Distributed Estimation and Control over Lossy Networks

Babak Hassibi and Ali Vakili

Department of Electrical Engineering

California Institute of Technology, Pasadena, CA 91125

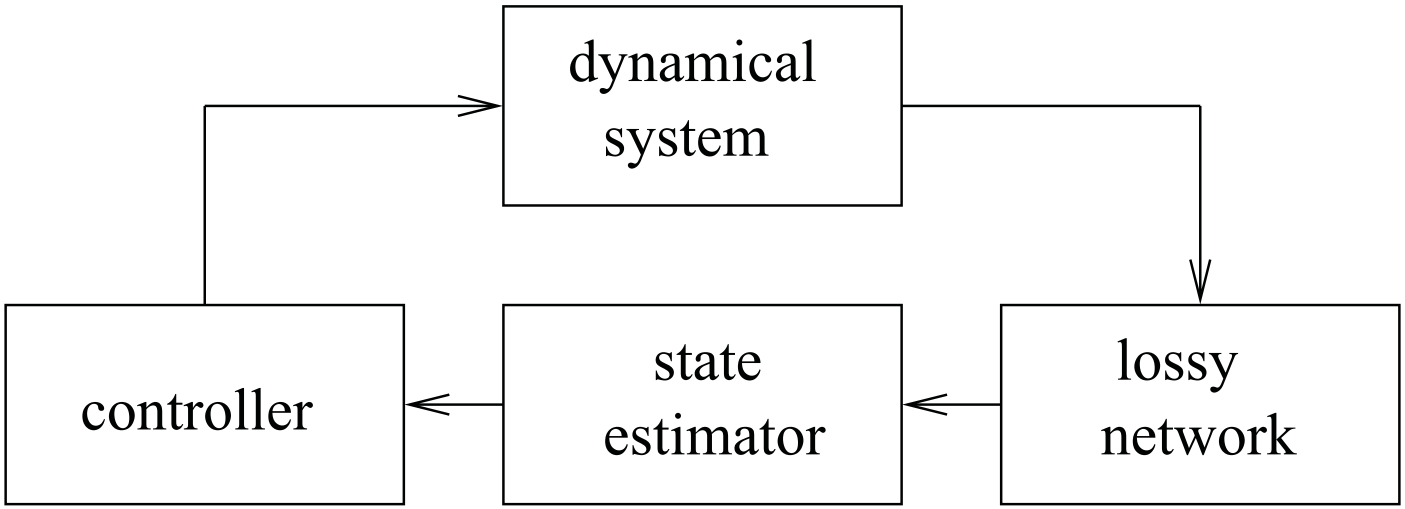
Ninth Annual Lee Center Workshop on Advanced Networking

May 22, 2009

Some Problems in Networks

- **Information Transmission over Networks**
  - network information theory
  - network models: wireless erasure networks
  - entropic vectors
  - optimization-based framework
- **Estimation and Control over Networks**
  - distributed algorithms
  - lossy networks
  - communication constraints, quantization
  - universal laws for networks

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

A Generic Example

Consider an LTI system

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

Each component of $y_i \in \mathcal{R}^p$ is a different sensor measurement.

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

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

Perhaps the greatest success story of control theory in the last 50 years has to do with the analysis of the *deterministic time-invariant Riccati recursion*:

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

- even though this is a nonlinear matrix recursion, conditions for stability and boundedness, convergence (and the rate of convergence) and the steady-state value $P = \lim_{i\to\infty} P_i$ are all well understood

. However, in network problems, in addition to being nonlinear, the Riccati recursion is *time-varying* and *random*, since the matrix R_i is

- 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 $E\mathrm{tr}P_i$?
- Unfortunately, very little is known.

How can we analyze these?

A Classical Problem: 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

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

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

LMS adaptive filtering:

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

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:

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

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**

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

and

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{\frac{r}{1-\gamma^{-2}} + h_i P_i h_i^T} + qI, \quad P_i^{-1} - \gamma^{-2} h_i^T h_i \geq 0.$$

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

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

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

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

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

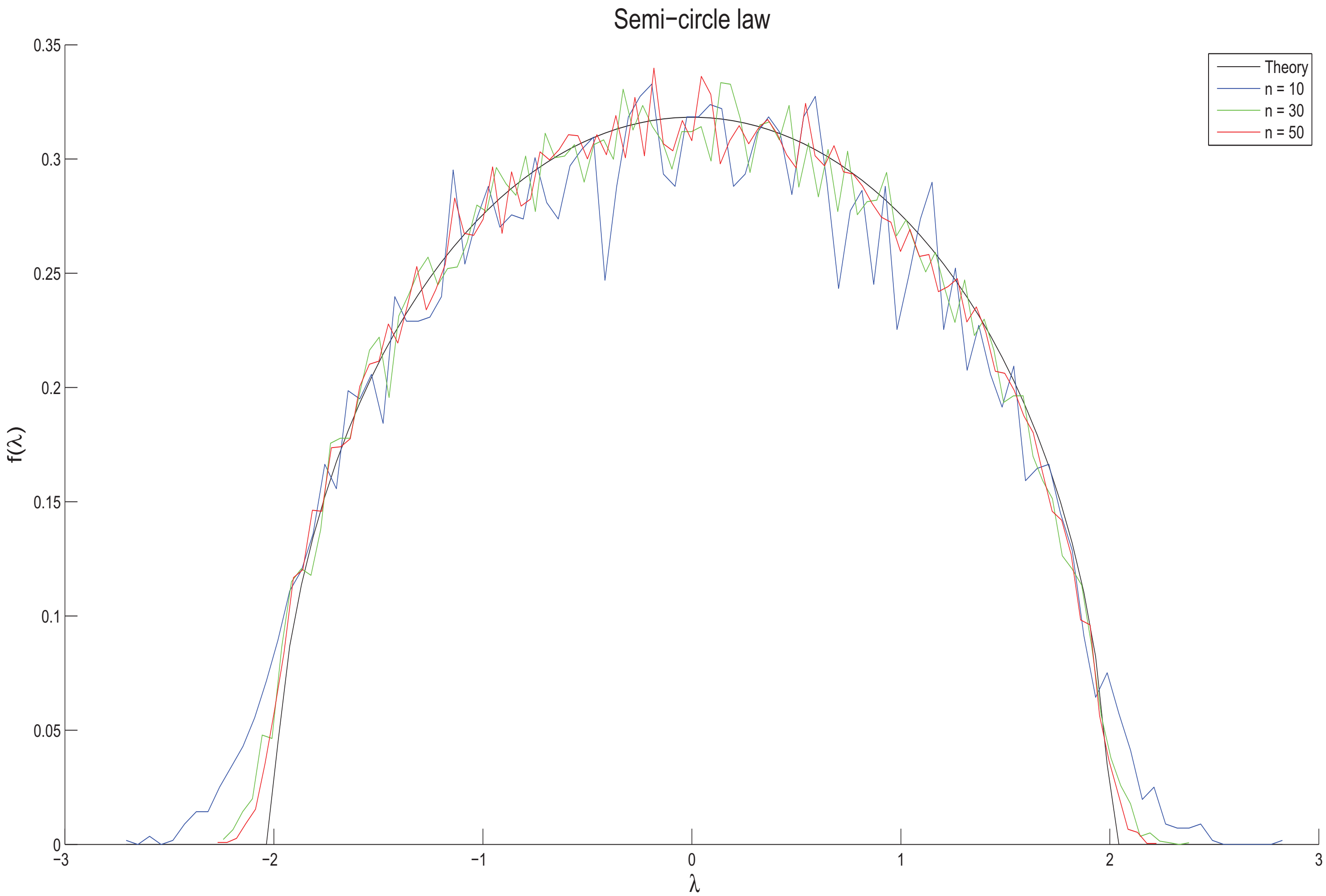


Figure 1: *Convergence to semi-circle law*

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

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

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

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

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.

