Random Matrix Recursions in Estimation, Control, and Adaptive Filtering

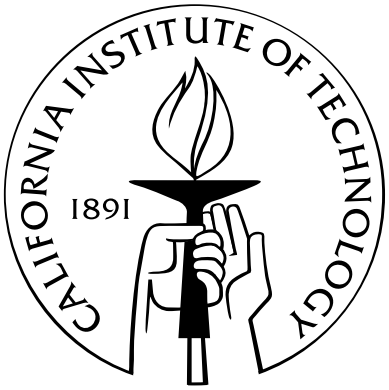
Thesis by

Ali Vakili

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2011

(Defended July 13, 2009)

To
my family

Acknowledgements

During the years of my study at Caltech, I have been privileged to work with and know many great people. Most important among them is my advisor, Prof. Babak Hassibi. Babak's deep understanding of the subject and his generosity in sharing his knowledge and spending time with me for lengthy discussions have been invaluable resources. His courage in tackling seemingly unsolvable problems and enthusiasm for exploring new subjects and approaches have always been, and will remain to be inspiring to me. I have learnt from him the importance of paying attention to small details and at the same time not missing the bigger picture. Without his constant support, this dissertation would have not been possible.

My sincere appreciations go to Professors Yaser S. Abu-Mostafa, Steven Low, Richard M. Murray, Joel A. Tropp, and P. P. Vaidyanathan, for being on my candidacy and thesis committees and providing me with great feedback. I have also had the opportunity of learning a lot from them throughout the courses and discussions in the past five years. I would like to thank Professor Bruno Sinopoli at Carnegie Mellon University for insightful discussions that we had on several occasions. I am also thankful to my co-authors Masoud Sharif and Amir Farajidana for the fruitful collaborations and intriguing conversations during our time as office mates at Caltech.

I am extremely thankful to Shirley Slattery for her perfect assistance on administrative matters and for her kindness and care. My colleagues, office mates, and friends have been an inseparable part of my life and work at Caltech. I will always be grateful to them for making my years as a graduate student pleasant and exciting.

I would like to thank my dearest friend Sormeh Shadbakht, for being the best and most colorful part of my life. Her friendship and support is truly irreplaceable for me.

And finally, my deepest gratitude goes to my family for years and years of love, support, and encouragement. My parents, Kamran Vakili and Hengameh Sepahsalari, with their fondness for science and education, have been the greatest inspiration in my life, and my brother, Armin, has been the coolest friend I have ever had. I would also like to thank my aunt, Taraneh Sepahsalari, for her kindness and constant support. I am truly grateful to my grandmother Parichehreh Derayeh, and my late grandfather, Esfandyar Sepahsalari, for being my first teachers from whom I learned how to read and write. I dedicate this thesis to my family as an inadequate appreciation of everything that they have done for me.

Abstract

This dissertation is devoted to the study of estimation and control over systems which can be described by linear time-varying state-space models. Examples of such systems are encountered frequently in systems theory, e.g., wireless sensor networks, adaptive filtering, distributed control, etc. While linear time-invariant systems are well-understood, there is no general theory that captures various aspects of time-varying systems, ranging from the optimum design of the system to the performance analysis. With little exception, tackling these problems normally boils down to studying time-varying linear or non-linear recursive matrix equations, known as Lyapunov and Riccati recursions. Unlike the time-invariant counterparts, these are notoriously hard to analyze. Here, we employ the theory of random matrices to elucidate different facets of these recursions and henceforth, answer several important questions about the performance, stability, and convergence of estimation and control over such systems.

The importance of such analysis cannot be overemphasized. In the light of recent advances in micro-sensor technologies, numerous new applications for wireless sensor networks have been made possible, including, but not limited to, distributed catastrophe surveillance, smart transportation, and power grid control systems. A vast amount of research has been dedicated to the study of wireless sensor networks in the past few years that has only revealed the insufficiency of our knowledge and the inefficiency of the existing tools in analyzing them. There are overwhelmingly new research opportunities in this field which are yet to be examined. In wireless sensor

networks the communications limitations, such as data loss and delay, become tightly connected to the control and estimation process. Classical estimation and control algorithms are immediately disqualified since the observation and/or control data may be lost in the system. Moreover, stingy power constraints imposed by the sensors call for a distributed rather than centralized approach to the problem. Handling sensor failures and security issues are other examples of obstacles which are yet to be overcome. A steady-state analysis of random matrix recursions that arise in this field not only determines the system performance, but can also guarantee the stability.

Another important example is adaptive filtering, which is a universal tool in many areas such as communications and control. Rigorous analysis of the steady-state and transient behavior of adaptive filters remains a formidable task in most of the cases. The so-called excess mean-square error (EMSE) in many classes of adaptive filters is not known except for approximations based on several unrealistic assumptions. Once again, the transient and steady-state behaviors are governed by random Lyapunov or Riccati recursions that bring about a great deal of technical difficulties.

We make two general assumptions throughout our approach. First, we assume that the coefficient matrices are drawn from jointly stationary matrix-valued random processes. The stationarity assumption hardly restricts the analysis since almost all cases of practical interest fall into this category. We further assume that the state vector size, n, is relatively large. This assumption will enable us to take advantage of many compelling results available in the theory of large random matrices. Although the derivations are carried out for $n \gg 1$, the law of large numbers guarantees fast convergence to the asymptotic results for n being as small as 10-20. Therefore the assumption is justifiable since in the majority of practical scenarios the state vector size satisfies this condition. Under these assumptions, we develop a framework which is capable of characterizing steady-state as well as transient behavior of adaptive

filters and control and estimation over communication networks. This framework has proven promising by successfully predicting universal laws for the first time in several problems.

In the following, we first look at random Lyapunov recursions and characterize their transient and steady-state behavior. This is motivated by several classes of adaptive filters, and through Lyapunov recursions appearing as lower bounds of random Riccati recursions in distributed Kalman filtering. Next we look at random Riccati recursions which manifest the system's performance in numerous important scenarios and therefore their analysis is of greater significance compared to the Lyapunov ones. At the same time, their nonlinearity makes them much more complicated to study. We begin with studying simple recursive-least-squares (RLS) filtering and extend our analysis beyond the standard case to filtering with multiple measurements, as well as the case of intermittent measurements. This is is motivated by the problem of a sensor sending its measurements through a lossy network. Finally, we study Kalman filtering with intermittent observations which is frequently used to model wireless sensor networks. In all of these cases we obtain interesting universal laws which depend on the structure of the problem, rather than specific model parameters. We verify the accuracy of our results through various simulations for systems with as few as 10 states.

Contents

List of Figures

Chapter 1

Introduction

The general purpose of this thesis is to study the performance of various algorithms used in estimation, control, and adaptive signal processing to estimate an unknown, not directly observable variable, through another dependent but observable variable. This broad definition encompasses a significant number of problems in engineering and science. Examples include signal detection in communications, locating an object using radar signals, orbit determination in astronomy, and estimating the response of a population through opinion polls. Although the first appearances of estimation problems in engineering, statistics, and science date back to hundreds of years ago, due to their ubiquitous applications they have continued to be a notorious subject of research in several fields.

The models and algorithms, and even the languages used in the analysis of these problems are as diverse as the problems themselves. Naturally the first step of every analysis in the estimation theory is to develop a model for the underlying real-world problem. The next step is to adopt an estimation algorithm, which should be followed by the performance study. If proved to be promising, the developed algorithm evolves into practical solutions. This thesis focuses on the third stage, which is the performance analysis of different estimation algorithms, in particular those related to estimation and control over lossy networks and adaptive filtering. The problems we consider are common in being modeled through a linear state-space model, and as a

result, in having their performance governed by the celebrated Lyapunov and Riccati matrix recursions.

The linear time-invariant state-space models have been studied extensively in the past decades. A huge host of algorithms have been developed for estimation in this framework. Kalman filtering is a famous example. At the same time, the time-varying counterpart has also been subject of a large body of research. As an example, time-varying linear state-space models have been the bread and butter of adaptive filtering. However there is a significant shortcoming when it comes to the analysis of time-varying state-space models: While the performance of estimation over LTI systems has been known for a long time and there exist powerful tools which can be employed successfully to this end, the performance analysis of time-varying counterparts has never been rigorous in most cases. For example in the adaptive filtering theory, as will be discussed later in the thesis, the performance of recursive-least-squares (RLS) filters is not known unless under often unrealistic assumptions. Another instance is the problem of estimation and control over lossy networks about which there has been a growing interest in recent years. A survey of the literature shows the impotency of the existing tools in the study of this important problem. As we already hinted, the underlying difficulty in both cases is dealing with time-varying (and often random) Lyapunov and Riccati recursions (especially the Riccati since it is nonlinear). This thesis is an attempt to establish a framework through which such random matrix-valued recursions can be analyzed. The approach that we take here is completely novel in the literature and as will be seen, proves to be successful in almost all scenarios of practical importance.

In this introductory chapter we will try to highlight the importance of studying random Lyapunov and Riccati recursions by giving an overview of some problems and applications both in recent areas of research such as wireless sensor/actuator

networks and in well-developed fields such as adaptive filtering. The motivation for taking the route of large random matrix theory and a high-level description of the proposed framework will also be presented. The chapter concludes with an account of the contributions of this thesis.

1.1 Linear State-Space Model

The standard state-space model is a very common model in systems theory for many reasons which are beyond this introduction [HSK99]. Consider

$$\begin{cases} x_{i+1} &= \mathrm{F}_i x_i + \mathrm{G}_i u_i, \qquad i \geq 0, \\ y_i &= \mathrm{H}_i x_i + v_i, \end{cases} \tag{1.1}$$

where $x_i \in \mathcal{R}^n$ is the state vector which is not observable and $y_i \in \mathcal{R}^m$ denotes the measurement vector at time i. The state and process noises (sometimes called driving sequences), $\{u_i\}$ and $\{v_i\}$, are assumed to be zero-mean white processes. Furthermore $\mathrm{F}_i \in \mathcal{R}^{n\times n}$, $\mathrm{G}_i \in \mathcal{R}^{n\times p}$, $\mathrm{H}_i \in \mathcal{R}^{m\times n}$ are system matrices and known to the observer. The initial state of the system, denoted by x_0, is also considered to be a random variable. The statistics of all the processes in the system can be summarized as

$$\mathbb{E}\begin{bmatrix} u_i \\ v_i \\ x_0 \end{bmatrix}\begin{bmatrix} u_j^T & v_j^T & x_0 & 1 \end{bmatrix} = \begin{bmatrix} \mathrm{Q}_i\delta_{ij} & \mathrm{S}_i\delta_{ij} & 0 & 0 \\ \mathrm{S}_i^T\delta_{ij} & \mathrm{R}_i\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}, \tag{1.2}$$

which insists that all the random variables are zero mean[1]. We will also assume that $\{u_i\}$ and $\{v_i\}$ are independent (in other words $\mathrm{S}_i = 0$).

[1]This does not affect the generality since it is straightforward to incorporate nonzero mean.

In the time-invariant case, i.e., when

$$\mathrm{F}_i \triangleq \mathrm{F}, \quad \mathrm{G}_i \triangleq \mathrm{G}, \quad \mathrm{H}_i \triangleq \mathrm{H}, \quad \mathrm{Q}_i \triangleq \mathrm{Q}, \quad \mathrm{R}_i \triangleq \mathrm{R}, \tag{1.3}$$

the problem is very well studied. It is known that the state covariance matrix satisfies a *Lyapunov recursion*

$$\Pi_{i+1} = \mathrm{F}\Pi_i\mathrm{F}^T + \mathrm{G}\mathrm{Q}\mathrm{G}^T, \qquad \Pi_0. \tag{1.4}$$

If F is stable (all eigenvalues strictly inside unit circle) and $\{\mathrm{F}, \mathrm{G}\mathrm{Q}^{\frac{1}{2}}\}$ is controllable[2] then Π_i converges as $i \to \infty$ to the unique solution of the *Lyapunov equation*

$$\Pi = \mathrm{F}\Pi\mathrm{F}^T + \mathrm{G}\mathrm{Q}\mathrm{G}^T. \tag{1.5}$$

The minimum mean square error (MMSE) recursive estimator of $\{x_i\}$ based on the $\{y_i\}$ is the very well-known Kalman filter. If the observer employs a Kalman filter, the estimation error covariance matrix, $\mathrm{P}_i = \mathbb{E}(x_i - \hat{x}_i)(x_i - \hat{x}_i)^T$, will satisfy a *Riccati recursion*

$$\mathrm{P}_{i+1} = \mathrm{F}\mathrm{P}_i\mathrm{F}^T + \mathrm{G}\mathrm{Q}\mathrm{G}^T - \mathrm{F}\mathrm{P}_i\mathrm{H}^T(\mathrm{R} + \mathrm{H}\mathrm{P}_i\mathrm{H}^T)^{-1}\mathrm{H}\mathrm{P}_i\mathrm{F}^T, \quad \mathrm{P}_0, \tag{1.6}$$

which is a nonlinear matrix-valued recursion. It is known that when $\{\mathrm{F}, \mathrm{G}\}$ is stabilizable[3] and $\{\mathrm{F}, \mathrm{H}\}$ is detectable[4] then P_i converges to the unique positive semidefinite solution of the discrete algebraic Riccati equation (DARE)

$$\mathrm{P} = \mathrm{F}\mathrm{P}\mathrm{F}^T + \mathrm{G}\mathrm{Q}\mathrm{G}^T - \mathrm{F}\mathrm{P}\mathrm{H}^T(\mathrm{R} + \mathrm{H}\mathrm{P}\mathrm{H}^T)^{-1}\mathrm{H}\mathrm{P}\mathrm{F}^T. \tag{1.7}$$

[2]There are several equivalent definitions of controllability. For example $\{\mathrm{F}, \mathrm{G}\}$ are controllable if and only if the controllability matrix $[\mathrm{G} \;\; \mathrm{F}\mathrm{G} \;\; \mathrm{F}^2\mathrm{G} \;\; \ldots \;\; \mathrm{F}^{n-1}\mathrm{G}]$ has rank n.

[3]The pair $\{\mathrm{F}, \mathrm{G}\}$ is called stabilizable if there is no left eigenvector of F, corresponding to an unstable eigenvalue of F, that is orthogonal to G. In other words if all unstable modes of F are controllable. There are several equivalent conditions in the literature [KSH00].

[4]The pair $\{\mathrm{F}, \mathrm{H}\}$ is called detectable if and only if $\{\mathrm{F}^T, \mathrm{H}^T\}$ is stabilizable.

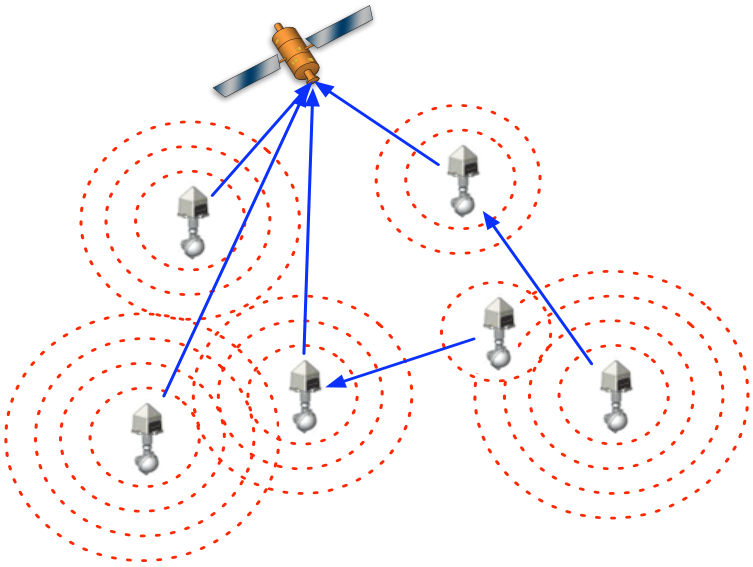


Figure 1.1. A sensor network is a collection of distributed sensors that make noisy observations of some dynamical system and forward these measurements either to another sensor or to a central location.

For an excellent discussion of the solutions of the DARE one can consult [KSH00].

However, things are not always as simple as the time-invariant case. As we will see, there are many important scenarios for which the time invariance assumption is no longer valid.

It is worth mentioning that there are other variants of this standard state-space model which arise in different contexts. For example x_{i+1} may be considered to have both u_i and u_{i+1} dependency; or the observation noise, v_i, can represent a nonwhite process which follows an update equation just like that of the state vector x_i. Although the analysis of these alternative models is possible through our framework, we will stay in the boundaries of the standard model because it is the most commonly used model in the literature, and furthermore the problems that have motivated the thesis are best modeled this way.

1.2 Estimation and Control over Lossy Networks

Recent advances in the microsensor technologies have facilitated numerous new applications of wireless sensor and/or actuator networks including but not limited to distributed catastrophe surveillance, smart transportation and power grid control systems. It is now widely recognized that many of the future applications of systems and control theory will pertain to problems of distributed estimation and control of multiple agents (both sensors and actuators) over networks.

A sensor network is a collection of distributed sensors that make noisy observations of some dynamical system (for example the environment temperature) and forward these measurements either to another sensor or to a central location in order to obtain an estimate of the state of the system. In some other systems the goal is to coordinate among several agents to perform a prespecified task collectively in which case the agents communicate through a network (usually wireless) to exchange various estimation and control data. What is common among all these scenarios is that the communication, or exchange of measurements, must be done in real time across a wireless network which is inherently unreliable.

A vast amount of research has been dedicated to the study of wireless sensor/actuator networks in the past few years that has only revealed the insufficiency of our knowledge and the inefficiency of the existing tools in analyzing them. There are many new research opportunities in this field which are yet to be examined. In such networks the communications limitations, such as data loss and delay, become tightly connected to the control and estimation process. Classical estimation and control algorithms are immediately disqualified since the observation and/or control data may be lost in the system. Moreover, stingy power constraints imposed by the sensors call for a distributed rather than centralized approach to the problem. Handling sensor failures and security issues are other examples of obstacles which are yet

to be overcome.

Several important questions may be asked about how best to operate across such unreliable networks. For example, what information should the agents transmit? What is the best coding scheme for measurements or control data to overcome the uncertainties in the network? How do we achieve all of this in a distributed fashion? And many more. The researchers in this field have taken many different directions, a complete account of which is well beyond the scope of this chapter. The interested reader may consult [Gup06] and the references therein for a literature review.

While complicated encoding schemes can be envisioned for these systems, they usually add too much overhead to the system which conflicts with the power constraints. The work along this direction is continuing to grow and may eventually lead to a breakthrough. From a practical point of view, however, most systems will not implement fancy encoding and decoding techniques, rather they will employ simple Kalman filtering and LQG control at least in the short to medium term. In other words, in the aforementioned sensory network problem, the central location in charge of constructing state estimates will simply implement a Kalman filter with both measurement and time updates every time it receives a measurement and only a time update when it does not receive a measurement. Similarly, in control problems, the agents will implement certainty equivalence control laws, i.e., they will use state feedback applied to their current estimate of the state (in fact, in many cases it can be shown that certainty equivalence or separation holds: the optimal control signal is optimal state feedback using the optimal state estimate).

Thus the main issue in such networks is to determine the actual performance of the system and its dependence on the network uncertainties. For example for a wireless sensor network we are interested in the estimation error of the state. In multi agent control problems we are interested in things such as stability, success

of the task and/or the optimal control cost. What makes the problem challenging, and different from the traditional theories, is that the underlying lossy (or unreliable) network introduces randomness into the system performance.

A very well-received model for the effect of unreliable links in these systems (see, e.g., Sinopoli et al. [SSF$^+$07, MS08], Epstein et al. [ESTM08], Imer et al. [IYB06] and the references therein) assumes that control and estimation data are in the form of intermittent packets. The packets (independently) may be received or not according to some probability of successful transmission. Adopting this model has made many problems mathematically tractable and infused great insights about the system behavior and performance. However this model brings up random Riccati recursions about which except for bounds on the stability conditions, very little is investigated. A steady-state analysis of such recursions not only determines the system performance, but can also guarantee the stability [KSM09].

More explicitly, assume that a wireless sensor is sending its measurements $\{y_i\}$ to the estimator as packets through a lossy network. Each packet, independently of other packets may be lost with probability $1-\bar{\gamma}$. Depending on whether a measurement is received or not, the Riccati recursion update is different (a time update when no observation is at hand and both time and measurement updates otherwise). Thus the lossy network makes the underlying Riccati recursion (which propagates the error covariance matrix of the state) random

$$\mathrm{P}_{i+1} = \mathrm{F}\mathrm{P}_i\mathrm{F}^T + \mathrm{G}\mathrm{Q}\mathrm{G}^T - \gamma_i \mathrm{F}\mathrm{P}_i\mathrm{H}^T(\mathrm{R}+\mathrm{H}\mathrm{P}_i\mathrm{H}^T)^{-1}\mathrm{H}\mathrm{P}_i\mathrm{F}, \quad \mathrm{P}_0, \tag{1.8}$$

where $\{\gamma_i\}$ is a Bernoulli process such that $\Pr(\gamma_i = 1) = \bar{\gamma}$.

Thus, unlike the standard LTI Kalman filter where the convergence properties of the Riccati recursion are well known, here we have to study the asymptotic behavior of a random Riccati recursion. We therefore envisage that, just as the study of

conventional Riccati recursions played a central role in conventional estimation and control, the study of random Riccati recursions will play a central role in the systems that arise from estimation and control over lossy networks. Since Riccati recursions are highly nonlinear, explicitly determining the distribution or mean value of the matrix being propagated seems hopelessly intractable. Therefore what we propose here is to leverage results from the theory of large random matrices (and, in particular, transform methods) to determine the asymptotic eigendistribution of the matrix being propagated.[5] As we will show in this thesis, under some fair assumptions our framework can successfully predict the steady-state eigendistribution of the error covariance matrix for estimation problems with intermittent observations. This is the first result of its kind in the field and at the same time can be shown to be very accurate for most cases of practical interest.

1.3 Adaptive Filtering

Adaptive filtering is a universal tool in many areas such as communications, control, and statistical signal processing. Basically adaptive algorithms appear whenever we encounter time-variant systems with little a priori information about the underlying signals. This is a topic with much practical applicability and interesting theoretical challenges. Although the ideas of adaptive signal reconstruction root back to Gauss and it can be considered a classical field with numerous textbooks and established practice, rigorous analysis of the steady-state and transient behavior of adaptive filters remains a formidable task in most of the cases. The reason is that adaptive filters are time-varying, often nonlinear and at the same time stochastic objects. When the regressors are random, as is the case in almost all applications, many classes

[5] As we shall see further below, the eigendistribution often contains all the useful information about the random matrix that we desire.

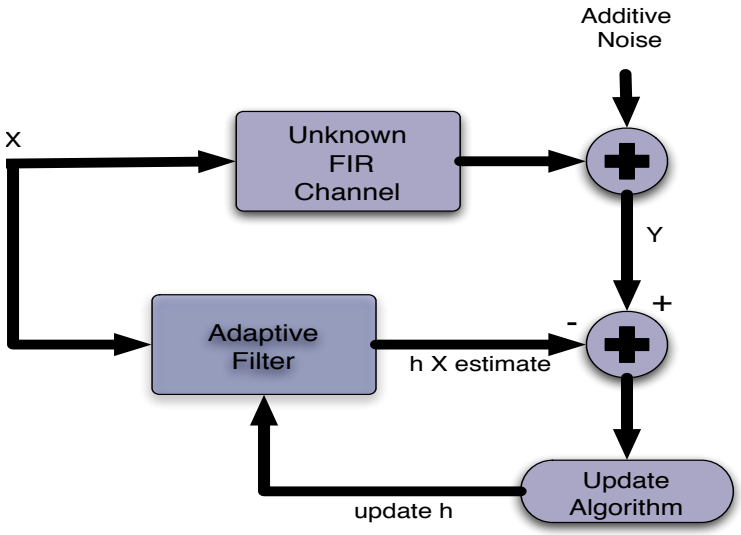


Figure 1.2. Schematic diagram of an adaptive filter used for estimating an unknown FIR channel.

of adaptive filters (e.g., recursive-least-squares [Hay01] or H^∞ [HSK99] filters) have turned out to be extremely difficult to analyze. In fact, the so-called excess mean-square error (EMSE) of these filters is not known except for approximations based on several unrealistic assumptions.

The adaptive filtering problem can be best represented as a special case of the state-space model we introduced earlier in (1.1)

$$\begin{cases} x_{i+1} &= x_i + \mathrm{G}_i u_i, \qquad i \geq 0, \\ y_i &= h_i x_i + v_i, \end{cases} \tag{1.9}$$

in which $y_i \in \mathcal{R}$ and the h_i are called regressor vectors. Two common adaptive algorithms for estimating x_i based on $\{y_{j\leq i}\}$ will be introduced and analyzed later in the thesis. Here we will just mention an example to show the ultimate connection with random Lyapunov and Riccati recursions. The RLS filter estimates x_i recursively

through an estimate update of the form

$$\hat{x}_{i+1} = \hat{x}_i + \frac{\mathrm{P}_i h_i^T}{r + h_i \mathrm{P}_i h_i^T}(y_i - h_i \hat{x}_i), \tag{1.10}$$

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

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T} + q\mathrm{I}, \quad \mathrm{P}_0, \tag{1.11}$$

where $\mathrm{Q}_i = q\mathrm{I}$ is assumed.

Once again, the transient and steady-state behaviors are governed by random Riccati (or Lyapunov as we will see) recursions that bring about a great deal of technical difficulties. A large part of the results in this field are based on an ergodicity assumption[6], which is known to be loose for non-Gaussian regressor vectors [Say03].

In this thesis we will show that any adaptive filtering problem (as far as we have studied) can be tackled successfully through our method. Even though adaptive filtering is a classical area, the viewpoint taken here is entirely novel and almost all the results are new. Moreover, unlike the traditional approaches where usually different analytical techniques are used for different classes of filters, our analysis provides one method applicable to all adaptive filters.

The above examples well emphasize the importance of characterizing the transient behavior and the steady-state statistics of these random matrix-valued recursions in analyzing and synthesizing the systems both currently in use and those rapidly emerging as future technologies. A global framework for studying these types of problems is developed in our work. This framework has proven promising by successfully predicting universal laws in several long-standing open problems.

[6]The ergodicity assumption requires the estimation error covariance matrix, P_i, to satisfy $\mathbb{E}[\mathrm{P}_i] = \left(\mathbb{E}[\mathrm{P}_i^{-1}]\right)^{-1}$.

1.4 Random Lyapunov and Riccati Recursions

As mentioned earlier, the above problems are deeply connected in the sense that they can be represented by a linear time-varying state-space model which ultimately brings up either a random Lyapunov recursion,

$$\Pi_{i+!} = \mathrm{F}_i \Pi_i \mathrm{F}_i^T + \mathrm{G}_i \mathrm{Q}_i \mathrm{G}_i^T, \tag{1.12}$$

or Riccati recursion

$$\mathrm{P}_{i+1} = \mathrm{F}_i \mathrm{P}_i \mathrm{F}_i^T + \mathrm{G}_i \mathrm{Q}_i \mathrm{G}_i^T - \mathrm{F}_i \mathrm{P}_i \mathrm{H}_i^T (\mathrm{R}_i + \mathrm{H}_i \mathrm{P}_i \mathrm{H}_i^T)^{-1} \mathrm{H}_i \mathrm{P}_i \mathrm{F}_i, \quad \mathrm{P}_0, \tag{1.13}$$

where all the coefficient matrices, F_i, G_i, H_i, R_i, and Q_i, can potentially be random.

There are several important questions that may be asked about these recursions: Does the process $\{\mathrm{P}_i\}$ (or $\{\Pi_i\}$) converge? And if it does, what are the statistics of the process at the steady state and the convergence rate? These questions directly relate to the questions about the stability and performance of the system.

Clearly, in contrast to the nonrandom case, $\{\mathrm{P}_i\}$ (or $\{\Pi_i\}$) do not converge to any specific matrix. However by assuming that the coefficient matrices are drawn from jointly stationary matrix-valued random processes, one might suspect that P_i may also converge to a stationary process.

In this thesis our goal is to investigate the possibility of existence of such a stationary process and to determine whether the system exhibits *universal laws*, i.e., whether the overall behavior of the system is independent of the microscopic details of the system and network (such as where measurements or packets are being dropped), but rather dependent only on macroscopic properties, such as the probability of packet drops. We believe such a theory can have a great impact on the analysis and design of estimation and control systems over lossy networks. As we shall see it also leads

to many new results in the theory of random matrices.

1.5 Our Approach

The stationarity assumption that we imposed on the problem hardly restricts the analysis since almost all cases of practical interest fall into this category. It is also reasonable to assume that the state vector size, n, is relatively large. This assumption will enable us to take advantage of many compelling results available in the theory of large random matrices. Although the derivations are carried out for $n \gg 1$, *the law of large numbers* guarantees fast convergence to the asymptotic results for n being as small as 10. Therefore the assumption is justifiable since in the majority of practical scenarios the state vector size satisfies this condition.

A natural way to study a random matrix is to look at its *eigendistribution*. The eigendistribution is defined as the marginal distribution of one randomly selected eigenvalue of the matrix, i.e., $f_{\mathrm{P}}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \Pr\{\lambda_i = \lambda\}$. For a random Lyapunov or Riccati recursion, studying the eigendistribution is important in several aspects. First, $\mathbb{E}[\lambda] = \frac{1}{n}\mathbb{E}[\mathrm{tr}\mathrm{P}_i]$ is nothing but the mean square error performance when P_i represents the error covariance matrix. Moreover, determining the support of the eigendistribution is crucial for finding various performance bounds and studying the system stability. On the other hand, the convergence properties of the eigendistribution in the transient phase directly establish the convergence properties of the recursion itself. Thus in our analysis we mainly focus on characterizing the eigendistributions.

In order to study the eigendistribution, we follow a common practice in the random matrix theory and look at certain transforms of the distribution. A celebrated transform used in the literature, *the Stieltjes transform*, is defined on the complex

plane as [TV04],

$$S_{\mathrm{P}}(z) = \mathbb{E}\left[\frac{1}{\lambda - z}\right] = \int \frac{f_{\mathrm{P}}(\lambda)}{\lambda - z} d\lambda. \tag{1.14}$$

The eigendistribution can be uniquely recovered from its Stieltjes transform through a well-known inversion formula [SC95],

$$f_{\mathrm{P}}(\lambda) = \lim_{\omega \to 0^+} \frac{1}{\pi} \mathrm{Im}\left[S_{\mathrm{P}}(\lambda + j\omega)\right]. \tag{1.15}$$

Therefore by determining the Stieltjes transform of a distribution, we have in fact determined the distribution itself. Moreover, using the Stieltjes transform allows one to directly find the moments of the distribution, $m_k = \mathbb{E}[\lambda^k]$, as the coefficients of the Laurent series expansion of $S(z)$. Therefore in our approach we try to compute the Stieltjes transforms of both sides of a random matrix-valued recursion using techniques from linear algebra and stochastic analysis. This way we find a recursion for the Stieltjes transform of the eigendistribution of P_i through which we can find the steady-state distribution and the transient-phase behavior.

In the remainder, we will present a brief overview of several problems related to the control and estimation over wireless sensor networks and adaptive filtering to which we have successfully applied our method.

1.6 Scope and Contributions of the Thesis

Random Lyapunov Recursions

Random Lyapunov recursions describe the state covariance matrix in a linear state-space model. They also appear in other cases, such as the error covariance matrix propagation in the least-mean-squares (LMS) adaptive filters. Random Lyapunov recursions also appear as lower bounds of random Riccati recursions in Kalman filtering

with intermittent observations [MS08, SSF$^+$04].

Although due to the linearity of the Lyapunov recursion it is, in principle, possible to find all the moments of the eigendistribution by averaging both sides of the recursion after some algebraic manipulations, this is an inefficient method and fails to work for the nonlinear Riccati recursions. Our approach, on the other hand, yields an efficient and systematic way of expressing the eigendistribution and the moments and establishes a framework through which the nonlinear Riccati counterparts can also be analyzed.

Two general forms of random Lyapunov recursions are studied in Chapter 3. For each case the Stieltjes transform at the steady state as well as its time recursion is determined. The moments of the eigendistribution are characterized and the convergence properties are derived from the convergence of the moments. We will also use an alternative R-transform approach to the problem. The simulation results suggest a very close match between the analytical results and the empirical data for n as small as 10 [VH08a].

It can be shown that the error covariance matrix in the least-mean-squares (LMS) filtering satisfies a special random Lyapunov recursion [Say03]. Several variants of LMS filters are studied in chapter 4 under different assumptions on the regressor vectors. While the LMS filter has been studied extensively in the literature, through our approach the stability and performance analysis of all LMS-like algorithms become straightforward [VH08b].

Random Riccati Recursions

Random Riccati recursions manifest the system's performance in numerous important scenarios, and therefore their analysis is of greater significance compared to the Lyapunov ones. At the same time, their nonlinearity makes them much more com-

plicated to study. In fact it is clear that averaging does not deliver the moments anymore. Our results on the eigendistributions of random Riccati recursions appear for the first time in the literature and predict the steady-state statistics accurately. Essentially we have been able to tackle several problems in systems theory which have been unsolved for many years.

Recursive-least-squares (RLS) algorithm is considered in chapter 5. We find the steady-state eigendistribution of the error covariance matrix for an RLS filter (and hence the mean-squared error performance) when the number of measurements is small compared to the state vector size, and with temporally white or shift-structured regressor vectors. Various assumptions on the state update will be considered. We also study the RLS filter with intermittent observations [VH08b] which is motivated by the problem of a sensor sending its measurements through a lossy network. The measurements of a bursty channel in communications can also be captured in this model.

In chapter 6 we extend our analysis to the more complicated case of an RLS filter with multiple measurements where the number of measurements, m, is comparable to the state size. We show that the Stieltjes transform of the eigendistribution at the steady-state satisfies a pair of implicit equations which can be numerically solved to find the distribution [VH09].

Generalized Kalman filtering with intermittent observations is frequently used to model wireless sensor networks. In chapter 7 we investigate this problem. We find the Stieltjes transform of the steady-state eigendistribution as the solution of a system of implicit equations. It is worth mentioning that recently the researchers in the field [KSM09] have realized that through steady-state analysis of these types of problems, one can also establish the stability conditions using *random dynamical systems* theory [Chu02].

Finally in chapter 8 we discuss a few interesting open problems that can potentially be handled in our framework. We also explain how the current analysis can be further extended to achieve a better understanding of the current systems and to design more robust networked control schemes for the emerging applications of such systems.

1.7 Conclusions

In the following chapters, we are going to propose a framework based on the large random matrix theory which is capable of characterizing the steady-state as well as transient behavior of adaptive filters and control and estimation over communication networks. The eigendistributions will be determined in several cases, and by comparison with the Monte Carlo simulation results, this approach can be verified to be accurate. The developed machinery is a considerable progress in the well-established theory of adaptive filters. In fact almost every adaptive filter can be analyzed through our method without any unrealistic assumptions. An important aspect of the results is the universality of them; they depend on the large-scale system parameters and structure rather than the modeling details (such as the distribution of the entries of system matrices). On the other hand, this work is a contribution to the random matrix theory since, to the best of our knowledge, the likes of the eigendistributions we find have never appeared in the literature before. As we will see, this method is successful in the analysis and synthesis of control and estimation over communication networks which has been an active area in recent years.

