

ON THE EIGENDISTRIBUTION OF THE STEADY-STATE ERROR COVARIANCE MATRIX FOR THE EXTENDED RLS ALGORITHM

Ali Vakili, Eythan Familier, and Babak Hassibi

Electrical Engineering Department
California Institute of Technology
Pasadena, California 91125
Email: avakili, eythan, hassibi@caltech.edu

ABSTRACT

In an earlier work [1], we used transform methods from the theory of random matrices to analytically compute the asymptotic eigendistribution of the error covariance matrix of the single-measurement RLS filter. When we have a multiplicity of measurements, as happens in extended RLS filtering, the analysis is much more complicated. In this paper we study the multiple measurement case and obtain a system of two coupled equations for the Stieltjes transform of the asymptotic eigendistribution. Numerical solutions of this system very well predict the actual asymptotic eigendistribution for systems with as low as $n = 10 - 20$ state dimensions.

Index Terms— RLS algorithm, Kalman filter, Stieltjes transform, eigendistribution, random matrix

1. INTRODUCTION

The Recursive-Least-Squares (RLS) algorithm recursively minimizes the sum of squared errors in estimating an unknown state vector. In fact, the RLS algorithm is nothing other than a special case of Kalman filtering. While the Kalman filter with time-invariant coefficient matrices is well-studied in the literature, not much is known about its steady-state and transient behavior in the time-varying case [2, 3]. Even for a simple RLS filter with one measurement per time instance, the mean square error performance is not known except under assumptions which are often not too reasonable.

In our earlier work [1], we used ideas from random matrix theory to establish a framework for analyzing the single measurement RLS filter with random regressor vectors, under two main assumptions: (a) the regressors represent a stationary distribution, and (b) the state dimension is large (although the results are valid for dimensions as low as 10). We found the steady-state eigendistribution of the error covariance matrix in various scenarios. However the results were

This work was supported in part by the National Science Foundation through grant CCF-0729203, by the David and Lucille Packard Foundation, by the Office of Naval Research through a MURI under contract no. N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

only valid when the number of measurements at each time step was small compared to the state vector size. In this paper, we present the more complicated analysis for the *extended RLS algorithm* [2] where the number of measurements is allowed to grow with n .

Throughout the paper, we remain in the linear state-space model framework [4],

$$\begin{cases} x_{i+1} &= x_i + u_i \\ y_i &= \mathbf{H}_i x_i + v_i \end{cases}, \quad (1)$$

in which $x_i \in \mathcal{R}^n$, the state vector we want to estimate, undergoes a random walk and u_i and v_i denote the zero-mean process and measurement noises, respectively. $y_i \in \mathcal{R}^m$ is the measured signal and \mathbf{H}_i is the $m \times n$ regressor matrix which is indeed time-variant and often random. We will assume that all the processes are stationary and their statistics follow,

$$\mathbb{E} \begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_j^T & v_j^T \end{bmatrix} = \begin{bmatrix} q\mathbf{I} & 0 \\ 0 & r\mathbf{I} \end{bmatrix} \delta_{ij}, \quad (2)$$

It is well-known that the RLS recursive estimate of x_i can be expressed as,

$$\hat{x}_{i+1} = \hat{x}_i + \mathbf{P}_i \mathbf{H}_i^* (r\mathbf{I} + \mathbf{H}_i \mathbf{P}_i \mathbf{H}_i^*)^{-1} (y_i - \mathbf{H}_i \hat{x}_i), \quad (3)$$

and the estimation error covariance, $\mathbf{P}_i = \mathbb{E}(x_i - \hat{x}_i)(x_i - \hat{x}_i)^T$, satisfies a Riccati recursion,

$$\mathbf{P}_{i+1} = \mathbf{P}_i - \mathbf{P}_i \mathbf{H}_i^* (r\mathbf{I} + \mathbf{H}_i \mathbf{P}_i \mathbf{H}_i^*)^{-1} \mathbf{H}_i \mathbf{P}_i + q\mathbf{I}, \quad \mathbf{P}_0. \quad (4)$$

The above Riccati recursion is nonlinear and time-variant and, in general, does not converge. Furthermore, when the \mathbf{H}_i are random, it is a random matrix recursion. When the \mathbf{H}_i are stationary—an assumption we shall make—it may be expected that \mathbf{P}_i will converge to a matrix-valued stationary random process. In this case, the stationary distribution of \mathbf{P}_i , and its statistics, such as $\mathbb{E}\mathbf{P}_i$ or $\mathbb{E}\text{tr}\mathbf{P}_i$, will be of interest. This is what we intend to analyze in this paper.

