Asymptotic Eigenanalysis of Large Random Lyapunov and Riccati Recursions

Babak Hassibi and Ali Vakili

Department of Electrical Engineering

California Institute of Technology, Pasadena, CA 91125

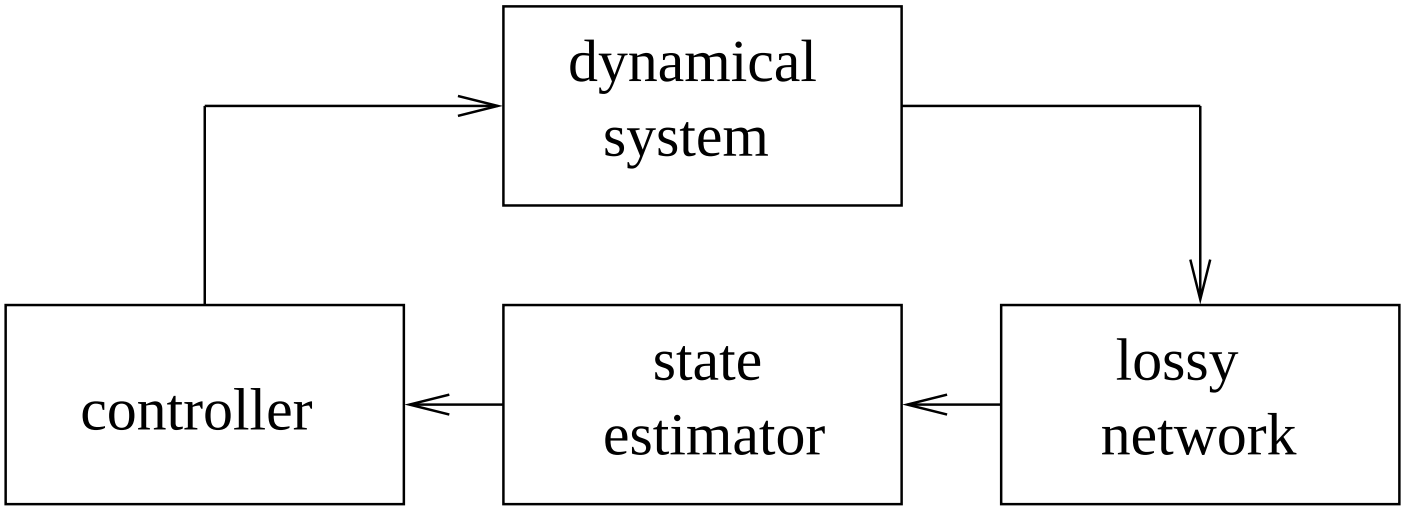
Frontiers in Distributed Communication, Sensing and Control

Yale University, Nov 1, 2008

Outline

- **Motivation**
  - estimation and control over lossy networks, adaptive filtering
- **Large Random Matrix Theory**
  - Stieltjes transform methods
- **Random Matrix Recursions**
  - Lyapunov recursions
  - Riccati recursions
- **Future Work and Open Problems**

Control and Estimation over Lossy Networks



- lossy network means that the measurements may be randomly dropped or arrive with random delays
- for LTI systems, the optimal estimator and controller are clear (Kalman filter, state feedback...)
- however, due to the randomness of the network, determining stability and analyzing the behavior of the system is very challenging

Consider an LTI system

$$\left\{ \begin{array}{rcl} x_{i+1} & = & Fx_i + Gu_i \\ y_i & = & Hx_i + v_i \end{array} \right. , \quad E \begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_j^* & v_j^* \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & rI_p \end{bmatrix} \delta_{ij},$$

where each component of $y_i \in \mathcal{R}^p$ corresponds to a different sensor measurement which is transmitted across a lossy network.

If these measurements are randomly dropped across the network, the error covariance of the state estimate, P_i, satisfies the Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - FP_iH^*(R_i + HP_iH^*)^{-1}HP_iF^*,$$

where $\{R_i\}$ is a sequence of iid diagonal matrices where

$$R_{i,jj} = \left\{ \begin{array}{ll} r & \text{with probability } 1 - \epsilon_j \\ \infty & \text{with probability } \epsilon_j \end{array} \right. .$$

More complicated random Riccati recursions arise if we add the possibility of measurements arriving with delay, rather than having them just dropped.

- The Riccati recursion given is nonlinear, time-variant and *random*, since the matrix R_i is random
- Clearly, unlike the deterministic time-invariant case, the P_i no longer converge
- Nonetheless, study of P_i is critical:
  - when is P_i bounded?
  - does P_i converge in distribution?
  - what is $E\mathrm{tr}P_i$?
- A special case of this problem (with a single sensor) has been studied by Sinopoli et al
  - (often loose) upper and lower bounds on P_i are given
  - not much else is known

How can we analyze these?

Adaptive Filtering

Consider the time-varying state-space model

$$\left\{ \begin{array}{rcl} x_{i+1} & = & x_i + u_i \\ y_i & = & h_i x_i + v_i \end{array} \right. , \quad E \begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_j^* & v_j^* \end{bmatrix} = \begin{bmatrix} qI & 0 \\ 0 & r \end{bmatrix} \delta_{ij}$$

Here, essentially, x_i undergoes a random walk. Moreover the row vector h_i is called the *regressor vector*. It is often *random*.

- the vectors h_i are occasionally spatially and temporally white
- if we are doing FIR adaptive filtering the h_i have a shift structure

  $$h_i = \begin{bmatrix} u_i & u_{i-1} & \dots & u_{i-n+1} \end{bmatrix}$$

  where the u_i are white.

In adaptive filtering the goal is to estimate x_i using the observations $\{y_j, j < i\}$.

LMS adaptive filtering:

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

and

$$P_{i+1} = (I - \mu h_i^T h_i) P_i (I - \mu h_i^T h_i) + r\mu^2 h_i^T h_i + qI.$$

RLS adaptive filtering:

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

and

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{r + h_i P_i h_i^T} + qI.$$

H^∞ **adaptive filtering**

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

and

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{\frac{r}{1-\gamma^{-2}} + h_i P_i h_i^T} + qI, \quad P_i^{-1} - \gamma^{-2} h_i^T h_i.$$

- The recursions given (Lyapunov for LMS and Riccati for RLS and H^∞) are all random, since the coefficient matrices depend on the h_i which are random.
- Again, the P_i no longer converge.
- Nonetheless, study of P_i is critical:
  - for LMS we need to know the range of the step size μ that guarantees the boundedness of P_i
  - for H^∞, we often care about the range of permissible γ (the H^∞) gain
  - in all cases we care about the mean square error, $E\mathrm{tr}P_i$.

How can we analyze all of these?

Random Lyapunov and Riccati Recursions

All of the above were examples of the random Lyapunov

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^*,$$

and random Riccati recursions

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - F_i P_i H_i^* (R_i + H_i P_i H_i^*)^{-1} H_i P_i F_i^*,$$

where the coefficient matrices $\{F_i, G_i, H_i, Q_i, R_i\}$ are possibly random.

Of course, this is too general a problem. We will make some assumptions:

1. the random matrices $\{F_i, G_i, H_i, Q_i, R_i\}$ are jointly stationary
   - even though P_i does not converge, its distribution may converge
2. the state dimension n is large
   - we thus hope to leverage results from large random matrix theory

We will provide a framework to tackle these problems.

Random Matrices

A $m \times n$ random matrix A is simply described by the joint distribution of its entries

$$p_A(A) = p_A(a_{ij}; i = 1, \ldots m; j = 1, \ldots n).$$

An example is the family of Gaussian random matrices, where the entries are jointly Gaussian.

- It turns out that for Gaussian matrices (and a large class of matrices derived from Gaussians) the joint (and marginal) distribution of the eigenvalues can be computed in closed form!
- Thus, much of the study of random matrices is devoted to the study of the distribution of the eigenvalues.
- When the matrices are not Gaussian (or Gaussian-derived), then things are much more complicated...
- ...unless the random matrices are large...

Wigner's Semi-Circle Law

Let A be a matrix whose entries are zero-mean, unit-variance and iid, with bounded fourth order moment. Consider the symmetric matrix

$$B = \frac{1}{\sqrt{2n}}(A + A^T).$$

Then as $n \to \infty$ the marginal distribution of the eigenvalues of B converges to the *semi-circle* distribution

$$p_\lambda(\lambda) = \begin{cases} \frac{1}{2\pi}\sqrt{4-\lambda^2} & \text{when } -2 \leq \lambda \leq 2 \\ 0 & \text{otherwise} \end{cases} .$$

This is an example of a universal law (akin to the law of large numbers).