Chapter 2

An Overview of Large Random Matrix Theory

2.1 Introduction

In the past century, random matrix theory has been used in various fields, such as physics, statistics, engineering, and finance. A complete list of all the problems which have successfully been tackled using these tools is well beyond this introduction. However, it suffices to say that it includes problems in information theory, neural networks, and condensed matter physics.

In this section we give an overview of pertinent definitions and results from random matrix theory. A much more comprehensive review of the subject toward applications in wireless communications theory can be found in [TV04]. Those interested in a more detailed and coherent analytical account of the theory should consult [Meh91].

This chapter is organized as follows. In section 2.2 we start by giving the basic definitions and notational conventions used throughout this and later chapters. Section 2.3 introduces several transforms of the eigenvalue distribution function which are used extensively in the literature, particularly the so-called Stieltjes transform. A very important property of large random matrices, the self-averaging property, is introduced in section 2.4, along with an example of its application. Finally in sec-

tion 2.5 we present the notion of freeness which replaces the independence notion for random variables which are noncommutative, such as random matrices.

2.2 Preliminaries

Just as a matrix is a collection of objects in a rectangular array, a random matrix is nothing other than the joint distribution of its elements,

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}, \tag{2.1}$$

$$f_{\mathrm{A}} = f_{\mathbf{a_{11}},\mathbf{a_{12}},\cdots,\mathbf{a_{mn}}}(a_{11}, a_{12}, \cdots, a_{mn}). \tag{2.2}$$

Often, the most important questions to be answered about a random matrix ensemble concern the distribution of the eigenvalues. For a realization of an $n \times n$ random matrix, A, the *empirical cumulative distribution* function of the eigenvalues is defined as,

$$F_{\mathrm{A}}(\lambda) = \frac{1}{n} \sum_{l=1}^{n} \mathbb{1} \left(\lambda_l(\mathrm{A}) \leq \lambda \right), \tag{2.3}$$

where $\lambda_l(\mathrm{M})$ denotes the l-th eigenvalue of M and $\mathbb{1}(\cdot)$ is the indicator function. If we replace $\mathbb{1}(\cdot)$ with $\Pr(\cdot)$, the probability measure over the random matrix ensemble, we will have the *expected CDF*. A *density function*, $f_A(\lambda)$, can be associated with this latter cumulative distribution. It would then represent the expected marginal distribution of one (randomly selected) eigenvalue of one sample of the random matrix ensemble,

$$f_{\mathrm{A}}(\lambda) = \frac{1}{n} \sum_{l=1}^{n} \Pr \left\{ \lambda_l(\mathrm{A}) = \lambda \right\}, \tag{2.4}$$

This density function is frequently referred to as the *eigendistribution* of A.

When considering a random matrix family, $\left(\mathcal{A}^{(n)}\right)$, where $\mathrm{A}^{(n)} \in \mathcal{A}^{(n)}$ are $n \times n$ matrices, one can ask whether the empirical eigendistributions converge to a nonrandom distribution as n grows. In other words,

$$f_{\mathrm{A}^{(n)}} \to f_{\mathrm{A}}. \tag{2.5}$$

In that case, we call f_{A} the asymptotic eigendistribution of $\left(\mathcal{A}^{(n)}\right)$.

Starting with the work of Wishart [Wis28] in the 1920s, random matrices were used in multivariate statistical analysis. For decades, the analysis was concentrated on fixed-dimensional matrices. The majority of results on the eigendistribution of fixed-dimensional random matrices are complicated and offer little insight, and are mainly limited to Gaussian random matrices or matrices derived from them [Meh91]. Starting with the work of Wigner in physics in the 1950s [Wig55, Wig58], it was realized that as soon as the matrix dimensions are allowed to grow, one usually finds simple, closed-form expressions for the eigendistribution that behave like *universal laws*, i.e., they depend on the matrix structure and statistics rather than the exact distribution of the entries. For example, the semicircle law of section (2.4) only requires the entries to be i.i.d. with zero mean and bounded second and fourth moments; the actual distribution of the elements does not matter.

Moreover, although the derivations are carried out for $n \gg 1$, *the law of large numbers* guarantees fast convergence of the expected eigendistribution to the asymptotic results for n being as small as $10 - 20$. The convergence rate can be found to be of order $O\left(n^{-\gamma/n}\right)$ for some $\gamma > 1$ (see, e.g., [Bai93a, Bai93b]). As a matter of fact, random matrices also show an ergodic behavior as $n \to \infty$ in the sense that the distribution of the eigenvalues of a single realization of the ensemble (empirical distribution) looks like the expected and consequently asymptotic eigendistribution.

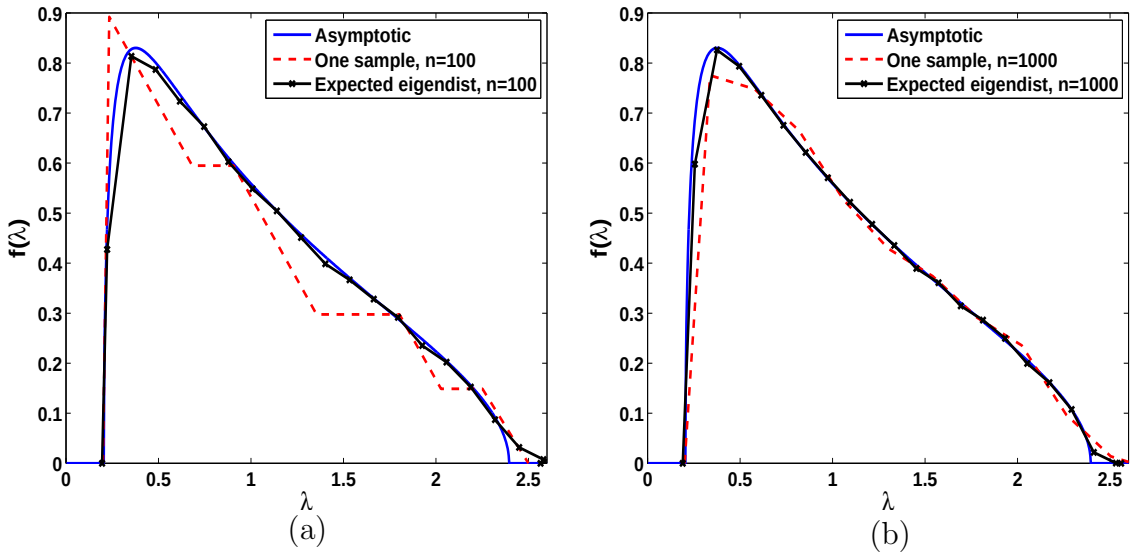


Figure 2.1. Comparison of the asymptotic eigendistribution with the empirical and expected distributions when $\beta = 0.3$ and $n = 100, 1000$.

In order to show this, let us consider a Wishart matrix,

$$\mathrm{W} = H^T H, \tag{2.6}$$

where H is an $m \times n$ matrix with independent $\mathcal{N}(0, \frac{1}{\sqrt{n}})$ entries. It can be shown that [TV04],

$$f_{\mathrm{W}}(\lambda) = \left(1 - \frac{1}{\beta}\right)^+ \delta(\lambda) + \frac{\sqrt{(\lambda - a)^+ (b - \lambda)^+}}{2\pi\beta\lambda}, \tag{2.7}$$

where $\beta = \frac{m}{n}$, $a = (1 - \sqrt{\beta})^2$, and $b = (1 + \sqrt{\beta})^2$. In figure 2.1 this asymptotic eigendistribution is plotted along with the distributions of the eigenvalues for only one matrix selected from the ensemble for $n = 100$ and $n = 1000$. We have also plotted the expected eigendistribution (approximately) by taking into account the eigenvalues of 100 samples. As can be seen, the expected eigendistribution converges quickly to the asymptotic one, as does the empirical one, although slower.

In the sequel, we denote both the asymptotic and empirical eigendistributions by simply eigendistribution for the sake of brevity whenever there is no ambiguity.

2.3 Transforms

It turns out that in the random matrix arena, results on the eigendistribution of random matrices can usually be expressed much more efficiently in terms of certain transforms of $f_{\mathrm{A}}(\lambda)$. There are several transforms used in the literature: the Shannon transform, Stieltjes transform, η-transform, etc. The choice of transform depends on the structure of the problem and the underlying matrices. For example in the multi input multi output communications, it is natural to use either the η-transform or the Shannon transform. In the problems that we are going to consider, the Stieltjes transform, R-transform, and S-transform turn out to be handy. We introduce these three transforms briefly in this section. A complete list of transforms and their relations with each other can be found in [TV04].

2.3.1 Stieltjes Transform

The most ubiquitous transform used in the random matrix literature is the so-called *Stieltjes transform* which was first used in the seminal work of Marčenko and Pastur [MP67]. The Stieltjes transform of the eigendistribution of a random matrix A– interchangeably referred to as the Stieltjes transform of the matrix itself–is defined on the complex plane as[1],

$$S_{\mathrm{A}}(z) = \mathbb{E}\left[\frac{1}{\lambda - z}\right] = \int \frac{f_{\mathrm{A}}(\lambda)}{\lambda - z} d\lambda. \tag{2.8}$$

for all z in the complex plane, except for the support of $f_{\mathrm{A}}(.)$. Having the Stieltjes transform, the eigendistribution can be uniquely retrieved through its inverse formula [SC95],

$$f_{\mathrm{A}}(\lambda) = \lim_{\omega \to 0^+} \frac{1}{\pi} \mathrm{Im}\left[S_{\mathrm{A}}(\lambda + j\omega)\right]. \tag{2.9}$$

[1]Clearly this definition is not limited to the eigendistribution of random matrices and one can consider the Stieltjes transform of any real random variable.

In other words, one should look at the imaginary part of the Stieltjes transform very close to the real line. In fact (2.9) is valid whenever the cumulative distribution function is differentiable. Otherwise a more general inverse can be used [Bai93a],

$$F_{\mathrm{A}}(\lambda_2) - F_{\mathrm{A}}(\lambda_1) = \lim_{\omega \to 0^+} \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} \mathrm{Im}\left[S_{\mathrm{A}}(\lambda + j\omega)\right] d\lambda. \tag{2.10}$$

The main reason for the Stieltjes transform to be a handy tool in random matrix theory is that it can often be directly computed from the random matrix itself, rather than determining the eigenvalues first and then finding their distribution.

Lemma 2.3.1 *The Stieltjes transform of the expected eigendistribution of a Hermitian random matrix A can be written as,*

$$S_A(z) = \mathbb{E}\frac{1}{n} tr\left(A - zI\right)^{-1}. \tag{2.11}$$

Proof: Since A is Hermitian, we can diagonalize it as $\mathrm{A} = \mathrm{U}\Lambda U^*$. Then,

$$\mathbb{E}\frac{1}{n}\mathrm{tr}\left(\mathrm{U}\Lambda U^* - z\mathrm{I}\right)^{-1} \;=\; \mathbb{E}\frac{1}{n}\mathrm{tr}\left(\Lambda - z\mathrm{I}\right)^{-1} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z}.$$

On the other hand,

$$S_{\mathrm{A}}(z) \;=\; \int \frac{dF_{\mathrm{A}}(\lambda)}{\lambda - z} = \int d\left[\mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left(\lambda_i \leq \lambda\right)\right]\frac{1}{\lambda - z} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z},$$

which completes the proof. □

Equation (2.11) can equivalently be written in another useful form,

$$S_{\mathrm{A}}(z) = -\frac{d}{dz}\mathbb{E}\frac{1}{n}\log\det(\mathrm{A} - z\mathrm{I}). \tag{2.12}$$

An example of how these expressions can be used to determine the eigendistribu-

tion can be found in subsection 2.4.

Another important property of the Stieltjes transform is how we can obtain the moments of the eigendistribution from the power series expansion of $S_{\mathrm{A}}(z)$. The Stieltjes transform of an eigendistribution with bounded support can be expanded as a Laurent series[2] in terms of z^{-1},

$$S_{\mathrm{A}}(z) = \mathbb{E}\left[\frac{1}{\lambda - z}\right] = -\frac{1}{z}\mathbb{E}\left[\frac{1}{1 - \frac{\lambda}{z}}\right] = -\frac{1}{z}\mathbb{E}\sum_{i=0}^{\infty}\frac{\lambda^i}{z^i} = -\sum_{i=0}^{\infty}\frac{m_i}{z^{i+1}}, \tag{2.13}$$

where $m_i = \mathbb{E}\left[\lambda^i\right]$ is the i^{th} moment of the distribution and $m_0 = 1$. Similarly, when the eigendistribution is bounded away from zero, a Taylor series expansion yields the moments $m_{-i} = \mathbb{E}\left[\lambda^{-i}\right]$,

$$S_{\mathrm{A}}(z) = \sum_{i=0}^{\infty} m_{-i} z^{i-1}. \tag{2.14}$$

Throughout the thesis, we commonly refer to the Stieltjes transform of the asymptotic eigendistribution of a matrix as the Stieltjes transform of the matrix itself.

2.3.2 R-Transform

The R-transform is defined on the complex plane through the functional inverse of the Stieltjes transform,

$$R_{\mathrm{A}}(z) = S_{\mathrm{A}}^{<-1>}(-z) - \frac{1}{z}, \tag{2.15}$$

where $< -1 >$ denotes the functional inverse. This equivalently can be written as,

$$S_{\mathrm{A}}(z) = \frac{1}{R_{\mathrm{A}}\left(-S_{\mathrm{A}}(z)\right) - z}. \tag{2.16}$$

The R-transform is an important object when considering the notion of *free probability* (section 2.5). It can be shown that it replaces the cumulant generating function

[2]The series converges outside a circle which includes the support of $f_{\mathrm{A}}(\lambda)$.

(logarithm of the characteristic function) for free random variables.

Example 2.3.2 *The Stieltjes transform of the eigendistribution of the Wishart matrix in (2.7) (also known as the Marčenko-Pastur law) can be found as,*

$$S_W(z) = \frac{1-\beta-z+\sqrt{z^2-2(\beta+1)z+(\beta-1)^2}}{2\beta z}, \tag{2.17}$$

with the R-transform being,

$$R_W(z) = \frac{1}{1-\beta z}. \tag{2.18}$$

For an eigendistribution with bounded support on the real line, the R-transform will have a Taylor series representation,

$$R_{\mathrm{A}}(z) = \sum_{i=1}^{\infty} k_i z^{i-1}, \tag{2.19}$$

where the k_i are referred to as the *free cumulants*[3]. The moments and free cumulants are related to each other as described in the following lemma.

Lemma 2.3.3 (Free Cumulants and the Moments) *The moments of the distribution can be recursively obtained through its free cumulants via the so-called free cumulants formula,*

$$m_n = \sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r \geq 0 \\ i_1+\cdots+i_r=n-r}} k_r m_{i_1} \cdots m_{i_r}. \tag{2.20}$$

Proof: A combinatorial proof which offers several insights can be found in [Spe97]. The lemma can be also proved by power series expansion of (2.16). □

2.3.3 S-Transform

Another important transform especially in the free probability domain is the S-transform. The S-transform can also be defined through the functional inverse of

[3]Counterpart of the cumulants for commutative random variables.

the Stieltjes transform as [Voi87],

$$\Sigma_{\mathrm{A}}(z) = \frac{z+1}{z}\left(\sum_{i=1}^{\infty} m_i z^i\right)^{<-1>} = \frac{z+1}{z}\left(-\frac{1}{z}S_{\mathrm{A}}\left(\frac{1}{z}\right) - 1\right)^{<-1>}, \tag{2.21}$$

where the m_i are the moments of the eigendistribution and $< -1 >$ denotes the functional inverse. It can be directly shown that the Stieltjes transform and the S-transform satisfy

$$\Sigma_{\mathrm{A}}(z) = -\frac{1}{z}S_{\mathrm{A}}\left(\frac{1+z}{z\Sigma_{\mathrm{A}}(z)}\right). \tag{2.22}$$

The S-transform is handy when looking at the eigendistribution of the multiplication of two random matrices. It does not have a counterpart in the realm of commutative random variables since for those variables one can simply translate the problem of multiplication of two random variables to an addition problem by using exponentiating.

Example 2.3.4 *The S-transform of the eigendistribution of the Wishart matrix in (2.7) can be found as*

$$\Sigma_W(z) = \frac{1}{1+\beta z}. \tag{2.23}$$

2.4 Self-Averaging Property

A useful property of the Stieltjes transform in the asymptotic regime is the *self-averaging* property which is stated in the follwing Lemma [TV04]:

Lemma 2.4.1 (Self-Averaging) *Let A be an $n \times n$ positive semidefinite random matrix. If the empirical eigendistribution of A almost surely converges to its mean value as $n \to \infty$, i.e.,*

$$\lim_{n\to\infty} \frac{1}{n} tr(A - zI)^{-1} = S_A(z) \quad a.s. \tag{2.24}$$

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

$$\lim_{n\to\infty} \frac{1}{n} x^T (A - zI)^{-1} x = S_A(z) \quad a.s. \tag{2.25}$$

Sketch of proof: The proof relies on lemma 2.7 of [BS98], which states that for x as described above and for a Hemitian positive semidefinite matrix C, we have for any $p \geq 2$

$$\mathbb{E}|x^T \mathrm{C} x - \mathrm{trC}|^p \leq k_p \left(\left(\mathbb{E}|x_1|^4 \mathrm{trCC}^T \right)^{\frac{p}{2}} + \mathbb{E}|x_1|^{2p} \mathrm{tr}(\mathrm{CC}^T)^{\frac{p}{2}} \right), \tag{2.26}$$

where k_p is a constant, x_1 represents one element of x and expectation is over x. Since all the higher moments of x are assumed to be bounded and $\mathrm{tr}(\mathrm{CC}^T)^{\frac{p}{2}} \leq (\mathrm{trCC}^T)^{\frac{p}{2}}$, we can rewrite (2.26) as

$$\mathbb{E}|x^T \mathrm{C} x - \mathrm{trC}|^p \leq k'_p \left(\mathrm{trCC}^T \right)^{\frac{p}{2}}. \tag{2.27}$$

Dividing both sides by n,

$$\mathbb{E}|\frac{1}{n} x^T \mathrm{C} x - \frac{1}{n}\mathrm{trC}|^p \leq \frac{1}{n^{p/2}} k'_p \left(\frac{1}{n} \mathrm{trCC}^T \right)^{\frac{p}{2}}. \tag{2.28}$$

In our case, $\mathrm{C} = (\mathrm{A} - z\mathrm{I})^{-1}$ for which one can show that,

$$\frac{d}{dz}(\mathrm{A} - z\mathrm{I})^{-1} = -(\mathrm{A} - z\mathrm{I})^{-1} \left(\frac{d}{dz}(\mathrm{A} - z\mathrm{I}) \right) (\mathrm{A} - z\mathrm{I})^{-1} = (\mathrm{A} - z\mathrm{I})^{-2}. \tag{2.29}$$

Therefore,

$$\frac{1}{n}\mathrm{tr}(\mathrm{A} - z\mathrm{I})^{-2} = \frac{d}{dz}\frac{1}{n}\mathrm{tr}(\mathrm{A} - z\mathrm{I})^{-1} = \frac{d}{dz} S_{\mathrm{A}_n}(z), \tag{2.30}$$

which together with (2.28) results in,

$$\mathbb{E}|\frac{1}{n}x^T(\mathrm{A}-z\mathrm{I})^{-1}x-\frac{1}{n}\mathrm{tr}(\mathrm{A}-z\mathrm{I})^{-1}|^p\leq\frac{k_p''}{n^{p/2}}. \tag{2.31}$$

The first moment of $\frac{1}{n}x^T(\mathrm{A}-z\mathrm{I})^{-1}x$ can be easily determined as,

$$\mathbb{E}\frac{1}{n}x^T(\mathrm{A}-z\mathrm{I})^{-1}x=\mathbb{E}\,\mathrm{tr}\frac{1}{n}(\mathrm{A}-z\mathrm{I})^{-1}xx^T=\mathbb{E}\frac{1}{n}\sum_{i,j=1}^{n}\left[(\mathrm{A}-z\mathrm{I})^{-1}\right]_{i,j}x_ix_j=\frac{1}{n}\mathrm{tr}(\mathrm{A}-z\mathrm{I})^{-1}. \tag{2.32}$$

This shows the convergence in moments for any $p\geq 1$. See [BS98] and the references therein for the details of the proof of almost sure convergence. □

As we will see throughout this manuscript, the self-averaging lemma proves to be a very strong tool in analyzing the asymptotic eigendistribution of random matrices. In the following subsection we show how one can find the celebrated semicircle law for the Wigner matrices using this lemma and some standard matrix algebra.

An Example: The Semicircle Law

The first successful result on the asymptotic eigendistribution of large random matrices goes back to the work of Wigner in the 1950s [Wig55, Wig58, Wig59] when he studied self-adjoint Hamiltonian operators. In his initial work he looked at symmetric matrices with zero diagonal entries and independent–save for the symmetry–off-diagonal entries which can be $\{-1,1\}$ with equal probability. He later extended his results for entries having a Gaussian distribution and further to matrices which are now called Wigner matrices. A Wigner matrix is a Hermitian matrix whose upper-triangular entries are independent with zero-mean and equal variance. If the variance of the entries is $\frac{1}{n}$ then the matrix is called a standard Wigner matrix.

Consider a standard $n\times n$ Wigner matrix, W. The Stieltjes transform of W can

be found using (2.11),

$$S_{\mathrm{W}}(z) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\mathrm{tr}\left(\mathrm{W} - z\mathrm{I}\right)^{-1}. \tag{2.33}$$

We break W as,

$$\mathrm{W} = \begin{bmatrix} w_{11} & w_{21}^T \\ w_{21} & \mathrm{W}_{22} \end{bmatrix}, \tag{2.34}$$

where w_{21} is an $(n-1)$-dimensional column vector. Now,

$$S_{\mathrm{W}}(z) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\mathrm{tr}\left(\begin{bmatrix} w_{11} - z & w_{21}^T \\ w_{21} & \mathrm{W}_{22} - z\mathrm{I} \end{bmatrix}\right)^{-1}. \tag{2.35}$$

Since the distribution of W is isotropic, without loss of generality we can look at only one diagonal entry of the above inverse, say the $(1,1)$ entry. It can be easily found in terms of the Schur complement of $(\mathrm{W}_{22} - z\mathrm{I})$, i.e.,

$$S_{\mathrm{W}}(z) = \lim_{n\to\infty} \mathbb{E}\frac{1}{w_{11} - z - w_{21}^T(\mathrm{W}_{22} - z\mathrm{I})^{-1}w_{21}}. \tag{2.36}$$

Using the self-averaging lemma, the third term in the denominator can be simplified to

$$\lim_{n\to\infty} w_{21}^T(\mathrm{W}_{22} - z\mathrm{I})^{-1}w_{21} = S_{\mathrm{W}}(z). \tag{2.37}$$

(Note that W_{22} is an $(n-1)\times(n-1)$ standard Wigner matrix and the entries of w_{21} have variance of $\frac{1}{n}$.) Moreover, since w_{11} is a zero-mean random variable with a vanishing variance in the limit, (2.36) can be written as,

$$S_{\mathrm{W}}(z) = \frac{1}{-z - S_{\mathrm{W}}(z)}. \tag{2.38}$$

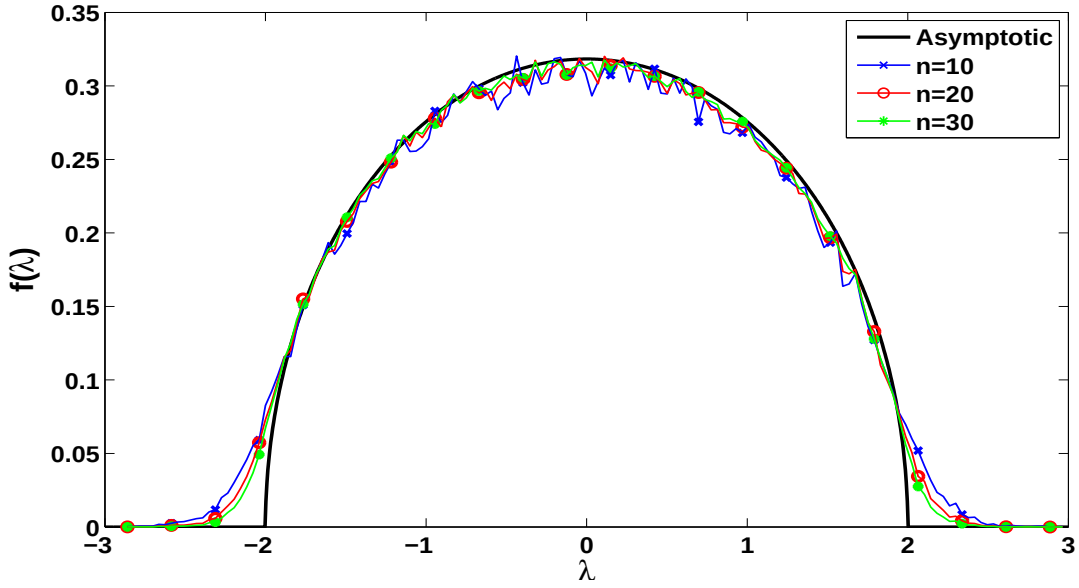


Figure 2.2. Semicircle law compared with the empirical expected eigendistribution for $n = 10, 20, 30$.

Thus,

$$S_{\mathrm{W}}(z) = \frac{-z \pm \sqrt{z^2 - 4}}{2}. \tag{2.39}$$

In order to find the asymptotic eigendistribution through the inversion formula (2.9), we should look at the imaginary part of the Stieltjes transform very close to the real line. Clearly the imaginary part of $S_{\mathrm{W}}(z \to \lambda + j0^+)$ is nonzero only for $-2 \leq \lambda \leq 2$, which immediately determines the support. The sign of the square root should be selected such that the distribution takes positive values. The resulting eigendistribution is the celebrated semicircle law,

$$f_{\mathrm{W}}(\lambda) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - \lambda^2} & \text{when } -2 \leq \lambda \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.40}$$

In figure 2.2 we have plotted $f_{\mathrm{W}}(\lambda)$ along with the expected eigendistribution for different values of n. The empirical curves are obtained by looking at the eigenvalues

of 1000 realizations of the Wigner matrix. As can be seen, even for $n = 10$ the asymptotic result is quite accurate in predicting the distribution.

2.5 Free Probability

Freeness replaces the notion of independence for noncommutative random variables, such as random matrices. Although defined in a more involved fashion compared to independence, once the freeness of two random variables is established, there exist systematic methods to compute the distribution of their summation and product. Free probability was developed by D. Voiculescu in the 1980s [Voi83] in his work on the operator theory. Later on it was realized that the notion could be extended to random matrices [4] [Voi87]. In order to define freeness, let us first define an asymptotic expectation functional for Hermitian matrices,

$$\phi(\mathrm{A}) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\mathrm{trA}. \tag{2.41}$$

Now the asymptotic freeness of two random matrices is defined as follows [TV04].

Definition 2.5.1 (Asymptotic Freeness) *Two Hermitian random matrices A and B are asymptotically free if for any r and all polynomials $p_i(\cdot)$ and $q_i(\cdot)$ $(1 \leq i \leq r)$ such that,*

$$\phi\left(p_i(A)\right) = \phi(q_i(B)) = 0, \tag{2.42}$$

the expected trace of the ordered multiplication of these matrix polynomials also vanishes, i.e.,

$$\phi(p_1(A)q_1(B)\cdots p_r(A)q_r(B)) = 0. \tag{2.43}$$

[4] Technically we consider asymptotic freeness for random matrices.

This definition can be generalized to several random matrices by putting a condition on the multiplication of the polynomials of alternating matrices [TV04]. There exist numerous results in the literature which state the freeness of certain classes of matrices. For example all random matrices and the identity matrix are asymptotically free. However it is not usually easy to prove the freeness of two random matrices.

It is worth mentioning that independent random variables are not necessarily free and vice versa. For example $x\mathrm{I}_n$ and $y\mathrm{I}_n$ where x and y are independent zero-mean random variables are not asymptotically free.

As mentioned earlier, the notion of freeness facilitates computing of the asymptotic eigendistribution of the summation and product of two random matrices. In fact one can find these asymptotic eigendistributions directly from the eigendistributions of the original matrices without getting involved in the eigenvector structure.

Theorem 2.5.2 (Sum of Free Random Matrices) *Let X and Y be two asymptotically free random matrices and $Z = X + Y$. The R-transform of the eigendistribution of Z satisfies*

$$R_Z(z) = R_X(z) + R_Y(z). \tag{2.44}$$

Proof: See [Voi86] or [Spe97] for a detailed proof of the theorem. □

Recalling (2.19), the free cumulants of $f_{\mathrm{Z}}(\lambda)$ can be written as the sum of the corresponding free cumulants of $f_{\mathrm{X}}(\lambda)$ and $f_{\mathrm{Y}}(\lambda)$. This in fact is the reason behind the naming since the coefficients in the power series expansion of the logarithm of the moment-generating function (cumulants) have the same property for independent commuting random variables.

Similarly the following theorem regarding the eigendistribution of multiplication of free random matrices is well known in the literature.

Theorem 2.5.3 (Multiplication of Free Random Matrices) *Let X and Y be two nonnegative asymptotically free random matrices and $Z = XY$. The S-transform of*

the eigendistribution of Z satisfies

$$\Sigma_Z(z) = \Sigma_X(z)\Sigma_Y(z). \tag{2.45}$$

Proof: See [Voi87] or [Spe97] for a detailed proof of the theorem. □

For a detailed list of free random matrices and simple examples of how the above theorems can be employed to derive interesting results look at [TV04]. Free probability has been an active area of research in the past decade. It has proven to be not only a powerful tool, but also a great source of intuition whenever we are dealing with large random matrices. The next chapter shows how the theory of large random matrices can be used to analyze the steady-state and transient behavior of random Lyapunov recursions.

Chapter 3

Random Lyapunov Recursions

3.1 Introduction

Lyapunov matrix recursions are named after the Russian mathematician Alexander Mikhailovitch Lyapunov who introduced the stability theory of linear and nonlinear systems in his doctoral dissertation (see [Lya92] for English translation). Although Lyapunov's work was published in 1892, his theory did not receive much attention until the 1960s. In the past 40 years, Lyapunov recursions have appeared frequently in many fields such as control theory, system theory, signal processing, and boundary value problems. Due to their numerous applications, time-invariant and time-varying Lyapunov recursions are very well studied [GQ95]. Their stability and explicit steady-state solutions are known in the literature. Moreover, a great body of research was devoted to studying efficient numerical methods for solving Lyapunov and Lyapunov-like equations.

One of the most well-known appearances of Lyapunov recursions is in describing the state covariance matrix in linear state-space models. Consider,

$$\left\{ \begin{array}{rcl} x_{i+1} & = & \mathrm{F}_i x_i + \mathrm{G}_i u_i \\ y_i & = & \mathrm{H}_i x_i + v_i \end{array} \right. , \quad \mathbb{E} \left[\begin{array}{c} u_i \\ v_i \end{array} \right] \left[\begin{array}{cc} u_j^T & v_j^T \end{array} \right] = \left[\begin{array}{cc} \mathrm{Q}_i & 0 \\ 0 & \mathrm{R}_i \end{array} \right] \delta_{ij}, \qquad (3.1)$$

in which $x_i \in \mathcal{R}^n$ is the unknown state vector to be estimated through the measurements $y_i \in \mathcal{R}^m$. Here u_i and v_i denote zero-mean input process and measurement noises respectively. Then it is straightforward to show that the state covariance matrix, $\Pi_i = \mathbb{E}x_i x_i^T$ satisfies

$$\Pi_{i+1} = \mathrm{F}_i \Pi_i \mathrm{F}_i^T + \mathrm{G}_i \mathrm{Q}_i \mathrm{G}_i^T. \tag{3.2}$$

For a complete analysis of the above recursion for time-invariant systems, see [LR95]. In this chapter we will study such recursions when the coefficient matrices are not only time varying, but indeed random. We will assume that those random coefficients represent jointly stationary matrix-valued random processes. There are several motivations to study such recursions. In estimation and control over communication networks, we frequently use linear random state-space models of the form (3.1). Clearly in this case the state covariance matrix undergoes a random Lyapunov recursion. On the other hand, random Lyapunov recursions also appear as lower bounds of random Riccati recursions in distributed Kalman filtering [MS08], which is a problem of great importance. Random Lyapunov-like recursions also arise in other applications such as least-mean-squares (LMS) adaptive filtering. The latter case is studied in the next chapter.

In this chapter we analyze the general form of random Lyapunov recursions under various assumptions on the model parameters, such as F_i being a diagonal or a multiple of identity. By computing the Stieltjes transform of the eigendistribution we find the eigendistribution and its moments at the steady state, as well as the convergence properties of the distribution of eigenvalues.

It is worth mentioning that due to the linearity of the Lyapunov recursion, it is in principle possible to find all the moments of the eigendistribution by averaging both sides of (3.2) after some algebraic manipulations. However this is an inefficient

method with little insight and fails to work for the nonlinear Riccati recursions. Our approach, on the other hand, yields an efficient and systematic way of expressing the eigendistribution and the moments and establishes a framework through which the nonlinear Riccati counterparts can also be analyzed.

This chapter is organized as follows. In section 3.2 we will look at the steady-state as well as the transient behavior of a Lyapunov recursion with $\mathrm{F} = \sqrt{\alpha}\mathrm{I}$. We will derive the moments of the eigendistribution as well. The transient behavior of the recursion is studied in section 3.3. In section 3.4 we will analyze the general case with F_i being an arbitrary matrix. Simulation results are provided which show the close prediction of both the moments and the eigendistribution even for relatively small values of n. Finally section 3.5 concludes the chapter.

3.2 Random Lyapunov Recursions with $\mathbf{F} = \sqrt{\alpha}\mathbf{I}$

In this section we will assume that F_i is simply a scaled version of identity, i.e., $\mathrm{F}_i = \sqrt{\alpha}\mathrm{I}$ and $\mathrm{Q}_i = \mathrm{I}$. Therefore the random Lyapunov recursion will be

$$\Pi_{i+1} = \alpha\Pi_i + \mathrm{G}_i\mathrm{G}_i^T, \tag{3.3}$$

where $\alpha \in (0, 1)$ and the G_i are independently drawn $n \times m$ matrices with i.i.d. entries having zero mean, $\frac{1}{\sqrt{m}}$ variance and bounded higher moments. We can find a recursion for the Stieltjes transform of Π_i as $n \to \infty$.