The rest of the paper is organized as follows. In Section 2 we present a few results from random matrix theory which

have been used in our work. In Section 3 we briefly review our earlier results on RLS filtering with single measurement in various scenarios. Section 4 contains our new result for the RLS algorithm with multiple measurements. Finally simulation results are presented in Section 5 which demonstrate a very close prediction of the eigendistribution through our method.

2. PERTINENT RESULTS FROM RANDOM MATRIX THEORY

In this section we give a brief overview of pertinent results from random matrix theory which will be used in this work. More comprehensive reviews on the topic can be found in [1, 5]

For an $n \times n$ random matrix M , whose elements have some joint probability distribution, the cumulative eigenvalue distribution can be defined as,

$$F_M(x) = \frac{1}{n} \sum_{l=1}^n \Pr \{ \lambda_l(M) \leq x \}, \quad (5)$$

where $\lambda_l(M)$ is the l -th eigenvalue of M . A probability density function can be assigned to the cumulative function defined as above. We will refer to this density function, $f_M(\lambda)$, as the *eigendistribution* of the random matrix M . Basically, $f_M(\lambda)$ is the marginal distribution of one randomly selected eigenvalue of M .

In many cases, the eigendistribution of a random matrix demonstrates interesting properties. When the matrix dimensions are allowed to grow, one can often obtain simple closed form results for the eigendistribution which can be interpreted as universal laws which are independent of the underlying distributions—counterparts to the central limit theorem for scalar random variables.

It should also be mentioned that in the case of random matrices, even for dimensions as low as $n = 10$, these asymptotic results are very close predictions of the eigendistribution (This is due to the fact that we are essentially averaging over n^2 entries.)

Many results in the theory of large random matrices are expressed in terms of some *transform* of the eigendistribution. A complete list of these transforms and examples of their applications can be found in [5]. The most frequently appearing one, the Stieltjes transform, was first used in the seminal work of Marcenko and Pastur [6]. Given a probability density function $f_\lambda(\lambda)$, its Stieltjes transform is defined as,

$$S_\lambda(z) = \mathbb{E} \left[\frac{1}{\lambda - z} \right]. \quad (6)$$

When $f_\lambda(\lambda)$ represents the eigendistribution of a random matrix, there exist equivalent definitions of the Stieltjes transform which turn out to be extremely useful. In fact, given a

random matrix M , the Stieltjes transform of its eigendistribution (which we will also denote as the Stieltjes transform of the matrix itself) can be expressed as,

$$S_M(z) = -\frac{d}{dz} \mathbb{E} \frac{1}{n} \log \det(M - zI) \quad (7)$$

or equivalently,

$$S_M(z) = \mathbb{E} \frac{1}{n} \text{tr} (M - zI)^{-1} \quad (8)$$

Having the Stieltjes transform, we can retrieve the distribution uniquely through the well-known inversion formula [7],

$$f_M(\lambda) = \lim_{\omega \rightarrow 0^+} \frac{1}{\pi} \text{Im} [S_M(\lambda + j\omega)]. \quad (9)$$

An important property of the Stieltjes transform which makes it very useful in random matrix theory is the self-averaging property which is stated in the following lemma [8].

Lemma 1 (Self-averaging property). *Let A be an $n \times n$ positive semidefinite random matrix. As $n \rightarrow \infty$ the eigendistribution of A converges almost surely to its mean value, i.e.,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(A - zI)^{-1} = S_A(z) \quad \text{a.s.} \quad (10)$$

(Note the absence of the expectation.) Furthermore, for any n -dimensional vector x with i.i.d. zero-mean, unit-variance elements which is independent of A , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} x^T (A - zI)x = S_A(z) \quad \text{a.s.} \quad (11)$$

The above lemma can be intuitively verified by comparing it to the definition (8).

3. RESULTS ON SINGLE-MEASUREMENT RLS FILTERS

The steady-state eigendistribution of the error covariance matrix for RLS adaptive filters with a single measurement at each time step has been analyzed in [1] for three different cases. Namely, temporally white regressors, shift-structured regressors, and RLS filter with intermittent observations, which is inspired by the recent interest in the estimation and control over lossy networks [9, 10].

It is shown that the Riccati recursion of (4) with $m = 1$ and h_i 's being independent of each other, having i.i.d. zero-mean unit-variance entries, converges for $q = \frac{\gamma}{n}$ and $n \gg 1$ to a stationary matrix-valued random process whose eigendistribution's Stieltjes transform satisfies,

$$\gamma S(z) + c = -\log(r - z - z^2 S(z)), \quad (12)$$

where c is a constant which can be found numerically by insisting that the distribution integrates to one. It is further

shown that the same is true when the regressors have a shift structure,

$$h_i = [u_i \quad u_{i-1} \quad \cdots \quad u_{i-n+1}], \quad (13)$$

where the u_i are drawn from a zero-mean, unit-variance white process.