One can show that

$$s_{i+1}(z) = \frac{1}{\alpha} s_i \left(\frac{z}{\alpha} - \frac{\beta/\alpha}{\beta - s_{i+1}(z)} \right).$$

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

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

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

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

then one can obtain the moments term by term as

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

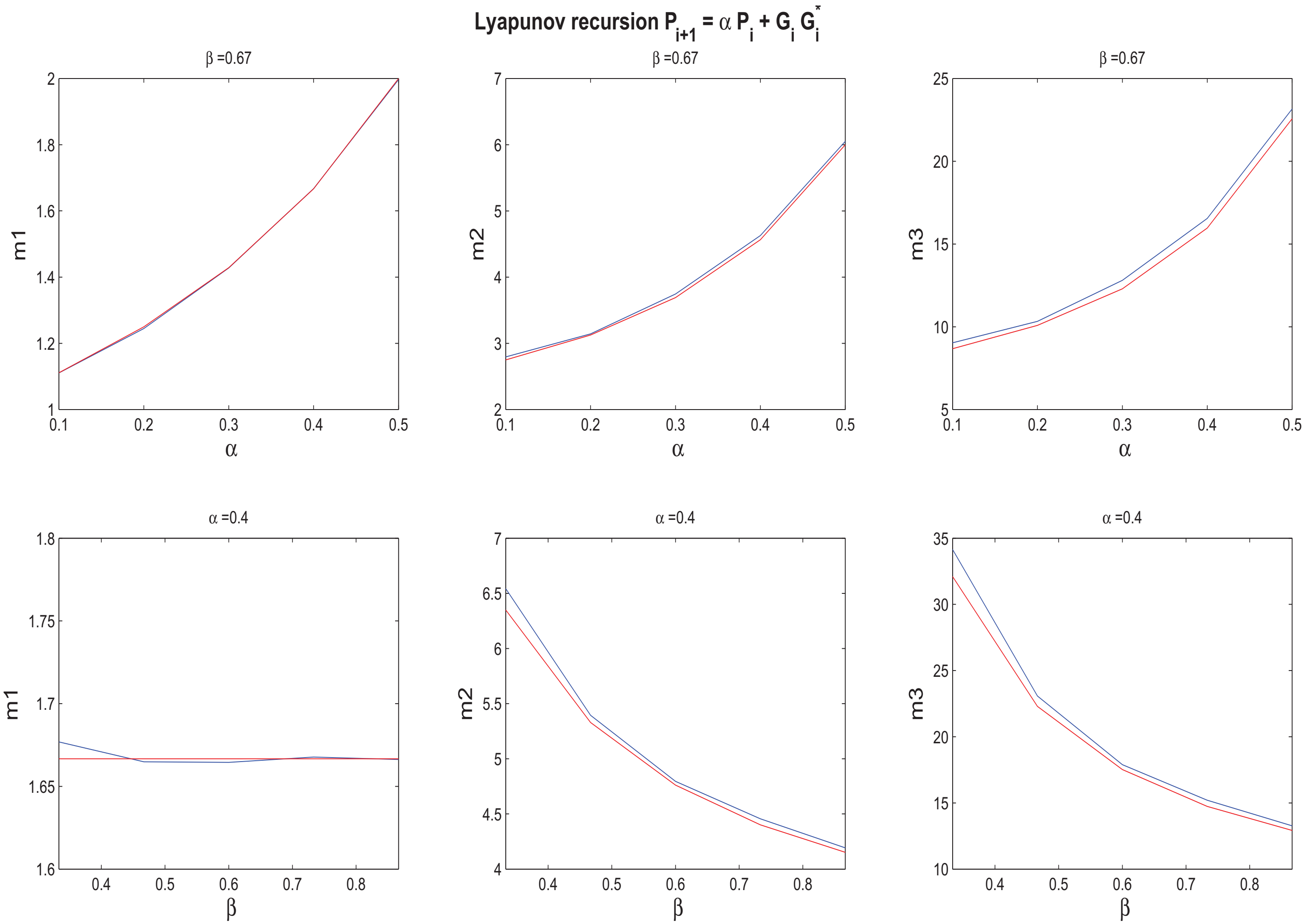


Figure 2: *Empirical moments in blue, theoretical in red ($n = 30$).*

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

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

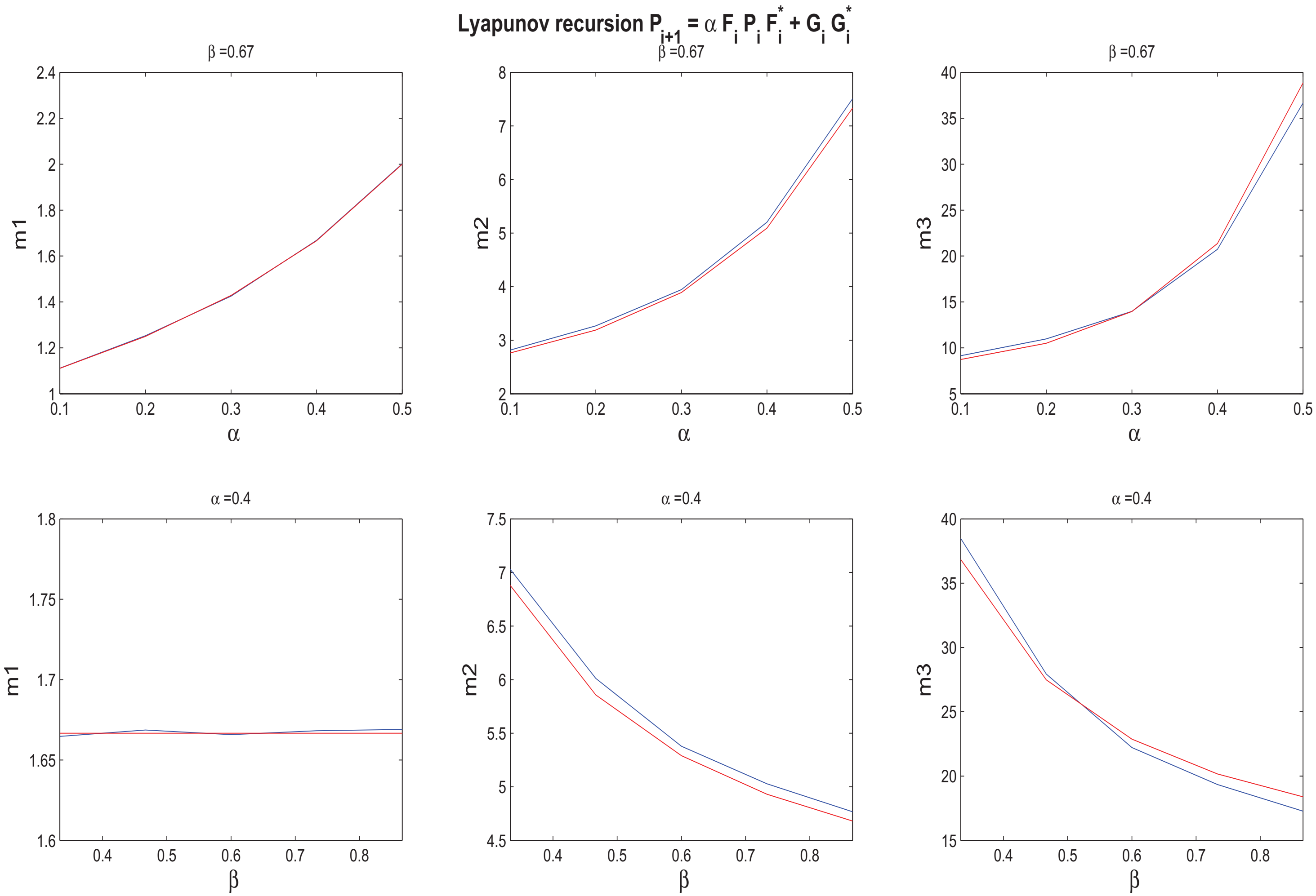


Figure 3: *Empirical moments in blue, theoretical in red* ($n = 30$).

LMS Adaptive Filtering