Theorem 3.2.1 *The Stieltjes transform of* Π_i *in (3.3) satisfies*

$$S_{i+1}(z) = \frac{1}{\alpha} S_i\left(\frac{z}{\alpha} - \frac{\frac{\beta}{\alpha}}{\beta + S_{i+1}(z)}\right), \tag{3.4}$$

where $\beta = \frac{m}{n}$.

Proof: We first apply the definition of the Stieltjes transform as in (2.11) to both sides of (3.3),

$$S_{i+1}(z) = \frac{1}{n}\mathbb{E}\ \mathrm{tr}\left(\alpha\Pi_i + \mathrm{G}_i\mathrm{G}_i^T - z\mathrm{I}\right)^{-1}. \tag{3.5}$$

Since G_i has an isotropic distribution, we can safely diagonalize Π_i. Thus $\Pi_i = \Lambda_i$. Now break Π_i and G_i as follows,

$$\Pi_i = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \quad , \quad \mathrm{G}_i = \begin{bmatrix} g_1 \\ \mathrm{G}_2 \end{bmatrix}, \tag{3.6}$$

where g_1 is an m-dimensional row vector and λ_1 is the first eigenvalue. We have also dropped the index i for simplicity. Now we can rewrite (3.5) as,

$$S_{i+1}(z) = \frac{1}{n}\mathbb{E}\ \mathrm{tr}\begin{bmatrix} \alpha\lambda_1 + g_1g_1^T - z & g_1\mathrm{G}_2^T \\ \mathrm{G}_2g_1^T & \alpha\Lambda_2 + \mathrm{G}_2\mathrm{G}_2^T - z\mathrm{I} \end{bmatrix}^{-1}. \tag{3.7}$$

We will not index the identity matrix with its size as long as it is clear from the context. Since we are looking for the diagonal entries of the above inverse matrix and they are all statistically the same, without loss of generality we consider the first entry. The first entry of the inverse can be easily written through the Schur complement of the $\alpha\Lambda_2 + \mathrm{G}_2\mathrm{G}_2^T - z\mathrm{I}$ as follows,

$$S_{i+1}(z) = \mathbb{E}\ \frac{1}{\alpha\lambda_1 + g_1g_1^T - z - g_1\mathrm{G}_2^T(\alpha\Lambda_2 + \mathrm{G}_2\mathrm{G}_2^T - z\mathrm{I})^{-1}\mathrm{G}_2g_1^T}. \tag{3.8}$$

Using matrix inversion lemma, we can rewrite the denominator as,

$$S_{i+1}(z) = \mathbb{E}\ \frac{1}{\alpha\lambda_1 - z + \underbrace{g_1(\mathrm{I}_m + \mathrm{G}_2^T(\alpha\Lambda_2 - z\mathrm{I}_{n-1})^{-1}\mathrm{G}_2)^{-1}g_1^T}_{\triangleq t(z)}}. \tag{3.9}$$

Now let us focus on computing $t(z)$. First of all, g_1 is independent of the inverse matrix in the middle and it has zero mean and $\frac{1}{\sqrt{m}}$ variance. Therefore using the self-averaging lemma,

$$t(z) = \frac{1}{m}\mathbb{E}\ \mathrm{tr}\ (\mathrm{I}_m + \mathrm{G}_2^T(\alpha\Lambda_2 - z\mathrm{I}_{n-1})^{-1}\mathrm{G}_2)^{-1}. \tag{3.10}$$

Following the technique used in (3.6), we break G_2 as,

$$\mathrm{G}_2 = \begin{bmatrix} g_{21} & \mathrm{G}_{22} \end{bmatrix}, \tag{3.11}$$

where $g_{21} \in \mathcal{R}^{n-1}$ is a column vector and $\mathrm{G}_{22} \in \mathcal{R}^{n-1\times m-1}$. Then,

$$t(z) = \frac{1}{m}\mathbb{E}\ \mathrm{tr}\begin{bmatrix} 1 + g_{21}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}g_{21} & g_{21}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}\mathrm{G}_{22} \\ \mathrm{G}_{22}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}g_{21} & \mathrm{I}_{m-1} + \mathrm{G}_{22}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}\mathrm{G}_{22} \end{bmatrix}^{-1}. \tag{3.12}$$

Once again, we look at the first diagonal entry of the above inverse matrix,

$$t(z) = \mathbb{E}\frac{1}{t_1(z)} \tag{3.13}$$

where,

$$\begin{aligned} t_1(z) \ &= \ 1 + g_{21}^T\left[(\alpha\Lambda_2 - z\mathrm{I})^{-1}\right. \\ &\quad \left. -(\alpha\Lambda_2 - z\mathrm{I})^{-1}\mathrm{G}_{22}\left(\mathrm{I} + \mathrm{G}_{22}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}\mathrm{G}_{22}\right)^{-1}\mathrm{G}_{22}^T(\alpha\Lambda_2 - z\mathrm{I})^{-1}\right] g_{21}, \end{aligned} \tag{3.14}$$

which can be easily simplified through the matrix inversion lemma to,

$$t(z) = \mathbb{E}\frac{1}{1 + g_{21}^T\left[\alpha\Lambda_2 - z\mathrm{I}_n + \mathrm{G}_{22}\mathrm{G}_{22}^T\right]^{-1} g_{21}}. \tag{3.15}$$

Since g_{21} is independent of the inverse matrix in the middle and it has i.i.d. zero-mean, $\frac{1}{\sqrt{m}}$ entries, we have,

$$t(z) = \mathbb{E}\frac{1}{1 + \frac{1}{m}\mathbb{E}\ \mathrm{tr}\left(\alpha\Lambda_2 - z\mathrm{I}_n + \mathrm{G}_{22}\mathrm{G}_{22}^T\right)^{-1}}. \tag{3.16}$$

Note that Λ_2 is an $(n-1)\times(n-1)$ version of Π_i. Since n is assumed to be large, in the limit the eigendistribution and henceforth the Stieltjes transform of $(\alpha\Lambda_2 + \mathrm{G}_{22}\mathrm{G}_{22}^T)$ converges to that of the right-hand side of (3.3), i.e., Π_{i+1}. Thus

$$t(z) = \mathbb{E}\frac{1}{1 + \frac{n}{m}S_{i+1}(z)} = \frac{1}{1 + \frac{1}{\beta}S_{i+1}(z)}. \tag{3.17}$$

Due to the fact that no randomness is left, the notion of expectation is dropped in the RHS. By replacing $t(z)$ in (3.9), and using the definition of Stieltjes transform (2.8), (3.4) will be obtained. □

As we will see in subsection 3.2.1, one can find a recursion for the moments of the eigendistribution from (3.4). Those recursions show the convergence (at least convergence in moments of any order) to a steady-state $S(z)$ as $i \to \infty$. The steady-state Stieltjes transform clearly satisfies

$$S(z) = \frac{1}{\alpha}S\left(\frac{z}{\alpha} - \frac{\frac{\beta}{\alpha}}{\beta + S(z)}\right). \tag{3.18}$$

It is not possible to analytically find $S(z)$ from (3.18). However, (3.18) can be solved numerically to find the Stieltjes transform and hence the eigendistribution at the steady state. Figure 3.1 compares the theoretical prediction with the empirically found eigendistribution for $n = 10$. The empirical curve is found by running the iteration 500 times (hence 5000 eigenvalues). The theoretical curve is found by starting with a uniform distribution and iteratively running (3.18). It can be seen that

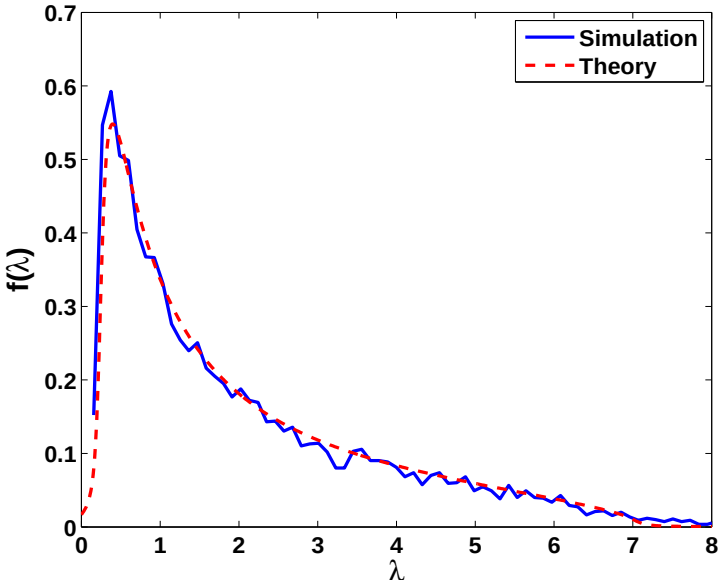


Figure 3.1. Steady-state eigendistribution of recursion (3.3) for $n = 10$, $m = 5$, $\alpha = 0.5$..

for n being as small as 10 the asymptotic result is pretty accurate in predicting the eigendistribution.

3.2.1 Finding the Moments

As mentioned earlier, the moments of the eigendistribution are the coefficients of the power series expansion of the Stieltjes transform. We can use this fact to obtain a recursion for the moments of the eigendistribution from (3.4). Consider

$$S_i(z) = -\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots \ . \tag{3.19}$$

Replacing the above expression in (3.4) we obtain,

$$-\frac{1}{z}-\frac{m_1^{i+1}}{z^2}-\frac{m_2^{i+1}}{z^3}-\frac{m_3^{i+1}}{z^4}-\cdots=-\frac{1}{z-\frac{1}{1+\frac{1}{\beta}S_{i+1}(z)}}-\frac{\alpha m_2^i}{\left(z-\frac{1}{1+\frac{1}{\beta}S_{i+1}(z)}\right)^2}-\cdots\ . \tag{3.20}$$

The RHS can be rewritten as,

$$-\frac{1}{z}\frac{1}{1-\frac{1/z}{1+\frac{1}{\beta}S_{i+1}(z)}}-\frac{\alpha m_1^i}{z^2}\frac{1}{\left(1-\frac{1/z}{1+\frac{1}{\beta}S_{i+1}(z)}\right)^2}-\frac{\alpha^2 m_2^i}{z^3}\frac{1}{\left(1-\frac{1/z}{1+\frac{1}{\beta}S_{i+1}(z)}\right)^3}-\cdots \tag{3.21}$$

We can expand $\frac{1/z}{1+\frac{1}{\beta}S_{i+1}(z)}$ in turn as a power series,

$$\frac{1/z}{1+\frac{1}{\beta}S_{i+1}(z)}=\frac{1}{z}\left(1-\frac{1}{\beta}S_{i+1}(z)+\frac{1}{\beta^2}S_{i+1}^2(z)-\cdots\right). \tag{3.22}$$

Therefore the RHS of (3.20) can be expanded as,

$$\begin{aligned}
RHS= &-\frac{1}{z}\left(1+\frac{1}{z}+\frac{1}{\beta}\frac{1}{z^2}+\left(\frac{m_1^{i+1}}{\beta}+\frac{1}{\beta^2}\right)\frac{1}{z^3}+\cdots+\frac{1}{z^2}+\frac{2}{\beta}\frac{1}{z^3}+\cdots+\frac{1}{z^3}\right)\\
&-\frac{\alpha m_1^i}{z^2}\left(1+\frac{2}{z}+\frac{2}{\beta}\frac{1}{z^2}+\cdots+\frac{3}{z^2}+\cdots\right)\\
&-\frac{\alpha^2 m_2^i}{z^3}\left(1+\frac{3}{z}+\cdots\right)\\
&-\frac{\alpha^3 m_3^i}{z^4}(1+\cdots)-\cdots\ .
\end{aligned} \tag{3.23}$$

where we have kept the terms up to $\frac{1}{z^4}$. Equating the coefficients of different powers of z on both sides, we will obtain the recursions of the moments as time increases. The coefficients of $\frac{1}{z}$ on both sides are trivially -1. Looking at $\frac{1}{z^2}$,

$$-\frac{m_1^{i+1}}{z^2}=-\frac{1}{z^2}-\alpha m_1^i\frac{1}{z^2}\quad\Rightarrow\quad m_1^{i+1}=1+\alpha m_1^i. \tag{3.24}$$

This immediately tells us that a necessary condition for the convergence of the first moment is $\alpha < 1$, since

$$\Delta m_1^{i+1} = m_1^{i+1} - m_1^i = \alpha(m_1^i - m_1^{i-1}) = \alpha \Delta m_1^i. \tag{3.25}$$

Looking at the recursion for other moments tells us that this is also a sufficient condition for the convergence of all moments. In fact in a similar fashion as we derived (3.24), we proceed to obtain,

$$m_2^{i+1} = \left(1 + \frac{1}{\beta}\right) + 2\alpha m_1^i + \alpha^2 m_2^i, \tag{3.26}$$

$$m_3^{i+1} = \left(1 + \frac{2}{\beta} + \frac{1}{\beta^2}\right) + \frac{1}{\beta} m_1^{i+1} + \alpha\left(\frac{2}{\beta} + 3\right) m_1^i + 3\alpha^2 m_2^i + \alpha^3 m_3^i. \tag{3.27}$$

To see the convergence rate of the second moment, consider the following similar to (3.25)

$$\Delta m_2^{i+1} = 2\alpha \Delta m_1^i + \alpha^2 \Delta m_2^i. \tag{3.28}$$

Since $\alpha < 1$, the slowest converging term will be Δm_1^i. A similar result can be obtained for all the higher order moments and therefore $\alpha < 1$ is the necessary and sufficient condition for convergence to the steady-state eigendistribution. It is straightforward to compute the steady-state moments from the recursions above,

$$m_1 = \frac{1}{1-\alpha}, \tag{3.29}$$

$$m_2 = \frac{\frac{1}{\beta}}{1-\alpha^2} + \frac{1}{(1-\alpha)^2}, \tag{3.30}$$

$$m_3 = \frac{\frac{1}{\beta}}{1-\alpha^3} + \frac{\frac{3}{\beta}}{(1-\alpha)(1-\alpha^2)} + \frac{1}{(1-\alpha)^3}. \tag{3.31}$$

Later when we talk about the R-transform approach to the problem, it will become clear why the moments have this structure. Figure 3.2 shows the first three moments

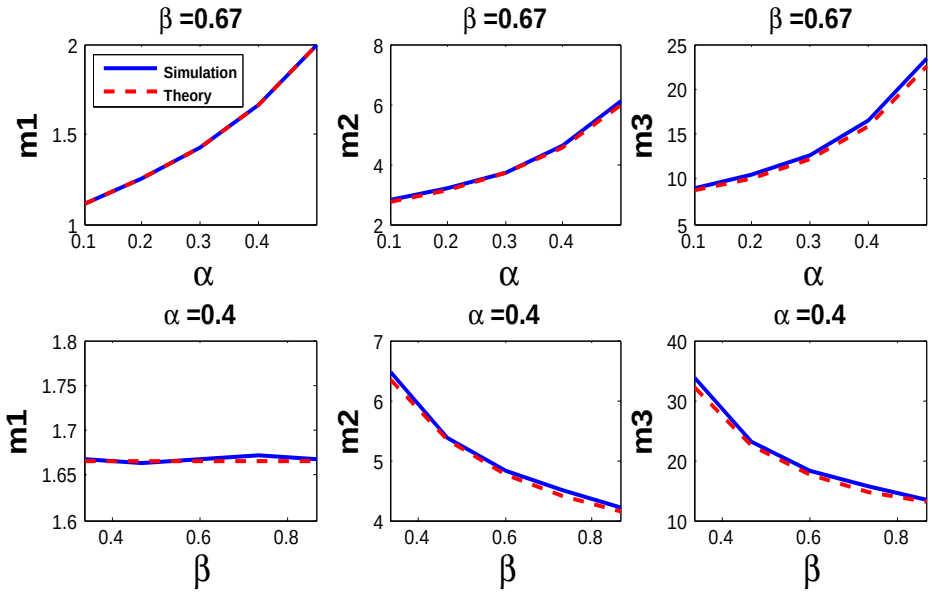


Figure 3.2. Moments of the steady-state eigendistribution of recursion (3.3) for $n = 20$ and $m = \beta n$. .

at the steady state for different values of α and β, as obtained from the simulations and the theory for $n = 10$. Finally it is worth mentioning that although the moment expressions were linear so far, m_4 and higher-order moments have nonlinear dependencies on lower-order moments.

3.3 Transient Behavior

The approach that was taken in subsection 3.2.1 to study the transient behavior of the moments and hence the eigendistribution is not applicable to the transient behavior for nonlinear Riccati recursion. Therefore in an attempt to develop more effective frameworks for transient analysis, we consider other methods for handling the Lyapunov recursion. With this goal in mind, we find two other methods specially useful. One is presented below and the other one is studied in the next subsection.

According to the moment analysis for the simple Lyapunov recursion, the k^{th}

moment at time $i+1$ can be regarded as,

$$m_k^{i+1} = f_k(m_1^i, \cdots, m_k^{i-1}) + \alpha^k m_k^i. \tag{3.32}$$

Now define $\mathbf{m}^i$ as the vector of all moments,

$$\mathbf{m}^i = \begin{bmatrix} m_1^i \\ m_2^i \\ \vdots \end{bmatrix}. \tag{3.33}$$

Therefore there is a function $\Phi(\cdot)$ such that,

$$\mathbf{m}^{i+1} = \Phi(\ \mathbf{m}^i). \tag{3.34}$$

If we assume that a steady state, $\mathbf{m}^*$, exists,

$$\mathbf{m}^* = \Phi(\ \mathbf{m}^*), \tag{3.35}$$

then we can look at a time i that the moment vector $\mathbf{m}^i$ is close enough to $\mathbf{m}^*$, such that a linear approximation would be possible, i.e.,

$$\Delta\ \mathbf{m}^{i+1} = \ \mathbf{m}^{i+1} - \ \mathbf{m}^* = \Phi(\ \mathbf{m}^i) - \Phi(\ \mathbf{m}^*) = \nabla\Phi|_{\ \mathbf{m}^*} \cdot (\ \mathbf{m}^i - \ \mathbf{m}^*) = \nabla\Phi|_{\ \mathbf{m}^*} \cdot \Delta\ \mathbf{m}^i. \tag{3.36}$$

According to (3.32), the operator $\nabla\Phi|_{\ \mathbf{m}^*}$ is lower triangular and its diagonal entries are powers of α,

$$\nabla\Phi|_{\ \mathbf{m}^*} = \begin{bmatrix} \alpha & & \cdots & 0 \\ * & \alpha^2 & \ddots & \vdots \\ * & * & \alpha^3 & \vdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{3.37}$$

Therefore, $(\nabla\Phi|_{\mathbf{m}^*})^l$ remains lower triangular with diagonal entries as,

$$\left[(\nabla\Phi|_{\mathbf{m}^*})^l\right]_{kk} = \alpha^{kl}. \tag{3.38}$$

This shows that the spectrum of the operator consists of powers of α with the smallest power being the dominant one. Therefore $\alpha < 1$ is a necessary condition for the convergence, and moreover, α is the convergence rate itself.

3.3.1 R-transform Approach

Alternatively, one can do all the analysis we have done so far using the R-transform.

Theorem 3.3.1 *The steady-state R-transform of Π_i which satisfies (3.3) can be found as,*

$$R(z) = \sum_{j=1}^{\infty} \frac{1}{1-\alpha^j} \frac{z^{j-1}}{\beta^{j-1}}. \tag{3.39}$$

Proof: As mentioned in Section 2.5, the R-transform of the sum of two asymptotically free random variables is simply the sum of their R-transforms. Given the independence of the G_i's for different i's, $\alpha\Pi_i$ and $\mathrm{G}_i\mathrm{G}_i^T$ are independent unitarily invariant matrices, and hence free [Voi00]. The R-transform of $\mathrm{G}_i\mathrm{G}_i^T$ can be found as [TV04],

$$R_{\mathrm{G}_i\mathrm{G}_i^T} = \frac{\beta}{\beta - z}. \tag{3.40}$$

Moreover according to the definition of the Stieltjes transform,

$$S_{\alpha\mathrm{A}}(z) = \mathbb{E}\left[\frac{1}{\lambda_{\alpha\mathrm{A}} - z}\right] = \mathbb{E}\left[\frac{1}{\alpha\lambda_{\mathrm{A}} - z}\right] = \frac{1}{\alpha}\mathbb{E}\left[\frac{1}{\lambda_{\mathrm{A}} - \frac{z}{\alpha}}\right] = \frac{1}{\alpha} S_{\mathrm{A}}\left(\frac{z}{\alpha}\right). \tag{3.41}$$

Therefore,

$$\begin{aligned} S_{\alpha\mathrm{A}}^{<-1>}\left(S_{\alpha\mathrm{A}}(z)\right) &= S_{\alpha\mathrm{A}}^{<-1>}\left(\frac{1}{\alpha}S_{\mathrm{A}}\left(\frac{z}{\alpha}\right)\right) = \alpha S_{\mathrm{A}}^{<-1>}\left(S_{\mathrm{A}}(\frac{z}{\alpha})\right) \\ &\Rightarrow \quad S_{\alpha\mathrm{A}}^{<-1>}(z) = \alpha S_{\mathrm{A}}^{<-1>}(\alpha z). \end{aligned} \tag{3.42}$$

The R-transform of αA can also be written in terms of the R-transform of A as,

$$R_{\alpha\mathrm{A}}(z) = S_{\alpha\mathrm{A}}^{<-1>}(-z) - \frac{1}{z} = \alpha\left(S_{\mathrm{A}}^{<-1>}(-\alpha z) - \frac{1}{\alpha z}\right) = \alpha R_{\mathrm{A}}(\alpha z). \tag{3.43}$$

Thus the recursion on the R-transofrm of Π_i is readily obtained

$$R_{i+1}(z) = \alpha R_i(\alpha z) + \frac{\beta}{\beta - z}. \tag{3.44}$$

The steady-state R-transform, $R(z)$, can be found by noting that,

$$\begin{aligned} R(z) &= \alpha R(\alpha z) + \frac{\beta}{\beta - z} \\ \alpha R(\alpha z) &= \alpha^2 R(\alpha^2 z) + \frac{\alpha\beta}{\beta - \alpha z} \\ \alpha^2 R(\alpha^2 z) &= \alpha^3 R(\alpha^3 z) + \frac{\alpha^2\beta}{\beta - \alpha^2 z} \\ &\vdots \end{aligned}$$

We can find $R(z)$ by adding up all the terms. Note that since $\alpha < 1$, this is a converging series,

$$R(z) = \frac{\beta}{\beta - z} + \frac{\alpha\beta}{\beta - \alpha z} + \frac{\alpha^2\beta}{\beta - \alpha^2 z} + \cdots$$

$$
\begin{aligned}
= & \left(1+\frac{z}{\beta}+\frac{z^2}{\beta^2}+\cdots\right)+\alpha\left(1+\alpha\frac{z}{\beta}+\alpha^2\frac{z^2}{\beta^2}+\cdots\right) \\
& s+\alpha^2\left(1+\alpha^2\frac{z}{\beta}+\alpha^4\frac{z^2}{\beta^2}+\cdots\right)+\cdots \\
= & \left(1+\alpha+\alpha^2+\cdots\right)+\frac{z}{\beta}\left(1+\alpha^2+\alpha^4+\cdots\right) \\
& s+\frac{z^2}{\beta^2}\left(1+\alpha^3+\alpha^6+\cdots\right)+\cdots\ ; \qquad (3.45)
\end{aligned}
$$

simplifying this expression yields (3.39). □

As mentioned earlier, the coefficients of the power series expansion of R-transform are the free cumulants of the eigendistribution from which the moments can be found. In this case the cumulants are

$$k_j = \frac{1}{1-\alpha^j}\frac{1}{\beta^{j-1}}. \qquad (3.46)$$

Using the relation between the moments and the free cumulants in (2.20), we obtain the moments,

$$
\begin{aligned}
m_1 &= k_1 = \frac{1}{1-\alpha} \\
m_2 &= k_1^2 + k_2 = \frac{1}{(1-\alpha)^2} + \frac{\frac{1}{\beta}}{(1-\alpha^2)} \\
&\vdots
\end{aligned}
$$

Not surprisingly these are the same expressions as found in subsection 3.2.1. This approach however, justifies the special structure of the moments that we observed earlier and provides a systematic method for computing the moments.

More important however, is the convergence rate that can be deduced from the recursion of the R-transform. Going back to (3.44), and expanding both sides as a

power series in terms of z,

$$\sum_{j=1}^{\infty} k_j^{i+1} z^{j-1} = \alpha \sum_{j=1}^{\infty} k_j^i \alpha^{j-1} z^{j-1} + \sum_{j=1}^{\infty} \frac{z^{j-1}}{\beta^{j-1}}, \tag{3.47}$$

where k_j^i is the j^{th} free cumulant at time i. By equating the coefficients of each power of z on both sides we find,

$$k_j^{i+1} = \alpha^j k_j^i + \frac{1}{\beta^{j-1}}, \tag{3.48}$$

which once again verifies that the convergence rate is α.

In this section we established two alternative methods for investigating the convergence properties of the recursion. In the next section we will look at a more general random Lyapunov recursion.

3.4 A More General Random Lyapunov Recursion

Our analysis for $\mathrm{F} = \sqrt{\alpha}\mathrm{I}$ can be extended to the recursions with more general F's. In particular, here we will assume that F is drawn from an $n \times n$ random matrix distribution with i.i.d. entries having zero mean, $\frac{1}{\sqrt{n}}$ variance, and bounded higher moments. Consider,

$$\Pi_{i+1} = \alpha \mathrm{F} \Pi_i \mathrm{F}^T + \mathrm{G}_i \mathrm{G}_i^T, \tag{3.49}$$

where α is selected such that $\sqrt{\alpha}\mathrm{F}$ is stable. The G_i's are assumed to be the same as before. Once again we are interested in finding the eigendistribution of Π_i, specially as $i \to \infty$. Note that we could as well assume that $\mathrm{F} = \mathrm{F}_i$, i.e., changing at each time step, and the analysis would have been the same. However a fixed F is a more realistic assumption. We only assume that at the beginning F is selected from a random matrix ensemble. According to the ergodicity of random matrices for sufficiently large n the steady-state eigendistribution of every single recursion will look like the

asymptotic one. For moderate values of n however the asymptotic result will have closer resemblance to the expected eigendistribution (over F) rather than individual realizations.

The eigendistribution of Π_i can be found as the solution of a pair of implicit equations.

Theorem 3.4.1 *The Stieltjes transform of* Π_i *which undergoes the Lyapunov recursion of (3.49) satisfies,*

$$S_{i+1}(z) = \frac{1}{\alpha}\Omega_i\left(\frac{z}{\alpha} - \frac{\frac{\beta}{\alpha}}{\beta + S_{i+1}(z)}\right), \tag{3.50}$$

$$\Omega_i^2(z) = -\frac{1}{z}S_i\left(-\frac{1}{\Omega_i(z)}\right), \tag{3.51}$$

where $\beta = \frac{m}{n}$.

Proof: The first part of the proof exactly follows that of theorem 3.2.1. Define $\mathrm{A}_i = \mathrm{F}\Pi_i\mathrm{F}^T$. Now since G_i is unitarily invariant, as in the proof of theorem 3.2.1, we can assume that A_i is diagonal, say $\mathrm{A}_i = \bar{\Lambda}$. Taking the Stieltjes transform of both sides of (3.49) gives,

$$S_{i+1}(z) = \frac{1}{n}\mathbb{E}\ \mathrm{tr}\left(\alpha\bar{\Lambda} + \mathrm{G}_i\mathrm{G}_i^T - z\mathrm{I}\right)^{-1}. \tag{3.52}$$

By breaking Λ and G_i as in (3.6), and following the consecutive step therein, we obtain

$$S_{i+1}(z) = \mathbb{E}\ \frac{1}{\alpha\bar{\lambda}_1 - z + \underbrace{g_1(\mathrm{I}_m + \mathrm{G}_2^T(\alpha\bar{\Lambda}_2 - z\mathrm{I}_{n-1})^{-1}\mathrm{G}_2)^{-1}g_1^T}_{\triangleq t(z)}}, \tag{3.53}$$

where $\bar{\lambda}_i$ are the eigenvalues of A_i. In order to compute $t(z)$, once again we note that

it can be written as the Stieltjes transform of the inverted matrix in the denominator,

$$t(z) = \frac{1}{m}\mathbb{E}\ \mathrm{tr}\ (\mathrm{I}_m + \mathrm{G}_2^T(\alpha\bar{\Lambda}_2 - z\mathrm{I}_{n-1})^{-1}\mathrm{G}_2)^{-1}. \tag{3.54}$$

We break G_2 as in (3.11) and repeat the calculations which follow (3.11) to obtain

$$t(z) = \frac{1}{1 + \frac{1}{m}\mathbb{E}\ \mathrm{tr}\left(\alpha\bar{\Lambda}_2 - z\mathrm{I}_n + \mathrm{G}_{22}\mathrm{G}_{22}^T\right)^{-1}}. \tag{3.55}$$

Since $\bar{\Lambda}_2$ is an $(n-1)\times(n-1)$ version of A_i and n is large, the Stieltjes transform of $(\alpha\Lambda_2 + \mathrm{G}_{22}\mathrm{G}_{22}^T)$ converges to $S_{i+1}(z)$, and we have

$$t(z) = \frac{1}{1 + \frac{1}{\beta}S_{i+1}(z)}. \tag{3.56}$$

By replacing (3.56) in (3.53) we have,

$$S_{i+1}(z) = \frac{1}{\alpha}S_{\mathrm{A}_i}\left(\frac{z}{\alpha} - \frac{\frac{\beta}{\alpha}}{\beta + S_{i+1}(z)}\right). \tag{3.57}$$

Therefore it remains to compute S_{A_i} in terms of $S_i(z)$ (the Stieltjes transform of Π_i). We first note that the eigenvalues of $\mathrm{F}\Pi_i\mathrm{F}^T$ are the same as those of $\mathrm{F}^T\mathrm{F}\Pi_i$. The Wishart matrix $\mathrm{F}^T\mathrm{F}$ is unitarily invariant. The same is true for Π_i. Therefore the two are asymptotically free [Voi00] and we can use theorem 2.5.3 which states that the S-transform of the product of two asymptotically free random matrices is the product of the individual S-transforms. Thus,

$$\Sigma_{\mathrm{A}_i}(z) = \Sigma_{\mathrm{F}^T\mathrm{F}}(z)\cdot\Sigma_{\Pi_i}(z). \tag{3.58}$$

The implicit equation between the Stieltjes transform and the S-transform states that,

$$\Sigma_{\mathrm{A}}(z) = -\frac{1}{z} S_{\mathrm{A}}\left(\frac{1+z}{z\Sigma_{\mathrm{A}}(z)}\right). \tag{3.59}$$

Replacing in (3.58),

$$S_{\mathrm{A}_i}\left(\frac{1+z}{z\Sigma_{\mathrm{A}_i}(z)}\right) = \Sigma_{\mathrm{F}^T\mathrm{F}}(z) S_{\Pi_i}\left(\frac{1+z}{z\Sigma_{\Pi_i}(z)}\right) = \Sigma_{\mathrm{F}^T\mathrm{F}}(z) S_{\Pi_i}\left(\frac{1+z}{z\Sigma_{\mathrm{A}_i}(z)} \cdot \Sigma_{\mathrm{F}^T\mathrm{F}}(z)\right). \tag{3.60}$$

Let

$$x = \frac{1+z}{z\Sigma_{\mathrm{A}_i}(z)}. \tag{3.61}$$

Then

$$x\Sigma_{\mathrm{A}_i}(z) = \frac{1+z}{z} \quad \Rightarrow \quad x \cdot (-\frac{1}{z}) S_{\mathrm{A}_i}(x) = \frac{1+z}{z} \quad \Rightarrow \quad z = -1 - xS_{\mathrm{A}_i}(x). \tag{3.62}$$

Replacing in (3.60), we obtain

$$S_{\mathrm{A}_i}(x) = \Sigma_{\mathrm{F}^T\mathrm{F}}\left(-1 - xS_{\mathrm{A}_i}(x)\right) \cdot S_{\Pi_i}\left(x\Sigma_{\mathrm{F}^T\mathrm{F}}\left(-1 - xS_{\mathrm{A}_i}(x)\right)\right). \tag{3.63}$$

The above implicit equation will prove to be very useful and we will recurrently use it throughout this manuscript. According to example 2.3.4, the S-transform of $\mathrm{F}^T\mathrm{F}$ can be written as,

$$\Sigma_{\mathrm{F}^T\mathrm{F}}(z) = \frac{1}{1+z}. \tag{3.64}$$

Now we can simplify (3.63) to

$$S_{\mathrm{A}_i}(x) = -\frac{1}{xS_{\mathrm{A}_i}(x)} S_{\Pi_i}\left(-\frac{1}{S_{\mathrm{A}_i}(x)}\right), \tag{3.65}$$

which completes the proof of the theorem by considering $\Omega_i(\cdot) = S_{\mathrm{A}_i}(\cdot)$. □

Clearly the steady-state eigendistribution satisfies

$$S(z) = \frac{1}{\alpha}\Omega\left(\frac{z}{\alpha} - \frac{\frac{\beta}{\alpha}}{\beta + S(z)}\right), \tag{3.66}$$

$$\Omega(z) = -\frac{1}{z\Omega(z)}S\left(-\frac{1}{\Omega(z)}\right). \tag{3.67}$$

It is not feasible to find an analytical solution to the above pair of equations. However they can be numerically solved in an efficient fashion. In fact since $\Omega(z)$ is a Stieltjes transform itself, we can start with two uniform eigendistributions and iteratively find a numerical solution. At each step we use the current value of the eigendistributions to compute the right-hand sides of the implicit equation system and update the left-hand side. In most cases that we have studied less than 20 steps of iteration has been enough to reach a sufficiently exact solution. Figure 3.3 shows the eigendistribution which is found in this manner. The Simulation results are generated from 1000 samples for $n = 10$. As can be seen, the asymptotic result is very accurate for n as small as 10.

3.4.1 Finding the Moments

As we did in subsection 3.2.1 for the simple Lyapunov recursion, we can use the recursion on the Stieltjes transform to obtain a recursion for each moment of the eigendistribution. It is straightforward to show that $\Omega_i(z)$ is a Stieltjes transform whose eigendistribution has bounded support[1]. Therefore we can expand it as a Laurent series in terms of z, i.e.,

$$\Omega_i(z) = -\frac{1}{z} - \frac{b_1^i}{z^2} - \frac{b_2^i}{z^3} - \cdots. \tag{3.68}$$

[1]This is true when the eigendistribution of Π_i has bounded support.

