat{\mathbf{P}}_t^\delta\mathbf{A}^*$, this determines an upper bound on the rank of minimal generators for this displacement. Any matrix $\mathbf{X}_t$ such that

$$\hat{\mathbf{P}}_t^\delta - \delta^{-1}\mathbf{A}\hat{\mathbf{P}}_t^\delta\mathbf{A}^* = \mathbf{X}_t\mathbf{S}\mathbf{X}_t^*. \tag{3.1}$$

is called a generator of the displacement, and $\mathbf{S}$ is a corresponding signature. We assume that a generator $\mathbf{X}_0$ and signature $\mathbf{S}_0$ for the initial covariance, $\mathbf{X}_0\mathbf{S}_0\mathbf{X}_0^* = \hat{\mathbf{P}}_0^\delta - \delta^{-1}\mathbf{A}\hat{\mathbf{P}}_0^\delta\mathbf{A}^*$, have been computed from $\hat{\mathbf{P}}_0^\delta$ by the singular value decomposition or by more specialized methods [SK1, SK2], see also Section 5. By substituting (2.4) into (3.1) and collapsing the summation using (2.1), we have

$$\begin{aligned}
\mathbf{X}_t\mathbf{S}\mathbf{X}_t^* &= -\delta^{-1}\mathbf{A}\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t^*\mathbf{A}^* + \delta^{t-1}\hat{\mathbf{z}}_1\hat{\mathbf{z}}_1^* + \delta^t\mathbf{X}_0\mathbf{S}\mathbf{X}_0^* + \sum_{j=1}^{t-1}\delta^{t-j-1}\left[\mathbf{b}\hat{\boldsymbol{\epsilon}}_j\hat{\mathbf{z}}_j^*\mathbf{A}^* + \mathbf{A}\hat{\mathbf{z}}_j\hat{\boldsymbol{\epsilon}}_j^*\mathbf{b}^* + \mathbf{b}\hat{\boldsymbol{\epsilon}}_j\hat{\boldsymbol{\epsilon}}_j^*\mathbf{b}^*\right] \\
&= -\delta^{-1}\mathbf{A}\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t^*\mathbf{A}^* + \delta^{t-1}\hat{\mathbf{z}}_1\hat{\mathbf{z}}_1^* + \hat{\mathbf{g}}_{t-1}\hat{\mathbf{g}}_{t-1}^* - \hat{\mathbf{h}}_{t-1}\hat{\mathbf{h}}_{t-1}^* + \delta^t\mathbf{X}_0\mathbf{S}_0\mathbf{X}_0^* \ ,
\end{aligned} \tag{3.2}$$

where we have defined the $n \times d$-dimensional matrices:

$$\hat{\mathbf{f}}_t \equiv \sum_{j=1}^{t}\delta^{t-j}\left(\mathbf{A}\hat{\mathbf{z}}_j + \frac{1}{2}\mathbf{b}\hat{\boldsymbol{\epsilon}}_j\right)\hat{\boldsymbol{\epsilon}}_j^*, \quad \hat{\mathbf{g}}_t \equiv 2^{-1/2}\left(\hat{\mathbf{f}}_t + \mathbf{b}\right), \quad \hat{\mathbf{h}}_t \equiv 2^{-1/2}\left(\hat{\mathbf{f}}_t - \mathbf{b}\right). \tag{3.3}$$

Let $\mathbb{I}_d$ be the d-dimensional identity matrix. We define the signature matrix $\mathbf{S} = (-1) \oplus 1 \oplus \mathbb{I}_d \oplus (-\mathbb{I}_d) \oplus \mathbf{S}_0$, and define the *displacement rank*, $\alpha \equiv \text{rank}(\mathbf{S})$. One particular choice of generator is

$$\mathbf{X}_t = \left(\delta^{-1/2} \mathbf{A} \hat{\mathbf{z}}_t | \delta^{(t-1)/2} \hat{\mathbf{z}}_1 | \hat{\mathbf{g}}_{t-1} | \hat{\mathbf{h}}_{t-1} | \delta^{t/2} \mathbf{X}_0 \right) . \tag{3.4}$$

This shows that $\mathbf{X}_t$ may be chosen with $\text{rank}(\mathbf{X}_t) \leq \alpha \equiv \text{rank}(\mathbf{S})$. Note that α depends on the initialization of the filter and it is desirable that the initial displacement have have minimal rank. The case $\hat{\mathbf{z}}_1 = \mathbf{0}$ is called the *prewindowed* case. As described in Section 5, we recommend choosing the initial conditions such that $\text{rank}(\mathbf{X}_0) = 1$.