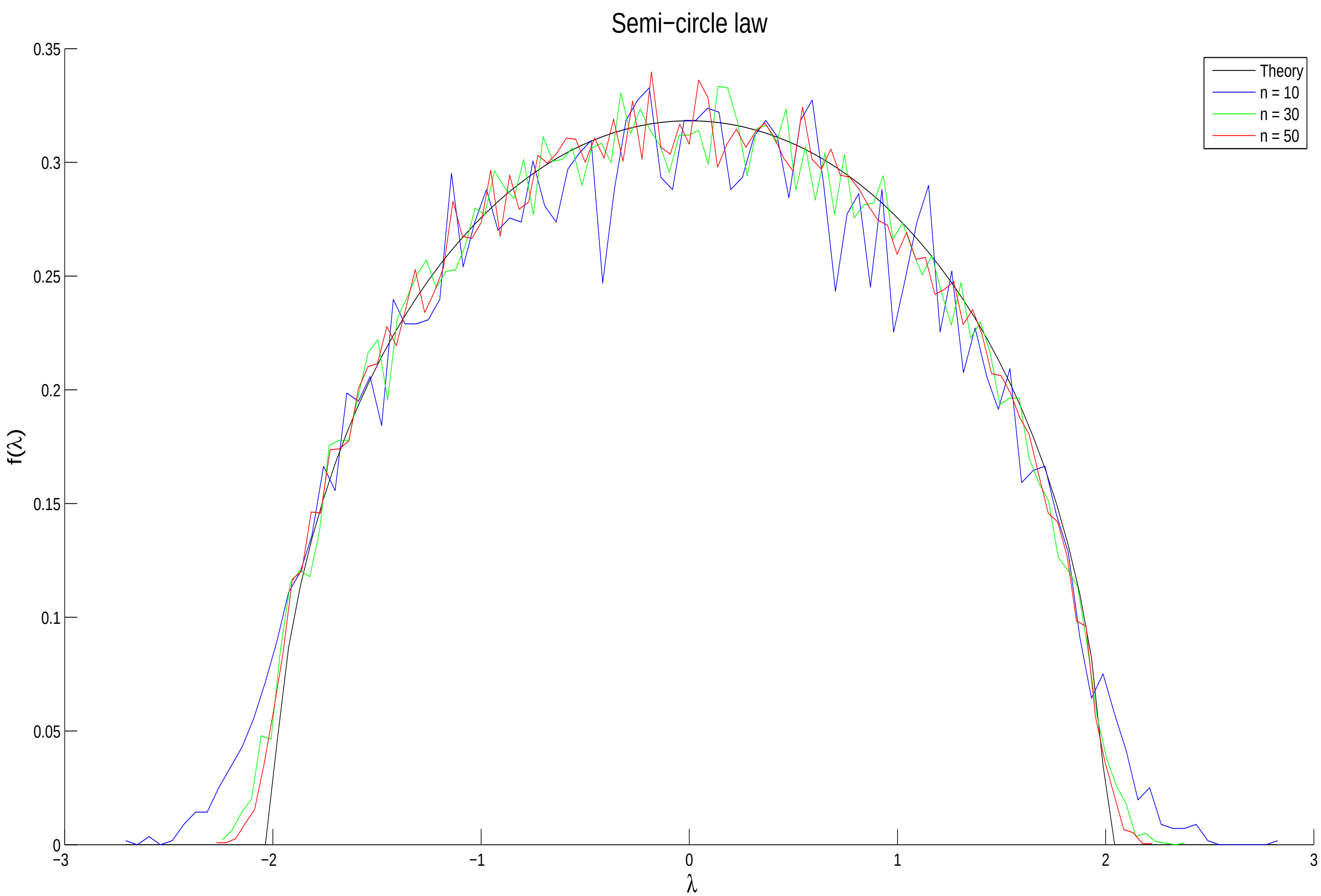


Figure 1: *Convergence to semi-circle law*

The Stieltjes Transform

First used to study the asymptotic eigendistribution for large random matrices by Marcenko and Pastur (1967). Given a distribution, $p_\lambda(\cdot)$, the *Steitjes transform* is the function of the complex variable z, defined as

$$s(z) = E\frac{1}{z-\lambda} = \int \frac{p_\lambda(\lambda)}{z-\lambda} d\lambda.$$

The distribution can be recovered from its transform using the formula

$$p_\lambda(\lambda) = \frac{1}{\pi}\mathrm{Im} s(\lambda + j0^+).$$

The moments $m_i = E\lambda^i$ can be found from expanding the Stieltjes around infinity:

$$s(z) = E\frac{1}{z-\lambda} = E\frac{1}{z}(1-\frac{\lambda}{z})^{-1} = E\frac{1}{z}\sum_{i=0}^{\infty}\frac{\lambda^i}{z^i} = \sum_{i=0}^{\infty}\frac{m_i}{z^{i+1}}.$$

Likewise, expanding around zero gives the moments $m_{-i} = E\lambda^{-i}$:

$$s(z) = \sum_{i=1}^{\infty} m_{-i} z^i.$$

When A is a symmetric random matrix, it is clear that the Stieltjes transform of its eigenvalue distribution is given by

$$s(z) = \frac{1}{n} E \mathrm{tr}(zI - A)^{-1}.$$

An alternative representation which will prove useful is

$$s(z) = \frac{1}{n} E \frac{d}{dz} \log \det(zI - A).$$

Back to the Semi-Circle Law

Let A be a matrix whose entries are zero-mean, unit-variance and iid, with bounded fourth order moment. Consider the symmetric matrix

$$B = \frac{1}{\sqrt{2n}}(A + A^T).$$

Then

$$\begin{aligned}
s(z) &= \lim_{n\to\infty} \frac{1}{n} E\mathrm{tr}(zI - B)^{-1} \\
&= \lim_{n\to\infty} \frac{1}{n} E\mathrm{tr}\left(zI - \frac{1}{\sqrt{2n}} \begin{bmatrix} b_{11} & B_{12} \\ B_{12}^T & B_{22} \end{bmatrix} \right)^{-1} \\
&= \lim_{n\to\infty} E \frac{1}{z - \underbrace{\frac{b_{11}}{\sqrt{2n}}}_{\to 0} - \underbrace{\frac{1}{2n} B_{12}(zI - B_{22})^{-1} B_{12}^T}_{\to s(z)}} = \frac{1}{z-s}.
\end{aligned}$$

This implies that $s^2 - zs + 1 = 0$, which means

$$s = \frac{z \pm \sqrt{z^2 - 4}}{2}.$$

Using the inverse transform

$$p_\lambda(\lambda) = \frac{1}{\pi} \mathrm{Im} s(\lambda + j0^+),$$

we obtain

$$p_\lambda(\lambda) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - \lambda^2} & \text{when } -2 \leq \lambda \leq 2 \\ 0 & \text{otherwise} \end{cases} .$$

Warm Up - A Random Lyapunov Recursion

Consider the Lyapunov recursion

$$P_{i+1} = \alpha P_i + G_i G_i^T,$$

where the G_i are independent $n \times m$ ($\frac{m}{n} = \beta < 1$) matrices with iid zero-mean, $\frac{1}{m}$-variance entries.

Let us work in a basis where P_i is diagonal. For simplicity drop the time index on P_i and G_i and partition things as

$$P_i = \begin{bmatrix} \lambda_1 & \\ & \Lambda_2 \end{bmatrix} \quad , \quad G_i = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}.$$

$$
\begin{aligned}
s_{i+1}(z) &= \frac{1}{n}E\mathrm{tr}(zI-\alpha P_i - G_iG_i^T)^{-1} \\
&= \frac{1}{n}E\mathrm{tr}\left[\begin{array}{cc} z-\alpha\lambda_1 - G_1G_1^T & -G_1G_2^T \\ -G_2G_1^T & zI-\alpha\Lambda_2 - G2G_2^T \end{array}\right]^{-1} \\
&= E\frac{1}{z-\alpha\lambda_1 - G_1G_1^T - G_1G_2^T(zI-\alpha\Lambda_2-G2G_2^T)^{-1}G_2G_1^T} \\
&= E\frac{1}{z-\alpha\lambda_1 - G_1(I_m - G_2^T(zI-\alpha\Lambda_2)^{-1}G_2)^{-1}G_1} \\
&= E\frac{1}{z-\alpha\lambda_1 - \underbrace{\frac{1}{m}E\mathrm{tr}(I_m - G_2^T(zI-\alpha\Lambda_2)^{-1}G_2)^{-1}}_{\triangleq t}}
\end{aligned}
$$

To compute t, let us further partition G_2:

$$
G_2 = \left[\begin{array}{cc} G_{21} & G_{22} \end{array}\right].
$$

Thus,

$$\begin{aligned}
t &= \frac{1}{m} E\mathrm{tr}(I_m - G_2^T(zI - \alpha\Lambda_2)^{-1}G_2)^{-1} \\
&= \frac{1}{m} E\mathrm{tr}\begin{bmatrix} 1 - G_{21}^T(zI - \alpha\Lambda_2)^{-1}G_{21} & -G_{21}^T(zI - \alpha\Lambda_2)^{-1}G_{22} \\ -G_{22}^T(zI - \alpha\Lambda_2)^{-1}G_{21} & I - G_{22}^T(zI - \alpha\Lambda_2)^{-1}G_{22} \end{bmatrix}^{-1} \\
&= E\frac{1}{1 - \underbrace{G_{21}^T(zI - \alpha\Lambda_2 - G_{22}G_{22}^T)^{-1}G_{21}}_{\frac{n}{m}s_i(z) = \frac{1}{\beta}s_i(z)}}
\end{aligned}$$

We therefore have

$$\begin{aligned}
s_{i+1}(z) &= E\frac{1}{z - \alpha\lambda_1 - \frac{\beta}{\beta - s_i(z)}} \\
&= \frac{1}{\alpha}E\frac{1}{\frac{z}{\alpha} - \frac{\beta/\alpha}{\beta - s_i(z)} - \lambda_1} = \frac{1}{\alpha}s_i\left(\frac{z}{\alpha} - \frac{\beta/\alpha}{\beta - s_i(z)}\right).
\end{aligned}$$

When $\alpha < 1$, the recursion is stable and therefore $s_i(z)$ converges to the solution of the implicit equation

$$s(z) = \frac{1}{\alpha} s\left(\frac{z}{\alpha} - \frac{\beta/\alpha}{\beta - s(z)}\right). \tag{1}$$

We do not know how to solve this equation in closed form. However, if one applies the expansion

$$s(z) = \sum_{i=0}^{\infty} \frac{m_i}{z^{i+1}}$$

to (1) then one can obtain the moments term by term as

$$\begin{aligned}
m_1 &= \frac{1}{1-\alpha} \\
m_2 &= \frac{1}{(1-\alpha)^2} + \frac{1/\beta}{1-\alpha^2} \\
m_3 &= \frac{1}{(1-\alpha)^3} + \frac{3/\beta}{(1-\alpha)(1-\alpha^2)} + \frac{1/\beta^2}{1-\alpha^3}
\end{aligned}$$