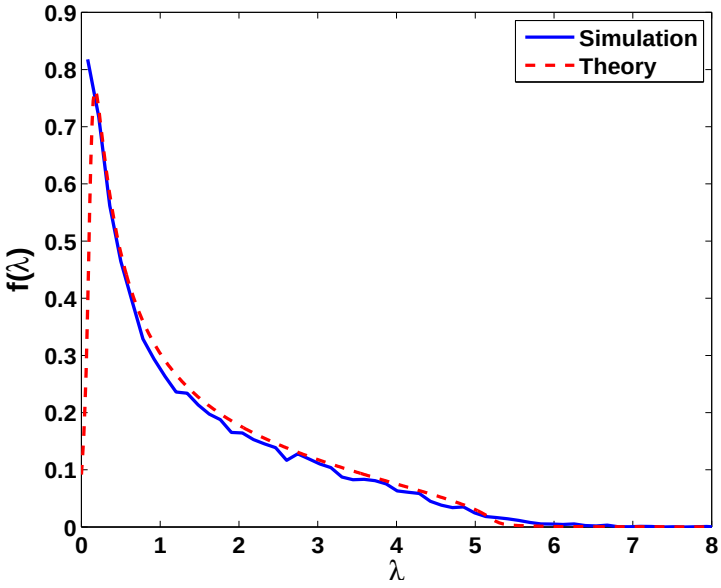


Figure 3.3. Steady-state eigendistribution of recursion (3.49) for $n = 10$, $m = 8$, $\alpha = 0.3$..

The power series expansion of (3.50) will be exactly the same as that of the simple Lyapunov recursion. Thus we do not repeat the details here,

$$\begin{aligned} -\frac{1}{z} - \frac{m_1^{i+1}}{z^2} - \frac{m_2^{i+1}}{z^3} - \cdots = & -\frac{1}{z}\left(1+\frac{1}{z}+\frac{1}{\beta}\frac{1}{z^2}+\cdots+\frac{1}{z^2}+\cdots\right) \\ & -\frac{\alpha b_1^i}{z^2}\left(1+\frac{2}{z}+\cdots\right) \\ & -\frac{\alpha^2 b_2^i}{z^3}(1+\cdots)-\cdots . \end{aligned} \tag{3.69}$$

For the sake of brevity, we have kept fewer terms than before to restrict the derivations to the first two moments. The power series expansion of (3.51) yields,

$$\begin{aligned} \left(-\frac{1}{z}-\frac{b_1^i}{z^2}-\frac{b_2^i}{z^3}-\cdots\right)^2 &= -\frac{1}{z}\left[\Omega_i(z)-m_1^i\Omega_i^2(z)+m_2^i\Omega_i^3(z)-\cdots\right] \\ &= -\frac{1}{z}\left[\left(-\frac{1}{z}-\frac{b_1^i}{z^2}-\frac{b_2^i}{z^3}-\cdots\right)-m_1^i\left(\frac{1}{z^2}+2\frac{b_1^i}{z^3}+\cdots\right)-\frac{m_2^i}{z^3}+\cdots\right]. \end{aligned} \tag{3.70}$$

Now looking at the coefficients of $\frac{1}{z^2}$ in (3.69) and $\frac{1}{z^3}$ in (3.70),

$$m_1^{i+1} = 1 + \alpha b_1^i, \tag{3.71}$$

$$b_1^i = m_1^i. \tag{3.72}$$

Therefore,

$$m_1^{i+1} = 1 + \alpha m_1^i. \tag{3.73}$$

And by looking at the coefficients of $\frac{1}{z^3}$ in (3.69) and $\frac{1}{z^4}$ in (3.70),

$$m_2^{i+1} = \left(1 + \frac{1}{\beta}\right) + 2\alpha b_1^i + \alpha^2 b_2^i, \tag{3.74}$$

$$b_2^i = 2b_1^i m_1^i - \left(b_1^i\right)^2 + m_2^i, \tag{3.75}$$

which simplify to,

$$m_2^{i+1} = \left(1 + \frac{1}{\beta}\right) + 2\alpha m_1^i + \alpha^2 \left(m_1^i\right)^2 + \alpha^2 m_2^i, \tag{3.76}$$

which unlike the first moment is different from that of the simple Lyapunov recursion. Once again, we can reason that the convergence rate is α and $\alpha < 1$ is a necessary condition for the convergence. The steady-state moments can be readily derived,

$$m_1 = \frac{1}{1-\alpha}, \tag{3.77}$$

$$m_2 = \frac{1 + \frac{1}{\beta}}{1-\alpha^2} + \frac{\alpha(2-\alpha)}{(1-\alpha^2)(1-\alpha)^2}. \tag{3.78}$$

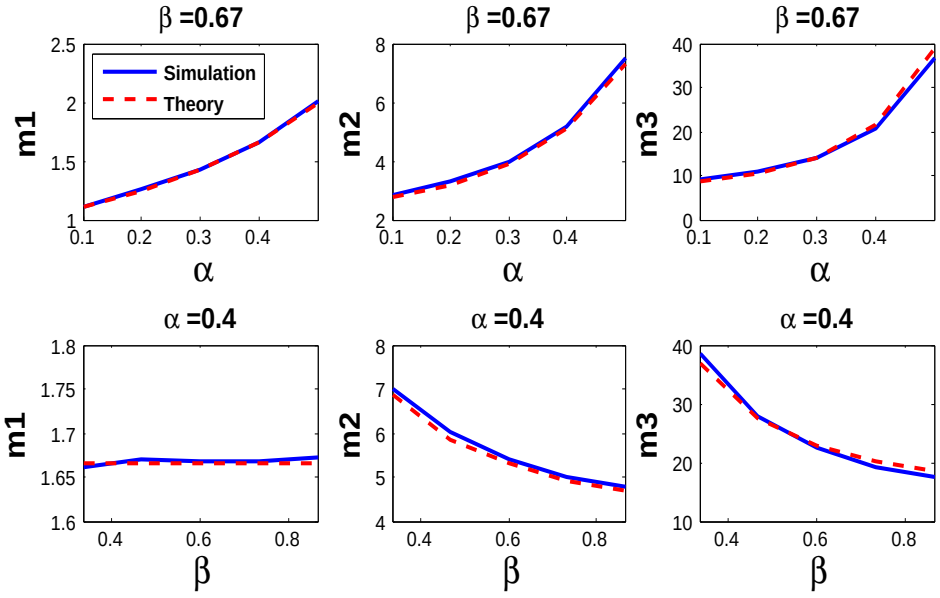


Figure 3.4. Moments of the steady-state eigendistribution of recursion (3.49) for $n = 20$ and $m = \beta n$. .

Although more involved than the simple Lyapunov case, the third moment can also be found with some further calculations.

$$\begin{aligned} m_3 &= \frac{1}{(1-\alpha^3)} \left\{ \left(\frac{2+m_1}{\beta} + \frac{1}{\beta^2} + 1 \right) \right. \\ &+ \left. \alpha m_1 \left(\frac{2}{\beta} + 3 \right) + \alpha^2 \left(3m_1^2 + m_2 \right) + \alpha^3 \left(m_1^3 + 3m_1 m_2 \right) \right\} \end{aligned} \tag{3.79}$$

Figure 3.4 shows the simulation results for $n = 20$ compared with the theoretical predictions of the first three moments for different values of α and β.

3.5 Conclusions

In this chapter we studied random Lyapunov recursions using tools from the theory of large random matrices. Random Lyapunov recursions describe the covariance of the state vector in a linear state-space model. Moreover they provide bounds on

the Riccati recursions which appear in the area of control and estimation over lossy networks. We analyzed a simple Lyapunov recursion where $\mathrm{F} = \sqrt{\alpha}\mathrm{I}$ and found that its Stieltjes transform satisfies a recursion that can be numerically solved in the steady state, from which the steady-state eigendistribution can be readily found. Using the recursion for the Stieltjes transform and power-series expansion we found recursions on the moments and deduced the convergence condition and the rate of convergence from them. We further employed two other methods to study the convergence. One method is based on the spectrum of the linear operator which describes the moment update close to the steady state. The other method uses the recursion of the R-transform and consequently free cumulants. The R-transform analysis also provides a systematic way of finding any moment of the eigendistribution.

We then extended our analysis to a more general case where F is a full matrix with i.i.d. entries. We obtained a pair of implicit equations which describe the Stieltjes transform both in transient and steady state and found the moments of the eigendistribution. Although the analysis is done for $n \to \infty$, the asymptotic results are shown to be quite accurate for n as small as 10 (due to the law of large numbers).

It is worth mentioning that one can think of more general forms of F and even G_i, provided that the freeness can be established. In those cases the system of implicit equations which describe the Stieltjes transform will be more involved. In some cases one may even need more than two equations. The only trick is that we should find the R-transform of $\mathrm{G}_i\mathrm{G}_i^T$ and the S-transform of F. Under our assumptions they had simple expressions which in turn simplified the derivations. The framework developed proves to be very promising and motivates the study of other Lyapunov-like (e.g., LMS filters) recursions or nonlinear Riccati recursion as we will do in the next chapters.

Chapter 4

The Least-Mean-Squares Filters

4.1 Introduction

The least-mean-squares adaptive filtering is the simplest, yet the most frequently used algorithm in adaptive signal processing due to its robustness and low computational cost [Say03, Hay01, HM84]. Introduced by Bernard Widrow and Ted Hoff in 1960 [WH60], LMS filtering is not the optimum algorithm for minimizing the mean-square error in estimating a parameter, but rather gives an approximation to the exact solution. However its simplicity and not asking for a stationary input or the statistics of the model, have made it very popular and kind of a standard to which other algorithms are normally compared.

Consider a special case linear state-space model introduced in the previous chapter [KSH00],

$$\left\{ \begin{array}{rcl} x_{i+1} & = & x_i + u_i \\ y_i & = & h_i^T x_i + v_i \end{array} \right. , \quad \mathbb{E} \left[\begin{array}{c} u_i \\ v_i \end{array} \right] \left[\begin{array}{cc} u_j^T & v_j^T \end{array} \right] = \left[\begin{array}{cc} \mathrm{Q}_i & 0 \\ 0 & r \end{array} \right] \delta_{ij}, \qquad (4.1)$$

in which $x_i \in \mathcal{R}^n$ is an unknown vector to be estimated through the measurements $y_i \in \mathcal{R}$. u_i and v_i denote zero-mean process and measurement noises respectively. The h_i are referred to as the regressor vectors. In some cases $u_i = 0$, but there are

other cases which we will talk about later in this chapter. For examples of state-space approach to adaptive filtering look at [SK94] and [HSK96] and the references therein. The LMS approximates the optimal x that minimizes $\mathbb{E}|y - h^T x|^2$ iteratively as,

$$\hat{x}_{i+1} = \hat{x}_i + \mu h_i (y_i - h_i^T \hat{x}_i), \tag{4.2}$$

where μ is called step size. It is straightforward to show that the error covariance matrix, i.e.,

$$\mathrm{P}_i = \mathbb{E}[(x_i - \hat{x}_i)(x_i - \hat{x}_i)^T], \tag{4.3}$$

satisfies a Lyapunov-type recursion. Note that,

$$\begin{aligned} x_{i+1} - \hat{x}_{i+1} &= x_i + u_i - \hat{x}_i - \mu h_i (y_i - h_i^T \hat{x}_i) = x_i + u_i - \hat{x}_i - \mu h_i (h_i^T x_i + v_i - h_i^T \hat{x}_i) \\ &= (\mathrm{I} - \mu h_i h_i^T)(x_i - \hat{x}_i) + u_i - \mu h_i v_i. \end{aligned} \tag{4.4}$$

Since v_i and u_i are independent of the the past and are zero mean,

$$\mathrm{P}_{i+1} = \left(\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T + \mathrm{Q}_i, \tag{4.5}$$

which is clearly a Lyapunov-type random recursion.

While the LMS filter has been studied extensively in the literature in the past 40 years, through our approach the stability and performance analysis of all LMS-like algorithms becomes straightforward. Moreover, the universality of the random matrix methods provides a unifying framework which does not depend at all on the details of the underlying model.

We will consider both temporally random regressor vectors (h_i being independent of $h_{j \neq i}$) and shift-structured regressor vectors (which is an important practical case, e.g., FIR channel estimation). Different assumptions on Q_i will also be studied.

This chapter is organized as follows. In section 4.2 we will study the very simple LMS estimation of a constant state vector. The analysis will be extended to the case of time-varying state in section 4.3. The moments will be found and the numerical method employed to find the eigendistribution will be explained. In section 4.4 we will generalize the results to the shift-structured regressor vectors. There are other variants of the LMS filtering problem for different applications which will be briefly considered in section 4.5, and finally section 4.6 will conclude the chapter.

4.2 A Simple Case: Constant Channel Estimator

When estimating a constant state vector, then $x_{i+1} = x_i = \cdots = x_0 = c$. This is the case for example in estimating an FIR channel [Say03]. This clearly means that the state-space model of (4.1) is simplified to,

$$\begin{cases} x_{i+1} &= x_i, \\ y_i &= h_i^T x_i + v_i, \end{cases} \qquad \mathbb{E}[v_i v_j^T] = r\delta_{ij}, \tag{4.6}$$

and the estimation error covariance matrix satisfies,

$$\mathrm{P}_{i+1} = \left(\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T. \tag{4.7}$$

The h_i (the regressors) are independently drawn $n \times 1$ vectors, with i.i.d. entries having zero mean, $\frac{1}{\sqrt{n}}$ variance, and bounded higher moments. The independence of the entries can be relaxed as we will see later in this section by assuming a shift structure on the regressor vector; a standard model used in the literature.

Once again we are interested in finding the empirical eigenvalue distribution of P_i, as $i \to \infty$. At its very least, the steady-state eigendistribution's first moment is the mean-square-error performance of our LMS filter. In this simple case we can actually

determine the eigendistribution through the Stieltjes transform method.

Theorem 4.2.1 (LMS Estimation of a Constant Vector) *When $\mu < \frac{2}{n}$ the eigendistribution of the P_i in (4.7) converges to a steady-state distribution whose Stieltjes transform is*

$$S(z) = \frac{1}{\gamma - z}, \tag{4.8}$$

where,

$$\gamma = \frac{\frac{r}{n}}{\frac{2}{n\mu} - 1}. \tag{4.9}$$

This implies that as $n \to \infty$ P_i converges to γI.

Proof: We first apply the definition of the Stieltjes transform as in (2.12) to both sides of the recursion (4.7),

$$S_{i+1}(z) = -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\left(\mathrm{I} - \mu h_i h_i^T\right)\mathrm{P}_i\left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T - z\mathrm{I}\right), \tag{4.10}$$

which can be rewritten in the form,

$$\begin{aligned} S_{i+1}(z) &= -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det \\ &\quad \left(\begin{array}{c} (\mathrm{P}_i - z\mathrm{I}) + \left[\begin{array}{cc} h_i & \mathrm{P}_i h_i \end{array}\right] \underbrace{\left[\begin{array}{cc} \mu^2(r + h_i^T\mathrm{P}_i h_i) & -\mu \\ -\mu & 0 \end{array}\right]}_{\triangleq \mathrm{A}_i} \left[\begin{array}{c} h_i^T \\ h_i^T\mathrm{P}_i \end{array}\right] \end{array}\right) \\ &= -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det(\mathrm{P}_i - z\mathrm{I}) \\ &\quad -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I} + (\mathrm{P}_i - z\mathrm{I})^{-1}\left[\begin{array}{cc} h_i & \mathrm{P}_i h_i \end{array}\right]\mathrm{A}_i\left[\begin{array}{c} h_i^T \\ h_i^T\mathrm{P}_i \end{array}\right]\right). \end{aligned} \tag{4.11}$$

The first term in the right hand side is just the Stieltjes transform at time instance

i. Moreover, since $\det(\mathrm{I}+\mathrm{AB}) = \det(\mathrm{I}+\mathrm{BA})$, we have

$$S_{i+1}(z) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I} + \begin{bmatrix} h_i^T \\ h_i^T\mathrm{P}_i \end{bmatrix}(\mathrm{P}_i - z\mathrm{I})^{-1}\begin{bmatrix} h_i & \mathrm{P}_i h_i \end{bmatrix}\mathrm{A}_i\right). \tag{4.12}$$

Since A_i does not depend on z, the derivative of its determinant with respect to z vanishes. Thus,

$$S_{i+1}(z) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{A}_i^{-1} + \begin{bmatrix} h_i^T \\ h_i^T\mathrm{P}_i \end{bmatrix}(\mathrm{P}_i - z\mathrm{I})^{-1}\begin{bmatrix} h_i & \mathrm{P}_i h_i \end{bmatrix}\right). \tag{4.13}$$

Therefore,

$$\begin{aligned} S_{i+1}(z) &= S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det \\ &\qquad \left(\mathrm{A}_i^{-1} + \begin{bmatrix} h_i^T(\mathrm{P}_i - z\mathrm{I})^{-1}h_i & h_i^T\mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}h_i \\ h_i^T\mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}h_i & h_i^T\mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}\mathrm{P}_i h_i \end{bmatrix}\right). \end{aligned} \tag{4.14}$$

We can now invoke the self-averaging lemma. The h_i are assumed to be independent of each other, and P_i on all the previous $h_{j<i}$ and hence independent of h_i. The upper-left term of the second matrix on the right-hand side can be written as,

$$h_i^T(\mathrm{P}_i - z\mathrm{I})^{-1}h_i \xrightarrow[n\to\infty]{} S_i(z). \tag{4.15}$$

Similarly the other terms can be simplified,

$$\begin{aligned} h_i \mathrm{P}_i (\mathrm{P}_i - z\mathrm{I})^{-1} h_i^T &= h_i (\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})(\mathrm{P}_i - z\mathrm{I})^{-1} h_i^T \\ &= h_i h_i^T + z h_i (\mathrm{P}_i - z\mathrm{I})^{-1} h_i^T \\ &\xrightarrow[n\to\infty]{} 1 + z S_i(z), \end{aligned} \tag{4.16}$$

and,

$$\begin{aligned} h_i \mathrm{P}_i (\mathrm{P}_i - z\mathrm{I})^{-1} \mathrm{P}_i h_i^T &= h_i \mathrm{P}_i (\mathrm{P}_i - z\mathrm{I})^{-1} (\mathrm{P}_i - z\mathrm{I} + z\mathrm{I}) h_i^T \\ &= h_i \mathrm{P}_i h_i^T + z h_i (\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})(\mathrm{P}_i - z\mathrm{I})^{-1} h_i^T \\ &= h_i \mathrm{P}_i h_i^T + z h_i h_i^T + z^2 h_i (\mathrm{P}_i - z\mathrm{I})^{-1} h_i^T \\ &\xrightarrow[n\to\infty]{} h_i \mathrm{P}_i h_i^T + z + z^2 S_i. \end{aligned} \tag{4.17}$$

On the other hand, the inverse of A_i as a 2×2 matrix can be readily found as,

$$\mathrm{A}_i^{-1} = \begin{bmatrix} 0 & -\mu^{-1} \\ -\mu^{-1} & -(r + h_i^T \mathrm{P}_i h_i) \end{bmatrix}. \tag{4.18}$$

Replacing (4.15)–(4.18) in (4.14) yields

$$S_{i+1}(z) = S_i(z) - \frac{1}{n} \frac{d}{dz} \mathbb{E} \log \det \begin{bmatrix} S_i(z) & 1 + S_i(z) - \mu^{-1} \\ 1 + S_i(z) - \mu^{-1} & -r + z + z^2 S_i(z) \end{bmatrix}. \tag{4.19}$$

Note that in the above expression all the randomness has disappeared after using the self-averaging lemma and there is no need for $\mathbb{E}$ anymore. The determinant of the 2×2 matrix in the above expression can be easily computed, and we will have

$$S_{i+1}(z) - S_i(z) = -\frac{1}{n} \frac{d}{dz} \log \left[\left(-r + \left(\frac{2}{\mu} - 1 \right) z \right) S_i(z) - \left(1 - \frac{1}{\mu} \right)^2 \right]. \tag{4.20}$$

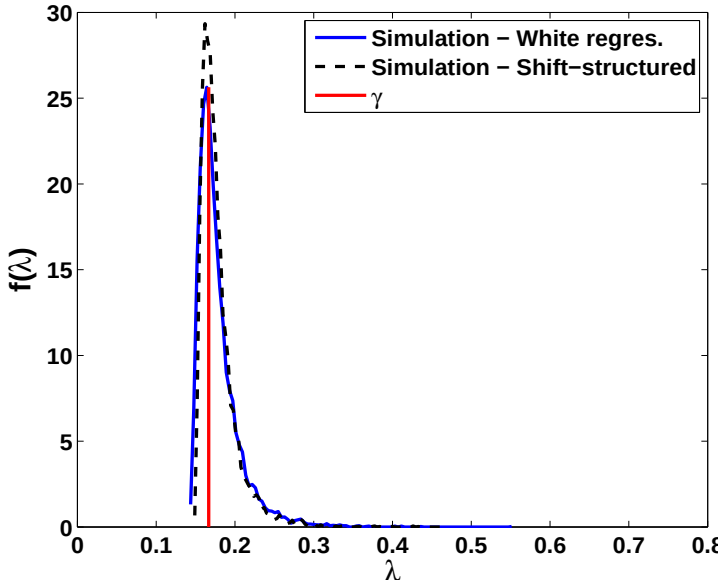


Figure 4.1. Steady-state eigendistribution of recursion (4.7) for $n = 10$, $r = 0.5$, and $\mu = 0.5$..

At the steady-state, this expression reduces to

$$\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S(z) - \left(1 - \frac{1}{\mu}\right)^2 = c \Longrightarrow S(z) = \frac{c}{z - \frac{r}{\frac{2}{\mu} - 1}}, \tag{4.21}$$

where c is some constant. Since $S(z)$ is a Stieltjes transform, the coefficient of $\frac{1}{z}$ in its power series expansion should be -1. In fact this condition says that the eigendistribution should integrate to 1. Clearly $c = -1$ here and the proof is complete. □

Theorem 4.2.1 correctly predicts the accumulation of all the eigenvalues of P_i at a single point γ. In figure 4.1 the empirical eigendistribution is plotted for $n = 10$ which shows how the eigenvalues are concentrated around γ.

We can also use the power expansion of the Stieltjes transform to obtain recursions

on the moments of the distribution from recursion (4.20)

$$S_{i+1}(z) - S_i(z) = -\frac{1}{n}\frac{\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S_i'(z) + \left(\frac{2}{\mu} - 1\right) S_i(z)}{\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S_i(z) - \left(1 - \frac{1}{\mu}\right)^2}. \tag{4.22}$$

Recall that

$$S_i(z) = -\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots .$$

Replacing in (4.22) we have

$$-\frac{m_1^{i+1} - m_1^i}{z^2} - \frac{m_2^{i+1} - m_2^i}{z^3} - \frac{m_3^{i+1} - m_3^i}{z^4} - \cdots =$$
$$\frac{1}{n}\frac{\left(-r + \left(\frac{2}{\mu} - 1\right) z\right)\left(\frac{1}{z^2} + \frac{2m_1^i}{z^3} + \frac{3m_2^i}{z^4} + \cdots\right) + \left(\frac{2}{\mu} - 1\right)\left(-\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots\right)}{(1 - \frac{1}{\mu})^2 - \left(-r + \left(\frac{2}{\mu} - 1\right) z\right)\left(-\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots\right)}. \tag{4.23}$$

We will not go into details of computing the moments in this case. It suffices to say that after equating the coefficients of different power of z on both sides, we obtain,

$$m_1^{i+1} = \left(1 - \frac{\mu^2}{n}\left(\frac{2}{\mu} - 1\right)\right) m_1^i + \mu^2 \frac{r}{n}, \tag{4.24}$$

with the steady-state solution of

$$m_1 = \frac{r}{\frac{2}{\mu} - 1} = \gamma, \tag{4.25}$$

as expected. One can observe from the recursion of m_1^i that a necessary convergence condition requires

$$\left(\frac{2}{\mu} - 1\right) > 0 \Longrightarrow \mu < 2. \tag{4.26}$$

Moreover, as n grows, the convergence becomes slower. A similar expression can

be found for all the other moment of the distribution. For example for the second moment,

$$m_2^{i+1} = \left(1 - \frac{2\mu^2}{n}\left(\frac{2}{\mu} - 1\right)\right) m_2^i + \frac{1}{n}\left(2\mu^2 r m_1^i + \mu^4 \left(r - (\frac{2}{\mu} - 1)m_1^i\right)^2\right). \quad (4.27)$$

Other moments' recursions show that the convergence rate of m_1^i is the dominant rate of convergence. One can also easily check that $m_2 = m_1^2 = \gamma^2$ as one expects from a mass probability at γ.

The temporally whiteness condition on the regressor vectors can be relaxed and one can extend the results of this section to include the shift-structured regressors as is assumed in most applications,

$$h_i = \begin{bmatrix} u_i & u_{i-1} & \cdots & u_{i-n+1} \end{bmatrix}. \quad (4.28)$$

We will leave the technical details of the argument for the subsequent section. In figure 4.1 we have also plotted the empirical eigendistribution for the shift-structured h_i which follows closely that of temporally white regressors.

4.3 LMS Estimation of a Time-Varying State Variable

In the previous section, we looked at the LMS estimation of a constant state vector. In other words the process noise was assumed to be zero. In this section we look at other applications of LMS filtering for which the state vector gets updated according to the linear state-space model (4.1) with the variance of the process noise being

$\mathrm{Q}_i = q\mathrm{I}$, i.e.,

$$\left\{ \begin{array}{rcl} x_{i+1} & = & x_i + u_i \\ y_i & = & h_i^T x_i + v_i \end{array} \right. , \quad \mathbb{E} \left[\begin{array}{c} u_i \\ v_i \end{array} \right] \left[\begin{array}{cc} u_j^T & v_j^T \end{array} \right] = \left[\begin{array}{cc} q\mathrm{I} & 0 \\ 0 & r \end{array} \right] \delta_{ij}.$$

where $x_i \in \mathcal{R}^n$ is an unknown vector to be estimated through the measurements $y_i \in \mathcal{R}$. Here u_i and v_i denote zero-mean process and measurement noises respectively. The state vector essentially undergoes a random walk through u_i. We will first assume that the regressor vectors are temporally and spatially white. In other words, they are independent of each other and have independent entries. The entries are assumed to have identical distributions with zero mean and $\frac{1}{\sqrt{n}}$ variance, and all the higher moments are sought to be bounded. According to (4.5), the estimation error covariance undergoes a random Lyapunov-like recursion

$$\mathrm{P}_{i+1} = \left(\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T + q\mathrm{I}. \tag{4.29}$$

As a linear recursion, it is possible to directly compute the moments of the eigendistribution of P_i by averaging different powers of both sides of the recursion. Here we will apply our method to find a recursion for the Stieltjes transform of P_i which provides a much simpler and more systematic approach to analyzing the steady-state and transient behavior of the LMS estimation process.

Theorem 4.3.1 *When $\mu < \frac{2}{n}$, as $n \to \infty$, the eigendistribution of the Stieltjes transform of P_i in (4.29) satisfies*

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz} \log \left[\left(-r + \left(\frac{2}{\mu} - 1 \right) z \right) S_i(z) - \left(1 - \frac{1}{\mu} \right)^2 \right]. \tag{4.30}$$

When $q = \frac{\kappa}{n}$, where $\kappa = O(1)$, the steady-state Stieltjes transform is the solution of

$$\kappa S(z) + c = -\log\left[\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S(z) - \left(1 - \frac{1}{\mu}\right)^2\right], \tag{4.31}$$

where $c = \log(-\mu^2)$ is a constant.

Proof: Most of the proof follows that of theorem 4.2.1. We start with the recursion (4.29) and apply the definition of the Stieltjes transform, (2.12), to both sides of the recursion,

$$\begin{aligned} -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det(\mathrm{P}_{i+1} - q\mathrm{I} - z\mathrm{I}) \quad &= \\ -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\left(\mathrm{I} - \mu h_i h_i^T\right)\mathrm{P}_i\left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T - z\mathrm{I}\right). \end{aligned} \tag{4.32}$$

The left hand side is just $S_{i+1}(z+q)$. Now we simply repeat the steps in the proof of theorem 4.2.1 for the right hand side to obtain (4.30). When $q = \frac{\kappa}{n}$, at the steady state,

$$S(z+q) - S(z) = \frac{\kappa}{n} S'(z). \tag{4.33}$$

Therefore

$$\frac{\kappa}{n} S'(z) = -\frac{1}{n}\frac{d}{dz}\log\left[\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S_i(z) - \left(1 - \frac{1}{\mu}\right)^2\right], \tag{4.34}$$

which directly leads to (4.31) and c is the integration constant. In order to determine c, we look at the expression as $\Re(z) \to \infty$. The Stieltjes transform belongs to an eigendistribution with bounded support and therefore its real part is zero at infinity, while $zS(z) \to -1$ (see the power series expansion of $S(z)$). From the other terms,

$$c = -\log\left[-\left(\frac{2}{\mu} - 1\right) - \left(1 - \frac{1}{\mu}\right)^2\right] = -\log\left(-\frac{1}{\mu^2}\right), \tag{4.35}$$

which completes the proof. □

It is not hard to see that if q does not scale like $\frac{1}{n}$, the eigendistribution does not have bounded support. Because of the dependence of $S(z+q)$ on $S(z)$ and the difference term being of order $\frac{1}{n}$, the support will be of order nq. Thus for assuring the convergence, we should assume $q = \frac{\kappa}{n}$. This can also be verified intuitively through the recursion itself. We have one single measurement of x_i at each time step. If the changes in x_i have more energy than $\mathrm{tr}q\mathrm{I} = nq = \kappa$, then the estimator will not be able to follow these changes. Therefore throughout the rest of this chapter we will hold to this assumption.

4.3.1 Finding the Eigendistribution

Clearly (4.31) cannot be solved analytically for $S(z)$, and we have to find the eigendistribution numerically (of course an approximation of it). Here we will use Newton step method to find the eigendistribution. Let us write

$$S(z) = u(z) + jv(z), \tag{4.36}$$

$$z = \lambda + j0^{+}. \tag{4.37}$$

We define a function $f(u, v; z)$ based on (4.31),

$$\begin{aligned} f(u,v;z) &= \kappa u + j\kappa v + c_r + jc_i + \\ &\log\left[\left(-r + \left(\frac{2}{\mu} - 1\right)\lambda\right)u - \left(1 - \frac{1}{\mu}\right)^2 + j\left(-r + \left(\frac{2}{\mu} - 1\right)\lambda\right)v\right], \end{aligned} \tag{4.38}$$

where c_r and c_i are the real and imaginary parts of c respectively. Clearly our objective is finding the zeros of $f(u, v; z)$. If we separate the real and imaginary parts of

$f(u, v; z)$,

$$\begin{aligned} f_r(u, v; z) &= \kappa u + c_r \\ &\quad + \frac{1}{2} \log \left[\left(\left(-r + \left(\frac{2}{\mu} - 1 \right) \lambda \right) u - \left(1 - \frac{1}{\mu} \right)^2 \right)^2 + \left(-r + \left(\frac{2}{\mu} - 1 \right) \lambda \right)^2 v^2 \right] \end{aligned} \tag{4.39}$$

$$f_i(u, v; z) = \kappa v + c_i + \tan^{-1} \frac{\left(-r + \left(\frac{2}{\mu} - 1 \right) \lambda \right) v}{\left(-r + \left(\frac{2}{\mu} - 1 \right) \lambda \right) u - \left(1 - \frac{1}{\mu} \right)^2}. \tag{4.40}$$

It is worth noting that the $\tan^{-1}$ inherently has a $2k\pi j$.

In order to numerically find the eigendistribution, we should start at some initial point, $S^0 = u^0 + jv^0$. At each step, $l + 1$, we update the value of S as

$$\begin{bmatrix} u^{l+1} \\ v^{l+1} \end{bmatrix} = \begin{bmatrix} u^l \\ v^l \end{bmatrix} - \left(D(f)|_{(u^l, v^l; z)} \right)^{-1} f(u^l, v^l; z), \tag{4.41}$$

where

$$D(f) = \begin{bmatrix} \frac{\partial f_r}{\partial u} & \frac{\partial f_r}{\partial v} \\ \frac{\partial f_i}{\partial u} & \frac{\partial f_i}{\partial v} \end{bmatrix}. \tag{4.42}$$

In order to simplify the expressions, let us define A and B as

$$A = -r + \left(\frac{2}{\mu} - 1 \right) \lambda, \quad B = \left(1 - \frac{1}{\mu} \right)^2. \tag{4.43}$$

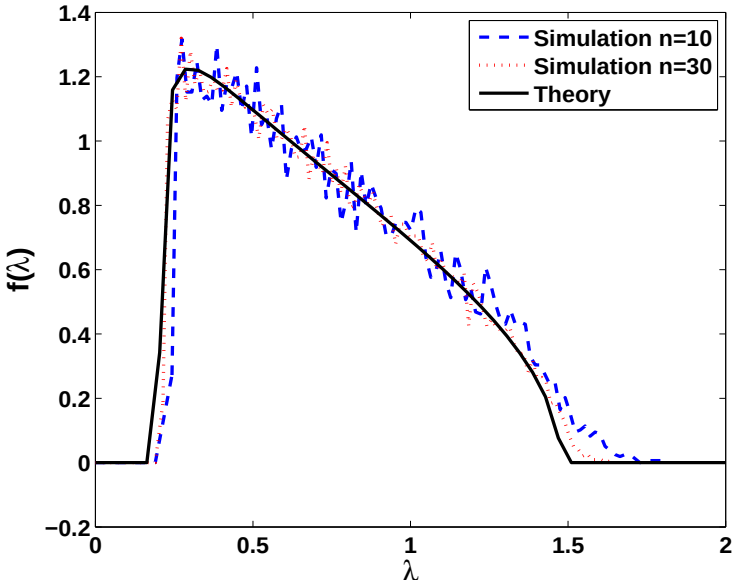


Figure 4.2. Steady-state eigendistribution of recursion (4.29) for $n = 10, 30$, $r = 0.5$, and $\mu = 0.5$..

The entries of $D(f)$ can be readily computed as

$$\frac{\partial f_r}{\partial u} = \kappa + \frac{A(Au - B)}{(Au - B)^2 + A^2v^2}, \tag{4.44}$$

$$\frac{\partial f_r}{\partial v} = \frac{A^2v}{(Au - B)^2 + A^2v^2}, \tag{4.45}$$

$$\frac{\partial f_i}{\partial u} = \frac{-A^2v}{(Au - B)^2 + A^2v^2}, \tag{4.46}$$

$$\frac{\partial f_i}{\partial v} = \kappa + \frac{A^2u - AB}{(Au - B)^2 + A^2v^2}. \tag{4.47}$$

Now we have all the necessary expressions to run the Newton method and numerically find u and v which satisfy $f(u, v; \lambda + j0^+) = 0$. The eigendistribution at λ will be $\frac{1}{\pi}v$ according to the definition. We have plotted the theoretical curve which is obtained as explained versus the empirical eigendistribution for $n = 10$ and $n = 30$. We have used 500 samples of the steady-state error covariance to produce the empirical curves. The tightness of the asymptotic results for n even being 10 can be verified.