¿From (2.8), any pair of generators, $\mathbf{X}_t$ and $\mathbf{X}_{t+1}$, satisfy the update equation:

$$\mathbf{X}_{t+1} \mathbf{S} \mathbf{X}^*_{t+1} = \delta \mathbf{X}_t \mathbf{S} \mathbf{X}^*_t + \hat{\mathbf{z}}_{t+1} \hat{\mathbf{z}}^*_{t+1} - \mathbf{A} \hat{\mathbf{z}}_{t+1} \hat{\mathbf{z}}^*_{t+1} \mathbf{A}^* . \tag{3.5}$$

To obtain a minimal rank generator of the displacement, we perform an eigenvector decomposition (EVD) of the updated displacement:

$$\mathbf{U}_{t+1} \Lambda_{t+1} \mathbf{U}^*_{t+1} = \delta \mathbf{X}_t \mathbf{S} \mathbf{X}^*_t + \hat{\mathbf{z}}_{t+1} \hat{\mathbf{z}}^*_{t+1} - \mathbf{A} \hat{\mathbf{z}}_{t+1} \hat{\mathbf{z}}^*_{t+1} \mathbf{A}^* , \tag{3.6}$$

where $\mathbf{U}_{t+1}$ is unitary and Λ_{t+1} is diagonal. We choose $\mathbf{X}_{t+1} = \mathbf{U}_{t+1} |\Lambda|^{1/2}_{t+1}$ where the eigenvalues of Λ_{t+1} are ordered to respect the signature matrix, $\mathbf{S}$. We refer to this choice of generator as the $EVD-induced$ generator. Given the EVD of $\mathbf{X}_t \mathbf{S} \mathbf{X}^*_t$, the EVD of (3.6) may be computed in $\mathcal{O}(\alpha^2 n)$ operations using the update/downdate algorithm of [GE]. If $\mathbf{A}$ is a structured matrix such that $\mathbf{A}\mathbf{z}_t$ is computable in $\mathcal{O}(n)$ operations, then the displacement generator, $\mathbf{X}_t$, can be updated in $\mathcal{O}(\alpha^2 n)$ operations.

The EVD update of $\mathbf{X}_{t+1}$ may be replaced by any alternative update of the generator which is computable in $\mathcal{O}(\alpha^2 n)$ flops and which guarantees that rank $\mathbf{X}_{t+1} \leq \alpha$. We recommend the EVD update of $\mathbf{X}_{t+1}$ to minimize $\|\mathbf{X}_{t+1}\|$ and preserve the orthogonality of the columns of $\mathbf{X}_{t+1}$. The EVD update/downdate of [GE] is backwards stable and may be computed to machine precision.

4 SQUARE ROOT DISPLACEMENT FILTER

We present a square root displacement version of the PLR update of $\mathbf{c}_t$ when $\mathbf{A}$ is upper triangular (UT). The lower triangular case is similar. The standard Cholesky update of $\mathbf{c}_t$ using (2.9) for a system requires $\mathcal{O}(n^2 d)$ time and $\mathcal{O}(n^2)$ space, while our version has $\mathcal{O}(n\alpha^2)$ time updates. However, the straightforward initialization of the SRDF is computationally intensive, $\mathcal{O}(n^3)$, in the general case. In the next section, we describe a novel filter architecture that can be initialized in $\mathcal{O}(\alpha^2 n)$ operations and that is always well-conditioned.

We define $\mathbf{R}_t$ and $\mathbf{W}_t$ by

$$\hat{\mathbf{P}}_t^\delta = \mathbf{R}_t \mathbf{R}_t^*, \quad \mathbf{W}_t = \mathbf{R}_t^{-1} \mathbf{R}_{t-1}, \tag{4.1}$$

where $\mathbf{R}_t$ is *upper triangular* and $R_{t,jj} > 0$ for $1 \le j \le n$. We make the time dependent change of coordinates: $\mathbf{u}_t = \mathbf{R}_{t-1}^{-1} \hat{\mathbf{z}}_t$, $\acute{\mathbf{c}}_t = \mathbf{R}_{t-1}^* \hat{\mathbf{c}}_t$. The transformed state space model is

$$\mathbf{u}_{t+1} = \left(\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t\right) \mathbf{W}_t \mathbf{u}_t + \mathbf{R}_t^{-1} \mathbf{b} \hat{\boldsymbol{\epsilon}}_t, \tag{4.2}$$

$$\mathbf{y}_t = \acute{\mathbf{c}}_t^* \mathbf{u}_t + \hat{\boldsymbol{\epsilon}}_t . \tag{4.3}$$