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

In this case, we obtain

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

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

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

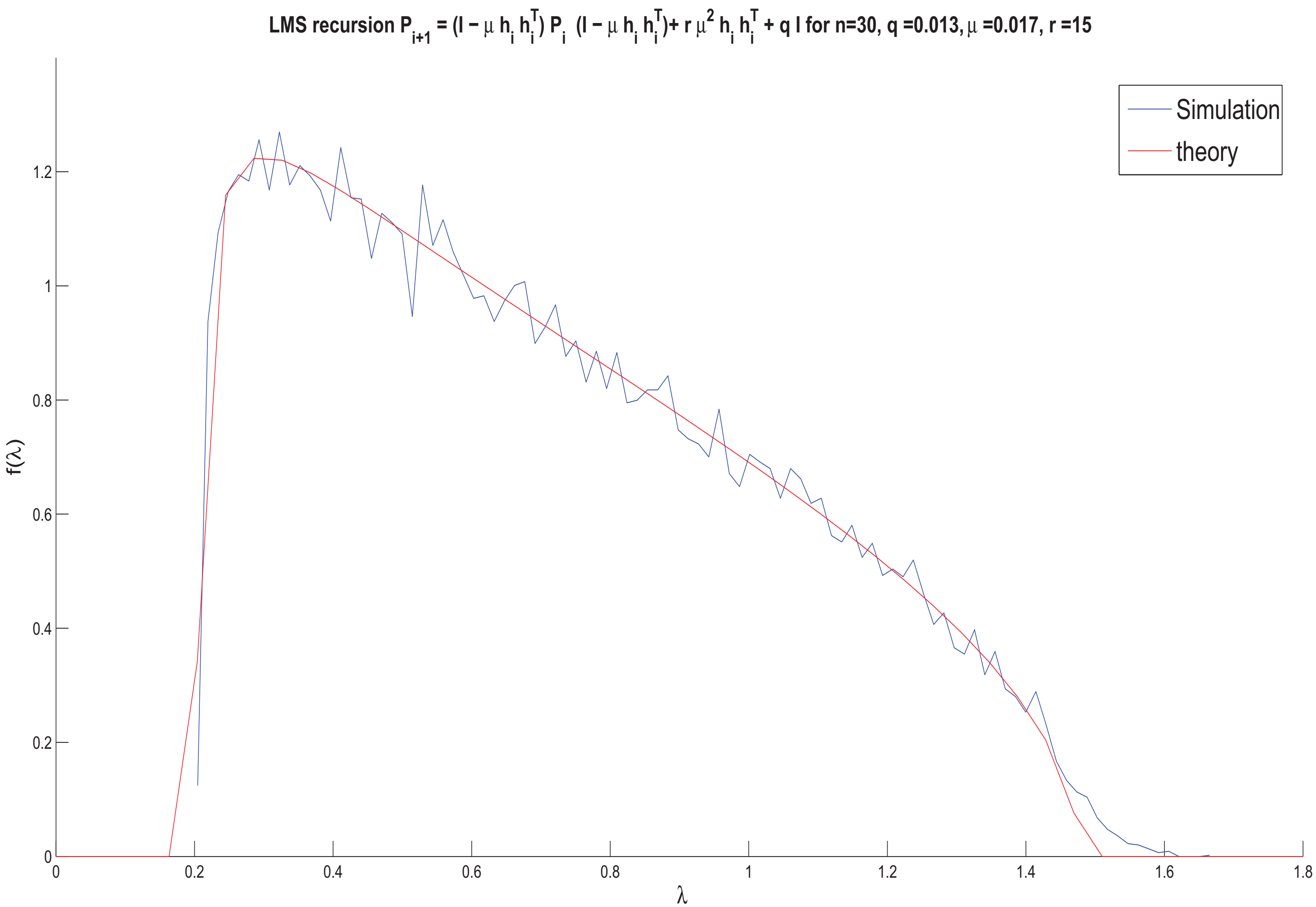


Figure 4: *Eigendistribution for LMS filter ($n = 30$).*

RLS Adaptive Filtering

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

A similar analysis yields

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

where c is a constant given by

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

Thus, c cannot be separately determined. (In fact, determining $m_1 = \frac{1}{n} E \mathrm{tr} P_i$ has been a long standing open problem.) However, we can simply find c by numerically solving (1) and insisting that the inverse Stieltjes transform integrate to one.