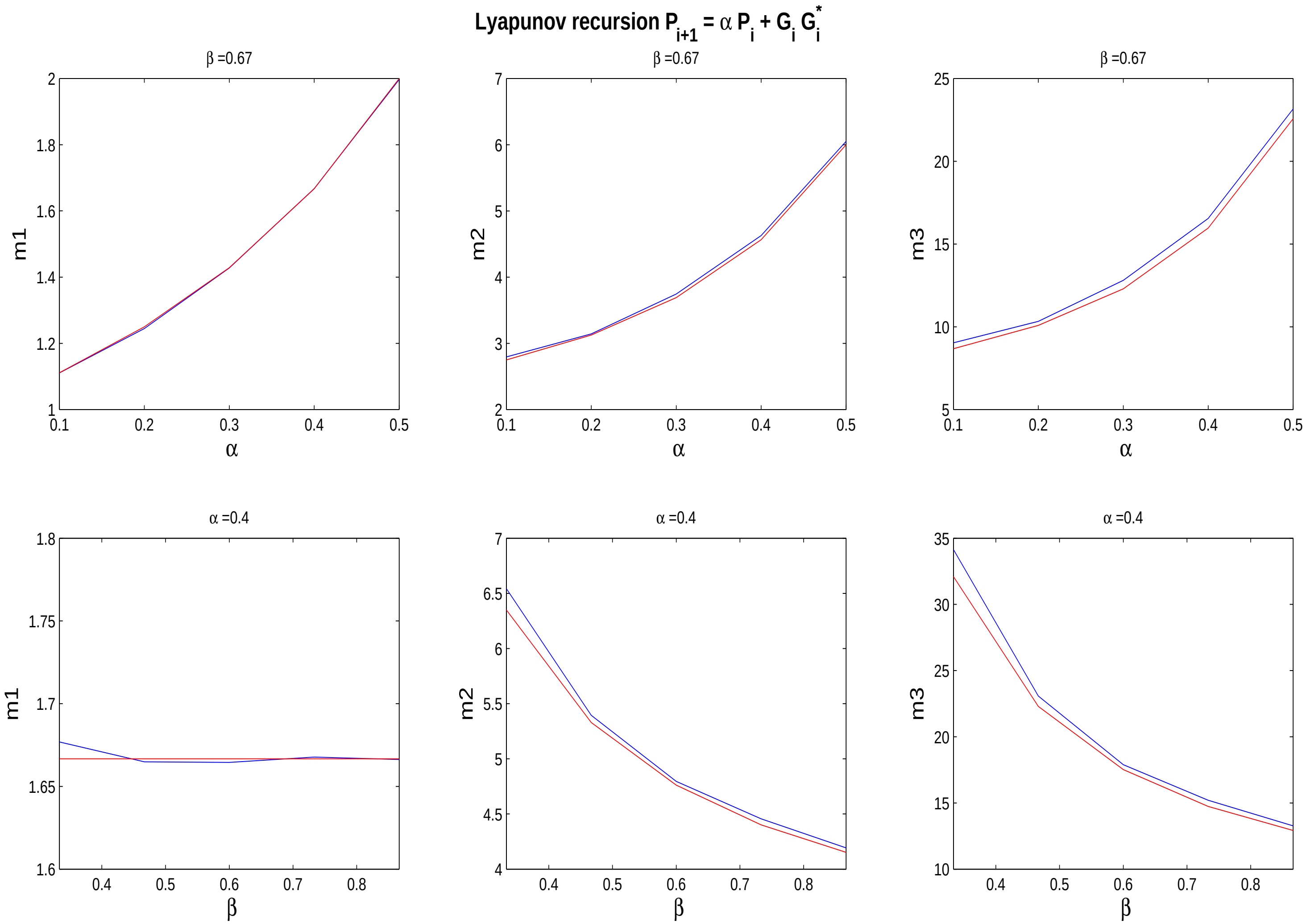


Figure 2: *Empirical moments in blue, theoretical in red* $(n = 30)$.

A More General Lyapunov Recursion

Consider now the Lyapunov recursion

$$P_{i+1} = \alpha F_i P_i F_i^T + G_i G_i^T,$$

where the G_i are as before and the F_i are independent $n \times n$ matrices with iid zero-mean, $\frac{1}{n}$ variance entries. A similar analysis shows that for $\alpha < 1$ the Stieltjes transform converges to

$$\begin{aligned} s(z) &= \frac{1}{\alpha} t\left(\frac{z}{\alpha} - \frac{\beta/\alpha}{\beta - s(z)}\right) \\ -zt^2(z) &= s\left(\frac{1}{t(z)}\right) \end{aligned}$$

Using the Laurent expansion around infinity one may obtain the moments as

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

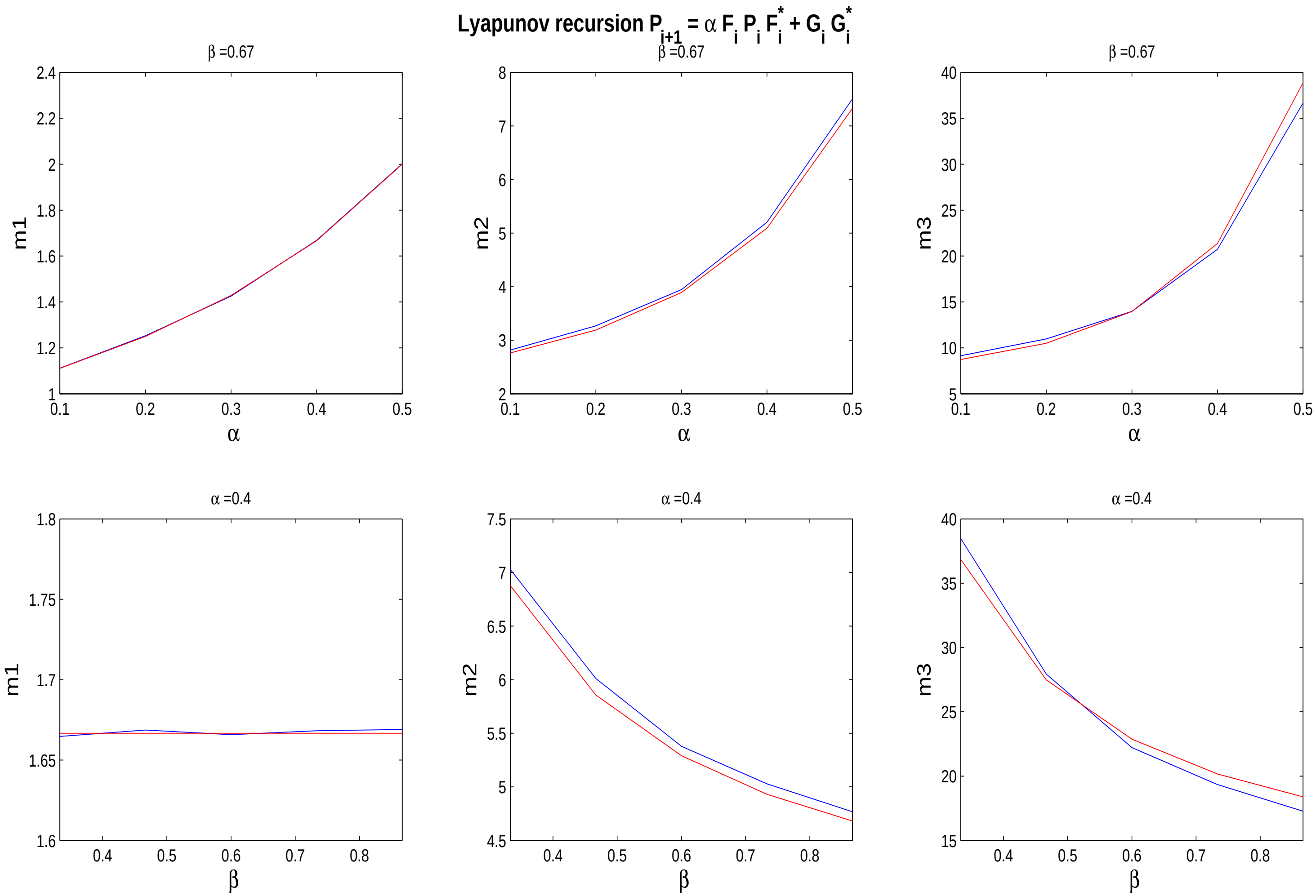


Figure 3: *Empirical moments in blue, theoretical in red ($n = 30$).*

LMS Adaptive Filtering

$$P_{i+1} = (I - \mu h_i^T h_i) P_i (I - \mu h_i^T h_i) + r\mu^2 h_i^T h_i + qI.$$

Assume that the h_i are independent random vetors with iid zero-mean $\frac{1}{n}$ variance entries.

$$\begin{aligned}
s_{i+1}(z) &= \frac{1}{n} E \frac{d}{dz} \log\det\Big(zI - (I - \mu h_i^T h_i) P_i (I - \mu h_i^T h_i) - r\mu^2 h_i^T h_i - qI\Big) \\
&= \frac{1}{n} E \frac{d}{dz} \log\det\Big((z-q)I - P_i + \begin{bmatrix} h_i^T & P_i h_i^T \end{bmatrix} \\
&\qquad \begin{bmatrix} -\mu^2(r + h_i P_i h_i^T) & \mu \\ \mu & 0 \end{bmatrix} \begin{bmatrix} h_i \\ h_i P_i \end{bmatrix}\Big) \\
&= s_i(z-q) + \frac{1}{n} E \frac{d}{dz} \log\det\Big(I + ((z-q)I - P_i)^{-1} \begin{bmatrix} h_i^T & P_i h_i^T \end{bmatrix} \\
&\qquad \begin{bmatrix} -\mu^2(r + h_i P_i h_i^T) & \mu \\ \mu & 0 \end{bmatrix} \begin{bmatrix} h_i \\ h_i P_i \end{bmatrix}\Big)
\end{aligned}$$

Therefore

$$s_{i+1}(z+q) \quad = \quad s_i(z) + \frac{1}{n} E \frac{d}{dz} \log\det\left(\begin{bmatrix} 0 & \mu^{-1} \\ \mu^{-1} & r + h_i P_i h_i^T \end{bmatrix} + \begin{bmatrix} h_i \\ h_i P_i \end{bmatrix} (zI - P_i)^{-1} \begin{bmatrix} h_i^T & P_i h_i^T \end{bmatrix} \right)$$

Now, as $n \to \infty$, we have

$$\begin{aligned} h_i(zI - P_i)^{-1} h_i^T &\quad\to\quad s_i \\ h_i(zI - P_i)^{-1} P_i h_i^T &\quad=\quad h_i(zI - P_i)^{-1}(-zI + P_i + zI) h_i^T \to -1 + z s_i \\ h_i P_i (zI - P_i)^{-1} P_i h_i^T &\quad=\quad h_i(-zI + P_i + zI)(zI - P_i)^{-1}(-zI + P_i + zI) h_i^T \\ &\quad\to\quad z - h_i P_i h_i^T - 2z + z^2 s_i \end{aligned}$$

and so

$$\begin{aligned} s_{i+1}(z+q) &\quad=\quad s_i(z) + \frac{1}{n} \frac{d}{dz} \log\det \begin{bmatrix} s_i & \mu^{-1} - 1 + z s_i \\ \mu^{-1} - 1 + z s_i & r - z + z^2 s_i \end{bmatrix} \\ &\quad=\quad s_i(z) + \frac{1}{n} \frac{d}{dz} \log(s_i(r + (1 - 2\mu^{-1})z) - (1 - \mu^{-1})^2). \end{aligned}$$