By hypothesis, $\mathbf{R}_t$ and $\mathbf{A}$ are UT, $\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t$ is UT and the diagonal elements of $\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t$ match those of $\mathbf{A}$: $\left(\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t\right)_{jj} = \mathbf{A}_{jj}$. The coefficient update is

$$\acute{\mathbf{c}}_t = \mathbf{W}_{t-1}^{-*} \acute{\mathbf{c}}_{t-1} + \frac{\mathbf{u}_t \left(\mathbf{y}_t - \acute{\mathbf{c}}_{t-1}^* \mathbf{u}_t\right)^*}{\delta + \mathbf{u}_t^* \mathbf{u}_t}. \tag{4.4}$$

These updates can be carried out in $\mathcal{O}(n)$ time and space as follows:

Conjugating equation (2.8) by $\mathbf{R}_{t-1}^{-1}$ yields

$$\mathbf{W}_t^* \mathbf{W}_t = \delta^{-1} \left(\mathbb{I} - \frac{\mathbf{u}_t \mathbf{u}_t^*}{\delta + \mathbf{u}_t^* \mathbf{u}_t}\right) , \quad \mathbf{W}_t^{-1} \mathbf{W}_t^{-*} = \delta \mathbb{I} + \mathbf{u}_t \mathbf{u}_t^* , \tag{4.5}$$

where $\mathbb{I}$ is the $n \times n$ identity matrix. Since $\mathbf{W}_t$ is UT with positive diagonal elements, it is a TIB matrix in the sense of Sec. 5 and it is uniquely determined by (4.5). We define $\boldsymbol{\beta}_t = \mathbf{R}_t^{-1} \mathbf{b}$ and update it by $\boldsymbol{\beta}_t = \mathbf{W}_t \boldsymbol{\beta}_{t-1}$. Conjugating equation (3.1) by $\mathbf{R}_t^{-1}$ and defining $\mathbf{Y}_t = \mathbf{R}_t^{-1} \mathbf{X}_t$ yields

$$\left(\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t\right) \left(\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t\right)^* = \delta \left(\mathbb{I} - \mathbf{Y}_t \mathbf{S} \mathbf{Y}_t^*\right). \tag{4.6}$$

Given $\mathbf{Y}_t$ and the diagonal elements of $\mathbf{A}$ (or at least the complex phases of the diagonal elements), $\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t$ is uniquely determined by (4.6). In the appendix, we demonstrate that a representation of $\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t$ can be calculated from $\mathbf{Y}_t$ in $\mathcal{O}(\alpha^2 n)$. The resulting representation of $\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t$ allows the matrix vector product, $\left(\mathbf{R}_t^{-1} \mathbf{A} \mathbf{R}_t\right) \mathbf{v}$, to be computed in $\mathcal{O}(\alpha n)$ time, where $\mathbf{v}$ is an arbitrary n-vector.

Conjugating (3.5) by $\mathbf{R}_t^{-1}$ yields the displacement update in the transformed coordinates:

$$\mathbf{Y}_{t+1} \mathbf{S} \mathbf{Y}_{t+1}^* = \delta \mathbf{W}_{t+1} \mathbf{Y}_t \mathbf{S} \mathbf{Y}_t^* \mathbf{W}_{t+1}^* + \mathbf{W}_{t+1} \mathbf{u}_{t+1} \mathbf{u}_{t+1}^* \mathbf{W}_{t+1}^* - \delta^{-1} \ \boldsymbol{\mu}_{t+1} \ \boldsymbol{\mu}_{t+1}^* , \tag{4.7}$$