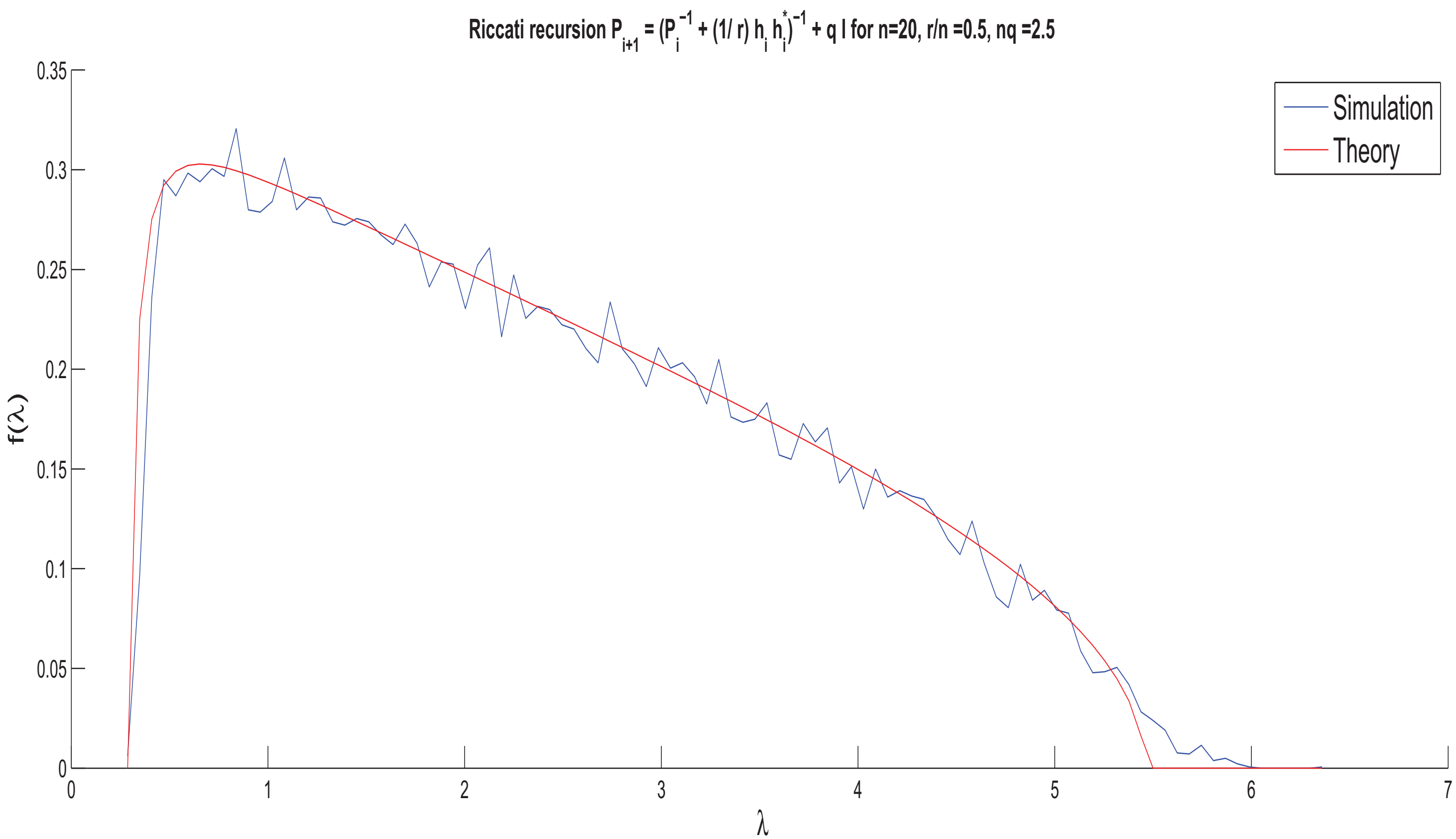


Figure 5: *Eigendistribution for RLS filter* $(n = 30)$.

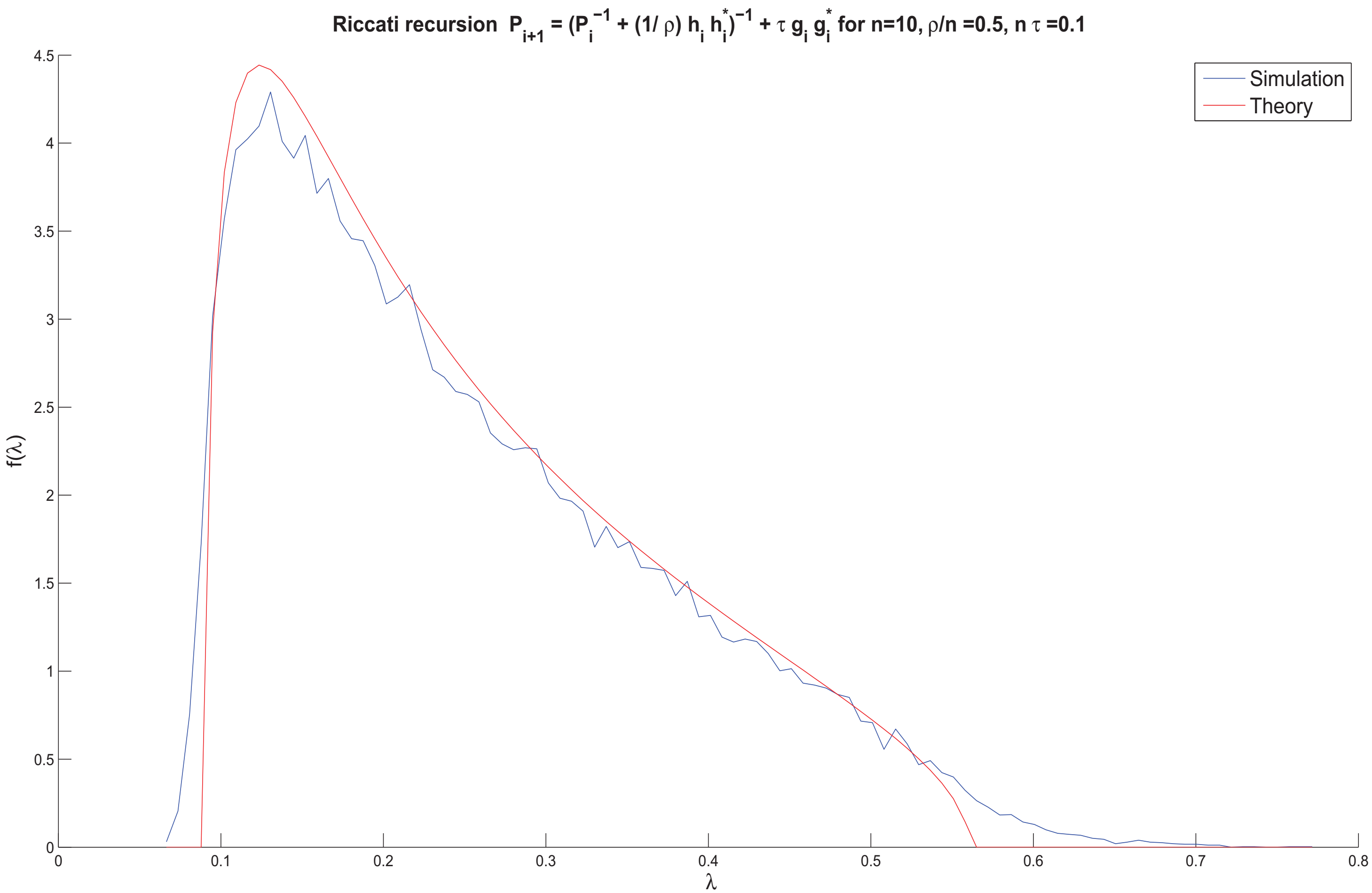


Figure 6: *Eigendistribution for RLS filter* ($n = 10$).

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

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

where the u_i are white.