4.3.2 Finding the Moments

As we saw in this manuscript so far, the Stieltjes transform's recursion can be used to find recursions of the moments of the eigendistribution. Apart from determining the steady-state moments which is an important issue itself –e.g., the first moment is the mean-squares error performance–, this will help in characterizing the transient behavior of the filtering process. Recall that

$$S_i(z) = -\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots \ .$$

In the regime of small q which as we discussed is the only scenario of interest, (4.30) can be written as,

$$S_{i+1}(z) - S_i(z) + \frac{\kappa}{n} S'_{i+1}(z) = -\frac{1}{n}\frac{d}{dz}\log\left[\left(-r + \left(\frac{2}{\mu} - 1\right)z\right) S_i(z) - \left(1 - \frac{1}{\mu}\right)^2\right]. \tag{4.48}$$

Replacing in (4.22) we have

$$-\frac{m_1^{i+1} - m_1^i}{z^2} - \frac{m_2^{i+1} - m_2^i}{z^3} - \frac{m_3^{i+1} - m_3^i}{z^4} - \cdots + \frac{\kappa}{n}\left(\frac{1}{z^2} + \frac{2m_1^{i+1}}{z^3} + \frac{3m_2^{i+1}}{z^4} + \cdots\right) = \frac{1}{n}\frac{\left(-r + \left(\frac{2}{\mu} - 1\right)z\right)\left(\frac{1}{z^2} + \frac{2m_1^i}{z^3} + \frac{3m_2^i}{z^4} + \cdots\right) + \left(\frac{2}{\mu} - 1\right)\left(-\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots\right)}{(1 - \frac{1}{\mu})^2 - \left(-r + \left(\frac{2}{\mu} - 1\right)z\right)\left(-\frac{1}{z} - \frac{m_1^i}{z^2} - \frac{m_2^i}{z^3} - \frac{m_3^i}{z^4} - \cdots\right)}, \tag{4.49}$$

where the right-hand side is directly written from (4.23). The terms in the denominator can be rearranged as,

$$(\frac{1}{\mu})^2 + \frac{1}{z}\left(-r + \left(\frac{2}{\mu} - 1\right)\right) + \frac{1}{z^2}\left(-rm_1^i + \left(\frac{2}{\mu} - 1\right)m_2^i\right) + \cdots \ . \tag{4.50}$$

The terms in the numerator of the right hand side can be easily found accordingly (since the numerator is the negative of the derivative of the denominator)

$$\frac{1}{z^2}\left(-r+\left(\frac{2}{\mu}-1\right)\right)+\frac{2}{z^3}\left(-rm_1^i+\left(\frac{2}{\mu}-1\right)m_2^i\right)+\cdots. \tag{4.51}$$

Replacing in (4.49) and noting that $(1+x)^{-1}=1-x+x^2-\cdots$, we can find the recursion of the moments. The first two moment follow

$$m_1^{i+1} = m_1^i\left(1-\frac{\mu^2}{n}\left(\frac{2}{\mu}-1\right)\right)+\frac{\mu^2 r+\kappa}{n}, \tag{4.52}$$

$$m_2^{i+1} = m_2^i\left(1-\frac{2\mu^2}{n}\left(\frac{2}{\mu}-1\right)\right)+\frac{(\kappa+2\mu^2 r)m_1^i+\kappa m_1^{i+1}+\mu^4\left(r-\left(\frac{2}{\mu}-1\right)m_1^i\right)^2}{n}. \tag{4.53}$$

Clearly the convergence rate and conditions follow those of the simple case analyzed in the previous section. The steady-state moments will be

$$m_1 = \frac{r+\frac{\kappa}{\mu^2}}{\frac{2}{\mu}-1}, \tag{4.54}$$

$$m_2 = \frac{rm_1+\frac{2\kappa m_1+\kappa^2}{2\mu^2}}{\frac{2}{\mu}-1}. \tag{4.55}$$

4.4 Generalization to Shift-Structured Regressors

So far we have considered regressor vectors which are spatially and temporally white. In this section we will generalize the results to include a more realistic model. We relax this assumption by showing that the results obtained earlier directly extend to the case where we have regressors with shift structure—a model that frequently arises in the literature when we have finite-impulse-response (FIR) adaptive filters

(see, e.g., [Hay01]). Thus, consider

$$h_i = \begin{bmatrix} u_i & u_{i-1} & \cdots & u_{i-n+1} \end{bmatrix}, \tag{4.56}$$

where the u_i are drawn from a zero mean, unit variance white process. The following lemma is the key in analyzing this case.

Lemma 4.4.1 *For a given vector h_i and a randomly chosen orthogonal matrix Θ, there exists a matrix Δ such that*

1. $h_i(\Theta + \Delta) = \|h_i\| e_1$
2. $(\Theta + \Delta)(\Theta + \Delta)^T = I$
3. $\mathbb{E}tr(\Delta\Delta^T) = O(\frac{1}{n})$

Proof: We omit the full proof of this lemma and refer the readers to the discussions in [MHH02, HMH01]. Here we offer some intuition as to why one may expect the result to hold. Note that any orthogonal matrix Θ is determined by $\frac{n(n-1)}{2}$ parameters, whereas a vector h_i is determined by only n parameters. Then it is plausible to think that the matrix Θ can be perturbed a *small* perturbation Δ so that $\Theta + \Delta$ remains orthogonal, yet it "rotates" h_i to lie along the first unit axis. The statement of the lemma gives a precise value on the "size" of Δ when Θ is generated randomly. □

Going back to the proof of theorem 4.3.1, we can see that all the steps were just algebraic manipulations except when we used the self-averaging property. This crucial step should be reexamined in the current scenario. In fact we can no longer assume that h_i and P_i are independent. However, using the aforementioned lemma, we can

show that (4.15) still holds.

$$\begin{aligned} h_i(\mathrm{P}_i - z\mathrm{I})^{-1}h_i^T &= h_i\Theta(\Theta\mathrm{P}_i\Theta^T - z\mathrm{I})^{-1}\Theta^T h_i^T \\ &= (\|h_i\|e_1 - h_i\Delta)(\bar{\mathrm{P}}_i - z\mathrm{I})^{-1}(\|h_i\|e_1 - h_i\Delta)^T \\ &= \|h_i\|^2 e_1(\bar{\mathrm{P}}_i - z\mathrm{I})^{-1}e_1^T - 2\|h_i\|e_1(\bar{\mathrm{P}}_i - z\mathrm{I})^{-1}\Delta^T h_i^T \\ &\qquad + h_i\Delta(\bar{\mathrm{P}}_i - z\mathrm{I})^{-1}\Delta^T h_i^T \\ &\xrightarrow[n\to\infty]{} S(z) + 0 + 0 \end{aligned} \tag{4.57}$$

Similar arguments can be used for $h_i\mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}h_i^T$ and $h_i\mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}\mathrm{P}_i h_i^T$ which show that the results so far will remain valid for the case of shift-structured regressor vectors.

We have to mention that the second term in the right-hand side of (4.57) behaves like $\frac{1}{\sqrt{n}}$. Therefore the convergence to the asymptotic result will be much slower in terms of n compared to the case of uncorrelated regressors. Figure 4.3 shows the simulation results for $n = 30$ and $\kappa = 0.5$. We have also plotted the empirical distribution for the cases where the successive regressors differ by two and three shifts, respectively (in which case more randomness is brought to the regressor vectors at each time step). As expected, with larger shifts the empirical results converge much faster to the theoretical one due to more freedom in choosing Δ.

4.5 Other Examples of LMS Filters

There are several variations of the LMS adaptive filtering which can be similarly considered in our framework. For example the leaky-LMS algorithm updates the estimates as,

$$\hat{x}_{i+1} = (1 - \mu\alpha)\hat{x}_i + \mu h_i(y_i - h_i^T\hat{x}_i), \tag{4.58}$$

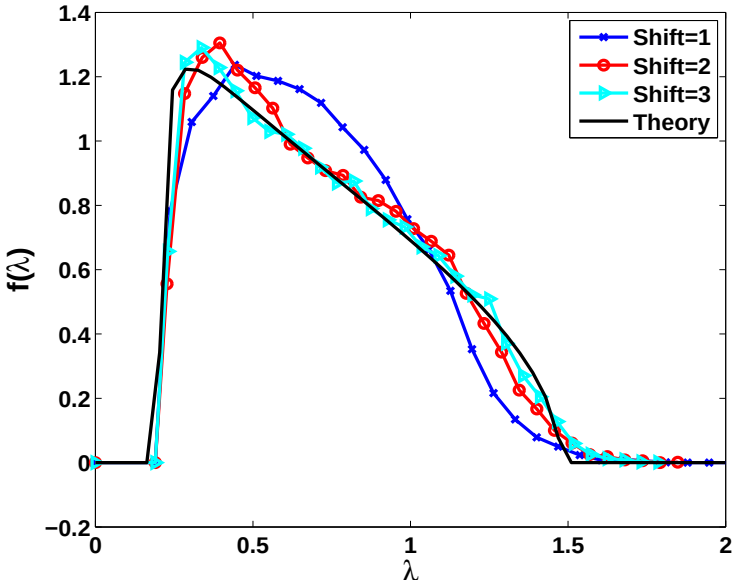


Figure 4.3. Steady-state eigendistribution of recursion (4.29) with shift-structured regressors for $n = 30$, $r = 0.5$, and $\mu = 0.5$, and $\kappa = 0.5$..

for some positive constant α. The leaky-LMS minimizes $\mathbb{E}|y - h^T x| + \alpha||x||^2$. In other words it penalizes the energy of x and is useful to avoid solutions with a large norm. It is not hard to show that the error covariance matrix satisfies a random Lyapunov-like recursion very much like the one we already considered

$$\mathrm{P}_{i+1} = \left((1-\mu\alpha)\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left((1-\mu\alpha)\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T + \mathrm{Q}_i. \tag{4.59}$$

The normalized versions of the LMS filtering, e.g., ϵ-NLMS, will effectively change the step size definition in our analysis,

$$\hat{x}_{i+1} = \hat{x}_i + \frac{\mu}{\epsilon + ||h_i||^2} h_i (y_i - h_i^T \hat{x}_i). \tag{4.60}$$

Also different assumptions may be made on the state update depending on the application. For example in some applications it is assumed that the direction of

the state updates is known, while its magnitude is unknown. In this case, the error covariance matrix satisfies,

$$\mathrm{P}_{i+1} = \left(\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T + q g_i g_i^T, \tag{4.61}$$

where $g_i \in \mathcal{R}^n$ is a column vector independent of h_i with i.i.d. entries having zero mean and $\frac{1}{\sqrt{n}}$ variance. We will analyze this case since it has some technical differences with the previous cases.

Theorem 4.5.1 *The eigendistribution of the P_i in (4.61) converges to a steady-state distribution whose Stieltjes transform is the solution of the quadratic equation,*

$$\left(-r + \left(\frac{2}{\mu} - 1\right) z\right) S^2(z) + \left(-\frac{r}{q} + \frac{1}{q}\left(\frac{2}{\mu} - 1\right) z - \left(1 - \frac{1}{\mu}\right)^2\right) S(z) + \frac{1}{q}\left(\frac{2}{\mu} - 1\right), \tag{4.62}$$

Proof: Applying the definition of the Stieltjes transform (2.12) to both sides of the recursion (4.61),

$$S_{i+1}(z) = -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\left(\mathrm{I} - \mu h_i h_i^T\right) \mathrm{P}_i \left(\mathrm{I} - \mu h_i h_i^T\right) + \mu^2 r h_i h_i^T + q g_i g_i^T - z\mathrm{I}\right), \tag{4.63}$$

The right hand side expression can be written as,

$$(\mathrm{P}_i - z\mathrm{I}) + \begin{bmatrix} g_i h_i & \mathrm{P}_i h_i \end{bmatrix} \underbrace{\begin{bmatrix} q & 0 & 0 \\ 0 & \mu^2(r + h_i^T \mathrm{P}_i h_i) & -\mu \\ 0 & -\mu & 0 \end{bmatrix}}_{\triangleq \mathrm{A}_i} \begin{bmatrix} g_i^T \\ h_i^T \\ h_i^T \mathrm{P}_i \end{bmatrix}.$$

From here on, the proof will be very much like the proof of theorem 4.2.1. The inverse

of A_i can found as

$$\mathrm{A}_i^{-1} = \begin{bmatrix} q^{-1} & 0 & 0 \\ 0 & 0 & -\mu^{-1} \\ 0 & -\mu^{-1} & -(r + h_i^T \mathrm{P}_i h_i) \end{bmatrix} \tag{4.64}$$

Moreover,

$$\begin{bmatrix} g_i^T \\ h_i^T \\ h_i^T \mathrm{P}_i \end{bmatrix} (\mathrm{P}_i - z\mathrm{I})^{-1} \begin{bmatrix} g_i h_i & \mathrm{P}_i h_i \end{bmatrix} = \begin{bmatrix} S_i(z) + q^{-1} & 0 & 0 \\ 0 & S_i(z) & -\mu^{-1} + 1 + zS_i(z) \\ 0 & -\mu^{-1} + 1 + zS_i(z) & -r + z + z^2 S_i(z) \end{bmatrix}. \tag{4.65}$$

Most of the terms in obtaining (4.65) are the same ones we had already calculated in the proof of theorem 4.2.1. The only difference is in terms of the form $g_i^T(\mathrm{P}_i - z\mathrm{I})^{-1}h_i$ which converge to zero since h_i and g_i are assumed independent and zero mean.

Replacing (4.65) and (4.64) in (4.63) yields

$$S_{i+1}(z) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det \begin{bmatrix} S_i(z) + q^{-1} & 0 & 0 \\ 0 & S_i(z) & -\mu^{-1} + 1 + zS_i(z) \\ 0 & -\mu^{-1} + 1 + zS_i(z) & -r + z + z^2 S_i(z) \end{bmatrix}, \tag{4.66}$$

where once again the expectation can be dropped since all the randomness has been used through the self-averaging lemma. The determinant of the 3×3 matrix in the

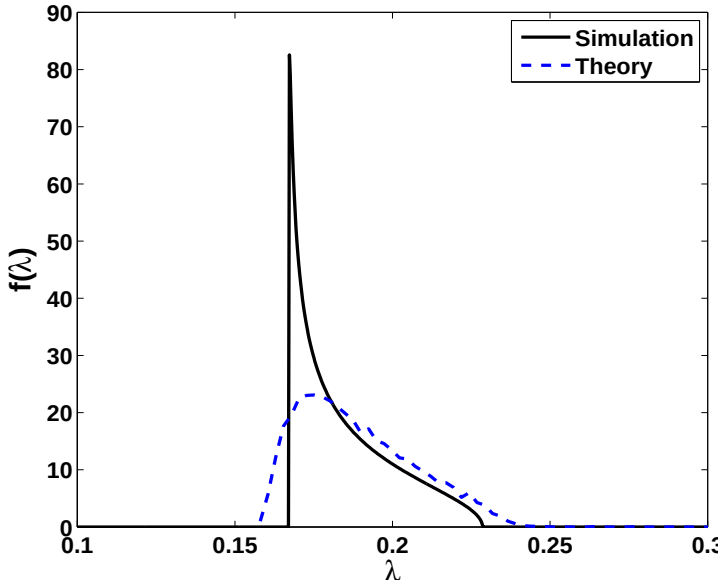


Figure 4.4. Steady-state eigendistribution of recursion (4.61) with shift-structured regressors for $n = 20$, $r = 0.5$, and $\mu = 0.5$, and $\kappa = 0.4$..

above expression is easy to compute, and we will have

$$S_{i+1}(z) - S_i(z) \;=\; -\frac{1}{n}\frac{d}{dz}\log\left\{\left[\left(-r+\left(\frac{2}{\mu}-1\right)z\right)S_i(z) - \left(1-\frac{1}{\mu}\right)^2\right]\left(S_i(z)+q^{-1}\right)\right\}. \tag{4.67}$$

At the steady state, this expression reduces to

$$\left(-r+\left(\frac{2}{\mu}-1\right)z\right)S^2(z)+\left(-\frac{r}{q}+\frac{1}{q}\left(\frac{2}{\mu}-1\right)z-\left(1-\frac{1}{\mu}\right)^2\right)S(z)+c=0, \tag{4.68}$$

where c is a constant. We can find $c = \frac{1}{q}\left(\frac{2}{\mu}-1\right)$ by looking at the expression for z such that $\Re(z) \to \infty$. □

In figure (4.4) we have plotted the empirical eigendistribution along with the one found by solving (4.62) for $S(z)$.

4.6 Conclusions

The least-mean-squares algorithm is the simplest, yet the most frequently used algorithm in adaptive filtering. In this chapter we showed how our developed framework can be applied to the analysis of such filters. The performance of LMS algorithm and its variations is mainly governed by random Lyapunov-like recursions. Since a linear recursion, different aspects of these filters are well studied and can be found in the literature. Our approach however introduces a systematic way of analyzing both the transient and the steady-state behavior of LMS filter. More importantly for us is that the framework developed can also be applied to the Riccati recursions about which very little is known unless with somewhat unrealistic assumptions.

Several variations of the LMS filters were considered, both in terms of the estimation update algorithm and the state's time behavior. The eigendistribution and the transient behavior of the moments were analyzed. Our results and method can help with finding many different performance metrics which are used in the field of adaptive filtering. Moreover, we have shown that the actual distribution of the underlying model is not important. For example we one only needs the conditions on the first and second moments of the entries of the regressor vectors in order for the analysis to hold.

We derived the results for temporally white regressors, i.e., when the regressors at different time instances are independent. We later proved that the results can be extended to the shift-structured regressors with a lower convergence to the asymptotic results for moderate values of n.

Throughout this chapter we have only considered single measurements. The results can 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 anymore. A more complicated analysis

can deal with that case. Since Recursive Least-Squares (RLS) filters are of more importance, we will leave such an analysis for the next chapter, with just remarking that the same extension can be done for LMS filters.

Chapter 5

The Recursive Least-Squares Filters

5.1 Introduction

One of the most important families of adaptive filters is the so-called *recursive least squares* (RLS) filter [Say03, Hay01, HM84]. As its name suggests, the RLS filter is an optimal filter that finds an estimate of a weight vector that recursively solves a least-square-error minimization problem. Moreover, it often demonstrates much faster convergence in its learning curve compared to other classes of adaptive filters such as LMS filter. There are also several variants of the RLS filter which are mathematically equivalent but differ considerably in terms of computational expense and ease of implementation. Due to these reasons, the RLS filter has received much attention in the literature and in practice. Although the idea of the RLS algorithm has its roots going back to Gauss, when the regressors are random, as is the case in many applications, the RLS filter has proven to be one of the most difficult adaptive algorithms to analyze. In fact, the so-called excess mean-square error (EMSE) of an RLS filter is not known except for approximations based on several unrealistic assumptions. The ubiquitous approximation in the literature is the *ergodic assumption* [Say03]. This assumption asks for the estimation error covariance matrix, P_i, to

satisfy

$$\mathbb{E}\mathrm{P}_i = \left(\mathbb{E}\mathrm{P}_i^{-1}\right)^{-1}. \tag{5.1}$$

This is clearly not a valid assumption. It can be shown (numerically) that for Gaussian regressor vectors it is a reasonable assumption (less than 10% error). The distortion becomes more and more severe for smaller forgetting factors.

The mean-square performance, as well as the tracking performance, of the RLS filter under the ergodic approximation and for temporally white regressors can be found in the literature [EF86, MIK00, Ewe94]. Our methodology on the other hand does not require such an assumption. Although our approach asks for the number of taps to be relatively large, as we will see, $n = 10$ is large enough to guarantee convergence to the asymptotic results.

We start by giving a brief overview of the recursive least-squares algorithm. We will stay in the framework of the generalized RLS filter [Say03, p. 772]. Recall the state-space model we considered in the previous chapters,

$$\left\{ \begin{array}{rcl} x_{i+1} &=& x_i + u_i \\ y_i &=& h_i x_i + v_i \end{array} \right. , \quad \mathbb{E} \left[\begin{array}{c} u_i \\ v_i \end{array} \right] \left[\begin{array}{cc} u_j^T & v_j^T \end{array} \right] = \left[\begin{array}{cc} \mathrm{Q}_i & 0 \\ 0 & r \end{array} \right] \delta_{ij}, \tag{5.2}$$

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. Here y_i is the measured signal and h_i is the so-called regressor vector. The regressor vectors are time-variant and often random.

The RLS recursively minimizes,

$$\min_x \left[x^T \Pi_0^{-1} x + ||y(n) - H(n)^T x||^2 \right], \tag{5.3}$$

where $x(n) = [x_0 \ x_1 \ \cdots \ x_n]^T$, $H(n) = [h_0^T \ h_1^T \ \cdots \ h_n^T]$ and $y(n) = [y_0 \ y_1 \ \cdots \ y_n]^T$,

and Π_0 is some positive definite matrix. It can be shown that [KSH00, SK94] the optimal estimate of x_i is

$$\hat{x}_{i+1} = \hat{x}_i + \frac{\mathrm{P}_i h_i^T}{r + h_i \mathrm{P}_i h_i^T}(y_i - h_i \hat{x}_i), \tag{5.4}$$

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

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T} + q\mathrm{I}, \quad \mathrm{P}_0. \tag{5.5}$$

The above Riccati recursion is nonlinear and time variant and, in general, does not converge. Furthermore, when the h_i are random, it is a random matrix recursion. As we explained earlier, when the h_i are stationary it may be expected that P_i will converge to a matrix-valued stationary random process. In this case, the stationary distribution of P_i, and its statistics, such as $\mathbb{E}\mathrm{P}_i$ or $\mathbb{E}\mathrm{tr}\mathrm{P}_i$, will be of interest. This is what we intend to analyze in this chapter for RLS adaptive filters.

We should also mention that the filter described by (5.2 - 5.5) is one of many variants of the RLS algorithm (such as those with a forgetting factor—see, e.g., [Say03, Hay01]). As we will show, our analysis easily extends to those variants, and in this chapter we mainly focus on the generalized filter (5.2 - 5.5). Finally it is also worth mentioning that the basic recursion (5.5) is closely related to that of the H^∞ adaptive filter [HSK99], [KSH00]. Therefore our approach can readily be extended to analyze the H^∞ filter. We will talk further about this connection in chapter 8.

This chapter is organized as follows. In section 5.2 we will analyze the basic RLS filter with temporally uncorrelated regressors and find the steady-state eigenvalue distribution of the error covariance matrix. Moreover, simulation results will be provided which demonstrate a very close match between the theoretical and empirical

curves. The analysis can be extended to include a more realistic structure for the regressor vectors, namely shift structure, in the same way we did for the LMS filters. In section 5.4 we further generalize the results to include the case of intermittent measurements where the measurements $\{y_i\}$ are transmitted through a network and may be dropped with some probability. This problem is motivated by the recently growing interest in estimation and control over lossy networks and creates the foundation for later chapters. We show how this scenario can be handled easily through our method and how closely we can predict the empirical curve. Finally section 5.5 concludes the chapter.

5.2 RLS Filter with Temporally Uncorrelated Regressor Vectors

First let us consider the RLS adaptive filter assuming that the regressor vectors are spatially and temporally uncorrelated. Although this assumption is widely used in the literature for analysis purposes, it is often not too reasonable. However it is not hard to generalize the results to more realistic cases such as shift-structured regressors. For the time being, any h_i is assumed to be comprised of zero-mean i.i.d. entries that are further independent of all other $h_{j\neq i}$'s. We will also assume a pure random walk for state updates. This model results in the following Riccati recursion for the error covariance,

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T} + q\mathrm{I}, \quad \mathrm{P}_0, \tag{5.6}$$

where h_i are independently drawn $n \times 1$ vectors having iid zero mean entries with variance $\frac{1}{\sqrt{n}}$ and finite higher order moments, while q is a constant representing the process noise variance. Clearly P_i will not converge to any steady-state P since the h_i are changing through time. However, its probability distribution does converge

to a steady state. The stability of the filter has been extensively analyzed in the literature [Say03, Hay01]. The convergence of P_i to a steady-state distribution can also be directly deduced from the work of Sinopoli and co-workers [KSM09].

One also should note that due to the highly nonlinear nature of this recursion, all the moments of the eigendistribution are coupled and therefore one cannot find the moments of the distribution by averaging both sides of (5.6) as was possible for Lyapunov-like recursions. However, using the theory of large random matrices, we can find the asymptotic behavior of the eigendistribution.

The following Theorem presents the main result.

Theorem 5.2.1 (RLS with temporally independent regressors) *Consider the random Riccati recursion in (5.6) where $n \to \infty$. The Stieltjes transform of the eigendistribution of P_i, $S_i(z)$, satisfies*

$$S_{i+1}(z+q) = S_i(z) + \frac{1}{n} \times \frac{1 + 2zS_i(z) + z^2 S_i'(z)}{r - z - z^2 S_i(z)}. \tag{5.7}$$

When $q = \frac{\kappa}{n}$ where $\kappa = O(1)$ (with respect to n), the steady-state Stieltjes transform $S(z)$ is the solution of

$$\kappa S(z) + c = -\log\left(r - z - z^2 S(z)\right), \tag{5.8}$$

where c is a constant to be determined.

Proof: We will use the definition of Stieltjes transform as in (2.12) to the recursion (5.6). The left hand side becomes

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{P}_{i+1} - q\mathrm{I} - z\mathrm{I}\right) = S_{i+1}(z+q). \tag{5.9}$$

Thus

$$\begin{aligned} S_{i+1}(z+q) &= -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T} - z\mathrm{I}\right) \\ &= -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\left\{\det\left(\mathrm{P}_i - z\mathrm{I}\right) \times \det\left(\mathrm{I} - (\mathrm{P}_i - z\mathrm{I})^{-1}\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T}\right)\right\} \\ &= -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{P}_i - z\mathrm{I}\right) \\ &\qquad -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I} - (\mathrm{P}_i - z\mathrm{I})^{-1}\,\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T}\right). \end{aligned} \tag{5.10}$$

The first term in the right hand side is just the Stieltjes transform at time i

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I} - (\mathrm{P}_i - z\mathrm{I})^{-1}\,\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T}\right). \tag{5.11}$$

Now since $\det(\mathrm{I} - \mathrm{AB}) = \det(\mathrm{I} - \mathrm{BA})$, we can rewrite (5.11) as,

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I} - h_i \mathrm{P}_i\,(\mathrm{P}_i - z\mathrm{I})^{-1}\,\frac{\mathrm{P}_i h_i^T}{r + h_i \mathrm{P}_i h_i^T}\right). \tag{5.12}$$

Since $\frac{1}{r+h_i \mathrm{P}_i h_i^T}$ does not depend on z,

$$\begin{aligned} S_{i+1}(z+q) &= S_i(z) \\ &\quad -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\left\{\det\left(r + h_i \mathrm{P}_i h_i^T\right)^{-1}\det\left(r + h_i \mathrm{P}_i h_i^T - h_i \mathrm{P}_i\,(\mathrm{P}_i - z\mathrm{I})^{-1}\,\mathrm{P}_i h_i^T\right)\right\} \\ &= S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log^{-1}\det\left(r + h_i \mathrm{P}_i h_i^T - h_i \mathrm{P}_i\,(\mathrm{P}_i - z\mathrm{I})^{-1}\,\mathrm{P}_i h_i^T\right). \end{aligned} \tag{5.13}$$

The self-averaging lemma for the Stieltjes transform can be invoked now.

$$\begin{aligned} h_i \mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}\mathrm{P}_i h_i^T &= h_i \mathrm{P}_i(\mathrm{P}_i - z\mathrm{I})^{-1}(\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})h_i^T \\ &= h_i \mathrm{P}_i h_i^T + z h_i(\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})(\mathrm{P}_i - z\mathrm{I})^{-1}h_i^T \\ &= h_i \mathrm{P}_i h_i^T + z h_i h_i^T + z^2 h_i(\mathrm{P}_i - z\mathrm{I})^{-1}h_i^T \\ &\xrightarrow[n\to\infty]{} h_i \mathrm{P}_i h_i^T + z + z^2 S_i(z) \end{aligned} \tag{5.14}$$

Replacing (5.14) in (5.13),

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log^{-1}\det\left(r - z - z^2 S_i(z)\right). \tag{5.15}$$

Noting that no randomness is left in the right hand side, we can drop the $\mathbb{E}$ notion and (5.7) is obtained. When $q = \frac{\kappa}{n}$,

$$S_{i+1}(z + \frac{\kappa}{n}) = S_{i+1}(z) + \frac{\kappa}{n} S'_{i+1}(z). \tag{5.16}$$

And in the steady-state

$$S(z) + \frac{\kappa}{n} S'(z) = S(z) - \frac{1}{n}\frac{d}{dz}\log(r - z - z^2 S(z)), \tag{5.17}$$

which yields (5.8) after integrating both sides. The constant c is the integration constant. □

We are assuming that the state vector's dimension is large. Since there is only one measurement at each time step, one can verify through (5.6) that the state vector update should have a variance which behaves like $q \propto \frac{1}{n}$, otherwise the RLS algorithm will have too much error to be said to be working. As a matter of fact, it is clear from (5.8) that the support of the eigendistribution will be $O(n)$ which is not desirable.

Therefore we should consider (5.6) for $q = \frac{\kappa}{n}$ where κ is some finite constant. This is why we have singled out this case in the above theorem.

There is an important difference between this case and the LMS recursions we studied in the previous chapter. We address this difference in the following corollary.

Corollary 5.2.2 *The constant c in (5.8) depends on the first moment of the eigendistribution as*

$$c = -\log(r + m_1). \tag{5.18}$$

Proof: We can use the power series expansion of the Stieltjes transform to find an expression for c. In fact (5.8) can be rewritten as

$$-\frac{\kappa}{z} - \frac{\kappa m_1}{z^2} + \cdots + c = -\log\left(r - z + z + m_1 + \frac{m_2}{z} + \cdots\right). \tag{5.19}$$

By letting z to assume a real part going to infinity, (5.18) will be immediate. □

Thus c depends on the first moment of the eigendistribution and cannot be determined explicitly as we did for the Lyapunov-like recursions. This was expected due to the moments being all coupled in a nonlinear recursion such random Riccati recursion; otherwise we could find all the moments readily through Laurent series expansion of $S(z)$. However, c can be numerically determined (with little effort) by insisting that the eigendistribution, which is directly related to the imaginary part of the Stieltjes transform, integrates to one.

5.2.1 Finding the Eigendistribution

Here we will develop a numerical method of finding the eigendistribution from the expression satisfied by the Stieltjes transform at the steady state. It will be similar

to the one we used in section 4.3.1 for the LMS algorithm. Denote

$$\begin{aligned} S(z) &= u(z) + jv(z), \quad (5.20) \\ z &= \lambda + j0^{+}. \quad (5.21) \end{aligned}$$

We define a function $f(u, v; z)$ based on (5.8),

$$f(u, v; z) = \kappa u + j\kappa v + c_r + jc_i + \log\left[\left(r - \lambda - \lambda^2 u\right) + j\left(-\lambda^2 v\right)\right], \quad (5.22)$$

where c_r and c_i are the real and imaginary parts of c respectively. We need to find the zeros of $f(u, v; z)$ for z very close to the real lne. Separating the real and imaginary parts of $f(u, v; z)$ yields the following two expressions

$$\begin{aligned} f_r(u, v; z) &= \kappa u + c_r + \frac{1}{2}\log\left[\left(r - \lambda - \lambda^2 u\right)^2 + \lambda^4 v^2\right], \\ f_i(u, v; z) &= \kappa v + c_i + \tan^{-1}\frac{\lambda^2 v}{\lambda + \lambda^2 u - r}. \quad (5.23) \end{aligned}$$

In order to numerically find the eigendistribution, we should start at some initial point, $S^0 = u^0 + jv^0$. At each step, $l + 1$, we update the value of S as

$$\begin{bmatrix} u^{l+1} \\ v^{l+1} \end{bmatrix} = \begin{bmatrix} u^{l} \\ v^{l} \end{bmatrix} - \left(D(f)|_{(u^l, v^l; z)}\right)^{-1} f(u^l, v^l; z), \quad (5.24)$$

where,

$$D(f) = \begin{bmatrix} \frac{\partial f_r}{\partial u} & \frac{\partial f_r}{\partial v} \\ \frac{\partial f_i}{\partial u} & \frac{\partial f_i}{\partial v} \end{bmatrix}. \quad (5.25)$$

The entries of $D(f)$ can be found by direct derivation as

$$\frac{\partial f_r}{\partial u} = \kappa - \frac{\lambda^2(r-\lambda-\lambda^2 u)}{(r-\lambda-\lambda^2 u)^2+\lambda^4 v^2}, \tag{5.26}$$

$$\frac{\partial f_r}{\partial v} = \frac{\lambda^4 v}{(r-\lambda-\lambda^2 u)^2+\lambda^4 v^2}, \tag{5.27}$$

$$\frac{\partial f_i}{\partial u} = \frac{\lambda^4 v}{(r-\lambda-\lambda^2 u)^2+\lambda^4 v^2}, \tag{5.28}$$

$$\frac{\partial f_i}{\partial v} = \kappa - \frac{\lambda^2(r-\lambda-\lambda^2 u)}{(r-\lambda-\lambda^2 u)^2+\lambda^4 v^2}. \tag{5.29}$$

We can now run the Newton method and numerically find u and v which satisfy $f(u,v;\lambda+j0^+)=0$. We start with some value of c and obtain an inverse Stieltjes transform (in fact, $\frac{1}{\pi}v$). This inverse should integrate to 1. Therefore we numerically solve for a c which satisfies this condition (another Newton method). In figure 5.1 this theoretically found curve is compared with the simulation results. We have used 500 samples of the steady-state error covariance to produce the empirical curves. The tightness of the asymptotic results for n even being 10 can be verified.

5.2.2 Moments of the Eigendistribution

As we have already remarked, unlike the Lyapunov-like cases the moments are all coupled for Riccati recursions. Therefore we cannot find the moments at time $i+1$ in terms of moments of the same or lower order at time i. To see this, recall the power series expansion of the Stieltjes transform.

$$S_i(z) = -\frac{1}{z} - \frac{m_1^i}{z^2} + \frac{m_2^i}{z^3} + \dots$$

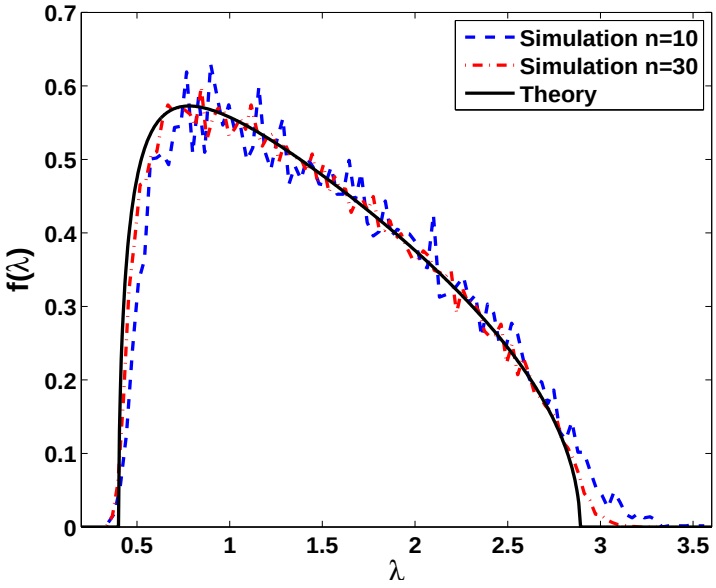


Figure 5.1. Steady-state eigendistribution of recursion (5.6) for $n = 10, 30$, $r = 1$, and $\kappa = 1$..