It is reasonable that we should take $q = \beta/n$. In this case,

$$s_{i+1} + \frac{\beta}{n}\frac{d}{dz}s_{i+1} = s_i(z) + \frac{1}{n}\frac{d}{dz}\log(s_i(r + (1-2\mu^{-1})z) - (1-\mu^{-1})^2).$$

It can be argued that, if $\mu < 2$, the Stieltjes transform converges, and we then have

$$\frac{\beta}{n}\frac{d}{dz}s = \frac{1}{n}\frac{d}{dz}\log(s(r + (1-2\mu^{-1})z) - (1-\mu^{-1})^2),$$

and so

$$\beta s + c = \log(s(r + (1-2\mu^{-1})z) - (1-\mu^{-1})^2),$$

for some constant c. In fact, c can be found by noting that $s(\infty) = 0$ and $\lim_{z\to\infty} zs = 1$, which yields

$$\beta s + j\pi - 2\log\mu = \log(s(r + (1-2\mu^{-1})z) - (1-\mu^{-1})^2), \tag{2}$$

While it is not possible to solve (2) in closed form it is easy to numerically solve it for each z by considering $s = u + jv$ and numerically solving for u and v.

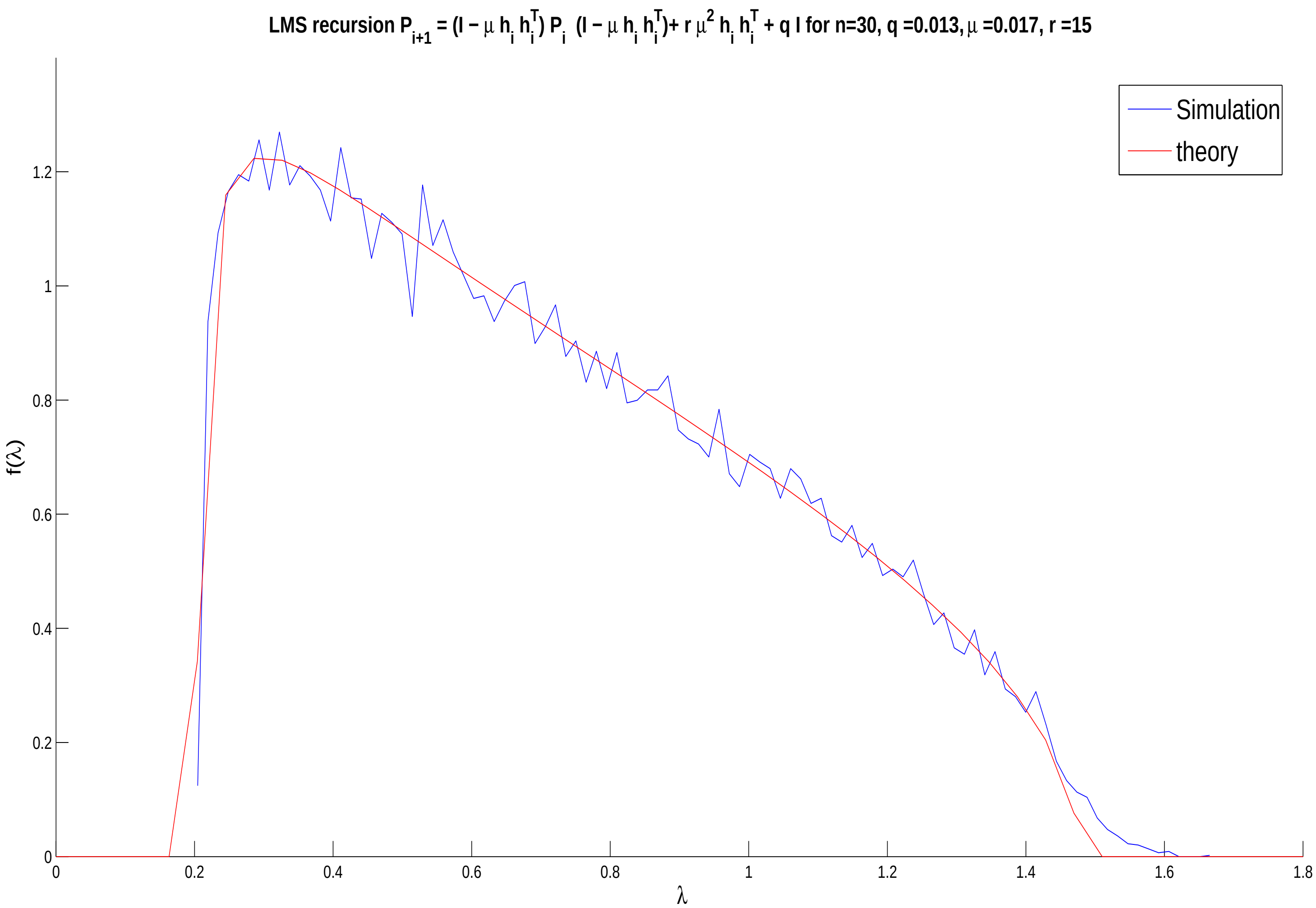


Figure 4: *Eigendistribution for LMS filter* $(n = 30)$.

RLS Adaptive Filtering

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{r + h_i P_i h_i^T} + qI.$$

We assume the same h_i and $q = \frac{\beta}{n}$. A similar analysis yields

$$\beta s + c = \log(r - z + z^2 s), \tag{3}$$

where c is a constant. Here the Stieltjes tranform always converges.

If we now use the fact that $s(z) = z^{-1} + m_1 z^{-2} + \ldots$, we obtain

$$c + \beta z^{-1} + \beta m_1 z^{-2} + \ldots = \log(r + m_1 + m_2 z^{-1} + \ldots),$$

from which we obtain

$$c = \log(r + m_1).$$

Thus, c cannot be separately determined. The reason is that the Riccati recursion is nonlinear and there is no simple algebraic equation for $m_1 = \frac{1}{n} E \mathrm{tr} P_i$. c is found by numerically solving (3) and insisting that the inverse Stieltjes transform integrate to one.

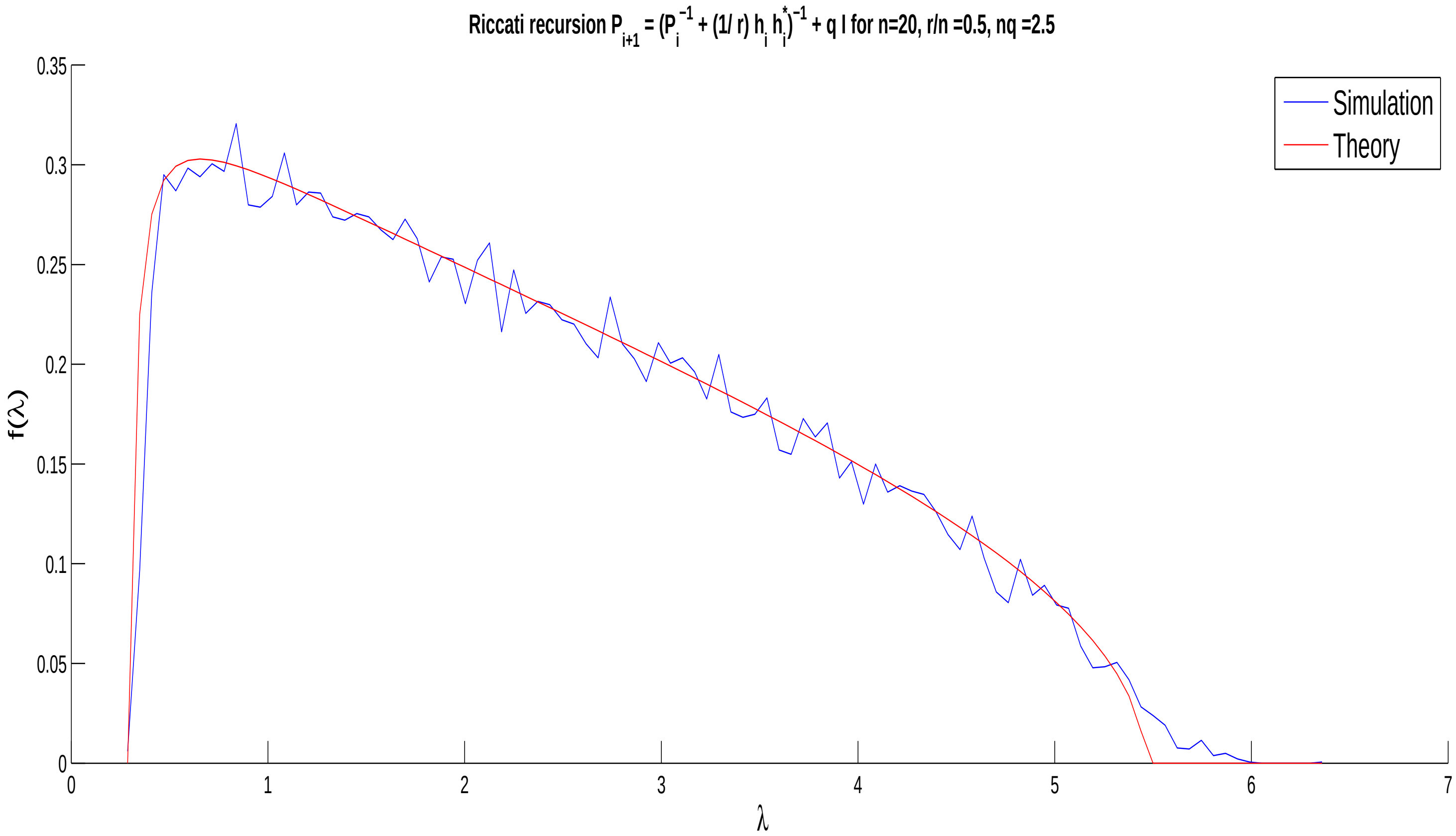


Figure 5: *Eigendistribution for RLS filter* $(n = 30)$.