- However, our results are *universal*, in the sense that they hold even for the above structure.
- The difference is the convergence rate
  - $O(\frac{1}{\sqrt{n}})$, as opposed to $O(\frac{1}{n})$

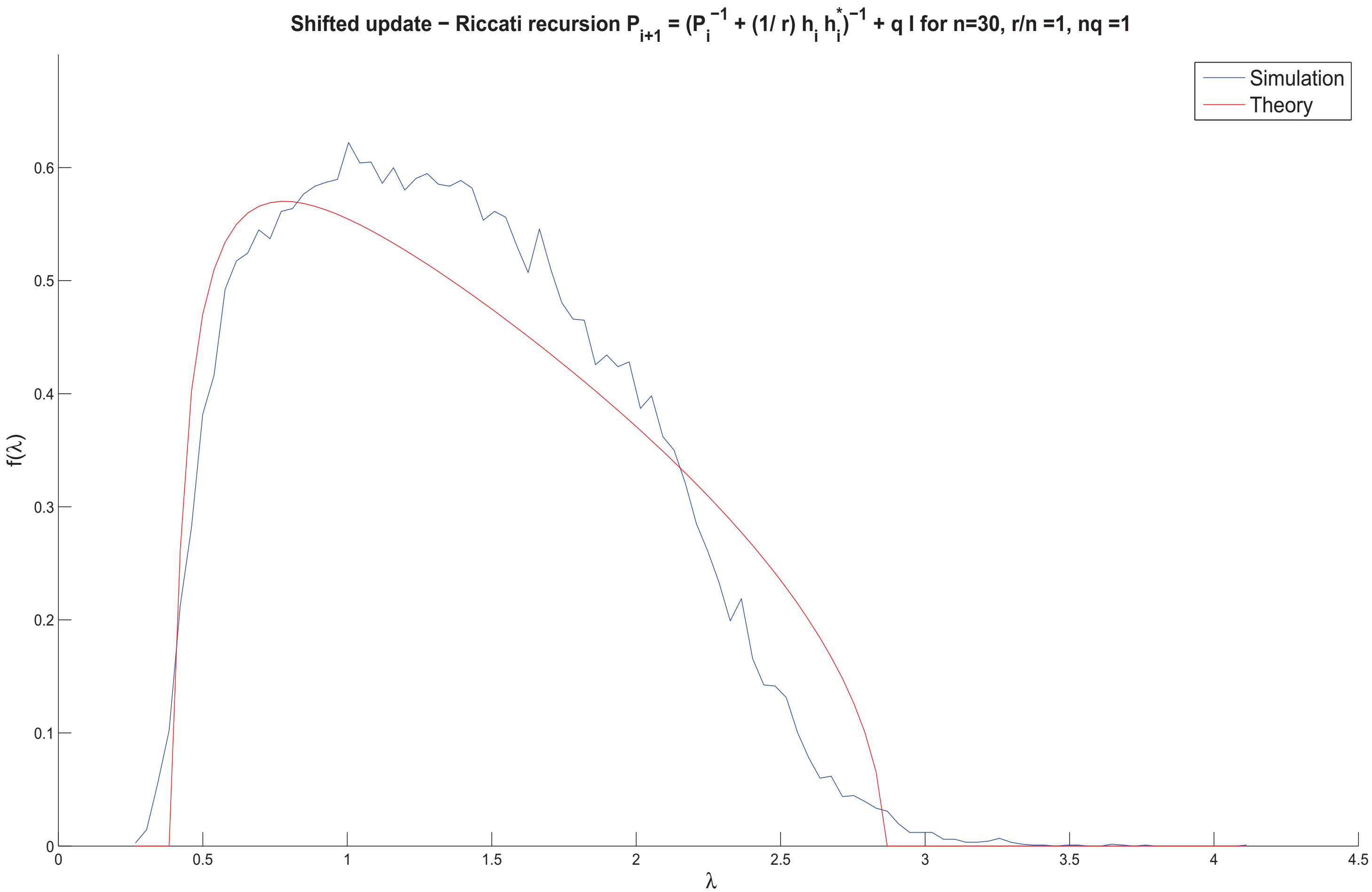


Figure 7: *Eigendistribution for RLS filter with shift-structured regressor*$(n = 30)$.

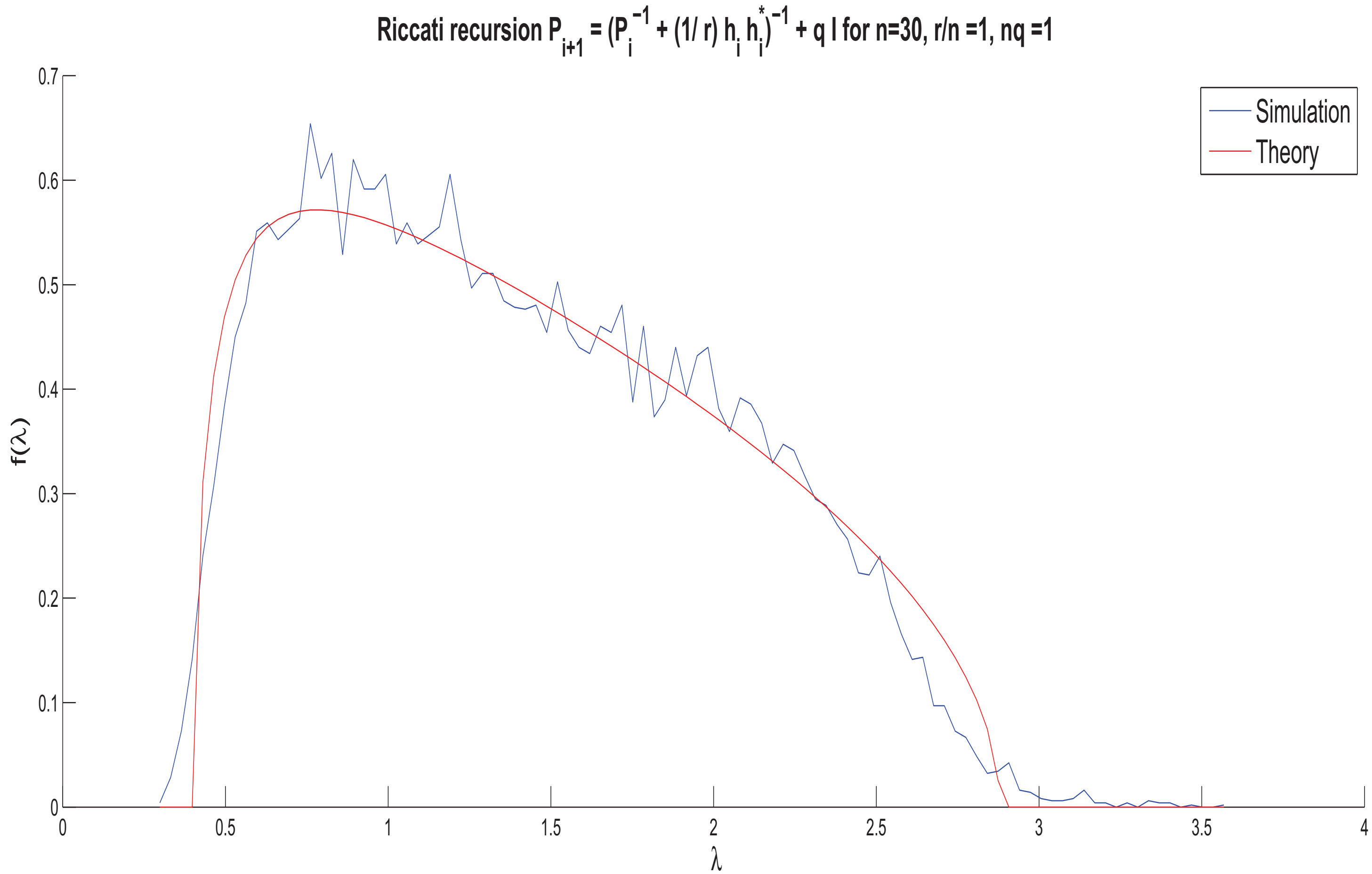


Figure 8: *Eigendistribution for RLS filter with shift-structured regressors and double shifts* $(n = 30)$.