Replacing in (5.7) while $q = \frac{\kappa}{n}$,

$$-\frac{m_1^{i+1} - m_1^i}{z^2} - \frac{m_2^{i+1} - m_2^i}{z^3} - \cdots + \frac{\kappa}{n}\left(\frac{1}{z^2} + \frac{2m_1^{i+1}}{z^3} + \cdots\right) = \frac{1}{n}\frac{\frac{m_2^i}{z^2} + \frac{2m_3^i}{z^3} + \cdots}{r + m_1^i + \frac{m_2^i}{z} + \frac{m_3^i}{z^2} + \cdots}. \tag{5.30}$$

Therefore,

$$\begin{aligned}\frac{1}{z^2}\left(m_1^i - m_1^{i+1} + \frac{\kappa}{n}\right) + \frac{1}{z^3}\left(m_2^i - m_2^{i+1} + 2\frac{\kappa}{n}m_1^{i+1}\right) + \cdots \\ = \frac{1}{nz^2(r + m_1^i)}\left(m_2^i + \frac{2m_3^i}{z} + \cdots\right) \times \left(1 - \frac{m_2^i}{(r + m_1^i)z} + \cdots\right)\end{aligned} \tag{5.31}$$

We can find that the first moments satisfies

$$m_1^{i+1} = m_1^i + \frac{\kappa}{n} - \frac{1}{n}\frac{m_2^i}{r + m_1^i}. \tag{5.32}$$

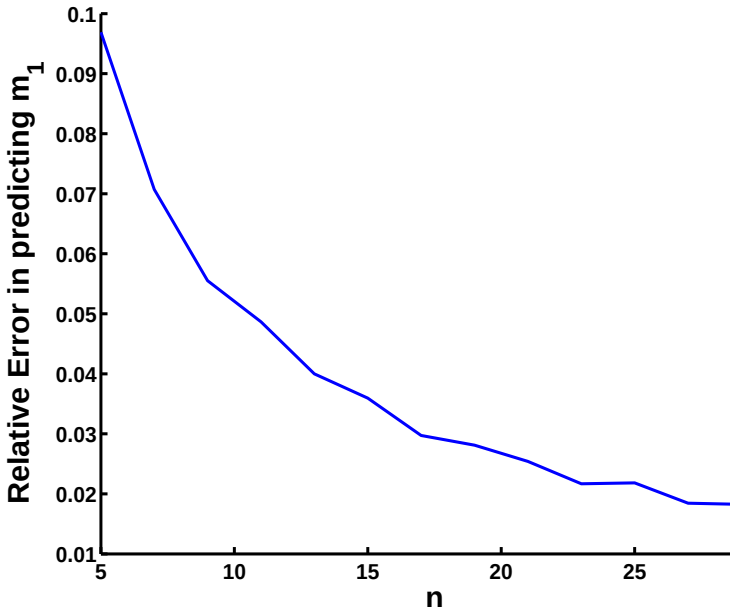


Figure 5.2. Relative error in prediction of the $m_1 = \mathbb{E}\lambda = \frac{1}{n}\mathbb{E}\text{trP}$ in recursion (5.6) for different values of n, the vector size, $r = 1$, and $\kappa = 1$..

Thus m_1^{i+1} is a function of m_2^i and analyzing the moments separately and sequentially will not be possible. To find the moments we should follow the recipe of the last subsection and find the eigendistribution first and obtain the moments through that quantity. In figure 5.2 we have plotted the error in predicting $m_1 = \mathbb{E}\lambda = \frac{1}{n}\mathbb{E}\text{trP}$ for different values of n. We find m_1 through simulation and then compare it to the m_1 found by insisting that the eigendistribution should integrate to one. As can be seen in the plot, the relative error is less than 3% for $n > 20$ and less than 10% for n as small as 5.

5.2.3 Generalization to the Shift-Structured Regressors

A similar argument as what we presented in section 4.3.2 can be done for the RLS filter we have considered here. In our analysis so far we have only focused on regressor vectors which are spatially and temporally white. In other words when not only the entries of h_i are independent, but so are h_i and any $h_{j\neq i}$. However the results are still

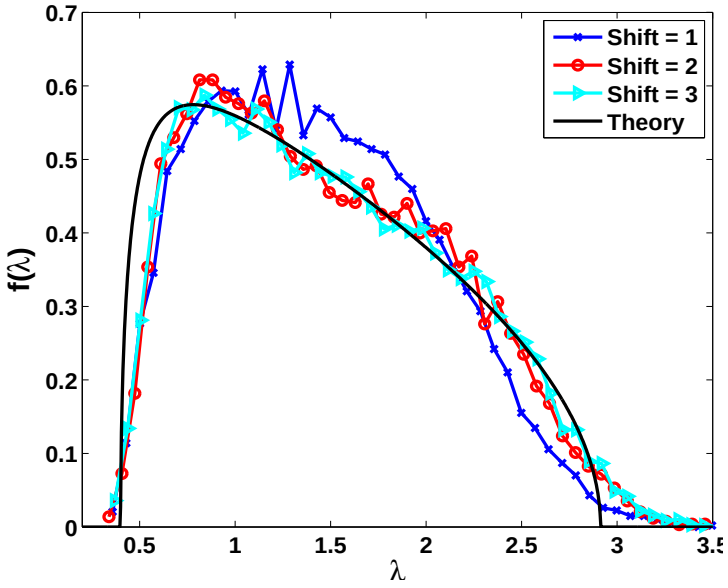


Figure 5.3. Steady-state eigendistribution of recursion (5.6) with shift-structured regressors for $n = 10$, $r = 1$ and $\kappa = 1$..

valid when looking at regressor vectors with a shift structure, i.e.,

$$h_i = \begin{bmatrix} u_i & u_{i-1} & \cdots & u_{i-n+1} \end{bmatrix}, \tag{5.33}$$

where the u_i are drawn from a zero mean, unit variance white process. As we mentioned earlier, this model is very common in the literature, particularly when considering finite-impulse-response (FIR) adaptive filters [Hay01].

One can use lemma 4.4.1 to show that while h_i and P_i are no longer independent, the crucial step at which the proof invokes the self-averaging lemma still holds. Therefore the Stieltjes transform's recursion and the steady-state expression behavior will remain the same.

The only difference is that the second-order term in the asymptotic analysis will be $\frac{1}{\sqrt{n}}$ instead of $\frac{1}{n}$, and consequently we have a slower convergence to the asymptotic results. Adding more randomness in the form of more than one shift at each time

step makes the actual eigendistribution for moderate n be closer to the theoretical prediction. Figure 5.3 shows the asymptotic eigendistribution along with the empirically found curves for $n = 10$, with different number of shifts at each time step. The empirical curves are found by looking at 1000 realizations of the recursion.

5.2.4 Incorporating the Forgetting Factor

Recursive Least-Squares filtering in its nature performs a least-squares minimization based on all the available data, from $i = 0$ to the current time. In order for the filter to be more dynamic and capture the changes in the statistics of the incoming data, it is common to assume some kind of forgetting procedure. In other words, giving less weight to the old measurements compared to the recent ones. The forgetting factor is the most common way of implementing this idea.

The RLS filter with a forgetting factor of $0 \ll \alpha < 1$ minimizes,

$$\min_x \left[x^T \left(\alpha^{-(i+1)} \Pi_0 \right)^{-1} x + \sum_{j=0}^{i} \alpha^{i-j} |y_j - h_j x|^2 \right] \tag{5.34}$$

It can be shown that [KSH00],

$$\hat{x}_{i+1} = \hat{x}_i + \alpha^{-1} \frac{\mathrm{P}_i h_i^T}{r + \alpha^{-1} h_i \mathrm{P}_i h_i^T} (y_i - h_i \hat{x}_i), \tag{5.35}$$

and

$$\mathrm{P}_{i+1} = \alpha^{-1} \mathrm{P}_i - \alpha^{-2} \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + \alpha^{-1} h_i \mathrm{P}_i h_i^T} + q\mathrm{I}, \quad \mathrm{P}_0 \tag{5.36}$$

In order to take into account the effect of the forgetting factor, we note that in the right hand side of (5.36) we can absorb all α^{-1} into P_i. Now since

$$S_{\alpha^{-1}\mathrm{P}_i} = \alpha S_{\mathrm{P}_i}(\alpha z), \tag{5.37}$$

we can replace $S_i(z)$ in (5.7) accordingly, i.e.,

$$S_{i+1}(z+q) = \alpha S_i(\alpha z) + \frac{1}{n}\frac{1 + 2z\alpha S_i(\alpha z) + z^2\alpha^2 S_i'(\alpha z)}{r - z - z^2\alpha S_i(\alpha z)}, \tag{5.38}$$

which can be numerically solved for $S(z)$ to find the eigendistribution.

5.3 RLS Estimation of a State Vector with Known Update Direction

In some applications of adaptive algorithms it is assumed that the direction of state updates is known. In other words, $\mathrm{Q}_i = q g_i g_i^T$, where the $g_i \in \mathcal{R}^n$ are independent of the h_i, and they are temporally and spatially white, which means that entries of g_i are i.i.d. and every g_i and $g_{j\neq i}$ are independent. The error covariance matrix undergoes a Riccati recursion of the form

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T} + q g_i g_i^T, \quad \mathrm{P}_0 \cdot \tag{5.39}$$

The entries of g_i have a zero-mean, $\frac{1}{\sqrt{n}}$-variance distribution with bounded higher moments. The Stieltjes transform of P_i satisfies the recursion presented in the next theorem.

Theorem 5.3.1 *Consider the random Riccati recursion in (5.39) where $n \to \infty$. The Stieltjes transform of the eigendistribution of P_i, $S_i(z)$, satisfies*

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz}\log\left[\left(S_i(z) + \frac{1}{q}\right)\left(-r + z + z^2 S_i(z)\right)\right]. \tag{5.40}$$

The steady-state eigendistribution is equal to

$$f_P(\lambda) = \frac{1}{\pi}\left|\text{Im}\,\frac{\sqrt{\left(q^{-1}\lambda^2+\lambda-r\right)^2-4\lambda^2\left(q^{-1}\lambda+c\right)}}{2\lambda^2}\right|, \tag{5.41}$$

where c is a constant to be determined and Im *denotes the imaginary part. The eigendistribution is zero whenever the imaginary part is zero.*

Proof: Using the definition of Stieltjes transform as in (2.12) in both sides of the recursion (5.39) we have

$$S_{i+1}(z+q) = -\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{P}_i - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r+h_i\mathrm{P}_i h_i^T} + q g_i g_i^T - z\mathrm{I}\right),$$

which can be written as,

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\left\{\det\left(\mathrm{P}_i - z\mathrm{I}\right) \;\times\; \det\left(\mathrm{I}+\left(\mathrm{P}_i-z\mathrm{I}\right)^{-1}\left(-\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r+h_i\mathrm{P}_i h_i^T}+q g_i g_i^T\right)\right)\right\},$$

and further,

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{P}_i - z\mathrm{I}\right) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I}+\left(\mathrm{P}_i-z\mathrm{I}\right)^{-1}\left(-\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r+h_i\mathrm{P}_i h_i^T}+q g_i g_i^T\right)\right).$$

The first term is just the Stieltjes transform at time i.

$$S_{i+1}(z+q) = S_i(z) - \frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I}+\left(\mathrm{P}_i-z\mathrm{I}\right)^{-1}\left(-\frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r+h_i\mathrm{P}_i h_i^T}+q g_i g_i^T\right)\right) \tag{5.42}$$

The second term can be rearranged as

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I}+\left(\mathrm{P}_i-z\mathrm{I}\right)^{-1}\begin{bmatrix}\mathrm{P}_i h_i^T & g_i\end{bmatrix}\begin{bmatrix}-\frac{1}{r+h_i\mathrm{P}_i h_i^T} & 0\\ 0 & q\end{bmatrix}\begin{bmatrix}h_i\mathrm{P}_i\\ g_i^T\end{bmatrix}\right). \tag{5.43}$$

Now since $\det(\mathrm{I}-\mathrm{AB})=\det(\mathrm{I}-\mathrm{BA})$, we can rewrite (5.43) as

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\mathrm{I}+\begin{bmatrix} h_i\mathrm{P}_i \\ g_i^T \end{bmatrix}(\mathrm{P}_i-z\mathrm{I})^{-1}\begin{bmatrix} \mathrm{P}_i h_i^T & g_i \end{bmatrix}\begin{bmatrix} -\frac{1}{r+h_i\mathrm{P}_i h_i^T} & 0 \\ 0 & q \end{bmatrix}\right). \tag{5.44}$$

The 2×2 matrix in the above expression does not have z dependency. Therefore we can readily bring it out of the $\frac{d}{dz}\mathbb{E}\log\det$, which leaves us with,

$$-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\left(\begin{bmatrix} -r-h_i\mathrm{P}_i h_i^T & 0 \\ 0 & q^{-1} \end{bmatrix}+\begin{bmatrix} h_i\mathrm{P}_i \\ g_i^T \end{bmatrix}(\mathrm{P}_i-z\mathrm{I})^{-1}\begin{bmatrix} \mathrm{P}_i h_i^T & g_i \end{bmatrix}\right). \tag{5.45}$$

Therefore,

$$\begin{aligned} S_{i+1}(z+q) &= S_i(z) \\ &-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\begin{bmatrix} -r-h_i\mathrm{P}_i h_i^T+h_i\mathrm{P}_i(\mathrm{P}_i-z\mathrm{I})^{-1}\mathrm{P}_i h_i^T & h_i\mathrm{P}_i(\mathrm{P}_i-z\mathrm{I})^{-1}g_i \\ g_i^T(\mathrm{P}_i-z\mathrm{I})^{-1}\mathrm{P}_i h_i^T & q^{-1}+g_i^T(\mathrm{P}_i-z\mathrm{I})^{-1}g_i \end{bmatrix}. \end{aligned} \tag{5.46}$$

Now we can use the self averaging lemma in the same way we did in proofs of theorem 4.5.1 and theorem 4.2.1, which yields

$$S_{i+1}(z+q)=S_i(z)-\frac{1}{n}\frac{d}{dz}\mathbb{E}\log\det\begin{bmatrix} -r+z+z^2S_i(z) & 0 \\ 0 & S_i(z)+q^{-1} \end{bmatrix}. \tag{5.47}$$

Noting the absence of randomness in the right hand side, we can omit the $\mathbb{E}$ notion and (5.40) is obtained.

At the steady state, (5.40) simplifies to

$$\frac{d}{dz}\log\left[\left(S(z)+\frac{1}{q}\right)\left(-r+z+z^2S(z)\right)\right]=0, \tag{5.48}$$

which means

$$\left(S(z)+\frac{1}{q}\right)\left(-r+z+z^2S(z)\right)=c', \tag{5.49}$$

where c' is some constant. We can solve the above quadratic equation for $S(z)$ with z being the parameter

$$S(z)=\frac{q^{-1}z^2+z-r\pm\sqrt{\left(q^{-1}z^2+z-r\right)^2-4z^2\left(q^{-1}z+c\right)}}{2z^2}, \tag{5.50}$$

where $c=-c'-q^{-1}r$. It is straightforward to check that in the limit the imaginary part of z does not contribute to the imaginary part of $S(z)$, and (5.41) will be immediate recalling that the eigendistribution is $\frac{1}{\pi}$ times the imaginary part of the Stieltjes transform very close to the real line. □

Once again, with no surprise, it is not possible to separately determine c. As a matter of fact one can show that

$$c=\frac{m_1}{q}, \tag{5.51}$$

using the same argument that we have used so far for $\mathrm{Re}(z)\to\infty$. However one can numerically find $f_{\mathrm{P}}(\lambda)$ and find c by insisting that the distribution should integrate to 1. In figure 5.4 the numerical solution for $f_{\mathrm{P}}(\lambda)$ is plotted versus the eigendistribution empirically found by running the recursion and looking at the steady-state eigenvalues. The simulation results are generated for 1000 samples of the recursion with $n=10$ and $n=30$. Once again it can be observed that although the analysis is done for $n\gg 1$, for state sizes as low as $n=10$ we can predict the the irregularly shaped eigendistribution very closely.

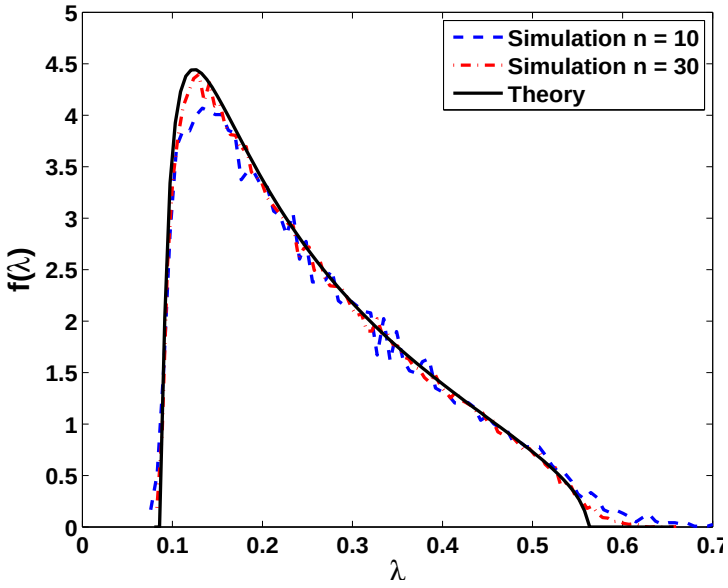


Figure 5.4. Steady-state eigendistribution of recursion (5.39) for $n = 10, 30$, $r = 0.5$ and $q = 0.1$..

5.4 RLS Filter with Intermittent Observations

As mentioned earlier, a great deal of research has been devoted recently to the study of estimation and control over lossy networks [SSF+04]. In this section we show how the problem of RLS filtering with intermittent observations can be handled in our framework.

Consider the state-space model of (5.2). Now assume that the measurements y_i are lost before reaching the estimator with some known probability ϵ. In other words with probability ϵ no y_i is available and with probability $1 - \epsilon$ a noisy measurement is received. It can be shown that (see, e.g., [SSF+04, SETM05, GDHM06]) the error covariance will have both time and measurement updates whenever a measurement is received, in which case the update equation will be the same as the one introduced earlier,

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^* h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^*} + \frac{\kappa}{n}\mathrm{I}, \tag{5.52}$$

(we assume $q = \frac{\kappa}{n}$ following the discussions in section 5.2). Otherwise, i.e., when no measurement at hand, the error covariance only undergoes a time update of the form,

$$\mathrm{P}_{i+1} = \mathrm{P}_i + \frac{\kappa}{n}\cdot \tag{5.53}$$

Assuming that the data loss process is a Bernoulli process independent through time, one can combine the two recursions above to get a single Riccati recursion. In other words, at each time step, with probability ϵ, equation (5.53) holds, and with probability $1-\epsilon$, the equation (5.52) applies. Therefore a, combined random Riccati recursion can be found for the RLS filtering with intermittent observations as

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^* h_i \mathrm{P}_i}{r_i + h_i \mathrm{P}_i h_i^*} + \frac{\beta}{n}\mathrm{I}, \tag{5.54}$$

where the $\{r_i\}$ are a collection of independent random variables such that,

$$r_i = \begin{cases} r & \text{with probability } 1-\epsilon, \\ \infty & \text{with probability } \epsilon. \end{cases} \tag{5.55}$$

Indeed whenever $r_i = \infty$ recursion (5.58) simplifies to (5.53). The following theorem presents the main result for this case.

Theorem 5.4.1 (RLS Filter with Intermittent Observations) *Consider an RLS adaptive algorithm that uses intermittent noisy observations of a state vector which undergoes a pure random walk. Assume that the measurements are lost independently of each other with probability $\epsilon < 1$. The error covariance matrix, P_i, of this RLS filter undergoes the recursion (5.58) and as $n \to \infty$ while κ is kept constant,*

the Stieltjes transform of P_i satisfies

$$S_{i+1}(z+\frac{\kappa}{n}) = S_i(z) + \frac{1-\epsilon}{n} \times \frac{1+2zS_i(z)+z^2S_i'(z)}{r-z-z^2S_i(z)}. \tag{5.56}$$

The steady-state Stieltjes transform satisfies

$$\kappa S(z) + c = -(1-\epsilon)\log\left(r - z + z^2 S(z)\right), \tag{5.57}$$

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

Proof: Rewrite (5.58) as

$$\mathrm{P}_{i+1} - q\mathrm{I} = \mathrm{P}_i - \frac{\mathrm{P}_i h_i^* h_i \mathrm{P}_i}{r_i + h_i \mathrm{P}_i h_i^*}. \tag{5.58}$$

Calculating the Stieltjes transform of both sides of recursion (5.58) yields

$$S_{i+1}(z+q) = -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r_i + h_i \mathrm{P}_i h_i^T}\right). \tag{5.59}$$

Since r_i is independent of all other random variables in the right-hand side, we can separate the expectation on r_i. Since it represents a Bernoulli process,

$$\begin{aligned} S_{i+1}(z+q) \;=\; & -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r_i + h_i \mathrm{P}_i h_i^T}\right)\Bigg|_{r_i=\infty} \Pr(r_i=\infty), \\ & -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r_i + h_i \mathrm{P}_i h_i^T}\right)\Bigg|_{r_i=r} \Pr(r_i=r)\cdot \end{aligned} \tag{5.60}$$

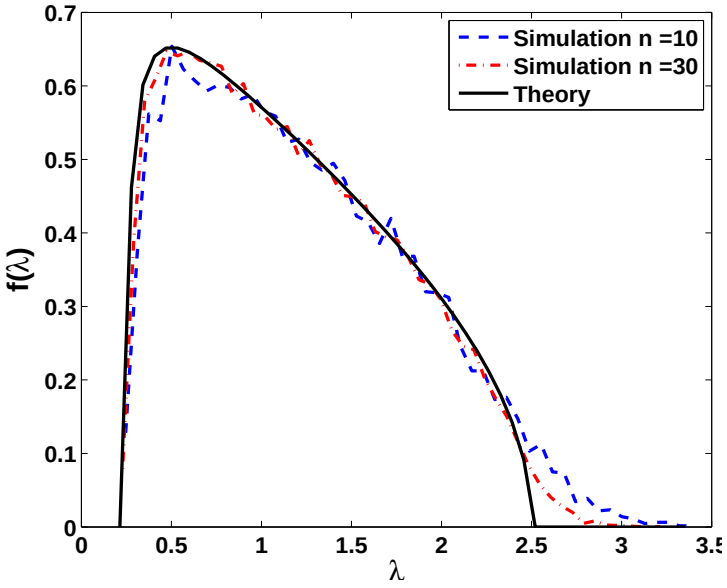


Figure 5.5. Steady-state eigendistribution of recursion (5.58) for $n = 10, 30$, $r = 0.5$, $\epsilon = 0.5$ and $\kappa = 0.1$..

Replacing (5.55) we have

$$\begin{aligned} & S_{i+1}(z+q) \\ &= -\frac{\epsilon}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{\infty + h_i \mathrm{P}_i h_i^T}\right) \\ &\qquad -\frac{1-\epsilon}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T}\right) \\ &= \epsilon S_i(z) - \frac{1-\epsilon}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I} - \frac{\mathrm{P}_i h_i^T h_i \mathrm{P}_i}{r + h_i \mathrm{P}_i h_i^T}\right). \end{aligned} \tag{5.61}$$

The second term in the right-hand side is the same as what we calculated in the proof of theorem 5.2.1 and we directly use it here,

$$S_{i+1}(z+q) = \epsilon S_i(z) - (1-\epsilon)\left(S_i(z) + \frac{1}{n} \times \frac{1 + 2zS_i(z) + z^2 S_i'(z)}{r - z - z^2 S_i(z)}\right), \tag{5.62}$$

which results in (5.56). Finally by replacing $S_{i+1}(z) = S_i(z) = S(z)$ and $q = \frac{\kappa}{n}$ in (5.56), one obtains (5.57). □

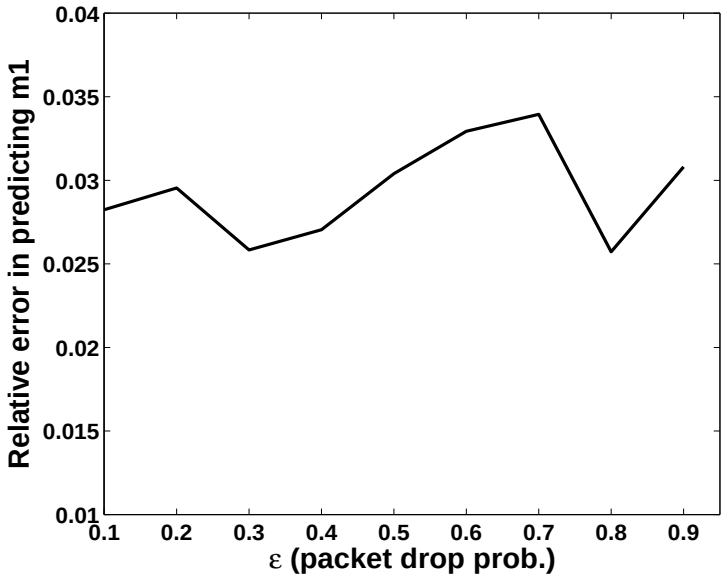


Figure 5.6. Relative error in prediction of the $m_1 = \mathbb{E}\lambda = \frac{1}{n}\mathbb{E}\mathrm{tr}\mathrm{P}$ in recursion (5.58) for different values of ϵ, $n = 20$, $r = 3.0$, and $\kappa = 3.0$..

As in the previous cases of Riccati recursions, the distribution as well as the constant c should be determined numerically. Figure 5.5 shows the theoretically predicted eigendistribution along with the empirical ones. The empirical curves are based on 1000 samples of the steady-state error covariance matrix for $n = 10$ and $n = 30$. In order to show how closely our method predicts $m_1 = \mathbb{E}\lambda = \frac{1}{n}\mathbb{E}\mathrm{tr}\mathrm{P}$ which is the mean-squares error performance, we have plotted the relative error between the theoretically found m_1 and the one obtained from Monte Carlo simulation for different values of packet drop probability. The state size is set at $n = 20$. A relative error of less than 4% is observed in the simulations.

5.5 Conclusions

The recursive-least-squares algorithm is one of the most important adaptive algorithms in use. While computationally more costly than LMS counterparts, RLS fil-

ters have several improved performance metrics in comparison which motivate their use. On the other hand, various implementation options provide liberty in selecting the appropriate model for specific applications. While the idea of RLS filters can rooted back to the early 1800s, an accurate performance analysis has been lacking in the literature. The fact that the performance of RLS filters is governed by a random Riccati recursion (a nonlinear recursion in its essence) plays the major role in making them hard to analyze. Most of the results on the RLS filters in the literature are based on loose approximations such as assuming $\mathbb{E}\mathrm{P}_i = \left(\mathbb{E}\mathrm{P}_i^{-1}\right)^{-1}$, where P_i is the error covariance matrix at time i.

In this chapter we analyzed RLS filters through our framework under two main assumptions: (a) The random regressors (and the packet drop process when applicable) are representing jointly stationary random processes, and (b) the state vector size, n, is large. Although the results are asymptotic for $n \to \infty$, as the simulations clearly show, even $n = 10$ is enough in order for the derivations to be accurate. Our results appear for the first time in the literature. We find explicit recursions satisfied by the eigendistribution of the error covariance matrix and characterize the steady-state distribution in the form of an implicit expression.

We studied several variations of the RLS filter. We started with RLS filtering of a state vector which undergoes a pure random walk with spatially and temporally white regressors. We developed the necessary tools for finding the eigendistribution numerically from the implicit expression that characterizes it. The results were shown to be extendable to shift-structured regressor vectors. Forgetting factor, an important concept in the adaptive filtering which essentially weighs out the older observations, was shown to be integrable into our analysis.

We then looked at two other variations of RLS. First we studied the RLS estimation of a process whose updates occur in known directions. We then studied the

case of RLS filtering with intermittent observations, i.e., when measurements may be lost according to some probability. For each case the Stieltjes transform's (hence eigendistribution's) time recursion was found and the steady-state eigendistribution was found as the solution of an implicit equation.

Apart from the novelty of the results, the universality that comes with the theory of large random matrices makes them powerful ones which do not depend on the details of the model.

Throughout the chapter we only considered single measurements. The results can be readily extended to multiple measurements as long as the number of measurements, m, is much less than the state vector size, n ($m \ll n$.) However this extension fails when m is comparable to n. A more involved analysis is required in this case. Due to the importance of RLS filters and their connection to Kalman filtering, we have devoted the next chapter to the so-called extended RLS filters for which $\frac{m}{n}$ does not tend to zero as n grows.

Chapter 6

The Extended RLS Algorithm

6.1 Introduction

In the previous chapter we studied variants of the recursive-least-squares (RLS) filtering under different assumptions on the regressor vectors and the state variable's updates. As a matter of fact, the RLS algorithm and Kalman filtering are deeply connected, and any problem solved in one framework can be translated into a solution in the other one. This relation was naturally noticed early in the literature [God74]. A detailed account of this equivalency can be found in [SK94]. The importance of this relation is that several ideas and implementation algorithms can be transfered between the two contexts. Moreover, since Kalman filtering is a more general concept, one may elaborate on this relation to find variations of the RLS filter with specialized properties and abilities. This chapter is dealing with the extended RLS algorithm which we will describe in the following, and this extended version, just like the simple case, is a special case of Kalman filtering, and in fact almost equal to it. Therefore the analysis in this chapter can be well considered as that of Kalman filtering under some specific assumptions[1].

[1]For a comprehensive account of how deterministic least-squares estimation can be interpreted as a stochastic estimation problem see [Say03, chapter 12]

We consider a special case of the standard state-space model,

$$\begin{cases} x_{i+1} &= x_i + u_i, \\ y_i &= \mathrm{H}_i x_i + v_i, \end{cases} \tag{6.1}$$

in which $x_i \in \mathcal{R}^n$, the state vector we want to estimate, undergoes a pure random walk, and u_i and v_i denote the zero-mean process and measurement noises, respectively. Here $y_i \in \mathcal{R}^m$ is the measured signal and H_i is the $m \times n$ regressor matrix which is indeed time-variant and often random. Therefore the measurements are no longer scalar variables, rather they are vectors. An immediate application for the extended RLS algorithm is the block RLS filtering [Say03, p. 752]. We will assume that all the processes are stationary and their statistics follow,

$$\mathbb{E} \begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_j^T & v_j^T \end{bmatrix} = \begin{bmatrix} q\mathrm{I} & 0 \\ 0 & r\mathrm{I} \end{bmatrix} \delta_{ij}. \tag{6.2}$$

Clearly this is a special case of the Kalman filtering with,

$$\mathrm{F}_i = \mathrm{I}, \quad \mathrm{G}_i = \mathrm{I}, \quad \mathrm{R}_i = r\mathrm{I}, \quad \mathrm{Q}_i = q\mathrm{I}. \tag{6.3}$$

It is well known that the RLS (equivalently, Kalman) recursive estimate of x_i can be expressed as [Say03, KSH00],

$$\hat{x}_{i+1} = \hat{x}_i + \mathrm{P}_i \mathrm{H}_i^T \left(r\mathrm{I} + \mathrm{H}_i \mathrm{P}_i \mathrm{H}_i^T \right)^{-1} (y_i - \mathrm{H}_i \hat{x}_i), \tag{6.4}$$

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

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \mathrm{P}_i \mathrm{H}_i^T \left(r\mathrm{I} + \mathrm{H}_i \mathrm{P}_i \mathrm{H}_i^T \right)^{-1} \mathrm{H}_i \mathrm{P}_i + q\mathrm{I}, \quad \mathrm{P}_0. \tag{6.5}$$

Therefore the randomness of H_i makes the error covariance propagation a random matrix-valued recursion. 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. Even for a simple RLS filter with one measurement per time instance, the mean square error performance is not known as we discussed in the previous chapter [Say03, Hay01].

Our approach which leverages ideas from large random matrix theory to establish a method for analyzing random Lyapunov and Riccati recursions proved to be successful for the single-measurement RLS and LMS filters[2]. This encourages further scrutiny about the extended RLS algorithm, when the number of measurements is allowed to grow with n. As we will see in this chapter, with some further technical complexities such an analysis can be carried out.

As we discussed in chapter 1, P_i clearly does not converge to any specific matrix as $i \to \infty$. However we may conjecture that if $\{\mathrm{H}_i\}$ is a stationary process (as is the case here), $\{\mathrm{P}_i\}$ may also converge to a stationary random process. Here we will investigate the existence and the statistics of such a stationary limit.

The main results will be presented in section 6.2 and the eigendistribution will be found as the solution of a pair of implicit equations. In section 6.3 we will explain the method of finding the steady-state eigendistribution from this implicit equation system. Simulation results presented demonstrate a very close prediction of the eigendistribution through our method. Finally, section 6.4 concludes the chapter.

[2]As a matter of fact, while the number of measurements is very small compared to the state vector size ($m \ll n$), the analysis in chapters 4 and 5 will remain valid.

6.2 Characterizing the Stieltjes Transform of the Error Covariance

Let us go back to the propagation of the error covariance matrix for an extended RLS algorithm with multiple measurements. The number of measurements, m, is assumed to be growing with n, i.e., we will assume that $n \to \infty$ and $\frac{m}{n} \to \beta$. As mentioned earlier, the error covariance matrix satisfies

$$\mathrm{P}_{i+1} = \mathrm{P}_i - \mathrm{P}_i \mathrm{H}_i^T \left(r\mathrm{I} + \mathrm{H}_i \mathrm{P}_i \mathrm{H}_i^T \right)^{-1} \mathrm{H}_i \mathrm{P}_i + q\mathrm{I}, \quad \mathrm{P}_0, \tag{6.6}$$

where H_i is an $m \times n$ random matrix that has i.i.d. entries having zero mean and $\frac{1}{\sqrt{n}}$ variance. Moreover, $\{\mathrm{H}_i\}$ is assumed to be temporally white. In other words every H_i and $\mathrm{H}_{j \neq i}$ are independent. We will follow our routine practice of first characterizing the Stieltjes transform and finding a time recursion for it. The following theorem encompasses the main results.