Another RLS Adaptive Filter

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{r + h_i P_i h_i^T} + q g_i g_i^T.$$

We assume the same h_i and and that the g_i are similarly distributed (iid zero-mean, $\frac{1}{n}$ variance). The analysis this time yields

$$z^2 s - (q^{-1} z^2 + z - r)s + q^{-1} z + c = 0, \tag{4}$$

where c is a constant. Once more the Stieltjes tranform always converges. If we attempt to find c using the series expansion of $s(z)$ we obtain

$$c = \frac{m_1}{q}.$$

We again cannot expect a simple algebraic equation for m_1. Solving the quadratic equation for s, the inverse Stieltjes transform yields

$$p_\lambda(\lambda) = \frac{1}{\pi} \mathrm{Im} \left(\frac{\sqrt{(q^{-1}\lambda^2 + \lambda - r)^2 - 4\lambda^2 (q^{-1}\lambda + c)}}{2\lambda^2} \right).$$

c is then found such that the resulting $p_\lambda(\cdot)$ integrates to one.

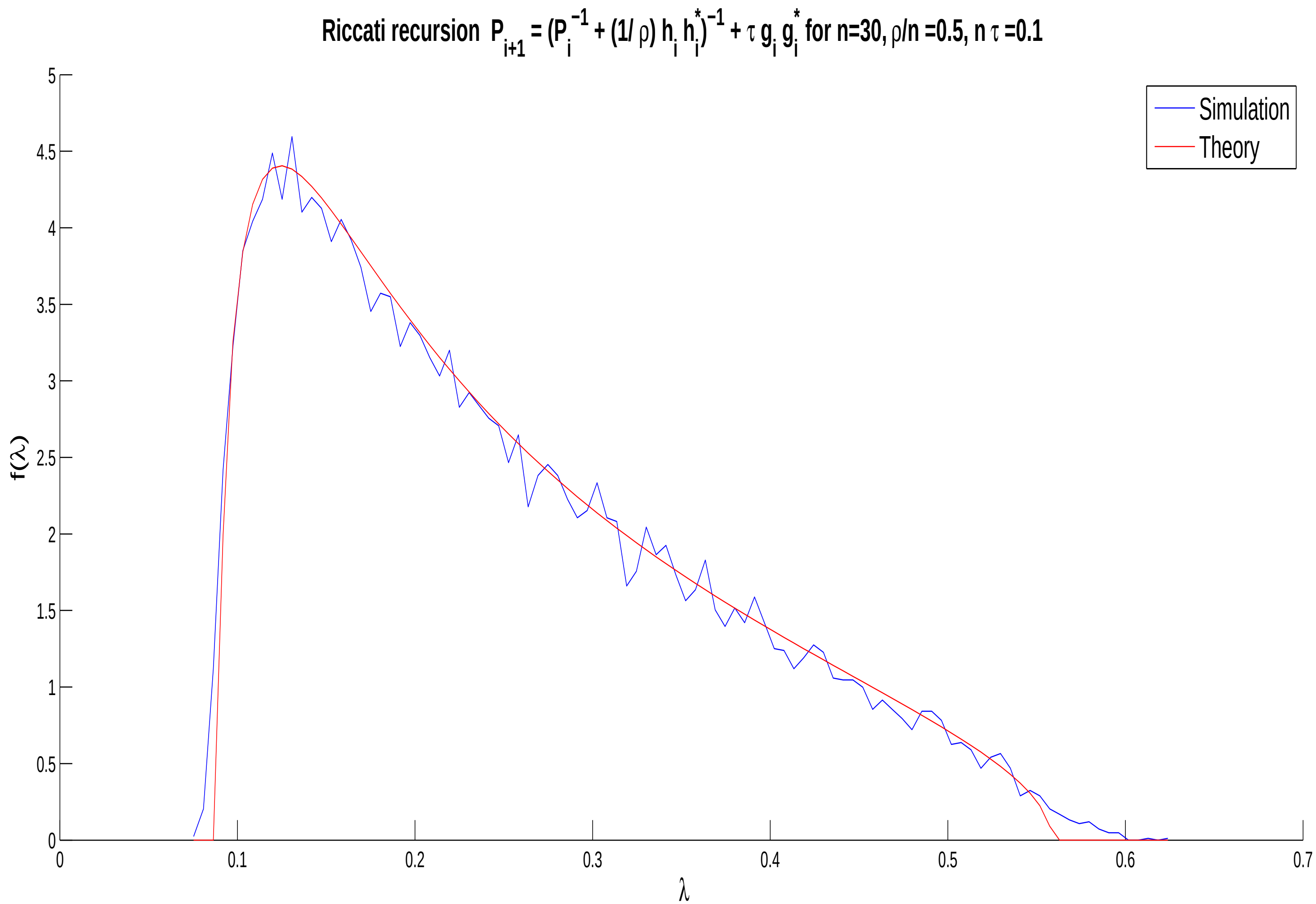


Figure 6: *Eigendistribution for RLS filter ($n = 30$).*

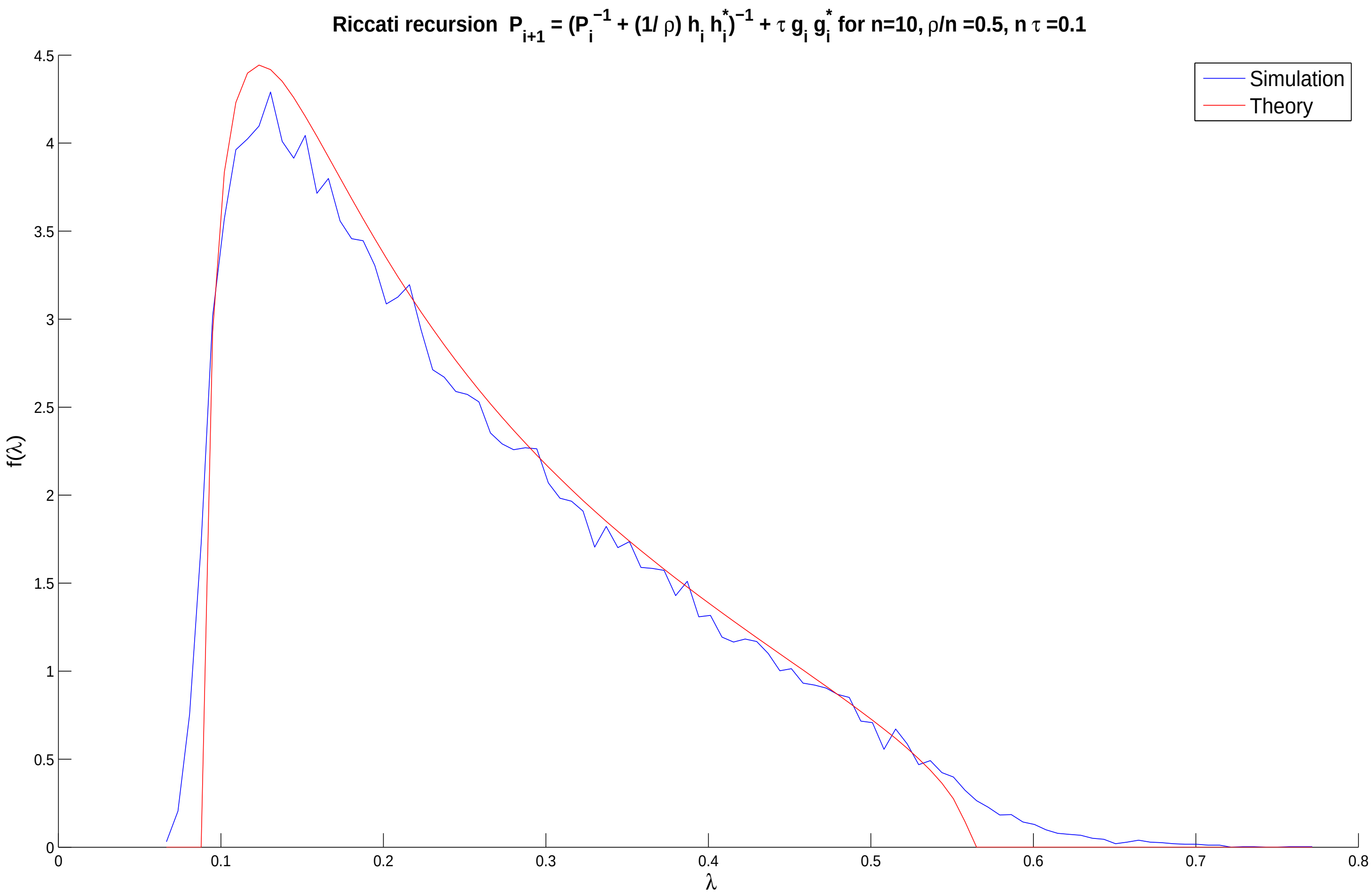


Figure 7: *Eigendistribution for RLS filter* $(n = 10)$.

Regressors with Shift Structure

So far we have assumed that the regressors h_i are spatially and temporally white. In many cases, they have a shift structure

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

where the u_i are white.

A crucial step in our derivations is that

$$h_i(zI - P_i)^{-1}h_i^T \rightarrow s(z).$$

where we relied on the fact that the h_i were independent of the P_i. However, this may not be necessary.

Lemma 1 *Let h_i be a given vector and Θ be a randomly chosen orthogonal matrix. Then there exists a matrix Δ such that*

$$h_i(\Theta + \Delta) = \|h_i\|e_1 \quad , \quad (\Theta + \Delta)(\Theta + \Delta)^T = I,$$

and $Etr(\Delta\Delta^*) = O(\frac{1}{n})$.