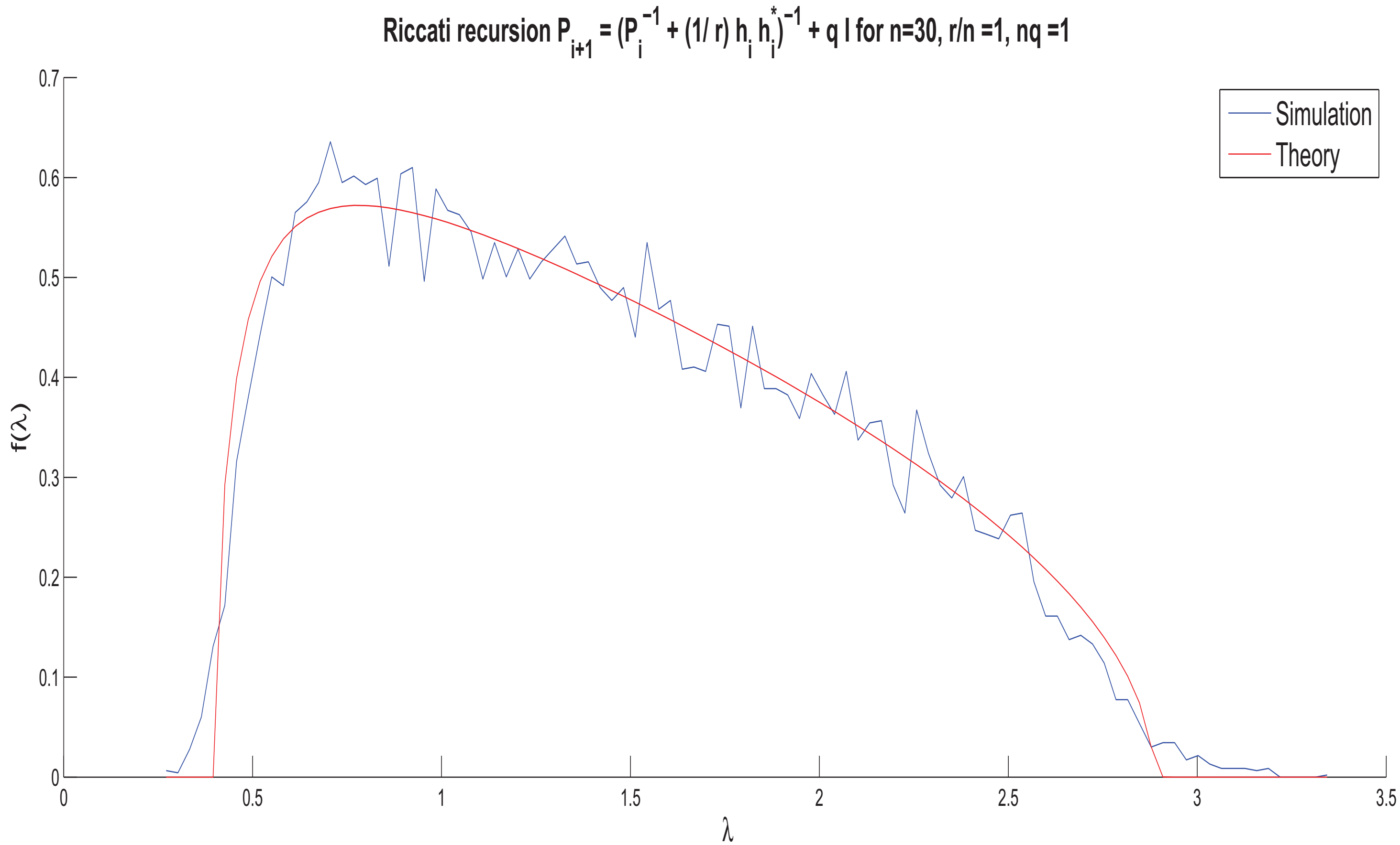


Figure 9: *Eigendistribution for RLS filter with shift-structured regressors and triple shifts ($n = 30$).*

RLS with Multiple Measurements

The results we have given so far assumed that $n \gg 1$, and also generalize to $n \gg m$, where m is the number of measurements.

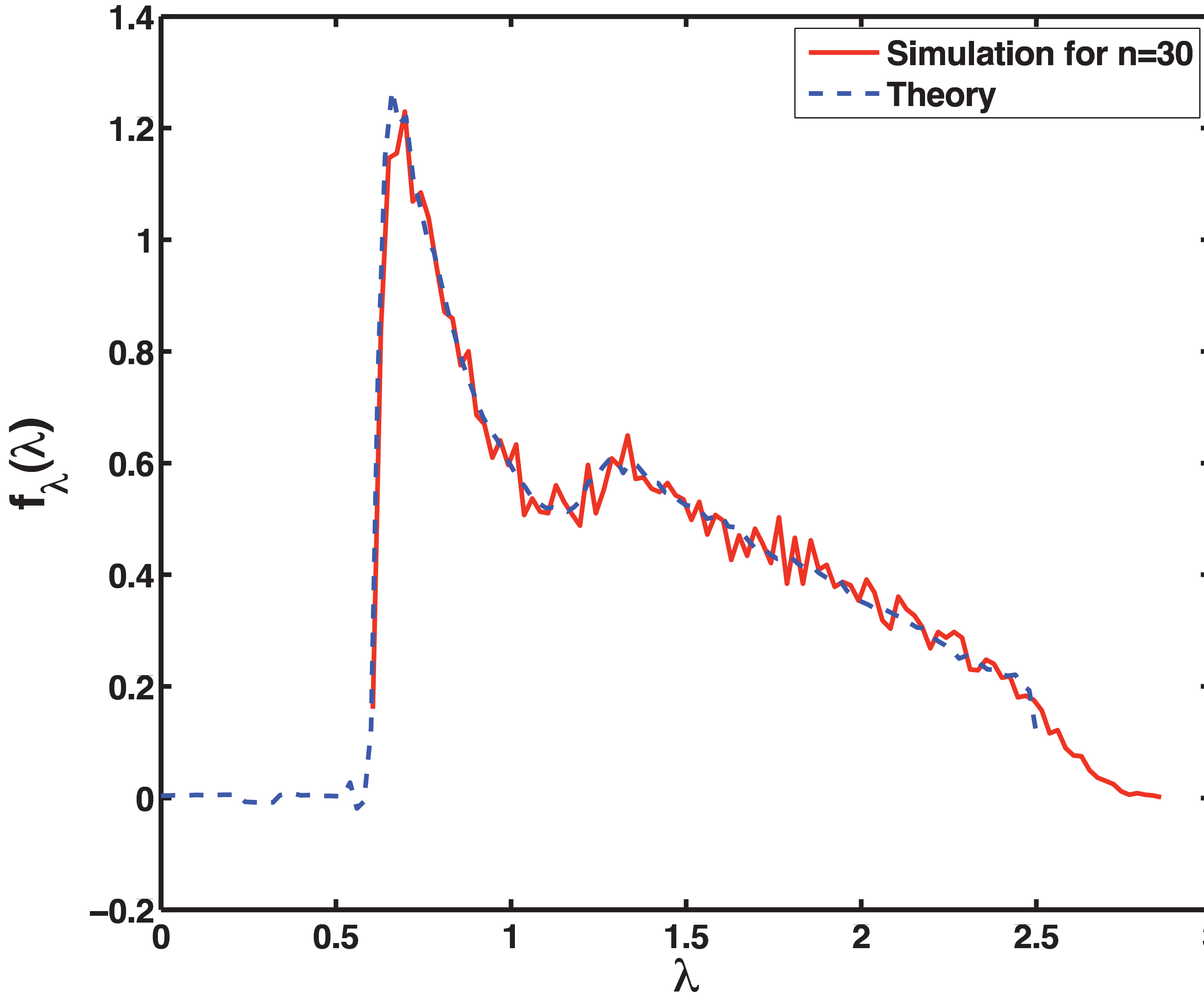
However, if $n \gg 1$ and $\frac{m}{n} = \beta$, a constant, then we need to analyze things anew. Here we have the random Riccati