where $\boldsymbol{\mu}_{t+1} \equiv \mathbf{W}_{t+1}\tilde{\mathbf{A}}_t\mathbf{u}_{t+1}$ with $\tilde{\mathbf{A}}_t \equiv \mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$, To enhance the numerical stability, we choose the EVD-induced generator: $\mathbf{Y}_t = \mathbf{V}_t|\mathbf{D}_t|^{1/2}$ where $\mathbf{V}_t$ is unitary and the eigendecomposition of the displacement is $\mathbf{Y}_t\mathbf{S}\mathbf{Y}_t^* = \mathbf{V}_t\mathbf{D}_t\mathbf{V}_t^*$. Similar to (3.6), we compute $\mathbf{Y}_{t+1}$ by updating the EVD of (4.7). We claim that $\mathbf{W}_t = \delta^{-1/2}\left(\mathbb{I} - \gamma_t\mathbf{u}_t\mathbf{u}_t^*\right)\mathbf{Q}_t$, where $\mathbf{Q}_t$ is an unitary matrix and γ_t satisfy $2\gamma_t - \gamma_t^2\|\mathbf{u}_t\|^2 = 1/\left(\delta + \|\mathbf{u}_t\|^2\right)$. (Proof: Since $\delta^{-1/2}\left(\mathbb{I} - \gamma_t\mathbf{u}_t\mathbf{u}_t^*\right)$ and $\mathbf{W}_t$ are both square roots of the same matrix defined in (4.5), $\mathbf{Q}_t$ is an unitary matrix.) Thus $\mathbf{W}_{t+1}\mathbf{Y}_t = \delta^{-1/2}\left(\mathbb{I} - \gamma_{t+1}\mathbf{u}_{t+1}\mathbf{u}_{t+1}^*\right)\mathbf{Q}_{t+1}\mathbf{Y}_t$. The EVD of $\mathbf{W}_{t+1}\mathbf{Y}_t\mathbf{S}\mathbf{Y}_t^*\mathbf{W}_{t+1}^*$ is a rank two perturbation of the known EVD of $\mathbf{Q}_{t+1}\mathbf{V}_t\mathbf{D}_t\mathbf{V}_t^*\mathbf{Q}_{t+1}^*$. Since $\mathbf{W}_{t+1}$ is a rank-one TIB matrix, $\mathbf{Q}_{t+1}\mathbf{V}_t$ may be rapidly evaluated in in $\mathcal{O}(\alpha n)$ operations as $\delta^{1/2}\left(\mathbb{I} - \gamma_{t+1}\mathbf{u}_{t+1}\mathbf{u}_{t+1}^*\right)^{-1}\mathbf{W}_{t+1}\mathbf{U}_t$. Thus the eigendecomposition of (4.7) is a rank four perturbation of the known eigenvectors, $\mathbf{Q}_{t+1}\mathbf{V}_t$ and may be evaluated in $\mathcal{O}(\alpha^2 n)$ operations (or fewer if the fast multipole algorithm is used) using the numerically stable algorithm of [GE]. The eigendecomposition-induced update minimizes the norm of $\mathbf{Y}_t$ and ensures that(4.6) has a positive semi-definite solution.

Although the time update of the SRDF is $\mathcal{O}\left(\alpha^2 n\right)$, the straightforward initialization requires $\mathcal{O}\left(n^3\right)$ operations to compute the initial value of $\mathbf{Y}_t$. The initialization reduces an arbitrary innovations filter to an equivalent problem of displacement rank α. Equation (3.2) shows that the mapping preserves this low displacement rank as the filter evolves. Initialization algorithms for related displacement structures are described in [SK1, SK2]. To simplify the initialization, we select the initial covariance, $\hat{\mathbf{P}}_{t=0}^{\delta}$, such that

$$\hat{\mathbf{P}}_0^{\delta} - \delta^{-1}\mathbf{A}\hat{\mathbf{P}}_0^{\delta}\mathbf{A}^* = \rho^2\mathbf{b}\mathbf{b}^* \ , \tag{4.8}$$

where ρ is a scalar free parameter which determines the size of the initial covariance. The resulting SRDF initialization is $\mathbf{Y}_0 = \rho\mathbf{b}$. Although $\mathbf{Y}_0$ does not require the solution of (4.8), the solution of (4.8) is required for $\boldsymbol{\beta}_0 = \mathbf{R}_0^{-1}\mathbf{b}$ and $\mathbf{u}_1 = \mathbf{R}_0^{-1}\mathbf{z}_1$.

Given the cost of the initialization, there are three cases where the SRDF will be of value: 1) when the number of time steps is large relative to n^2; 2) when the initialization may be done off-line and real time performance is important; 3) when the state space model has a special structure which simplifies the initialization. In the next section, we concentrate on filter architectures which simplify the initialization.