Using this lemma:

$$\begin{aligned}
h_i(zI - P_i)^{-1}h_i^T &= h_i\Theta(zI - \Theta P_i\Theta^T)^{-1}\Theta^T h_i^T \\
&= (\|h_i\|e_1 - h_i\Delta)(zI - \bar{P}_i)^{-1}(\|h_i\|e_1 - h_i\Delta)^T \\
&= \|h_i\|^2 e_1(zI - \bar{P}_i)^{-1}e_1^T - 2\|h_i\|e_1(zI - \bar{P}_i)^{-1}\Delta^T h_i^T \\
&\qquad + h_i\Delta(zI - \bar{P}_i)^{-1}\Delta^T h_i^T \\
&\rightarrow s(z) + 0 + 0
\end{aligned}$$

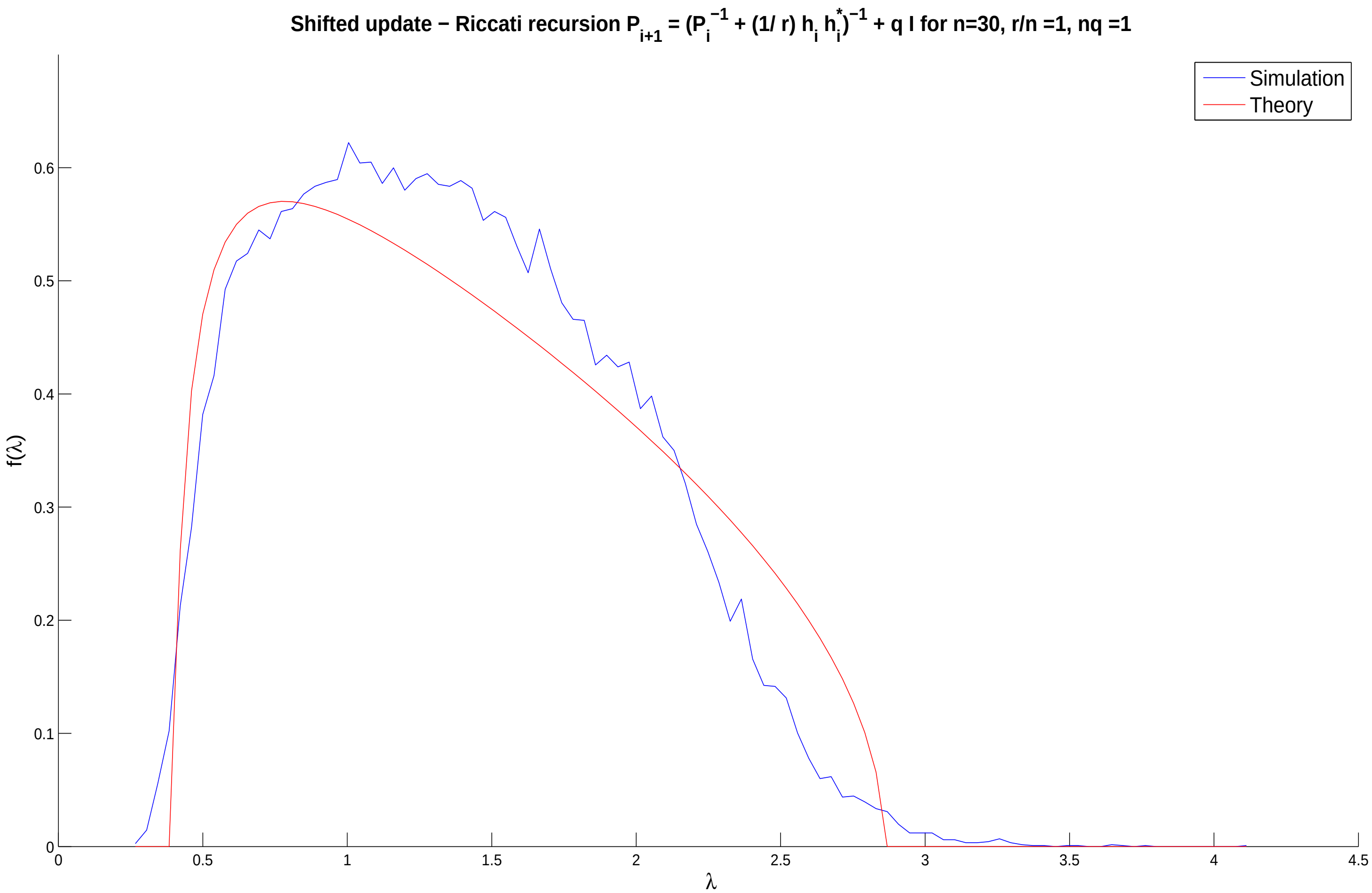


Figure 8: *Eigendistribution for RLS filter with shift-structured regressor*$(n = 30)$.

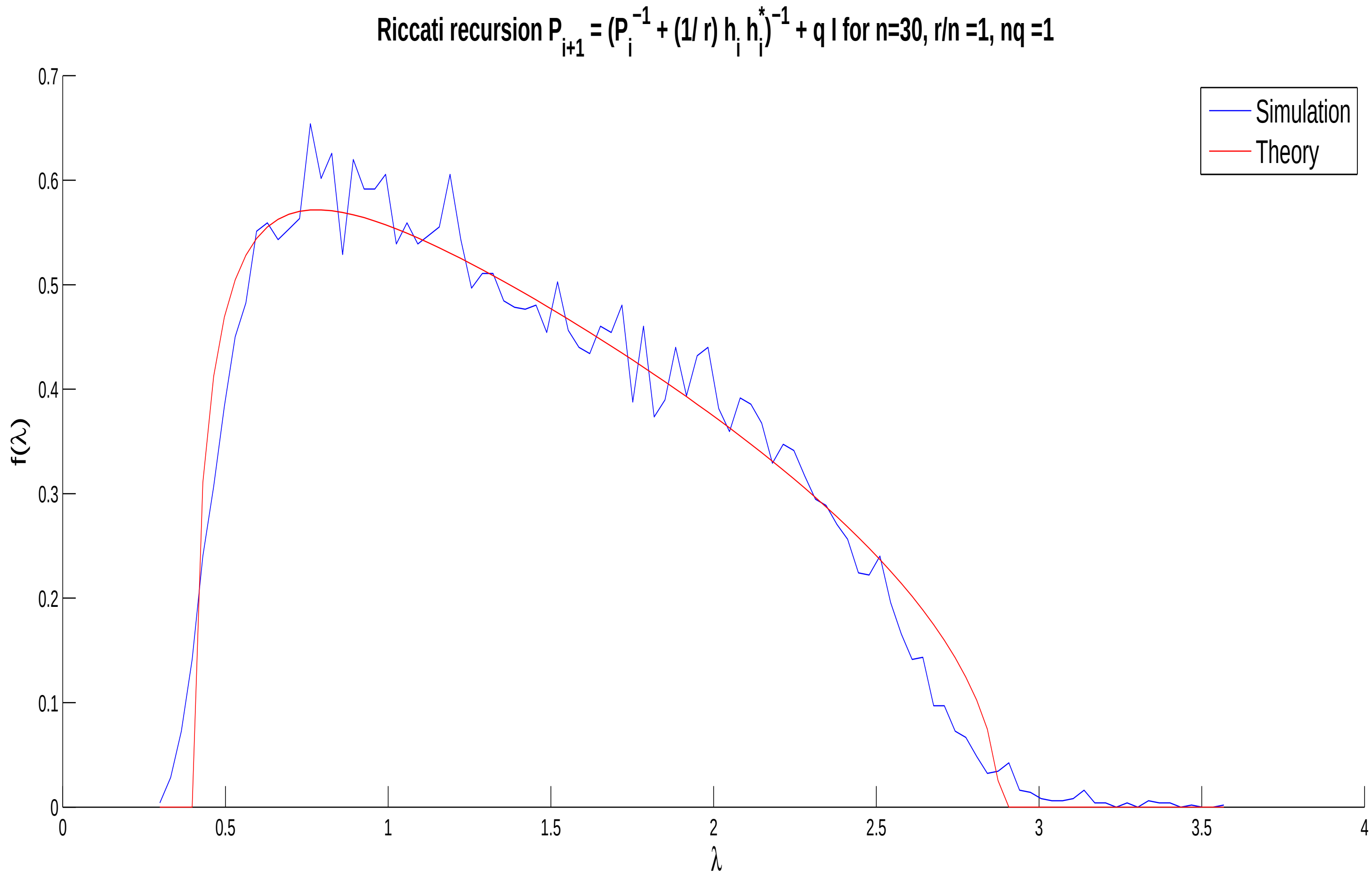


Figure 9: *Eigendistribution for RLS filter with shift-structured regressors and double shifts* $(n = 30)$.