For the RLS filter with temporally white regressor, and intermittent single measurements, where the y_i 's may be lost independently with probability ϵ , it is shown that the asymptotic steady-state Stieltjes transform satisfies

$$\gamma S(z) + c = -(1 - \epsilon) \log(r - z + z^2 S(z)), \quad (14)$$

where c is again a constant which is determined by insisting that the inverse Stieltjes transform should integrate to one.

These results can readily be extended to the case of multiple measurements as long as the number of measurements, m , is much less than the state vector size, n ($m \ll n$). When m is comparable to n , the results are not valid and a more complicated analysis is required which is the subject of this paper.

4. ANALYZING THE EXTENDED RLS FILTER

To this end, we find an expression for the Stieltjes transform of the eigendistribution of P_i , as $i \rightarrow \infty$, which in turn determines the eigendistribution and important statistical measures of the process P_i , such as $\mathbb{E} \text{tr} P_i$. As mentioned earlier, we will also assume a large dimensional state vector which allows us to benefit from the large random matrix theory. The following theorem presents the main result.

Theorem 1. *Consider the random Riccati recursion in (4) where $m, n \rightarrow \infty$ while $\frac{m}{n} = \beta$. As $i \rightarrow \infty$, the Stieltjes transform of the eigendistribution of P_i converges to the steady-state Stieltjes transform, $S(z)$, which satisfies the pair of implicit equations*

$$S(z+q) = \frac{t(z)}{1-zt(z)} - \frac{1}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right) \quad (15)$$

$$t(z) = \frac{\beta}{r - \frac{z}{1-zt(z)} + \frac{z^2}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right)}. \quad (16)$$

The above equations can be solved numerically to obtain $S(z)$ and eventually $f_\lambda(\lambda)$.

Proof. After bringing qI to the LHS of (4), we use the definition (7) to compute the Stieltjes transform of both sides,

$$S_{i+1}(z+q) = -\frac{1}{n} \mathbb{E} \frac{d}{dz} \log \det (P_i - P_i H_i^* (rI + H_i P_i H_i^*)^{-1} H_i P_i - zI) \quad (17)$$

After rearranging the terms, and using the fact that $\det(I - AB) = \det(I - BA)$, the RHS can be written as,

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n} \mathbb{E} \frac{d}{dz} \log \det (rI + H_i P_i H_i^* - H_i P_i (P_i - zI)^{-1} P_i H_i^*) \quad (18)$$

after some straightforward algebraic manipulations we obtain,

$$S_{i+1}(z+q) - S_i(z) = -\frac{1}{n} \mathbb{E} \frac{d}{dz} \log \det \left(rI - H_i \underbrace{(zI + z^2(P_i - zI)^{-1})}_{\triangleq D_i(z)} H_i^* \right), \quad (19)$$

Since H_i has an isotropic distribution, we can assume without loss of generality that P_i is diagonal. Therefore $D_i(z)$ which is defined above will be diagonal. By computing the derivative in (19) we have,

$$\text{RHS} = \frac{1}{n} \mathbb{E} \text{tr} \left[H_i D_i'(z) H_i^* (rI - H_i D_i(z) H_i^*)^{-1} \right], \quad (20)$$

which can be written as

$$\text{RHS} = \frac{1}{n} \mathbb{E} \text{tr} \left[D_i'(z) D_i^{-1}(z) \times \left(\left(D_i^{-1}(z) - \frac{H_i^* H_i}{r} \right)^{-1} D_i^{-1}(z) - I \right) \right]. \quad (21)$$

Note that the j -th diagonal entry of $D_i(z)$ is just

$$d_{i,j} = z + \frac{z^2}{\lambda_j - z} = \frac{\lambda_j z}{\lambda_j - z}, \quad (22)$$

and since we are looking for the marginal distribution of one eigenvalue of P_i , it does not matter which eigenvalue we are looking at. Therefore the second term in (21) can be expressed as,

$$-\frac{1}{n} \mathbb{E} \text{tr} D_i'(z) D_i^{-1}(z) = \frac{1}{z} \mathbb{E} \left(1 + \frac{z}{\lambda - z} \right) = \frac{1}{z} + S_i(z). \quad (23)$$