5 TRIANGULAR INPUT BALANCED FORM

The initialization of the square root displacement filter may be simplified by choosing the state space coordinates so that the system is in triangular input balanced (TIB) form. As described in [MR], the system, $(\mathbf{A}, \mathbf{b})$, is in UTIB form if

$$\mathbb{I} - \mathbf{A}\mathbf{A}^* = \frac{\sigma^2}{r}\mathbf{b}\mathbf{b}^* \ , \tag{5.1}$$

with $\mathbf{A}$ *upper triangular* and $\sigma^2/r > 0$ arbitrary subject to $\sigma^2\|\mathbf{b}\,\|^2/r \leq 1$. TIB systems have a number of advantages [MR]: First, the appendix shows how to construct a sequence of rank-one TIB systems $\mathbf{F}_k$ such $\mathbf{A} = \mathbf{F}_1 \dots \mathbf{F}_d$ in $\mathcal{O}\,(nd)$ operations. Given this representation, multiplying $\mathbf{A}$ by a vector or solving a matrix system is $\mathcal{O}\,(nd)$. Second, the solution of Stein's equation, (2.10), is a multiple of the identity, $\mathbf{P}_\infty = r\mathbb{I}$. Third, the solution of (4.8) is $\hat{\mathbf{P}}^\delta_{t=0} = (\mu^2 r/\sigma^2)\mathbb{I} \ + \ \mathcal{O}\,((1-\delta))$. This expansion allows a fast initialization via (4.8) if we drop the $\mathcal{O}\,((1-\delta))$ correction. Fourth, the expectation of empirical covariance tends to $\mathrm{E}[\hat{\mathbf{P}}^\delta_{t\to\infty}] = \mathbf{P}_\infty/(1-\delta) = r\mathbb{I}/(1-\delta)$. Thus, the identification of $\hat{\mathbf{c}}_t$ becomes well-conditioned asymptotically.

Theorem 1 [MR]: Every exponentially asymptotically stable, completely realizable system $(\mathbf{A}, \mathbf{b})$ is similar to a UTIB system $(\mathbf{A}', \mathbf{b}')$, in which the diagonal elements of $\mathbf{A}'$ can be chosen in any order.

The impulse response , $\mathbf{h}_j \equiv \mathbf{c}^*\mathbf{A}^j\mathbf{b}$, of the PLR innovations filter is a linear combination of the n rows of the $n \times \infty$ matrix $\mathbf{M} \equiv [\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, ...]$. For TIB systems, the n basis functions, $M_{i,\cdot}$, of the impulse response are orthonormal:

$$\mathbf{M}\mathbf{M}^* = \sum_{j=0}^{\infty} \mathbf{A}^j\mathbf{b}\mathbf{b}^*\mathbf{A}^{*j} \ = \ \frac{1}{\sigma^2}\mathbf{P}_\infty = r\mathbb{I}. \tag{5.2}$$

For SISO systems, when the TIB form $\mathbf{A}$ has only one eigenvalue (and therefore only one Jordan block), the rows of $\mathbf{M}$ are given by orthonormal functions of the form $\lambda^t p_k\,(t)$. The resulting polynomials, p_k, are the *Meixner* polynomials [NSU], a discrete analog of the Laguerre polynomials. This orthonormality improves the conditioning of the estimates of the coefficients, $\hat{\mathbf{c}}_t$.

Given the advantages of TIB form, we recommend choosing the TIB architecture when the filter designer knows the eigenvalues of the system response and is not constrained to a particular state matrix. The following theorem provides an explicit construction of TIB forms for the case of a SISO system:

Theorem 2[MR, GGMS]: Suppose $(\mathbf{A}, \mathbf{b})$ are UTIB and $\mathbf{A}$ is nonsingular then $\mathbf{A}_{jj} = \lambda_j$ and $\mathbf{A}_{ij} = g_i b_j$ for $i < j$ and $1 \leq i, j \leq n$, where $\lambda_j = |\lambda_j| e^{i\theta_j}$, $|\lambda_j|^2 = 1 + \alpha_j |b_j|^2$ and $g_j = e^{i\theta_j} \alpha_j \bar{b} |\lambda_j|^{-1}$. The α_j satisfy $\alpha_{j+1} = \alpha_j |\lambda_j|^{-1}$, $\alpha_j = \alpha_1 - \sum_{k=1}^{j-1} |g|_k^2$ and $\alpha_1 = -1$.

This theorem enables the user to specify the eigenvalues of $\mathbf{A}$ and choose the remainder of the filter architecture such that $\mathbf{A}$ is TIB. An alternative construction of the TIB form specifies the diagonal entries of $\mathbf{A}$ and then determines a sequence of n hyperbolic Givens rotations [GL] so that $\mathbf{A}$ satisfies (5.1) [MR]. This factorization can be obtained from a Potapov cascade embedding as in [LK]. We anticipate that the TIB architecture will have many additional applications in filter design. In our applications, our goal is to identify the impulse response, and the eigenvalues of $\mathbf{A}$ simply determine a basis for modeling the impulse response. Thus the filter designer need only choose the eigenvalues accurately enough to capture the characteristic time scales of the impulse response.