Theorem 6.2.1 (Extended RLS Adaptive Filtering) *Consider the random Riccati recursion in (6.6) where $m, n \to \infty$ while $\frac{m}{n} \to \beta$. As $i \to \infty$, the Stieltjes transform of the eigendistribution of P_i, $S_i(z)$, satisfies a pair of implicit equations,*

$$S_{i+1}(z+q) = \frac{t_i(z)}{1 - z t_i(z)} - \frac{1}{(1 - z t_i(z))^2} S_i \left(\frac{z}{1 - z t_i(z)} \right), \tag{6.7}$$

$$t_i(z) = \frac{\beta}{r - \frac{z}{1 - z t_i(z)} + \frac{z^2}{(1 - z t_i(z))^2} S_i \left(\frac{z}{1 - z t_i(z)} \right)}. \tag{6.8}$$

The eigendistribution of P_i converges to a steady-state distribution whose Stieltjes

transform, $S(z)$, satisfies the pair of implicit equations,

$$S(z+q) = \frac{t(z)}{1-zt(z)} - \frac{1}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right), \tag{6.9}$$

$$t(z) = \frac{\beta}{r - \frac{z}{1-zt(z)} + \frac{z^2}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right)}. \tag{6.10}$$

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

Proof: Rewrite recursion (6.6) as,

$$\mathrm{P}_{i+1} - q\mathrm{I} = \mathrm{P}_i - \mathrm{P}_i\mathrm{H}_i^T \left(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T\right)^{-1} \mathrm{H}_i\mathrm{P}_i. \tag{6.11}$$

We now apply the definition of the Stieltjes transform as in (2.12) to both sides of the above recursion. At the LHS,

$$-\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_{i+1} - q\mathrm{I} - z\mathrm{I}\right) = S_{i+1}(z+q). \tag{6.12}$$

Therefore,

$$S_{i+1}(z+q) = -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - \mathrm{P}_i\mathrm{H}_i^T(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)^{-1}\mathrm{H}_i\mathrm{P}_i - z\mathrm{I}\right). \tag{6.13}$$

Bringing $(\mathrm{P}_i - z\mathrm{I})$ out of the expression at the RHS,

$$-\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left[(\mathrm{P}_i - z\mathrm{I}) \times \left(\mathrm{I} - (\mathrm{P}_i - z\mathrm{I})^{-1}\,\mathrm{P}_i\mathrm{H}_i^T(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)^{-1}\mathrm{H}_i\mathrm{P}_i\right)\right]$$

or equivalently,

$$-\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{P}_i - z\mathrm{I}\right) - \frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{I} - (\mathrm{P}_i - z\mathrm{I})^{-1}\,\mathrm{P}_i\mathrm{H}_i^T(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)^{-1}\mathrm{H}_i\mathrm{P}_i\right).$$

The first term in the RHS is $S_i(z)$. In order to simplify the second term, we note that $\det(\mathrm{I} - \mathrm{AB}) = \det(\mathrm{I} - \mathrm{BA})$. Thus,

$$S_{i+1}(z+q) - S_i(z) = -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(\mathrm{I} - \mathrm{H}_i\mathrm{P}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)^{-1}\right)\cdot \tag{6.14}$$

Multiplying by $(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)$ we arrive at

$$\begin{aligned} S_{i+1}(z+q) - S_i(z) &= -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det(r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T)^{-1} \\ &\quad -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left((r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T) - \mathrm{H}_i\mathrm{P}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T\right)\cdot \end{aligned}$$

The first term in the right-hand side is equal to zero since it has no z dependency. Therefore we are left with,

$$S_{i+1}(z+q) - S_i(z) = -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left((r\mathrm{I} + \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T) - \mathrm{H}_i\mathrm{P}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T\right)\cdot \tag{6.15}$$

In order to transform the above expression to a form which helps us take the calculations further, we note that

$$\begin{aligned} \mathrm{H}_i\mathrm{P}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T &= \mathrm{H}_i(\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T \\ &= \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T + z\mathrm{H}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{P}_i\mathrm{H}_i^T \\ &= \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T + z\mathrm{H}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}(\mathrm{P}_i - z\mathrm{I} + z\mathrm{I})\mathrm{H}_i^T \\ &= \mathrm{H}_i\mathrm{P}_i\mathrm{H}_i^T + z\mathrm{H}_i\mathrm{H}_i^T + z^2\mathrm{H}_i\left(\mathrm{P}_i - z\mathrm{I}\right)^{-1}\mathrm{H}_i^T\cdot \end{aligned} \tag{6.16}$$

Replacing (6.16) in (6.15) yields

$$S_{i+1}(z+q) - S_i(z) = -\frac{1}{n}\mathbb{E}\frac{d}{dz}\log\det\left(r\mathrm{I} - \mathrm{H}_i(z\mathrm{I} + z^2(\mathrm{P}_i - z\mathrm{I})^{-1})\mathrm{H}_i^T\right). \tag{6.17}$$

Since H_i has an isotropic distribution, we can assume without loss of generality that P_i is diagonal. Therefore if we define the auxiliary matrix $\mathrm{D}_i(z)$ as

$$\mathrm{D}_i(z) = z\mathrm{I} + z^2(\mathrm{P}_i - z\mathrm{I})^{-1}, \tag{6.18}$$

it can be assumed diagonal. Computing the derivative in (6.17) we obtain

$$S_{i+1}(z+q) - S_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{H}_i\mathrm{D}_i'(z)\mathrm{H}_i^T\left(r\mathrm{I} - \mathrm{H}_i\mathrm{D}_i(z)\mathrm{H}_i^T\right)^{-1}\right], \tag{6.19}$$

since [Rog80]

$$\frac{d}{dz}\log\det\mathrm{A}(z) = \mathrm{tr}\left[\mathrm{A}'(z)\mathrm{A}^{-1}(z)\right]. \tag{6.20}$$

Let us rewrite (6.19) as

$$S_{i+1}(z+q) - S_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{H}_i\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\underbrace{\mathrm{D}_i(z)\mathrm{H}_i^T\left(r\mathrm{I} - \mathrm{H}_i\mathrm{D}_i(z)\mathrm{H}_i^T\right)^{-1}}\right], \tag{6.21}$$

The underbraced part on RHS can be rearranged as,

$$\mathrm{D}_i(z)\mathrm{H}_i^T\left(r\mathrm{I} - \mathrm{H}_i\mathrm{D}_i(z)\mathrm{H}_i^T\right)^{-1} = \frac{1}{r}\left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\mathrm{H}_i^T\cdot \tag{6.22}$$

By replacing (6.22) in (6.22) we obtain

$$S_{i+1}(z+q) - S_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{H}_i\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\frac{\mathrm{H}_i^T}{r}\right]. \tag{6.23}$$

Now by noting that $\mathrm{trAB} = \mathrm{trBA}$, we can move the leftmost H_i inside the trace to the right side of the expression. Therefore after adding and subtracting a $\mathrm{D}_i^{-1}(z)$ term

we arrive at

$$S_{i+1}(z+q) - S_i(z) \;=\; \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\left(\frac{\mathrm{H}_i^T\mathrm{H}_i}{r} - \mathrm{D}_i^{-1}(z) + \mathrm{D}_i^{-1}(z)\right)\right], \tag{6.24}$$

which can be readily simplified to

$$S_{i+1}(z+q) - S_i(z) \;=\; \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\left(-\mathrm{I} + \left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\mathrm{D}_i^{-1}(z)\right)\right] \tag{6.25}$$

We now deal with the two terms on the RHS separately. Let us define

$$p_i(z) = -\frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\right], \tag{6.26}$$

and

$$q_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\mathrm{D}_i^{-1}(z)\right]. \tag{6.27}$$

In order to calculate $p_i(z)$, note that according to (6.18) the entries of the diagonal matrix $\mathrm{D}_i(z)$ can be written as,

$$d_j^i = z + \frac{z^2}{\lambda_j^i - z} = \frac{\lambda_j^i z}{\lambda_j^i - z}, \tag{6.28}$$

where the superscript i denotes the time and λ_j^i denotes the j^{th} eigenvalue of P_i. The entries of $\mathrm{D}_i'(z)$ can now be found as

$$e_j^i = \frac{\lambda_j^i(\lambda_j^i - z) + \lambda_j^i z}{(\lambda_j^i - z)^2} = \frac{\left(\lambda_j^i\right)^2}{(\lambda_j^i - z)^2}. \tag{6.29}$$

Therefore we can find $p_i(z)$ as

$$\begin{aligned} p_i(z) &= -\frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-1}(z)\right] = -\mathbb{E}\left[\frac{(\lambda^i)^2}{(\lambda^i - z)^2}\Big/\frac{\lambda^i z}{\lambda^i - z}\right] \\ &= -\frac{1}{z}\mathbb{E}\left[1 + \frac{z}{\lambda^i - z}\right] = -\frac{1}{z} - S_i(z)\cdot \end{aligned} \tag{6.30}$$

So it remains to calculate (6.27) which is more challenging. Since trAB = trBA, and considering the fact that $\mathrm{D}_i(z)$ and hence $\mathrm{D}_i'(z)$ are diagonal and therefore they commute, we can write $q_i(z)$ as,

$$q_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left[\mathrm{D}_i'(z)\mathrm{D}_i^{-2}(z)\left(\mathrm{D}_i^{-1}(z) - \frac{1}{r}\mathrm{H}_i^T\mathrm{H}_i\right)^{-1}\right]. \tag{6.31}$$

Let us break $\mathrm{D}_i(z)$ and H_i in the form,

$$\mathrm{D}_i(z) = \begin{bmatrix} d_1 & 0 \\ 0 & \mathrm{D}_2 \end{bmatrix} \quad , \quad \mathrm{H}_i = \begin{bmatrix} h_1 & \mathrm{H}_2 \end{bmatrix}, \tag{6.32}$$

where h_1 is an $m \times 1$ vector. We have dropped the index i for simplicity. Replacing in (6.31) we obtain

$$q_i(z) = \frac{1}{n}\mathbb{E}\mathrm{tr}\left(\begin{bmatrix} \frac{d_1'}{d_1^2} & 0 \\ 0 & \mathrm{D}_2'\mathrm{D}_2^{-2} \end{bmatrix}\begin{bmatrix} d_1^{-1} - \frac{h_1^T h_1}{r} & -\frac{h_1^T \mathrm{H}_2}{r} \\ -\frac{\mathrm{H}_2^T h_1}{r} & \mathrm{D}_2^{-1} - \frac{\mathrm{H}_2^T \mathrm{H}_2}{r} \end{bmatrix}^{-1}\right). \tag{6.33}$$

Since everything is isotropic, without loss of generality we can look at the expected value of the first diagonal entry in the above matrix (instead of $\mathbb{E}\mathrm{tr}$). The first entry can be readily found through the Schur complement as

$$q_i(z) = \frac{d_1'}{d_1^2}\frac{1}{d_1^{-1} - \frac{h_1^T h_1}{r} - \frac{h_1^T \mathrm{H}_2}{r}\left(\mathrm{D}_2^{-1} - \frac{\mathrm{H}_2^T \mathrm{H}_2}{r}\right)^{-1}\frac{\mathrm{H}_2^T h_1}{r}}, \tag{6.34}$$

or equivalently

$$q_i(z) = \frac{d_1'}{d_1^2} \frac{1}{d_1^{-1} - h_1^T \left[\frac{1}{r}\mathrm{I} - \frac{\mathrm{H}_2}{r} \left(\mathrm{D}_2^{-1} - \frac{\mathrm{H}_2^T \mathrm{H}_2}{r} \right)^{-1} \frac{\mathrm{H}_2^T}{r} \right] h_1}. \tag{6.35}$$

We can now invoke the matrix inversion lemma to rewrite (6.35) as

$$q_i(z) = \mathbb{E} \frac{d_1'}{d_1^2} \frac{1}{d_1^{-1} - \underbrace{h_1^T \left(r\mathrm{I} - \mathrm{H}_2 \mathrm{D}_2 \mathrm{H}_2^T \right)^{-1} h_1}_{\triangleq t_i(z)}}, \tag{6.36}$$

where the subscript i in $t_i(z)$ insists its dependency on H_i and P_i. Moreover, since D_2 is in fact a function of z, so it $t_i(z)$. Assuming that $t_i(z)$ is known, we have

$$q_i(z) = \mathbb{E} \frac{d_1'}{d_1^2} \frac{1}{d_1^{-1} - t_i(z)}. \tag{6.37}$$

Replacing (6.28) and (6.29) in (6.37) we arrive at

$$q_i(z) = \mathbb{E} \frac{(\lambda^i - z)^2 (\lambda^i)^2}{(\lambda^i)^2 (\lambda^i - z)^2 z^2} \frac{1}{d_1^{-1} - t_i(z)} = \frac{1}{z^2} \mathbb{E} \frac{1}{d_1^{-1} - t_i(z)}, \tag{6.38}$$

which can be further simplified to

$$\begin{aligned} q_i(z) &= \frac{1}{z^2} \mathbb{E} \frac{1}{\frac{\lambda^i - z}{\lambda^i z} - t_i(z)} = \frac{1}{z} \mathbb{E} \frac{\lambda^i}{(1 - z t_i(z)) \lambda^i - z} \\ &= \frac{1}{z(1 - z t_i(z))} \mathbb{E} \frac{\lambda^i}{\lambda^i - \frac{z}{1 - z t_i(z)}} = \frac{1}{z(1 - z t_i(z))} \mathbb{E} \left[1 + \frac{\frac{z}{1 - z t_i(z)}}{\lambda^i - \frac{z}{1 - z t_i(z)}} \right] \end{aligned} \tag{6.39}$$

noting that λ^i is one randomly selected eigenvalue of P_i, the above expression can be written in terms of $S_i(z)$,

$$q_i(z) = \frac{1}{z(1 - z t_i(z))} + \frac{1}{(1 - z t_i(z))^2} S_i \left(\frac{z}{1 - z t_i(z)} \right). \tag{6.40}$$

Therefore, combining (6.40), (6.30), and (6.25) yields

$$S_{i+1}(z) = -\frac{1}{z} + \frac{1}{z(1 - zt_i(z))} + \frac{1}{(1 - zt_i(z))^2} S_i\left(\frac{z}{1 - zt_i(z)}\right), \tag{6.41}$$

which simplifies to (6.7).

It remains to calculate $t_i(z)$. We first note that using the self-averaging lemma, lemma 2.4.1, $t_i(z)$ can be written as

$$t_i(z) = \mathbb{E}\frac{1}{n}\mathrm{tr}\left(r\mathrm{I} - \mathrm{H}_2\mathrm{D}_2\mathrm{H}_2^T\right)^{-1}. \tag{6.42}$$

We break H_2 as

$$\mathrm{H}_2 = \begin{bmatrix} h_{21} \\ \mathrm{H}_{22} \end{bmatrix}, \tag{6.43}$$

where h_{21} is an $(n-1)$-dimensional row vector. Replacing it in (6.42),

$$t_i(z) = \mathbb{E}\frac{1}{n}\mathrm{tr}\begin{bmatrix} r - h_{21}\mathrm{D}_2 h_{21}^T & -h_{21}\mathrm{D}_2\mathrm{H}_{22}^T \\ -\mathrm{H}_{22}\mathrm{D}_2 h_{21}^T & r\mathrm{I} - \mathrm{H}_{22}\mathrm{D}_2\mathrm{H}_{22}^T \end{bmatrix}^{-1}. \tag{6.44}$$

Once again, as all the matrices are isotropically distributed, we can as well look at one diagonal entry instead of $\mathrm{tr}(\cdot)$. The $(1,1)$ entry of the inverse matrix in (6.44) can be found through the Schur complement,

$$t_i(z) = \frac{m}{n}\mathbb{E}\frac{1}{r - h_{21}\mathrm{D}_2 h_{21}^T - h_{21}\mathrm{D}_2\mathrm{H}_{22}^T\left(r\mathrm{I} - \mathrm{H}_{22}\mathrm{D}_2\mathrm{H}_{22}^T\right)^{-1}\mathrm{H}_{22}\mathrm{D}_2 h_{21}^T}, \tag{6.45}$$

through matrix inversion lemma we can simplify the denominator to

$$t_i(z) = \frac{m}{n}\mathbb{E}\frac{1}{r - h_{21}\left(\mathrm{D}_2^{-1} - \frac{1}{r}\mathrm{H}_{22}^T\mathrm{H}_{22}\right)^{-1} h_{21}^T}. \tag{6.46}$$

Applying the self-averaging lemma one more time to $h_{21}\left(\mathrm{D}_2^{-1}-\frac{1}{r}\mathrm{H}_{22}^T\mathrm{H}_{22}\right)^{-1}h_{21}^T$, and noting that the entries of h_{21} have variance $\frac{1}{\sqrt{n}}$, we obtain

$$t_i(z)=\frac{m}{n}\mathbb{E}\frac{1}{r-\frac{1}{n}\mathrm{tr}\left(\mathrm{D}_2^{-1}-\frac{1}{r}\mathrm{H}_{22}^T\mathrm{H}_{22}\right)^{-1}}. \tag{6.47}$$

One can immediately notice the similarity between the trace expression in (6.48) and the second term in (6.31). The only difference is in the size of D_{22}, which is a $(n-1)\times(n-1)$ matrix now. However we are considering the regime of large n and the difference between the two is at most $O\left(\frac{1}{n}\right)$. Therefore

$$t_i(z)=\frac{m}{n}\mathbb{E}\frac{1}{r-\mathbb{E}\frac{1}{d^{-1}-t_i(z)}}, \tag{6.48}$$

which can be simplified further by noting that

$$\mathbb{E}\frac{1}{d^{-1}-t(z)}=\mathbb{E}\frac{1}{\frac{\lambda-z}{\lambda z}-t(z)}=\frac{z}{1-zt}\mathbb{E}\frac{\lambda}{\lambda-\frac{z}{1-zt}}. \tag{6.49}$$

Therefore,

$$\mathbb{E}\frac{1}{d^{-1}-t(z)}=\frac{z}{1-zt}+\frac{z^2}{(1-zt)^2}S_i\left(\frac{z}{1-zt}\right). \tag{6.50}$$

Substituting in (6.48), and noting that $\frac{m}{n}=\beta$, we arrive at (6.8).

Whether or not a steady-state solution exists, is a different question. Showing the existence of the steady-state is another involved proof in itself. As a matter of fact, the authors in [KSM09] have recently used arguments from random dynamical systems [AC98, Chu02]. They show that a random recursion like the Riccati recursion we are considering will either diverge or will converge to a steady-state distribution. As they study the Kalman filtering with intermittent observations, they also show that there exists a threshold for packet drop probability above which the convergence is guaranteed. Here we have no missing measurements, therefore the convergence

can be deduced from their derivations. The only difference is that we are looking at random H_i, which can be integrated in their analysis in a straightforward manner. Therefore a steady-state eigendistribution do exists and its Stieltjes transform should (obviously) satisfy (6.9) and (6.10). □

We have to numerically solve the equations (6.9) and (6.10) in order to find the steady-state eigendistribution. In the next section we will explain how this numerical solution can be carried out efficiently.

6.3 Finding the Eigendistribution

Let us recall the implicit equation system satisfied by the Stieltjes transform at the steady state,

$$S(z+q) = \frac{t(z)}{1-zt(z)} - \frac{1}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right), \tag{6.51}$$

$$t(z) = \frac{\beta}{r - \frac{z}{1-zt(z)} + \frac{z^2}{(1-zt(z))^2} S\left(\frac{z}{1-zt(z)}\right)}. \tag{6.52}$$

Clearly it is unfeasible to find an analytical expression for the eigendistribution. However, we can find this distribution numerically, in an efficient way. The most coarse method would be iterating on the Stieltjes transform. One may start with an initial value for $S(z)$ and $t(z)$, namely $S_0(z)$ and $t_0(z)$, and at each iteration, $k = 1, 2, \cdots$, calculate

$$S_k(z+q) = \frac{t_{k-1}(z)}{1-zt_{k-1}(z)} - \frac{1}{(1-zt_{k-1}(z))^2} S_{k-1}\left(\frac{z}{1-zt_{k-1}(z)}\right), \tag{6.53}$$

$$t_k(z) = \frac{\beta}{r - \frac{z}{1-zt_{k-1}(z)} + \frac{z^2}{(1-zt_{k-1}(z))^2} S_{k-1}\left(\frac{z}{1-zt_{k-1}(z)}\right)}. \tag{6.54}$$

Please note the difference between the iteration on k in (6.53) and (6.54) and the time recursion of (6.7) and (6.8). Here k refers to the iteration step for solving the steady-state equations. Particularly, whi?e in (6.8) $t_i(z)$ depends on $S_i(z)$, here, $t_k(z)$ is calculated based on $S_{k-1}(z)$.

As mentioned above, it would be very inefficient to carry out this numerical solution because at each time step, one needs to find the values of $S_k(z)$ and $t_k(z)$ on a very large grid in the complex plane, and at the next time step, needs to employ some interpolation method to determine values of the two functions.

There exists a roundabout for this problem. The key is that we are looking for the eigendistribution which depends on $S(z = \lambda + j0^+)$. Thus the only important points of $S(z)$ for us are those very close to the real line. Let us fix the imaginary part of z's that we consider. We will have a line very close to the real line. We denote this line by ℓ. Looking at the equation system, we note that one only needs to have the values of $t_k(z)$ on ℓ. The $S(z+q)$ will also take arguments on ℓ since q is real. The only term that needs to be calculated away from ℓ is $S_{k-1}\left(\frac{z}{1-zt_{k-1}(z)}\right)$. However, recall the definition of the Stieltjes transform,

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

If at each time step $k-1$ we calculate $f_{k-1}(\lambda)$, the distribution related to the $S_{k-1}(z)$, at the next time step, k, any value of $S_{k-1}(z)$ can be found from $f_{k-1}(\lambda)$. Thus there is no need to propagate $S_{k-1}(z)$ on a grid on the whole complex plane.

More precisely, this is the method. Fix a small imaginary part, say ϵ_i. Assume

$$t_0(\lambda + j\epsilon_i) = c, \tag{6.55}$$

$$f_0(\lambda) = \mathcal{U}(0, M), \tag{6.56}$$

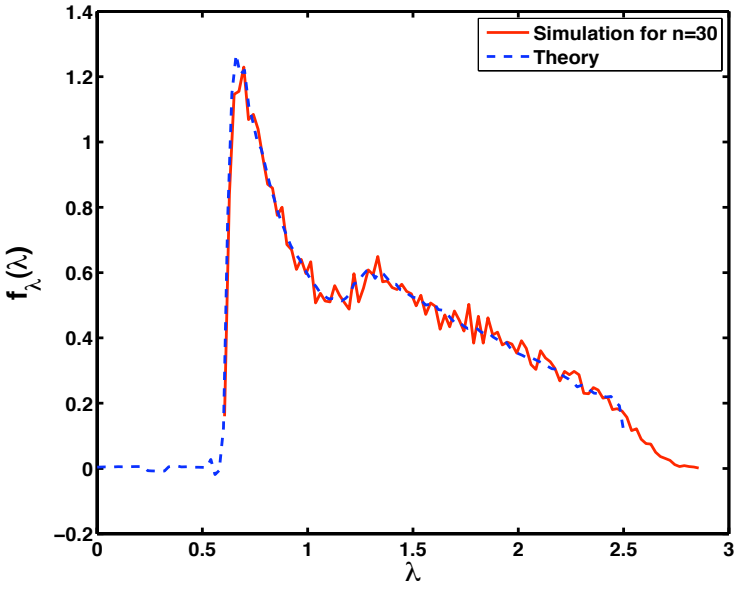


Figure 6.1. Steady-state eigendistribution of the extended RLS algorithm, recursion (6.6) for $n = 30$, $m = 15$, $r = 0.5$, and $q = 0.5$..

where $\mathcal{U}(0, M)$ denotes a uniform distribution on $(0, M)$ for some M which can be a program parameter and c is a constant. Now, at each time step, carry out (6.53) and (6.54) along with

$$S_{k-1}\left(\frac{z}{1 - zt_{k-1}(z)}\right) = \int_0^\infty \frac{f_{k-1}(\lambda)}{\lambda - \left(\frac{z}{1-zt_{k-1}(z)}\right)} d\lambda \cdot \tag{6.57}$$

At the end of the iteration step, $f_k(\lambda)$ can found as

$$f_k(\lambda) = \frac{1}{\pi} \operatorname{Im} S_k(z)\Big|_{z=\lambda+j\epsilon_i} \cdot \tag{6.58}$$

We have employed this method to find the eigendistribution at the steady-state for $r = 0.5$ and $q = 0.5$. The resulting eigendistribution is plotted in figure 6.1. The figure also shows the eigendistribution obtained empirically for $m = 15$ and $n = 30$ (for 500 samples). It can be seen that our framework closely captures the behavior

of the eigendistribution. The numerical iteration was run for $K = 20$ times.

6.4 Conclusions

This chapter was a complement of chapter 5 in which we had studied the error covariance propagation for RLS adaptive algorithms. The results in chapter 5 are valid for the regime of large n while the number of measurements, m, does not grow with n. Here we considered the case of $n \to \infty$ while $\frac{m}{n} \to \beta > 0$. This scenario fall into the category of extended RLS algorithm which is closely related to the Kalman filtering.

We assumed the the measurement matrix H_i has i.i.d. entries and every H_i and $\mathrm{H}_{j\neq i}$ are independent. Under this assumption we found the steady-state eigendistribution as the solution of a system of implicit equations. We also found the time recursion of the Stieltjes transform. This latter result can be used to further expand the analysis to the convergence properties of the filter. The proof is much more cumbersome in this case compared to the single measurement RLS.

A numerical method for efficiently solving the aforementioned system of implicit equation was explained too. We showed that only values of the Stieltjes transform on a single line need to be propagated, rather than on the whole complex plain, in order to iteratively solve for the steady-state distribution. Simulation results provided demonstrate the accuracy of our method in predicting the steady-state eigendistribution. Therefore our approach can be readily used to find various performance metrics of the extended RLS filters. The close relation between the current case and the Kalman filter motives us to apply this method to the problem of Kalman filtering with intermittent observations. A very popular framework in studying estimation and control over lossy networks. The next chapter is devoted to this analysis.

Chapter 7

Kalman Filtering with Intermittent Observations

7.1 Introduction

The development of LQG control and Kalman filtering in the framework of linear dynamical systems has played a crucial role in systems and control theory. Not only they have resulted in numerous practical schemes in areas ranging from aerospace to chemical industry, but they were also the foundation of subsequent advances in the field. The study of matrix-valued Riccati equations has been at the very heart of this development. The literature on Riccati equations is vast enough to show its importance.

The Riccati equation emerges as a result of a wide range of viewpoints, spanning dynamic programming to canonical spectral factorization. Furthermore, conditions such as stabilizability and detectability (or controllability and observability), now ubiquitously used in systems and control, were first introduced as conditions to guarantee the convergence of Riccati equations and recursions [KSH00].

In the past decade, groundbreaking advances in microsensor technology have made several previously out of reach applications feasible. The proposed and/or already deployed applications include but are not limited to distributed catastrophe surveil-

lance, smart transportation systems, and more efficient electric power grids. All these applications essentially rely on large-scale networks that incorporate communication, estimation, and control. Therefore a significant body of research in recent years has been devoted to the study of distributed estimation and control over networks of many sensors and actuators. An important aspect of these systems which has prohibited direct extension of the classical theories of control and estimation is the natural unreliability of the underlying communication links. The stingy power constraints of microsensors only worsens the situation by making reliable communication further unrealistic.

A very well-received model for studying the effect of unreliable links in distributed sensing and control problems ([SSF+04, SSF+07, ESTM08, IYB06, GDH+09]) assumes that the estimation and control data are in the form of packets which travel through an erasure network and each packet may be independently lost according to some probability. This model makes many problems mathematically tractable and at the same time has great resemblance with practice. In the very heart of this approach lies Kalman filtering with intermittent observations.

To be more explicit, the model (that probably appeared for the first time in [SSF+04]) considers the Kalman filtering with intermittent observations such that the measurements of the system are sent across a simple erasure channel and are dropped with a certain probability p_d. The main result is that, for unstable systems, there is a critical value of the drop probability, say p_d^*, such that for $p_d > p_d^*$ the expected estimation error covariance diverges to infinity, whereas for $p_d < p_d^*$ it remains bounded. For quite some time the value of p_d^* was unknwown in general. Recently the authors in [MS08] have found tight bounds on p_d^* for quite general settings. However, in the bounded case, the asymptotic expected error covariance matrix is not known, and what is known are certain (often loose) upper and lower bounds.

Recall the state-space model used throughout this thesis. Consider

$$\begin{cases} x_{i+1} &= \mathrm{F}x_i + u_i, \\ y_i &= \mathrm{H}x_i + v_i, \end{cases} \tag{7.1}$$

$$\mathbb{E}\begin{bmatrix} u_i \\ v_i \end{bmatrix}\begin{bmatrix} u_j^T & v_j^T \end{bmatrix} = \begin{bmatrix} \mathrm{Q} & 0 \\ 0 & \mathrm{R}_i \end{bmatrix}\delta_{ij}, \tag{7.2}$$

where x_i is the n-dimensional state vector, y_i is the m-dimensional measurement vector, and u_i and v_i are zero-mean process and measurement noises. All the measurements are assumed to have the same noise variance and to be independent of each other. Each component of the measurement vector may also be lost independently (over both measurements and time) with some fixed packet drop probability, p_d. Therefore the measurement noise covariance matrix should be represented by a matrix-valued i.i.d. random matrix process. In fact, the only thing time-varying about (7.1) and (7.2) is the noise covariance, R_i.

The estimation error covariance of the Kalman filter for the above system can be shown to satisfy a random Riccati recursion,

$$\mathrm{P}_{i+1} = \mathrm{F}\mathrm{P}_i\mathrm{F}^T + \mathrm{Q} - \mathrm{F}\mathrm{P}_i\mathrm{H}^T(\mathrm{R}_i + \mathrm{H}\mathrm{P}_i\mathrm{H}^T)^{-1}\mathrm{H}\mathrm{P}_i\mathrm{F}^T. \tag{7.3}$$

Clearly, in contrast to the classic case of time-invariant Kalman filtering, the above Riccati recursion does not converge to any specific value. The reason being that the covariance matrix R_i is indeed random and time varying. However, there are several important questions that may be asked about such a recursion, especially about the distribution of the eigenvalues of P_i.

Since R_i is a matrix-valued stationary random process, it may be expected that P_i also converges to a stationary process. Furthermore, one can argue that the state

vector size is usually large due to the fact that the dynamical systems under consideration are often complex. This justifies the use of the framework we have developed so far. In this chapter, we will try to find the eigenvalue distribution of the prediction error covariance under these two assumptions.

In the previous chapters, these were more or less all the assumptions we were imposing on the problem. However, for Kalman filtering with intermittent observations, due to some technicalities we will make two further assumptions. The first is that the measurement matrix H is also time varying and random. In this sense, we will be forced to depart from the model (7.1) and (7.2). The main reason is that we do not quite yet know how to extend our techniques to deal with a fixed H. Nonetheless, as will be explained later in the paper, this assumption, and the relation to the constant H case, can be justified through the ergodicity of random matrix ensembles. The second extra assumption is that the matrix F is stable. This implies that the matrix-valued process P_i is bounded and thus we do not need to worry about stability in our analysis.

In the literature, as mentioned earlier, the problem of Kalman filtering with intermittent observations was first considered in [SSF$^+$04]. The tightness of the lower bound is further investigated in [PB09]. In [MS08] the authors characterize the critical packet drop probability for boundedness of the error covariance for a wide range of systems. Other authors have considered various sensor data transmission scenarios [XH05, RK07, HD07] for this problem.

The rest of the chapter is organized as follows. In section 7.2 we study the problem of Kalman filtering with intermittent observations while the process noise covariance matrix has a specific form, namely $\mathrm{Q} = q\mathrm{I}$. We find the steady-state eigendistribution of P_i as the solution of a pair of implicit equations. Section 7.3 deals with the problem when Q has the form of a Wishart matrix. Simulation results and methodology will

be provided in section 7.4 which show the accuracy of our method in predicting the eigendistribution. Finally, section 7.5 concludes the chapter.

7.2 The case of Q = qI

In this section we consider Kalman filtering with intermittent measurements under the assumption of a stable system matrix and a time-varying observation matrix. Although this is not the convention in the literature, there are several reasons for doing so. The first is that we do not yet know how to deal with the time-invariant case in our method. The second reason is that in many recent applications of distributed estimation, the sensors' environment and/or their position is time-varying. Finally, as mentioned earlier, in the theory of large random matrices every single realization of an ensemble has an eigendistribution that converges almost surely to the asymptotic eigendistribution. Therefore one may expect that if the matrix H is large enough, then the eigendistribution of P_i may be the same irrespective of H and whether it is time varying. The second assumption is that the matrix F is stable. As mentioned in the introduction, this guarantees the boundedness of P_i and will absolve us of having to consider stability issues. Of course, relaxing these two assumptions is critical and can be considered as the future development of our framework. In this sense, the results of this chapter represent significant progress toward analyzing these more general and realistic cases.

In this section, we will consider the case of the state process noise covariance matrix being a multiple of identity. For the resulting random Riccati recursion, we will find the steady-state eigendistribution of P_i. The transient behavior is well worth future scrutiny and will not be considered here.

When $\mathrm{Q} = q\mathrm{I}$ in the linear time-varying state-space model under study, the error

covariance matrix undergoes a random Riccati recursion of the form,

$$\mathrm{P}_{i+1} = \alpha \mathrm{F}\left(\mathrm{P}_i^{-1} + \mathrm{H}_i^T \mathrm{R}_i^{-1} \mathrm{H}_i\right)^{-1} \mathrm{F}^T + q\mathrm{I}, \quad \mathrm{P}_0, \tag{7.4}$$

in which α is a scalar between 0 and 1, and F is assumed to be an $n \times n$ matrix with i.i.d. entries having zero mean and variance $\frac{1}{\sqrt{n}}$. The observation matrix, H_i, is an $m \times n$ matrix with i.i.d. zero-mean $\frac{1}{\sqrt{m}}$-variance entries and q is a constant denoting the variance of the state process noise. As mentioned earlier, we will assume that the observations may independently be missing with some probability p_d[1]. Thus R_i can be modeled as a diagonal matrix with independent entries such that,

$$(\mathrm{R}_i^{-1})_{jj} = \begin{cases} 0 & \text{with probability } p_d, \\ \frac{1}{r} & \text{with probability } 1 - p_d. \end{cases}$$

We are interested in finding the steady-state, i.e., when $i \to \infty$, eigendistribution of the error covariance matrix P_i, or equivalently its Stieltjes transform. It can be shown that the Stieltjes transform at the steady-state satisfies a set of implicit equation on the complex plane:

Theorem 7.2.1 *Let F be an $n \times n$ matrix with i.i.d. entries having zero mean and variance $\frac{1}{\sqrt{n}}$ and α to be such that $\sqrt{\alpha}F$ is stable. As $i \to \infty$ and $n \to \infty$, the eigendistribution of P_i in (7.4) converges to a stationary distribution whose Stieltjes transform, $S_P(z)$, satisfies,*

$$S_P(z) = -\frac{1}{z-q} + \frac{\alpha r' S_P(z)}{z-q} \Omega\big(-\alpha r' S_P(z)\big), \tag{7.5}$$

$$\Omega(z) = -\frac{1}{r'u(z)} - \frac{1}{r'u^2(z)} S_P\Big(\frac{1}{u(z)}\Big), \tag{7.6}$$

[1]Note that we are considering a more general setup than the work of Sinopoli et al. [SSF+04] where either all or none of the measurements are received.

in which $u(z)$ is an expression in terms of z and $\Omega(z)$,

$$u(z) = \frac{z}{r'} - \frac{\beta'/r'}{\beta' + \Omega(z)}, \tag{7.7}$$

which is used here for the sake of brevity in the expressions, $\beta' = \frac{(1-p_d)m}{n}$, and $r' = \frac{r}{1-p_d}$.