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

where the H_i are $n \times n$ independent matrices, whose entries are iid zero-mean, $\frac{1}{n}$ variance.

We now obtain

$$\begin{aligned} s(z+q) &= \frac{t(z)}{1 - zt(z)} - \frac{1}{(1 - zt(z))^2} s\left(\frac{z}{1 - zt(z)}\right) \\ t(z) &= \frac{1}{r - \frac{z}{1-zt(z)} + \frac{z^2}{(1-zt(z))^2} s\left(\frac{z}{1-zt(z)}\right)} \end{aligned}$$


Simulation for n=30
Theory
f_λ(λ)
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

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

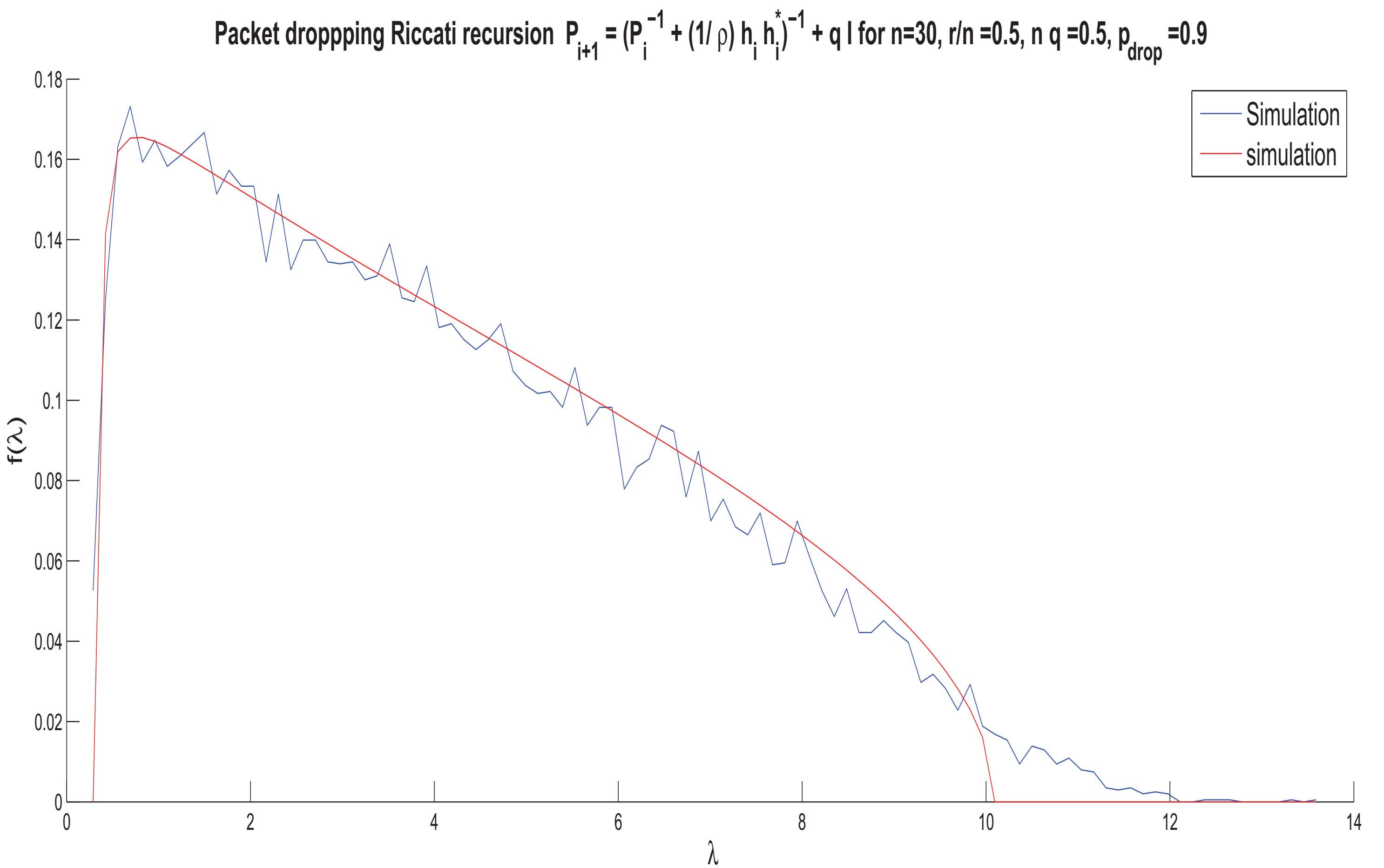


Figure 10: *Eigendistribution for RLS filter with packet drops* ($n = 30$).

Intermittent Observations

We can even consider a more general case, closer to what we started off with in the beginning:

$$P_{i+1} = FP_iF^* - FP_iH_i^*(R_i + H_iP_iH_i^*)^{-1}H_iP_iF^* + qI.$$

Note that the only difference is that the H_i are also random (rather than being fixed).