6 LMS and CHANDRASEKHAR FILTERS

The least mean square error (LMS) estimate replaces Φ_t by the identity matrix in (2.9): $\hat{\mathbf{c}}_t = \hat{\mathbf{c}}_{t-1} - \mu_t \mathbf{z}_t \ \boldsymbol{\epsilon}_t^*$. The performance of the LMS algorithm depends strongly on the condition number of the state covariance. If $\mathbf{A}$ is a TIB filter, then $E[\Phi_t]$ tends to a multiple of the identity matrix plus a random fluctuation of relative size $(1-\delta)/(1-\delta\|\mathbf{A}\|^2)$. Thus the LMS update sacrifices little accuracy since the second order correction is small. Thus the TIB architecture allows both the PLR and LMS updates to perform comparably with the advantages of both filters: well-conditioning, (near) second order accuracy, and $\mathcal{O}(n)$ time advances.

Assuming that the innovations were zero, the fast filter of Sayed and Kailath [SK1] can update the displacement of $\hat{\mathbf{P}}_{t+1}^{\delta} - \mathbf{A}\hat{\mathbf{P}}_t^{\delta}\mathbf{A}^*$. In contrast, our fast filters are based on updates to the displacement $\hat{\mathbf{P}}_t^{\delta} - \delta^{-1}\mathbf{A}\hat{\mathbf{P}}_t^{\delta}\mathbf{A}^*$. The two displacements are related through (3.1). If the innovations $\boldsymbol{\epsilon}_t$ were zero, then our filter would satisfy the structural conditions of [SK1]. The presence of the innovation, $\mathbf{b}\boldsymbol{\epsilon}_t$, in the state update introduces an additional low rank term. We have not compared our fast filter with the generalized Chandrasekhar filter of [SK1] since we do not know how to treat the innovations $\boldsymbol{\epsilon}_t\mathbf{b}$ in the generalized Chandrasekhar filter.

Chandrasekhar filters are commonly used in conjunction with finite impulse response (FIR) (time lag) models. FIR models have the disadvantage that the characteristic time of the impulse response is bounded by the number of time lags/free parameters. As the duration of the impulse response becomes longer, more time lag coefficients need to be identified. In contrast, the innovation filters (2.1) have the impulse response duration controlled by the eigenvalues of $\mathbf{A}$ while the number of free parameters is chosen by $\dim(\mathbf{A})$. This freedom allows the filter designer to specify the characteristic time scales of the impulse response and then to optimize the number of terms in the orthonormal basis expansion of the impulse response.

7 SUMMARY

We estimate the coefficients of the impulse response of the innovations form of a state space system using pseudolinear regression. In our case, this is identical to using recursive least squares when the innovations are estimated sequentially and not updated. The key advantage of adaptive estimation using the innovations filter is that the estimation of $\mathbf{c}_t$ is a linear problem.

The state space system has small displacement rank if the initial system does. The coefficient estimation is implemented with a square root displacement filter. The square root displacement filter constructs a state space coordinate change at each time step to approximate TIB form. The square root displacement filter depends on the state matrix, $\mathbf{A}$, through the current values of $\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$ and $\mathbf{u}_t$ or equivalently $\mathbf{Y}_t$ and $\mathbf{u}_t$. The time advance requires $\mathcal{O}(\alpha^2 n)$ operations. The difficulty in applying the SRDF is that the starting values of $\mathbf{Y}_t$ and $\mathbf{u}_t$ are expensive to compute for an arbitrary innovations filter. When the state system is in triangular input balanced form, an approximate initialization is available in $\mathcal{O}(n)$. If the filter designer knows the eigenvalues of $\mathbf{A}$, Theorem 2 constructs the TIB form for SISO systems. This corresponds to expanding the impulse response in an orthonormal basis of exponential decaying polynomials, thereby improving the conditioning of the coefficient estimate, $\hat{\mathbf{c}}_t$. In practice, the eigenvalues of $\mathbf{A}$ need not be known exactly, but only well enough that the series representation of the impulse response is reasonable.