Proof: First of all, note that with high probability the term $\mathrm{H}_i^T \mathrm{R}_i^{-1} \mathrm{H}_i$ can be written as $\frac{(1-p_d)}{r} \bar{\mathrm{H}}_i^T \bar{\mathrm{H}}_i$, where $\bar{\mathrm{H}}_i$ is an $m(1-p_d) \times n$ with i.i.d. entries having zero mean and $\frac{1}{\sqrt{m(1-p_d)}}$ variance. Now we can rewrite (7.4) as,

$$\mathrm{P}_{i+1} = \alpha r' \mathrm{F} \left(r' \mathrm{P}_i^{-1} + \bar{\mathrm{H}}_i^T \bar{\mathrm{H}}_i \right)^{-1} \mathrm{F}^T + q\mathrm{I}, \tag{7.8}$$

where,

$$r' = \frac{r}{1-p_d}. \tag{7.9}$$

Using the definition of the Stieltjes transform as in (2.8),

$$S_{\mathrm{P},i+1}(z) = \frac{1}{\alpha r'} S_{\mathrm{B},i}\left(\frac{z-q}{\alpha r'}\right), \tag{7.10}$$

in which we have used B_i to denote,

$$\mathrm{B}_i = \mathrm{F} \left(r' \mathrm{P}_i^{-1} + \bar{\mathrm{H}}_i^T \bar{\mathrm{H}}_i \right)^{-1} \mathrm{F}^T. \tag{7.11}$$

Recall the implicit relation between the Stieltjes transform and the S-transform,

$$\Sigma(\gamma) = -\frac{1}{\gamma} S\left(\frac{\gamma+1}{\gamma \Sigma(\gamma)}\right). \tag{7.12}$$

And the fact that for two free random variables X and Y,

$$\Sigma_{\mathrm{AB}}(\gamma) = \Sigma_{\mathrm{A}}(\gamma)\Sigma_{\mathrm{B}}(\gamma). \tag{7.13}$$

If one can establish the freeness of two noncommutative random variables, the Stieltjes transform of their product can be expressed in terms of the individual Stieltjes transforms in an implicit form. Now going back to (7.11), according to the definition (2.11), it makes no difference to look at the Stieltjes transform of $\mathrm{F}^T\mathrm{F}\left(r'\mathrm{P}_i^{-1} + \bar{\mathrm{H}}_i^T\bar{\mathrm{H}}_i\right)^{-1}$. Moreover, we can diagonalize P_i without loss of generality since the distributions of H_i and F are isotropic. Now the results of [Rya98] can be used to establish the freeness of $\mathrm{F}^T\mathrm{F}$ and $\left(r'\mathrm{P}_i^{-1} + \bar{\mathrm{H}}_i^T\bar{\mathrm{H}}_i\right)^{-1}$ in the steady-state. To do so it is necessary to use random dynamical systems theory [Chu02] to establish the existence of a unique steady-state eigendistribution for P_i [KSM09]. Then it is straightforward to show that for this choice of F,

$$\Sigma_{\mathrm{F}^T\mathrm{F}}(\gamma) = \frac{1}{1+\gamma}. \tag{7.14}$$

Therefore, using (7.13), (7.14), and the definition of the S-transform, we can show that,

$$S_{\mathrm{B}}(z) = -\frac{1}{z} - \frac{S_{\mathrm{B}}(z)}{z}\Omega(-S_{\mathrm{B}}(z)), \tag{7.15}$$

(please note the absence of the index i due to the steady-state analysis from now on,) where $\Omega_i(z)$ is just the Stieltjes transform of

$$\mathrm{A}_i = r'\mathrm{P}_i^{-1} + \bar{\mathrm{H}}_i^T\bar{\mathrm{H}}_i. \tag{7.16}$$

It only remains to find the relation between $\Omega_i(z)$ and the Stieltjes transform of P_i from the above equation. For simplicity, let us drop all the indexes and focus on

$\mathrm{A} = r'\mathrm{P}^{-1} + \bar{\mathrm{H}}^T\bar{\mathrm{H}}$. By applying the definition of the Stieltjes transform (2.11) to both sides we obtain,

$$\Omega(z) = \frac{1}{n}\mathbb{E}\ \mathrm{tr}\left(r'\mathrm{P}^{-1} + \bar{\mathrm{H}}^T\bar{\mathrm{H}} - z\mathrm{I}\right)^{-1}. \tag{7.17}$$

Since $\bar{\mathrm{H}}$ has an isotropic distribution, without loss of generality, P^{-1} can be assumed to be a diagonal matrix, Λ. Now we break Λ and $\bar{\mathrm{H}}$ in the form,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \qquad \bar{\mathrm{H}} = \begin{bmatrix} h_1 & \mathrm{H}_2 \end{bmatrix}, \tag{7.18}$$

and rewrite the RHS of (7.17) as,

$$\frac{1}{n}\mathbb{E}\ \mathrm{tr}\begin{bmatrix} r'\lambda_1 + h_1^T h_1 - z & h_1^T\mathrm{H}_2 \\ \mathrm{H}_2^T h_1 & r'\Lambda_2 + \mathrm{H}_2^T\mathrm{H}_2 - z\mathrm{I} \end{bmatrix}^{-1}. \tag{7.19}$$

Clearly, since we are interested in the marginal eigendistributions, it does not matter which diagonal entry of the above inverse we look at. Thus, without loss of generality, we can focus on the first diagonal entry which can be written as,

$$\mathbb{E}\frac{1}{r'\lambda_1 - z + h_1^T(\mathrm{I} - \mathrm{H}_2(r'\Lambda_2 - z\mathrm{I} + \mathrm{H}_2^T\mathrm{H}_2)^{-1}\mathrm{H}_2^T)h_1}. \tag{7.20}$$

Using the matrix inversion lemma, we can rewrite the above expression as,

$$\mathbb{E}\frac{1}{r'\lambda_1 - z + h_1^T(\mathrm{I} + \mathrm{H}_2(r'\Lambda_2 - z\mathrm{I})^{-1}\mathrm{H}_2^T)^{-1}h_1}. \tag{7.21}$$

Now one can invoke the self-averaging Lemma -using the independence of h_1 and the

inversed matrix in the denominator- to find that,

$$\Omega(z) = \mathbb{E}\frac{1}{r'\lambda_1 - z + \underbrace{\frac{1}{m'}\mathbb{E}\ \mathrm{tr}(\mathrm{I} + \mathrm{H}_2(r'\Lambda_2 - z\mathrm{I})^{-1}\mathrm{H}_2^T)^{-1}}_{\triangleq w(z)}}, \tag{7.22}$$

where $m' = m(1 - p_d)$. In order to find $w(z)$, we employ the same technique used to obtain (7.22) by breaking H_2 in the form,

$$\mathrm{H}_2 = \begin{bmatrix} h_{21} \\ \mathrm{H}_{22} \end{bmatrix}, \tag{7.23}$$

which eventually yields,

$$w(z) = \mathbb{E}\frac{1}{1 + \underbrace{h_{21}(r\Lambda_2 + \mathrm{H}_{22}^T\mathrm{H}_{22} - z\mathrm{I})^{-1}h_{21}^T}_{\triangleq w'(z)}}, \tag{7.24}$$

where, by invoking the self-averaging lemma, $w'(z)$ in turn can be rewritten as,

$$w'(z) = \frac{n}{m'} \times \frac{1}{n}\mathrm{tr}(r\Lambda_2 + \mathrm{H}_{22}^T\mathrm{H}_{22} - z\mathrm{I})^{-1}. \tag{7.25}$$

The second part of the expression above is nothing but an $(n-1) \times (n-1)$ version of (7.17). Since we are considering the large n regime, $w'(z)$ can be simplified as $\frac{n}{m'}\Omega(z)$. Therefore,

$$w(z) = \frac{1}{1 + \frac{n}{m'}\Omega(z)}. \tag{7.26}$$

We can now replace (7.26) in (7.22). Since λ_1 is a randomly selected eigenvalue of P^{-1}, (7.26) can be written in terms of the Stieltjes transform of P^{-1},

$$\Omega(z) = \frac{1}{r'}S_{\mathrm{P}^{-1}}\left(\frac{z}{r'} - \frac{(1-p_d)\beta/r'}{(1-p_d)\beta + \Omega(z)}\right), \tag{7.27}$$

which together with the relation between the Stieltjes transforms of a matrix and its inverse,

$$S_{\mathrm{P}^{-1}}(z) = -\frac{1}{z} - \frac{1}{z^2} S_{\mathrm{P}}\left(\frac{1}{z}\right), \tag{7.28}$$

result in (7.6). Substituting (7.15) in (7.10) yields (7.5) and completes the proof. □

The eigendistribution at the steady state can be found numerically from the implicit system of equations in theorem 7.2.1. We will present the simulation results in section 7.4.

It is also worth mentioning that the assumptions on the problem can be further relaxed by assuming that different measurements may have different packet drop probabilities, the average of which equals p_d. This is often a more realistic model since in practice different sensors may be deployed at different locations and consequently have different channel strengths when they communicate through a wireless network.

7.3 The case of $\mathbf{Q} = \mathbf{G}\mathbf{G}^{\mathbf{T}}$

In order to extend the results to the case of state noise process covariance being a full matrix, here we consider Q being a Wishart matrix, which means that,

$$\mathrm{Q} = \mathrm{G}\mathrm{G}^{T}, \tag{7.29}$$

where G is an $n \times m_G$ matrix with i.i.d. zero-mean, $\frac{1}{\sqrt{m_G}}$-variance entries. The random Riccati recursion of the state estimation error covariance will be

$$\mathrm{P}_{i+1} = \alpha \mathrm{F}\left(\mathrm{P}_i^{-1} + \mathrm{H}_i^T \mathrm{R}_i^{-1} \mathrm{H}_i\right)^{-1} \mathrm{F}^T + \mathrm{G}\mathrm{G}^T, \quad \mathrm{P}_0, \tag{7.30}$$

where all the other parameters are as defined in the previous section. All the coefficients are time invariant except for the observation matrix H_i and observation noise

covariance R_i - of course, due to the intermittent observations,- and once again, we are interested in finding the steady-state eigendistribution of the error covariance matrix. The following theorem describes its Stieltjes transform as the solution of a pair of implicit equations.

Theorem 7.3.1 *Let F and G be $n \times n$ and $n \times m_G$ matrices with i.i.d. entries having zero mean and variances $\frac{1}{\sqrt{n}}$ and $\frac{1}{\sqrt{m_G}}$, respectively, and α to be such that $\sqrt{\alpha}F$ is stable. As $i \to \infty$, the eigendistribution of P_i in (7.30) converges to a stationary distribution whose Stieltjes transform, $S_P(z)$, satisfies*

$$S_P(z) = -\frac{1}{\alpha r' v(z)} + \frac{S_P(z)}{v(z)}\Omega\big(-\alpha r' S_P(z)\big), \tag{7.31}$$

$$\Omega(z) = -\frac{1}{r' u(z)} - \frac{1}{r' u^2(z)} S_P(\frac{1}{u(z)}), \tag{7.32}$$

in which $v(z)$ and $u(z)$ are expressions in terms of z, $S_P(z)$, and $\Omega(z)$,

$$v(z) = \frac{z}{\alpha r'} - \frac{\beta_G/\alpha r'}{\beta_G + S_P(z)} \tag{7.33}$$

$$u(z) = \frac{z}{r'} - \frac{\beta'/r'}{\beta' + \Omega(z)}, \tag{7.34}$$

which are used in order to simplify the equations, $\beta' = \frac{(1-p_d)m}{n}$, $r' = \frac{r}{1-p_d}$, and $\beta_G = \frac{m_G}{n}$.

Sketch of proof: The proof essentially follows the proof of theorem 7.2.1. The only difference is that instead of having

$$\mathrm{P}_i = \alpha r' \mathrm{B}_i + q\mathrm{I}, \tag{7.35}$$

in this case we have

$$\mathrm{P}_i = \alpha r' \mathrm{B}_i + \mathrm{G}\mathrm{G}^T, \tag{7.36}$$

where B_i is as defined in (7.11). This is similar to the expression (7.16) that we dealt with in the proof of theorem 7.2.1. Once again, although G is not time-varying, it is selected from an isotropic random matrix ensemble and therefore we can diagonalize P_i and follow the same techniques that where used to obtain (7.27) to show that,

$$S_{\mathrm{P},i+1}(z) = \frac{1}{\alpha r'} S_{\mathrm{B},i}\left(\frac{z}{\alpha r'} - \frac{\beta_G/\alpha r'}{\beta_G + S_{\mathrm{P},i+1}(z)}\right). \tag{7.37}$$

Using (7.37) instead of (7.10), the rest of the proof will be similar to that of theorem 7.2.1.

As in theorem 7.2.1, this implicit pair of equations for $S_{\mathrm{P}}(z)$ can be rewritten as an expression for the steady-state eigendistribution involving an auxiliary distribution, which can be efficiently solved through numerical methods in order to determine the eigendistribution.

7.4 Finding the Eigendistribution

In order to find the steady-state eigendistribution, we have to numerically solve systems of implicit equations. This is reminiscent of the extended RLS filters. Here we will discuss a quite similar numerical method of calculating the distribution for the Kalman filtering with intermittent observations. Recalling the implicit equations of theorem 7.2.1,

$$S(z) = -\frac{1}{z-q} + \frac{\alpha r' S(z)}{z-q}\Omega\big(-\alpha r' S(z)\big), \tag{7.38}$$

$$\Omega(z) = -\frac{1}{r' u(z)} - \frac{1}{r' u^2(z)} S(\frac{1}{u(z)}), \tag{7.39}$$

we note that here, the auxiliary function $\Omega(z)$ is a Stieltjes transform itself. Therefore an eigendistribution can be uniquely assigned to it. Let us denote this distribution by

$g(\lambda)$ and the steady-state eigendistribution of P by $f(\lambda)$. Once again, calculating and propagating $S(z)$ and $\Omega(z)$ on a grid on the complex plain will be a very inefficient way of numerically solving these equations. Instead, we will propagate the distributions as they uniquely determine the Stieltjes functions.

We start with uniform distribution for both $f(\lambda)$ and $g(\lambda)$,

$$f_0(\lambda) = \mathcal{U}(0, M), \tag{7.40}$$

$$g_0(\lambda) = \mathcal{U}(0, M), \tag{7.41}$$

for some constant M. Now we will iteratively solve the system of equations. At each iteration step, $k = 1, 2, \cdots$, we have

$$\begin{aligned} S_k(z) &= -\frac{1}{z-q} + \frac{\alpha r' S_{k-1}(z)}{z-q}\Omega_{k-1}\big(-\alpha r' S_{k-1}(z)\big), && (7.42)\\ \Omega_k(z) &= -\frac{1}{r' u_{k-1}(z)} - \frac{1}{r' u_{k-1}^2(z)} S_{k-1}(\frac{1}{u(z)}), && (7.43)\\ u_{k-1}(z) &= \frac{z}{r'} - \frac{\beta'/r'}{\beta' + \Omega_{k-1}(z)}, && (7.44) \end{aligned}$$

where all the derivations are carried out using the distributions. For example, to compute $u_{k-1}(z)$,

$$u_{k-1}(z) = \frac{z}{r'} - \frac{\beta'/r'}{\beta' + \int_0^\infty \frac{g_{k-1}(\lambda)}{\lambda - z} d\lambda}. \tag{7.45}$$

As a matter of fact, the only z's considered in the recursion will be those necessary for finding $f_k(\lambda)$ and $g_k(\lambda)$. In other words, for any λ, we find $S(\lambda + j\epsilon_i)$ and $\Omega(\lambda + j\epsilon_i)$ as explained, for some small ϵ_i. Then according to the inverse Stieltjes transform,

$$f_k(\lambda) = \frac{1}{\pi} S_k(\lambda + j\epsilon_i), \tag{7.46}$$

$$g_k(\lambda) = \frac{1}{\pi} \Omega_k(\lambda + j\epsilon_i), \tag{7.47}$$

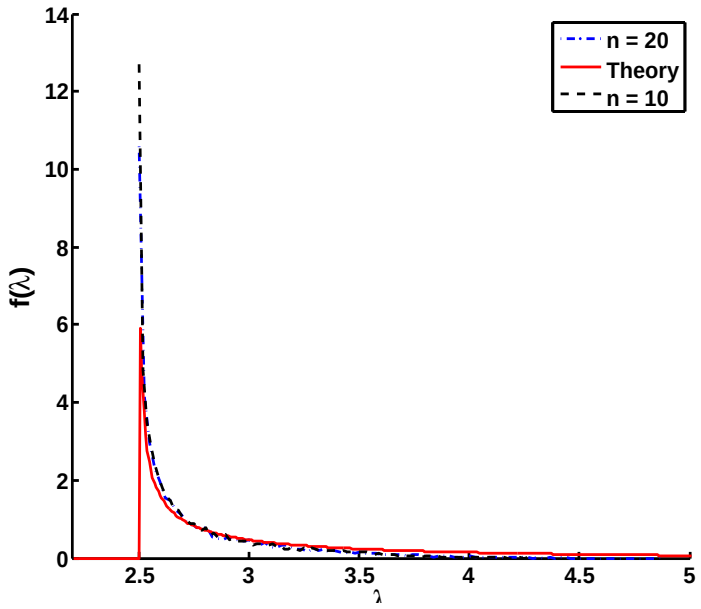


Figure 7.1. The empirical steady-state eigendistribution of the recursion (7.4) for $n = 10, 20$, $m = 30$, $r = 0.5$, $q = 2.5$, and $p_d = 0.2$ compared to the theoretically found eigendistribution..

we will run the iteration until the difference between $f_{k-1}(\cdot)$ and $f_k(\cdot)$ becomes negligible. A similar discussion can be made about the results of theorem 7.3.1.

Figure 7.1 shows the simulation results for $n = 10, 20$ and $m = 30$, for the case of $\mathrm{Q} = q\mathrm{I}$. The empirical curve is generated through Monte Carlo simulation of the recursion. It can be observed that the asymptotic theoretical prediction closely matches with the empirical curve for state vector size being as low as $n = 10$.

In figure 7.2, we have plotted the theoretical curve obtained by numerically solving the implicit equations for the case of $\mathrm{Q} = GG^T$ versus the empirical eigendistributions which are found through Monte Carlo simulation of the recursion (7.30) for various values of n. It can be seen that the theoretical curve captures the behavior of the empirical one very closely.

As mentioned in the previous subsection, our proof remains valid when assuming that the observations are dropped independently with different probabilities, while the

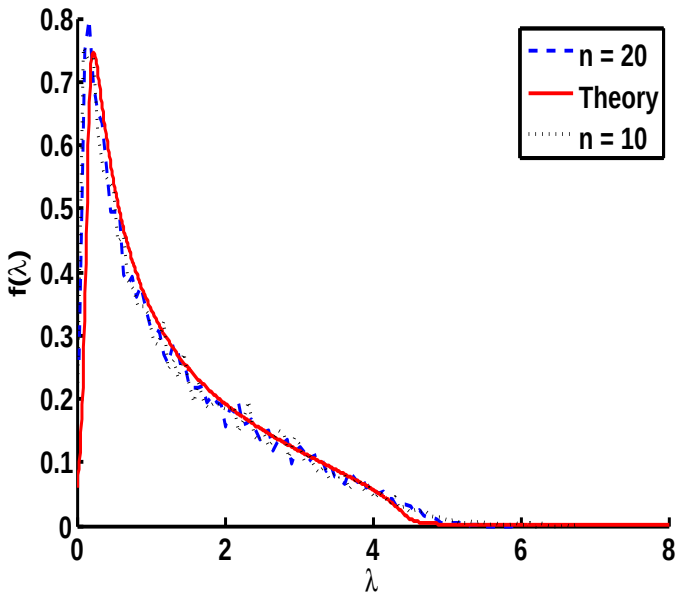


Figure 7.2. The empirical steady-state eigendistribution of the recursion (7.30) for $n = 10, 20$, $m = 30$, $r = 0.5$, $q = 2.5$, $m_G = n$, and $p_d = 0.2$ compared to the theoretically found eigendistribution.

average of these packet drop probabilities equals p_d. In figure 7.3, we have compared the theoretical curve with the Monte Carlo simulation results for this scenario. Each observation may be dropped with probability p_j, where p_j's are selected uniformly between zero and $2p_d$.

7.5 Conclusions

The problem of control and estimation over networks has received a lot of attention in recent years. The communications limitations and uncertainties make these networks of sensors and actuators very hard to analyze. A ubiquitous model for these networks assumes that the control and estimation data are in the form of packets which are passing through an erasure network. This model brings up Kalman filtering with intermittent observations whose error covariance matrix propagates through a random

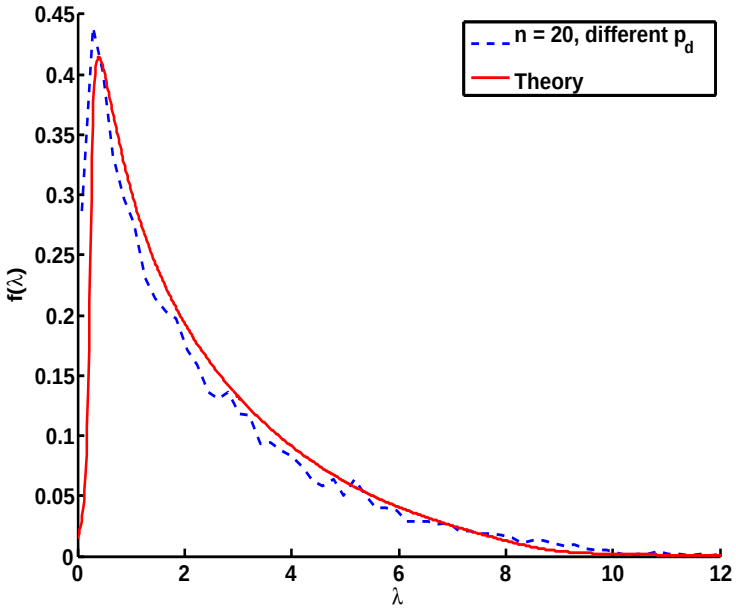


Figure 7.3. The empirical steady-state eigendistribution of the recursion (7.30) for $n = 10, 20$, $m = 30$, $r = 1.5$, $q = 2.5$, $m_G = n$, and p_d uniformly distributed on $[0, 0.4]$, compared to the theoretically found eigendistribution.

Riccati recursion.

Just as Riccati recursions have been in the core of the development of classical control and estimation theory, random Riccati recursions will play a central role in these problems. In this chapter we applied our framework to two examples of random Riccati recursions which arise in these networks. Due to technicalities, we made two assumptions. The first one was that the system matrix, F, is stable. The second assumption which is not quite the tradition in the literature was that the observation matrix H is time varying. This assumption can be justified in many applications, however it should definitely be relaxed in order to extend the generality of the results.

Through our framework, we showed how the steady-state eigendistribution of the error covariance matrix can be characterized. We carried out the analysis under two different structures for the process noise covariance. Namely, when it is a multiple of identity and when it is represented by a full-rank Wishart matrix. We obtained

implicit equation systems that allow one to compute the asymptotic eigendistribution when the state dimension is large. However, as the simulation results suggest, even $n = 10$ is large enough for the results to hold.

Our assumption on the packet drop probabilities are much less restrictive than what appears in the literature. While in the previous works the authors assume that either all or none of the measurements are received at each time step, in our work we allow the measurements to be intermittent independently. As we have mentioned in several occasions throughout the thesis, our framework delivers universal laws which are not tied to the details of the model. In this case we showed that different measurements do not even need to have the same packet drop probability and only the average of these probabilities is important.

Overall, the results of this chapter are novel. Although relaxing some assumptions is necessary, the results are of much significance as they are the first of their kind to appear in the literature.

Chapter 8

Future Work

The framework developed in this thesis proved to be successful in obtaining non-trivial results in numerous examples of random Lyapunov and Riccati recursions in the literature. The scope of the problems solved spans the steady-state analysis of all adaptive filters and many random Riccati recursions which are encountered in networked control and estimation, along with the transient behavior of random Lyapunov and Lyapunov-like recursions. The framework is well worth further scrutiny in order to be applied to other problems. As a matter of fact, several interesting open problems and future research opportunities are brought up by the current work which we will briefly address in what follows.

8.1 Transient Analysis

Despite being well developed for steady-state analysis, when it comes to the transient behavior this machinery is still in a primitive stage. We presented different methods for analyzing the transient behavior of random Lyapunov and Lyapunov-like recursions in chapter 3 and 4. However, those methods are not yet proven to be capable of handling nonlinear random Riccati counterparts. While we do obtain time recursions for the Stieltjes transform of P_i for random Riccati recursions, finding the convergence properties is not immediate from those recursions.

The moments' approach specifically fails for the Riccati recursions since as we showed, the moments are highly coupled together. A first step can be studying local convergence behavior near enough to the steady state.

It would be very interesting to be able to deduce the transient behavior directly from the Stieltjes transform's time updates. Having developed such a methodology, one can revisit other problems in the field which are unsolved or partially solved under unrealistic assumptions, ranging from the learning curves of different adaptive filters to the tracking performance analysis.

8.2 Relaxing the Assumptions

In some cases, a few restrictive assumptions were necessary to carry on with the proofs. One significant example was when we looked at random Riccati recursions which arise in networked estimation. There we had to impose the assumption of H being time variant. This is not a common assumption in the literature. However justifiable (as we discussed it in chapter 7), we still need to overcome technicalities and expand the results to fixed observation matrices. This was not an issue in the analysis of adaptive filtering since in that domain the regressor vectors (or matrices) are time varying and often random.

There are also other assumptions in the existing model that need to be relaxed in order to better match reality. For example, the intermittent observations are modeled by a binary process to make it mathematically tractable. The random matrix approach, on the other hand, normally delivers universal laws, independent of the model specifications such as the probability distribution of the elements and this gives hope about the possibility of relaxing such assumptions.

8.3 Characterizing the Support of the Eigendistribution

Another interesting direction which we did not investigate in this thesis is to directly determine the support of the eigendistribution from the Stieltjes transform's expression without finding the eigendistribution itself. As a matter of fact, usually the shape of the eigendistribution is not as important as its support since it bounds the maximum and the minimum of the eigenvalues. There might be fundamental bounds on the performance of networked estimation that can be exploited by the characterization of the support.

8.4 Other Applications

A better understanding of how complex networks that unify communications, control and estimation work will help us analyze the existing examples of them, such as the gene regulatory network in biological sciences, and also design novel systems for future applications such as power grid control. We believe that by creating a firm theoretical foundation such an understanding will not be out of reach.

On the other hand, there exist connections between random Riccati recursions and particle filtering. Particle filtering is a powerful practical tool that through a sequential Monte Carlo method finds the optimal estimate of an unknown state. However, it is formidable to analyze. Recently it has been conjectured that the performance of particle filtering is fundamentally bounded by a modified Riccati recursion [SH09]. If this conjecture proves to be true, our developed machinery can be used to determine the performance of particle filtering which is used in many fields such as petroleum engineering [CO08].

Bibliography

[AC98] L. Arnold and I. Chueshov. Order-preserving random dynamical systems: Equilibria, attractors, applications. *Dynamics and Stability of Systems*, 13:265–280, 1998.

[Bai93a] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *The Annals of Probability*, 21(2):625–648, 1993.

[Bai93b] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *The Annals of Probability*, 21(2):649–672, 1993.

[BS98] 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*, 26(1):316–345, 1998.

[Chu02] I. Chueshov. *Monotone random systems.* Springer, 2002.

[CO08] Y. Chen and D. S. Oliver. Efficient ensemble-based closed-loop production optimization. In *SPE/ODE Improved Oil Recovery Symposium*, April 2008.

[EF86] E. Eleftheriou and D. Falconer. Tracking properties and steady-state performance of RLS adaptive filter algorithms. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(5):1097–1110, October 1986.

[ESTM08] M. Epstein, L. Shi, A. Tiwari, and R. M. Murray. Probabilistic performance of state estimation across a lossy network. *Automatica*, 44(12):3046–3053, December 2008.

[Ewe94] E. Eweda. Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels. *IEEE Transactions on Signal Processing*, 42(11):2937–2944, November 1994.

[GDH+09] 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*, 54(8):1807–1819, August 2009.

[GDHM06] V. Gupta, A. F. Dana, J. P. Hespanha, and R. M. Murray. Data transmission over networks for estimation. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, 2006.

[God74] D. Godard. Channel equalization using a kalman filter for fast data transmission. *IBM Journal of Research and Development*, 18(3):267, 1974.

[GQ95] Z. Gajic and M. T. J. Qureshi. *Lyapunov matrix equation in system stability and control*. Academic Press, 1995.

[Gup06] V. Gupta. *Distributed estimation and control in networked systems*. PhD thesis, California Institute of Technology, June 2006.

[Hay01] S. Haykin. *Adaptive Filter Theory*. Academic Press, 2001.

[HD07] M. Huang and S. Dey. Stability of Kalman filtering with Markovian packet drop. *Automatica*, 43(4):598–607, 2007.

[HM84] M. L. Honig and D. G. Messerschmitt. *Adaptive Filters: Structure, Algorithms, and Applications*. Kluwer Academic Publishers, 1984.

[HMH01] B. M. Hochwald, T. L. Marzetta, and B. Hassibi. Space-time autocoding. *IEEE Transactions on Information Theory*, 47(7):2761–2781, November 2001.

[HSK96] B. Hassibi, A. H. Sayed, and T. Kailath. H^∞ optimality of the LMS algorithm. *IEEE Transactions on Signal Processing*, 44(2):267–280, February 1996.

[HSK99] B. Hassibi, A. H. Sayed, and T. Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*. SIAM, 1999.

[IYB06] O. C. Imer, S. Yuksel, and T. Basar. Optimal control of dynamical systems over unreliable communication links. *Automatica*, 42(9):1429–1439, September 2006.

[KSH00] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, 2000.

[KSM09] S. Kar, B. Sinopoli, and J. M. F. Moura. Kalman filtering with intermittent observations: Weak convergence to a stationary distribution. *CoRR*, abs/0903.2890, 2009.

[LR95] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Oxford University Press, 1995.

[Lya92] A. M. Lyapunov. The general problem of the stability of motion. *International Journal of Control*, March 1992.

[Meh91] M. L. Mehta. *Random Matrices*. Prentice-Hall, 1991.

[MHH02] T. L. Marzetta, B. Hassibi, and B. M. Hochwald. Structured unitary space-time autocoding constellations. *IEEE Transactions on Information Theory*, 48(4):942–950, April 2002.

[MIK00] D. G. Manolakis, V. K. Ingle, and S. M. Kogon. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*. McGraw-Hill, 2000.

[MP67] V. A. Marčenko and L. A. Pastur. Distributions of eigenvalues for some sets of random matrices. *Math USSR-Sbornik*, 1(4):457–483, 1967.

[MS08] Y. Mo and B. Sinopoli. A characterization of the critical value for kalman filtering with intermittent observations. In *Proceedings of 47th IEEE Conference on Decision and Control*, Dec. 2008.

[PB09] K. Plarre and F. Bullo. On Kalman filtering for detectable systems with intermittent observations. *IEEE Transactions on Automatic Control*, 54(2):386–390, February 2009.

[RK07] C. Robinson and P. R. Kumar. Sending the most recent observation is not optimal in networked control: Linear temporal coding and towards the design of a control specific transport protocol. In *Proceedings of IEEE Conference on Decision and Control*, pages 334–339, Dec. 2007.

[Rog80] G. S. Rogers. *Matrix Derivatives*. Dekker, 1980.

[Rya98] O. Ryan. On the limit distributions of random matrices with independent or free entries. *Communications in Mathematical Physics*, 193:595–626, 1998.

[Say03] A. H. Sayed. *Fundamentals of Adaptive Filtering*. Wiley-IEEE, 2003.

[SC95] J. W. Silverstein and S. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):295–309, 1995.

[SETM05] L. Shi, M. Epstein, A. Tiwari, and R.M. Murray. Estimation with information loss: Asymptotic analysis and error bounds. In *44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05*, pages 1215–1221, Dec. 2005.

[SH09] R. T. Sukhavasi and B. Hassibi. Sign-of-innovations particle filtering. In *34th IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2009.

[SK94] A. H. Sayed and T. Kailath. A state-space approach to adaptive RLS filtering. *Signal Processing Magazine, IEEE*, 11(3):18–60, Jul 1994.

[Spe97] R. Speicher. Free probability theory and non-crossing partitions. *Sminaire Lotharingien de Combinatoire*, 39, 1997.

[SSF+04] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, Sep 2004.

[SSF+07] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, Jan 2007.

[TV04] A. M. Tulino and S. Verdu. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1):1–182, 2004.

[VH08a] A. Vakili and B. Hassibi. On the asymptotic eigenvalue distribution of certain random Lyapunov and Riccati recursions. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, 2008.

[VH08b] A. Vakili and B. Hassibi. A Stieltjes transform approach for analyzing the RLS adaptive filter. In *Proceedings of 46th annual Allerton Conference on Communication, Control and Computing*, September 2008.

[VH09] A. Vakili and B. Hassibi. On the eigendistribution of the steady-state error covariance matrix for the extended RLS algorithm. In *Proceedings of 34th International Conference on Acoustics, Speech, and Signal Processing*, April 2009.

[Voi83] D. Voiculescu. Asymptotically commuting finite rank unitary operators without commuting approximants. *Acta Sci. Math.*, 45:429–431, 1983.

[Voi86] D. Voiculescu. Addition of certain non-commuting random variables. *Journal of Functional Analysis*, 66:323–346, 1986.

[Voi87] D. Voiculescu. Multiplication of certain non-commuting random variables. *J. Operator Theory*, 18:223–235, 1987.

[Voi00] D. Voiculescu. Lectures on free probability theory. *Lectures on Probability Theory and Statistics: Ecole d'Ete de Probabilites; Lecture Notes in Mathematics*, pages 283–349, Springer, 2000.

[WH60] B. Widrow and M. E. Hoff. Adaptve switching circuits. In *IRE WESCON Convention Record*, pages 96–104, 1960.

[Wig55] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics*, 62(3):548–564, 1955.

[Wig58] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *The Annals of Mathematics*, 67(2):325–327, 1958.

[Wig59] E. Wigner. Statistical properties of real symmetric matrices with many dimensions. In *Proceedings of 4th Canadian Mathematical Congress*, pages 174–176, 1959.

[Wis28] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.

[XH05] Y. Xu and J. Hespanha. Estimation under controlled and uncontrolled communications in networked control systems. In *Proceedings of IEEE CDC-ECC*, pages 842–847, Dec. 2005.