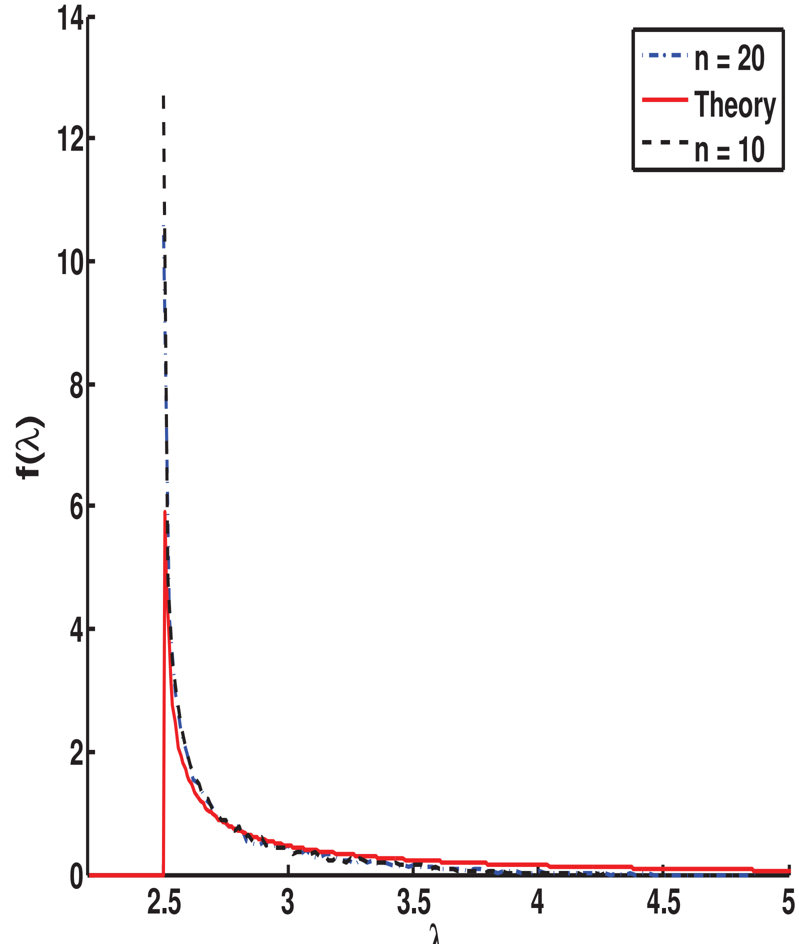


Figure 11: *Eigendistribution for Kalman filter with packet drops* ($n = 10, 20$).

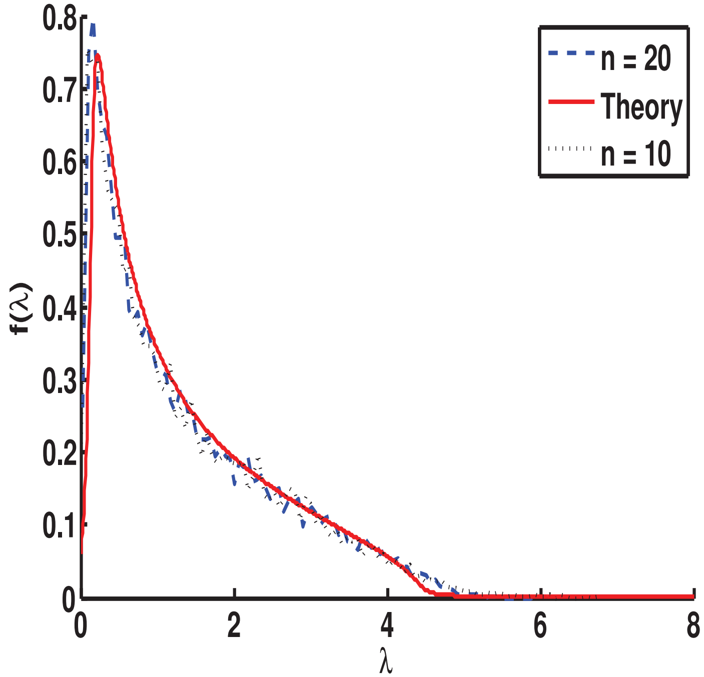


Figure 12: *Eigendistribution for Kalman filter with packet drops* ($n = 10, 20$).

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.

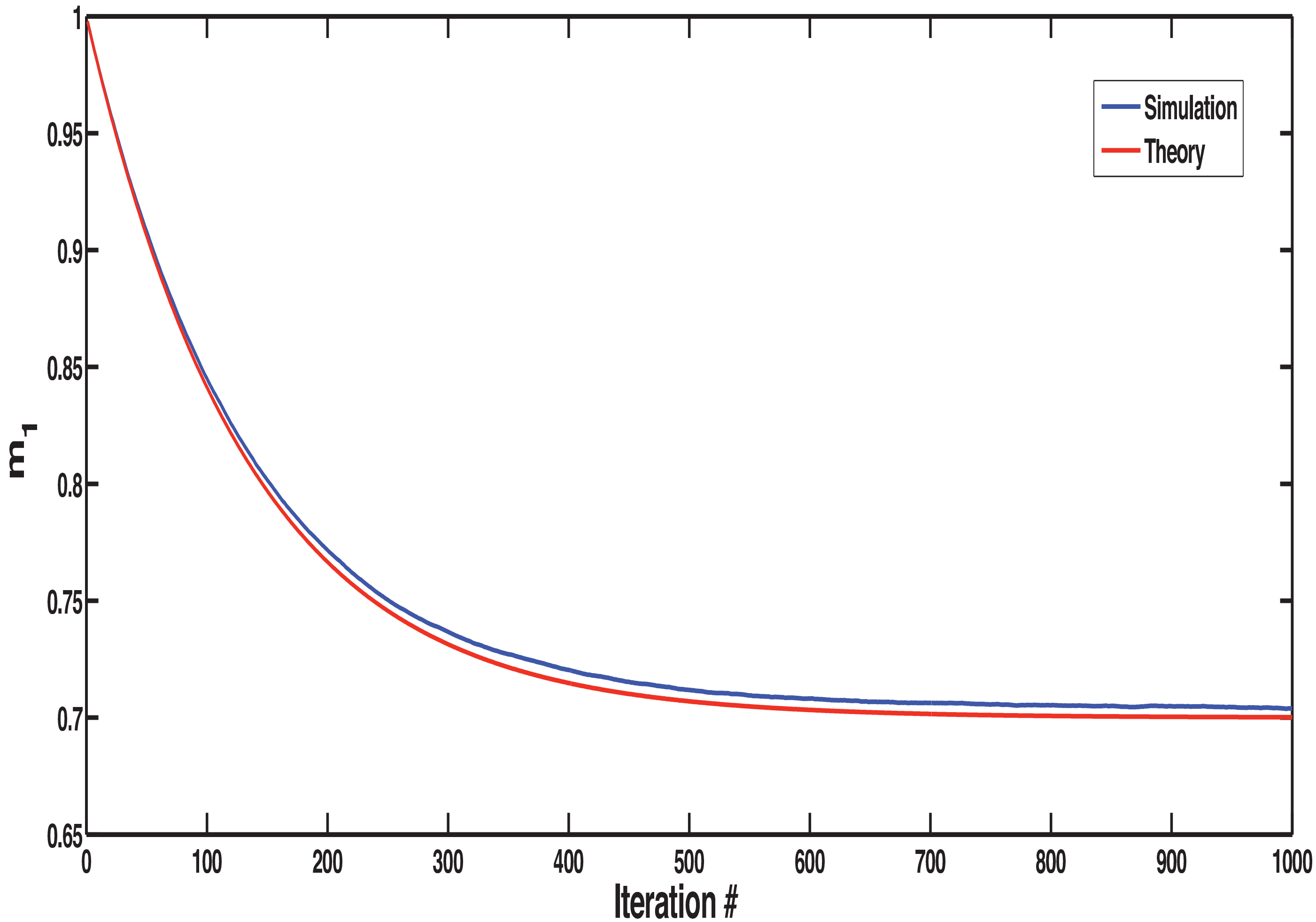


Figure 13: *Transient behavior of m_1 for LMS filter ($n = 30$).*

Comments, Conclusions and Future Work

- The approach we have developed seems quite promising, since we have obtained quite a few new and nontrivial results.
- In many cases of interest we do identify *universal laws.*
- 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 is either constant, 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
  - it is not yet clear how much randomness is required to exhibit universal behavior (is random packet drops enough? do we need many measurements? do we also need random delays?)