The fast square root displacement PLR filter is:

$$\hat{\boldsymbol{\epsilon}}_t = y_t - \hat{\mathbf{c}}_{t-1}^* \mathbf{u}_t \tag{7.1}$$

$$\hat{\mathbf{c}}_t = \mathbf{W}_{t-1}^{-*} \hat{\mathbf{c}}_{t-1} + \mathbf{u}_t \frac{\hat{\boldsymbol{\epsilon}}_t}{\delta + \mathbf{u}_t^* \mathbf{u}_t} \tag{7.2}$$

$$\boldsymbol{\beta}_t = \mathbf{W}_t \boldsymbol{\beta}_{t-1} \tag{7.3}$$

$$\mathbf{u}_{t+1} = \tilde{\mathbf{A}}_t \mathbf{W}_t \mathbf{u}_t + \boldsymbol{\beta}_t \hat{\boldsymbol{\epsilon}}_t \tag{7.4}$$

$$\mathbf{W}^*_{t+1}\mathbf{W}_{t+1} = \delta^{-1}\left(\mathbb{I} - \frac{\mathbf{u}_{t+1}\mathbf{u}^*_{t+1}}{\delta + \mathbf{u}^*_{t+1}\mathbf{u}_{t+1}}\right) , \tag{7.5}$$

$$\mathbf{W}^{-1}_{t+1}\mathbf{W}^{-*}_{t+1} = \delta\mathbb{I} + \mathbf{u}_{t+1}\mathbf{u}^*_{t+1} \tag{7.6}$$

$$\mathbf{V}_{t+1}\mathbf{D}\mathbf{V}^*_{t+1} = \delta\mathbf{W}_{t+1}\mathbf{V}_t\mathbf{D}\mathbf{V}^*_t\mathbf{W}^*_{t+1} + \mathbf{W}_{t+1}\mathbf{u}_{t+1}\mathbf{u}^*_{t+1}\mathbf{W}^*_{t+1} - \delta^{-1}\ \boldsymbol{\mu}_{t+1}\ \overset{*}{\boldsymbol{\mu}}_{t+1} \tag{7.7}$$

$$\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{A}}^*_{t+1} = \delta\left(\mathbb{I} - \mathbf{Y}_{t+1}\mathbf{S}\mathbf{Y}^*_{t+1}\right) \tag{7.8}$$

where $\mathbf{V}_t$ is unitary, $\mathbf{Y}_t = \mathbf{V}_t|\mathbf{D}_t|^{1/2}$, $\tilde{\mathbf{A}}_t = \mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$, $\boldsymbol{\mu}_{t+1} \equiv \mathbf{W}_{t+1}\tilde{\mathbf{A}}_t\mathbf{u}_{t+1}$. The update of the eigendecomposition is performed using $\mathbf{W}_{t+1}\mathbf{V}_t = \delta^{-1/2}\left(\mathbb{I} - \gamma_{t+1}\mathbf{u}_{t+1}\mathbf{u}^*_{t+1}\right)\tilde{\mathbf{V}}_t$ where $\tilde{\mathbf{V}}_t = \delta^{1/2}\left(\mathbb{I} - \gamma_t\mathbf{u}_t\mathbf{u}^*_t\right)^{-1}\mathbf{W}_{t+1}$ is unitary. This algorithm may be modified by replacing the prediction error, $\hat{\boldsymbol{\epsilon}}_t$, with *a posteriori* errors (residuals), $\boldsymbol{\epsilon}_t = y_t - \hat{\mathbf{c}}^*_t\mathbf{u}_t$, in the update of $\hat{c}_t$ [LS]. In this case, (7.2) is replaced by

$$\hat{\mathbf{c}}_t = \left(\mathbb{I} - \frac{\mathbf{u}_t\mathbf{u}^*_t}{\delta + \mathbf{u}^*_t\mathbf{u}_t}\right)^{-1}\left(\mathbf{W}^{-*}_{t-1}\hat{\mathbf{c}}_{t-1} - \frac{\mathbf{u}_t\mathbf{y}^*_t}{\delta + \mathbf{u}^*_t\mathbf{u}_t}\right) . \tag{7.9}$$

8 APPENDIX: FAST FACTORIZATION OF $\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$

To achieve an $\mathcal{O}(\alpha^2 n)$ factorization of $\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$, we rewrite (4.6)

$$\left(\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t\right)\left(\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t\right)^* = \delta\left(\mathbb{I} - \sum_{k=1}^{\alpha} s_k\mathbf{y}_{k,t}\mathbf{y}^*_{k,t}\right) , \tag{8.1}$$

where $\mathbf{y}_{k,t}$ is the kth column of $\mathbf{Y}_t$. We define a sequence of α UT factorization of the partial sums:

$$\tilde{\mathbf{A}}^{(k)}_t\tilde{\mathbf{A}}^{(k)\ *}_t = \tilde{\mathbf{A}}^{(k-1)}_t\tilde{\mathbf{A}}^{(k-1)\ *}_t - \delta s_k\mathbf{y}_{k,t}\mathbf{y}^*_{k,t} , \tag{8.2}$$