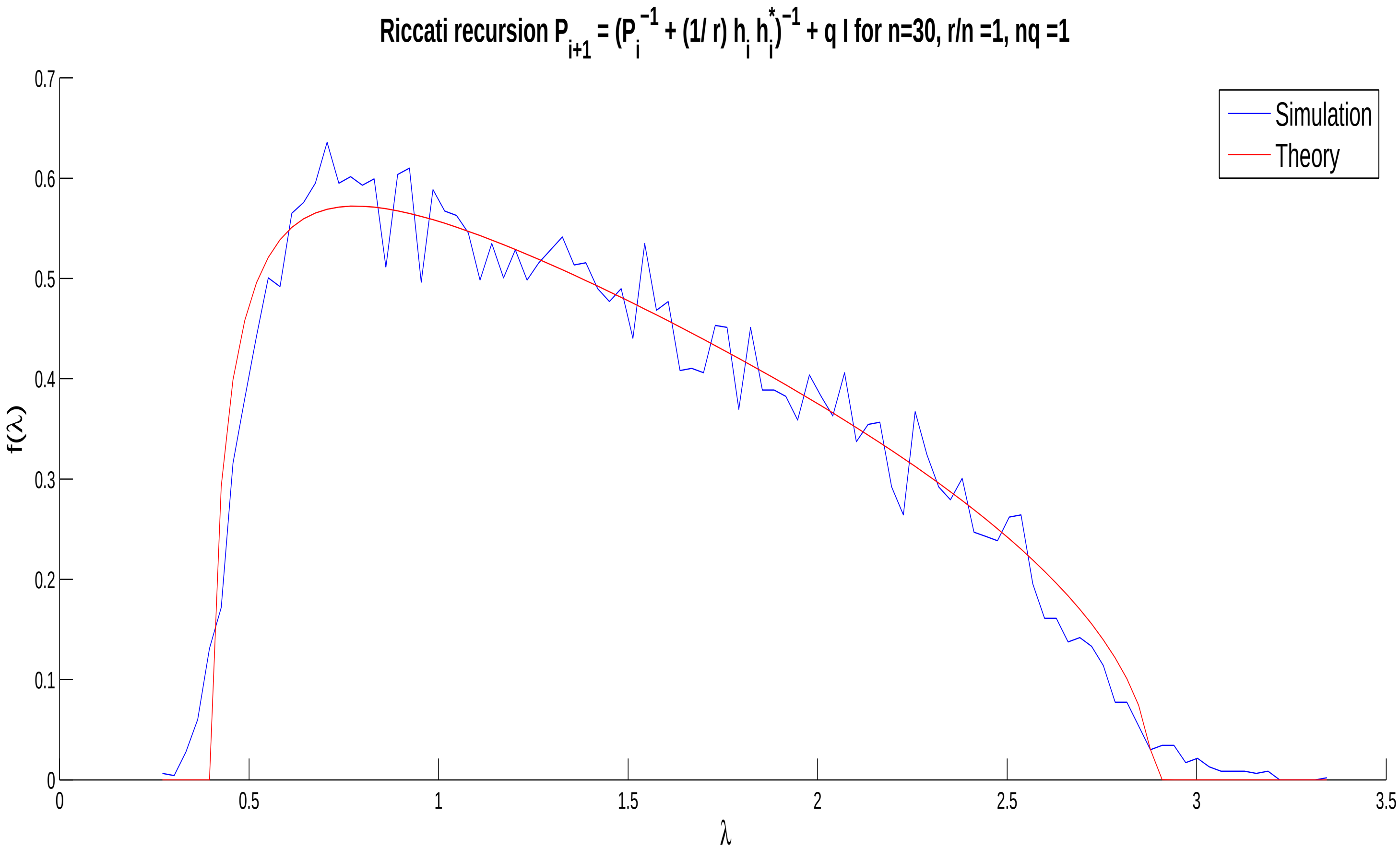


Figure 10: *Eigendistribution for RLS filter with shift-structured regressors and triple shifts* $(n = 30)$.

RLS with Multiple Measurements

The analysis we have done so far assumed that $n \gg 1$. If we have multiple measurements, say m of them, then as long as $n \gg m$ our analysis is still valid.

However, if $n \gg 1$ and $\frac{m}{n} = \beta$, a constant, then we need to analyze things anew. Here we go....

$$P_{i+1} = P_i - P_i H_i^* (rI + H_i P_i H_i^*)^{-1} H_i P_i + qI,$$

where the H_i are $n \times n$ (we have taken $\beta = 1$ for simplicity) independent matrices, whose entries are iid zero-mean, $\frac{1}{n}$ variance.

$$
\begin{aligned}
s_{i+1}(z+q) &= \frac{1}{n} E \frac{d}{dz} \log\det(zI + qI - P_{i+1}) \\
&= \frac{1}{n} E \frac{d}{dz} \log\det\left(zI - P_i + P_i H_i^* (rI + H_i P_i H_i^*)^{-1} H_i P_i\right) \\
&= s_i(z) + \frac{1}{n} E \frac{d}{dz} \log\det\left(I + (zI - P_i)^{-1} P_i H_i^* (rI + H_i P_i H_i^*)^{-1} H_i P_i\right) \\
&= s_i(z) + \frac{1}{n} E \frac{d}{dz} \log\det\left(rI + H_i P_i H_i^* + H_i P_i (zI - P_i)^{-1} P_i H_i *\right)
\end{aligned}
$$

The problem is that, since H_i is a square matrix, $H_i P_i (zI - P_i)^{-1} P_i H_i *$ does not self-average. (Which is why we need a different analysis.)

Let us write

$$
s_{i+1}(z+q) = s_i(z) + \frac{1}{n} E \frac{d}{dz} \log\det\left(rI + H_i \underbrace{\left(-zI + z^2 (zI - P_i)^{-1}\right)}_{\triangleq D(z)} H_i^* \right).
$$

Taking derivatives with respect to z:

$$\begin{aligned} s_{i+1}(z+q) &= s_i(z) + \frac{1}{n} E\mathrm{tr}\left(H_i D' H_i^* (rI + H_i D H_i^*)^{-1}\right) \\ &= s_i(z) + \frac{1}{n} E\mathrm{tr}\left(D'D^{-1} - D'D^{-2}\left(D^{-1} + \frac{H_i^* H_i}{r}\right)\right) \\ &= s_i(z) + E\frac{d'(z)}{d(z)} - \frac{1}{n} E\mathrm{tr} D'D^{-2}\left(D^{-1} + \frac{H_i^* H_i}{r}\right) \end{aligned}$$

Now $d(z) = -z + \frac{z^2}{z-\lambda} = \frac{\lambda z}{z-\lambda}$ and therefore $d'(z) = \frac{-\lambda^2}{(z-\lambda)^2}$. Hence

$$E\frac{d'(z)}{d(z)} = E\frac{-\lambda^2}{(z-\lambda)^2}\frac{z-\lambda}{\lambda z} = E\frac{-\lambda}{z(z-\lambda)} = \frac{1}{z}(1 - z s_i(z)) = \frac{1}{z} - s_i(z).$$

The other term

$$-\frac{1}{n} E\mathrm{tr} D'D^{-2}\left(D^{-1} + \frac{H_i^* H_i}{r}\right),$$

requires more work. We did a similar calculation when deriving (1).

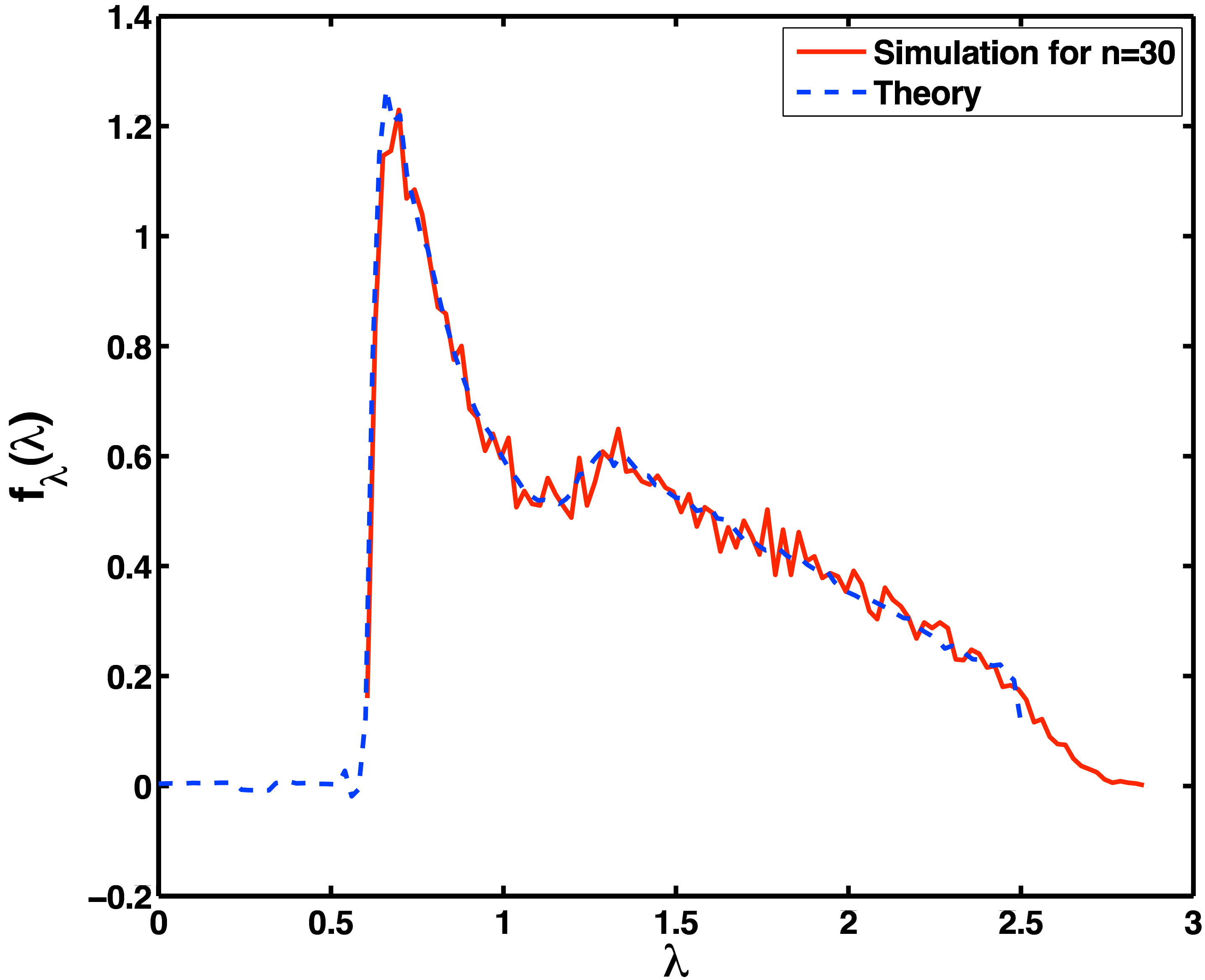
Using a technique similar to what was done there (essentially going through two successive partitions of H_i and two successive Schur complements—I will spare you the details), yields

$$s(z+q) = \frac{t(z)}{1-zt(z)} - \frac{1}{(1-zt(z))^2} s\left(\frac{z}{1-zt(z)}\right) \qquad (5)$$

$$t(z) = \frac{1}{r - \frac{z}{1-zt(z)} + \frac{z^2}{(1-zt(z))^2} s\left(\frac{z}{1-zt(z)}\right)} \qquad (6)$$

Unlike the Lyapunov case (1) an expansion of $s(z)$ and $t(z)$ does not give the moments m_i (the moments become heavily coupled in the coefficients upon expansion).

Since a closed form solution is not at all apparent, we can again numerically solve (5-6). This is a bit trickier than the earlier implicit equations, since we cannot solve things for each z, but rather need to solve for the whole complex plane at once.