For computing the first term, which is more challenging, we break $D_i(z)$ and H_i in the form,

$$D_i(z) = \begin{bmatrix} d_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad H_i = [h_1 \quad H_2], \quad (24)$$

where h_1 is an $m \times 1$ vector. We have dropped the index i for simplicity. Doing so, it is not hard to show that this term can be written as,

$$A = \mathbb{E} \frac{d_1'}{d_1^2} \frac{1}{d_1^{-1} - \underbrace{h_1^* (rI - H_2 D_2 H_2^*)^{-1} h_1}_{\triangleq t(z)}}, \quad (25)$$

which can be shown to be equal to,

$$A = \frac{1}{z(1-zt)} + \frac{1}{(1-zt)^2} S_i \left(\frac{z}{1-zt} \right). \quad (26)$$

Replacing (23) and (26) into (18) yields (15). Therefore it remains to compute,

$$t(z) = \frac{1}{n} \text{tr} (rI - H_2 D_2 H_2^*)^{-1}. \quad (27)$$

We break H_2 as,

$$H_2 = \begin{bmatrix} h_{21} \\ H_{22} \end{bmatrix}, \quad (28)$$

and using the same techniques by which (25) was obtained, $t(z)$ can be written as,

$$t(z) = \frac{m}{n} \frac{1}{r - h_{21} (D_2^{-1} - \frac{1}{r} H_{22}^* H_{22})^{-1} h_{21}^*}. \quad (29)$$

Invoking the self-averaging property, we can rewrite the above expression as,

$$t(z) = \frac{m}{n} \frac{1}{r - \mathbb{E} \frac{1}{d_2^{-1} - t(z)}}, \quad (30)$$

which can be simplified further by noting that

$$\mathbb{E} \frac{1}{d_2^{-1} - t(z)} = \mathbb{E} \frac{1}{\frac{\lambda - z}{\lambda z} - t(z)} = \frac{z}{1-zt} \mathbb{E} \frac{\lambda}{\lambda - \frac{z}{1-zt}}. \quad (31)$$

Therefore,

$$\mathbb{E} \frac{1}{d_2^{-1} - t(z)} = \frac{z}{1-zt} + \frac{z^2}{(1-zt)^2} S_i \left(\frac{z}{1-zt} \right), \quad (32)$$

which together with (30) result in (16). \square

Now one has to solve the equations (15) and (16) numerically in order to find $S(z)$. Then by using the inverse formula of the Stieltjes transform (9) the eigendistribution in the steady-state case can be obtained.

5. SIMULATIONS

Figure 1 compares the theoretically predicted curve (through Theorem 1) with the eigendistribution obtained empirically for $m = 15$ and $n = 30$. In order to find the theoretical curve we have numerically solved the coupled equations describing $S(z)$ and $t(z)$ on a grid in the complex plane. It can be seen that our framework closely captures the behavior of the eigendistribution. The results are plotted for $r = 0.5$ and $q = 0.48$.

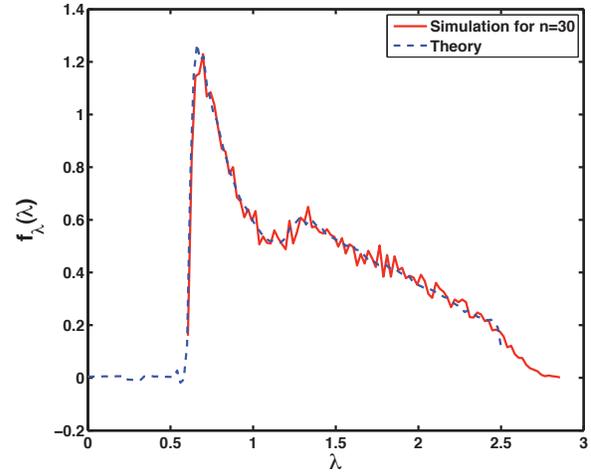


Fig. 1. The steady-state error covariance eigendistribution for the extended RLS filter with $m = 15$ for $n = 30$, $r = .5$, and $q = 0.48$

6. REFERENCES

- [1] A. Vakili and B. Hassibi, "A Stieltjes transform approach for analyzing the RLS adaptive filter," in *Proc. of 46th annual Allerton Conference*, September 2008.
- [2] A.H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-IEEE, 2003.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 2001.
- [4] B. Hassibi, A.H. Sayed, and T. Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*, SIAM, 1999.
- [5] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [6] V. A. Marcenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.
- [7] J. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.
- [8] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of Probability*, vol. 26, pp. 316–345, 1998.
- [9] B. Sinopoli, L. Schenato, M. Franceschetti, K. Pooll, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sep 2004.
- [10] V. Gupta, A. F. Dana, J. Hespanha, R. M. Murray, and B. Hassibi, "Data transmission over networks for estimation and control," *IEEE Transactions on Automatic Control*, submitted 2006.