with $\tilde{\mathbf{A}}^{(0)}_t = \delta^{1/2}\mathbb{I}$. We rewrite $\tilde{\mathbf{A}}^{(k)}\tilde{\mathbf{A}}^{(k)*}_t$ in the product form

$$\tilde{\mathbf{A}}^{(k)}_t = \tilde{\mathbf{A}}^{(k-1)}_t\mathbf{F}^{(k)}_t = \mathbf{F}^{(1)}_t\mathbf{F}^{(2)}_t\ldots\mathbf{F}^{(k)}_t . \tag{8.3}$$

Here $\mathbf{F}^{(k)}_t$ is the UT factor of $\mathbf{F}^{(k)}_t\mathbf{F}^{(k)\ *}_t = \mathbb{I} - \delta_k\xi_{k,t}\xi^*_{k,t}$, with $\xi_{k,t}$ defined by $\tilde{\mathbf{A}}^{(k-1)}_t\xi_{k,t} \equiv \mathbf{y}_{k,t}$. Each $\mathbf{F}^{(k)}_t$ is TIB and its inverse is computable in $\mathcal{O}(3n)$ operations given $\xi_{k,t}$. Given $\mathbf{F}^{(1)}_t\ldots\mathbf{F}^{(k-1)}_t$, then $\mathcal{O}(kn)$ operations are required to compute $\xi_{k,t}$. Thus $\mathcal{O}(\alpha^2 n)$ operations are required to complete the factorization. Given the factorization, matrix vector products are computable in $\mathcal{O}(\alpha n)$ operations.

At each step in this factorization of $\tilde{\mathbf{A}}^{(k)}_t$, we guarantee that $\mathbf{F}^{(k)}_t\mathbf{F}^{(k)\ *}_t$ is nonsingular by reordering the choice of columns in (8.2) such that all of the negative s_k are processed prior to the processing of the positive s_k.

Both $\tilde{\mathbf{A}}_t^{(\alpha)}$ and $\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$ are triangular and satisfy (8.1). Since $\mathbf{R}_t$ is also triangular, the similarity transformation preserves the order of the eigenvalues of $\mathbf{A}$. Thus we recover $\mathbf{R}_t^{-1}\mathbf{A}\mathbf{R}_t$ by right multiplying by the appropriate diagonal unitary matrix $\mathbf{D}$ to preserve the complex phases of the diagonal of $\mathbf{A}$.

ACKNOWLEDGEMENT

The authors thank Prof. A. Sayed and the referees for their helpful comments. The work of KSR was funded by U.S. Dept. of Energy Grant DE-FG02-86ER53223.

References

[GGMS] P. E. Gill, G. H. Golub, W. Murray and M. A. Saunders, "Methods for modifying matrix factorizations," *Math. Comp.*, vol. 28, pp. 505-535, 1974.

[GL] G. H. Golub and C. F. van Loan, *Matrix Computations*, second edition, Section 12.6.4, Baltimore: Johns Hopkins Universsity Press, 1989.

[GE] M. Gu and S. C. Eisenstat, "A Stable and Efficient Algorithm for the rank-one modification of the symmetric eigenvalue problem," *SIAM J. Matrix Anal. Appl.*, vol. 15, pp. 1266-1276, 1994.

[HD] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, p. 16, New York: John Wiley and Sons, 1988.

[KS] T. Kailath and A. H. Sayed, "Displacement structure: theory and applications," *SIAM Review*, vol. 37, pp. 297-386 1995.

[LT] P. Lancaster and M. Tismenetsky, *Theory of Matrices*, second edition, Boston : Academic Press, 1985.

[LK] H. Lev-Ari and T. Kailath, "State space approach to factorization of lossless transfer functions and structured matrices," *Linear Algebra and its Applications*, vol. 162,pp. 273-295, 1992.

[LS] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, Cambridge, MA: MIT Press, 1983.

[MR] A. P. Mullhaupt and K. S. Riedel, "Band matrix representation of triangular input balanced ," submitted.

[NSU] A. F. Nikiforov, S. K. Suslov and V. B. Uvarov, *Classical Orthogonal Polynomials of a Discrete Variable*, pp. 30, 44, New York: Springer Verlag, 1991.

[SK1] A. H. Sayed and T. Kailath, "Extended Chandrasekhar recursions," *IEEE Trans. Automatic Control*, vol. 39, pp. 619-623, 1994.

[SK2] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, pp. 18-60, 1994.