Simulation for n=30
Theory
$f_\lambda(\lambda)$
λ
1.4
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
0
0.5
1
1.5
2
2.5
3

Intermittent Observations

The case of intermittent observations can be handled in our framework. As an example, consider the the RLS filter where observations are dropped with probability ϵ, i.e., with probability ϵ

$$P_{i+1} = P_i + \frac{\beta}{n} I \quad , \quad s_{i+1}(z+q) = s_i(z) \text{ or } \beta s'(z) = 0$$

and with probability $1 - \epsilon$,

$$P_{i+1} = P_i - \frac{P_i h_i^* h_i P_i}{r + h_i P_i h_i^*} + \frac{\beta}{n} I \quad , \quad \beta s_{i+1} + c = \log(r - z + z^2 s_i)$$

Combining these two yields

$$\beta s + c = (1-\epsilon)\log(r - z + z^2 s).$$

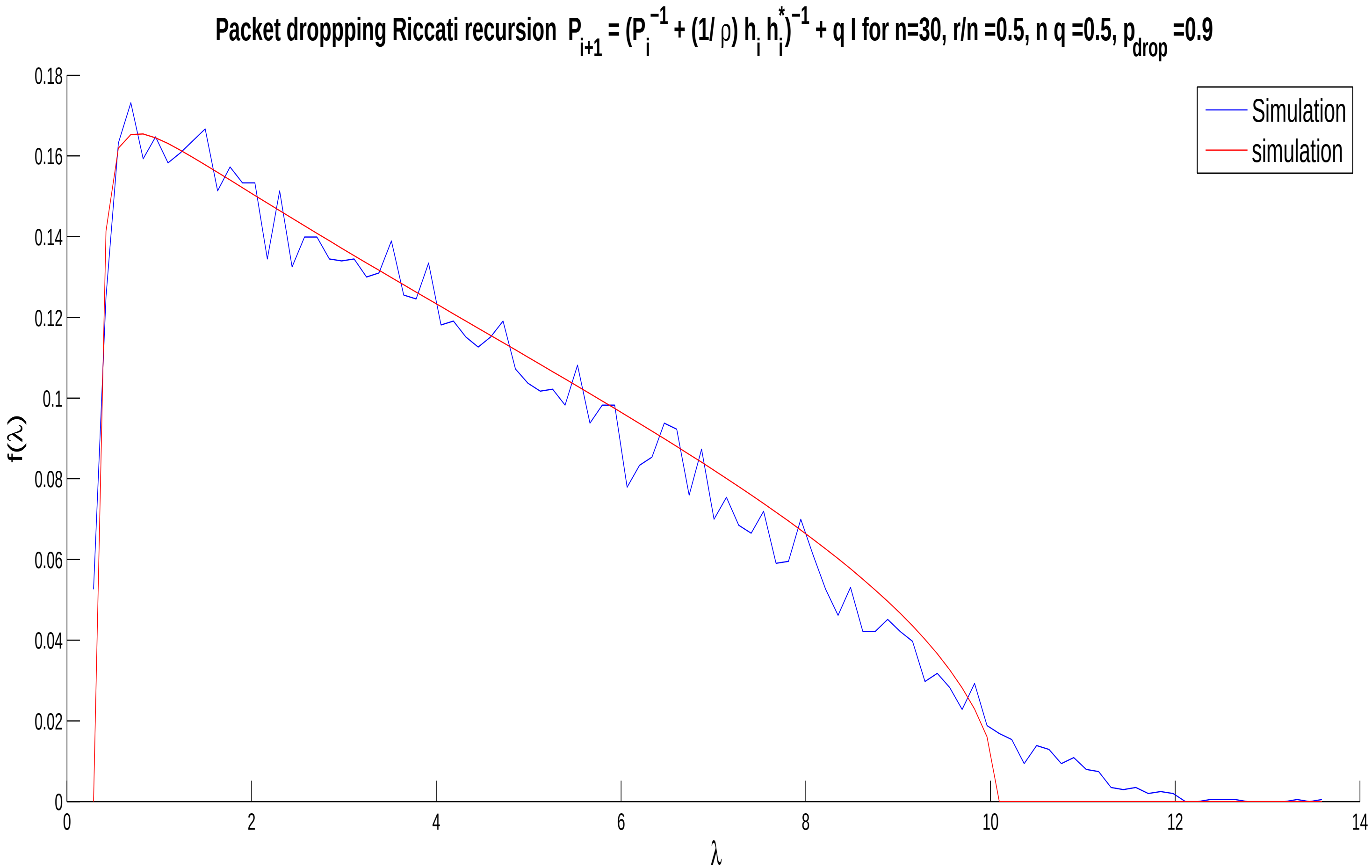


Figure 11: *Eigendistribution for RLS filter with packet drops* ($n = 30$).

Transient Analysis

It is also possible to do a transient analysis, i.e., to investigate the convergence rate of the eigendistributions.

- **For LMS**:

$$s_{i+1}(z) = s_i(z) - \frac{\beta}{n} s_i'(z) - \frac{1}{n}\frac{d}{dz}\log\left(s_i(z)(-r - z + \frac{2z}{\mu}) - (1 - \frac{1}{\mu})^2\right),$$

which, upon expanding the coefficients of the Stieltjes transform yields the following recursion for $m_1 = \frac{1}{n} E \mathrm{tr} P_i$:

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

which after γn steps suggests a convergence rate of

$$e^{-\mu(2-\mu)\gamma}.$$

- **For RLS**:

$$s_{i+1}(z) = s_i(z) - \frac{\beta}{n} s_i'(z) - \frac{1}{n} \frac{d}{dz} \log \left(r - z - z^2 s_i(z) \right).$$

While it is possible to show that $s_i(z)$ converges, identifying the convergence rate is more tricky.

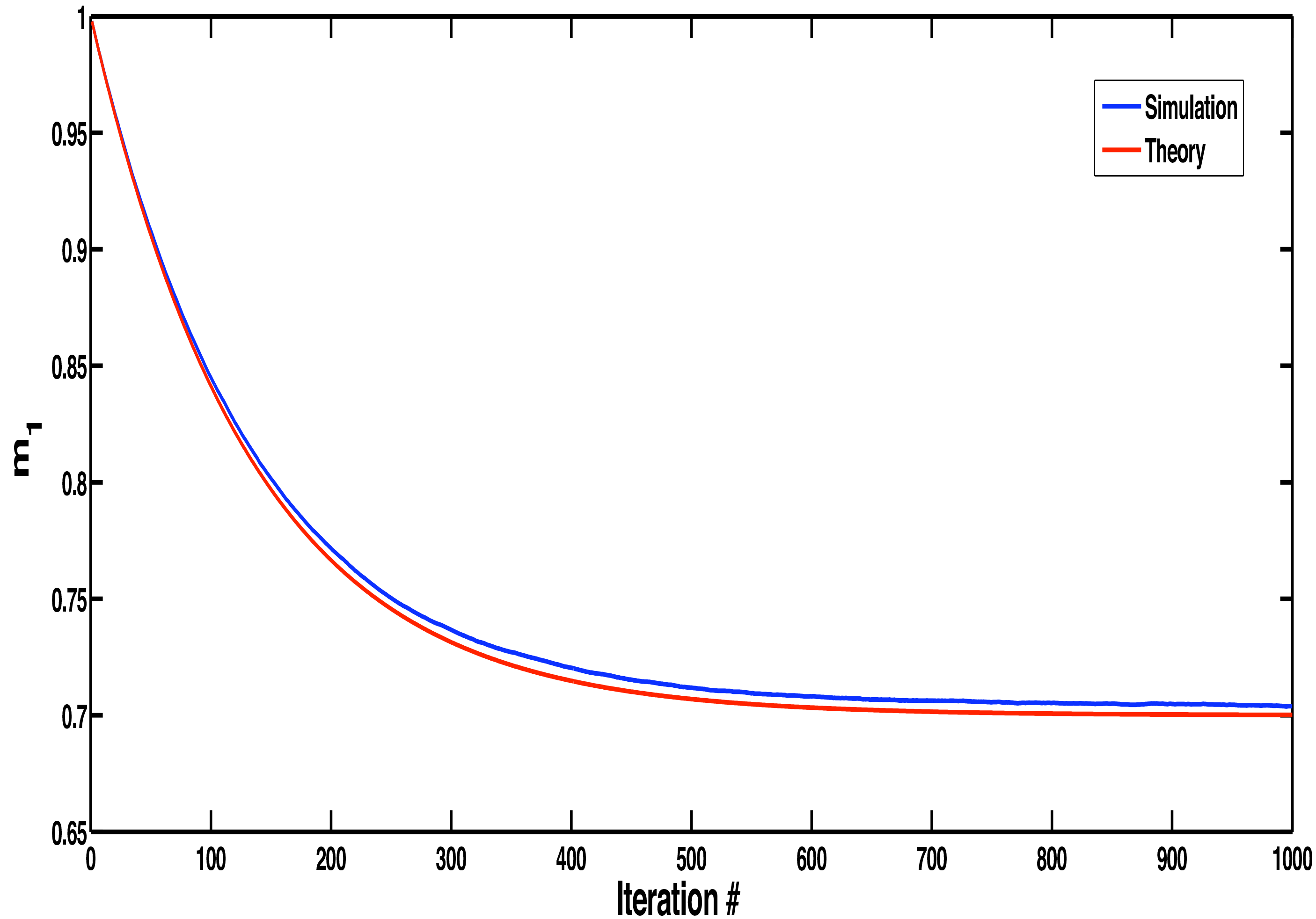


Figure 12: *Transient behavior of m_1 for LMS filter ($n = 30$).*

Comments, Conclusions and Future Work

- The approach we have developed seems quite promising, since we have obtained quite a few new and nontrivial results.
- Nonetheless, much is still left to do (we are not yet quite capable of analyzing Sinopoli-type packet loss systems)
- We have a more or less complete theory when the measurment matrices H_i are chosen randomly (even with shift structure), R_i is possibly random and the state transition matrix F_i is either a multiple of identity or the $\{F_i\}$ is an iid random matrix process of zero-mean iid entries.
- The theory very well matches the empirical results for systems with as low as $n = 10, 20$ states.
- The transient analysis is more or less complete for Lyapunov-type recursions. For Riccati recursions we still need to analyze the convergence rate of the corresponding Stieltjes transform recursions.

- The main open problem is to extend the theory to fixed F (and H), yet random R_i (this would cover the problem of estimation and control over lossy networks, e.g., the Sinopoli model and many others).