To appear in *Computers & Electrical Engineering.*

# A Simple FIR-Filter Interpretation of the Extreme Eigenvalues of a Toeplitz Autocorrelation Matrix

Jeffrey O. Coleman

***Abstract*— The convergence of LMS adaptive algorithms is typically limited by the eigenvalue spread of a Toeplitz autocorrelation matrix with elements from the central portion of an autocorrelation function. If that autocorrelation function describes a random process input to an FIR filter, the ratio of the filter output power to that obtained in response to a unit-power white input varies, as the filter response is changed, across the closed interval from the minimum eigenvalue to the maximum eigenvalue of the autocorrelation matrix. This simple fact permits important relationships between these extreme eigenvalues and the spectrum at the filter input to be understood easily and without resort to the classic asymptotic approximation with a cyclic matrix. In particular, (1) a pure line spectrum with fewer distinct lines than the matrix order leads to a singular matrix; (2) the spectral minimum/maximum is a lower/upper bound on the minimum/maximum eigenvalue; and (3) those bounds are approached asymptotically with increasing matrix order (the classic result). Further, filter-optimization experience may offer the system designer some intuition for the variation of extreme eigenvalues with matrix order and key spectral parameters.**

## I. Introduction

The eigenvalue spread of Toeplitz autocorrelation matrices is of central importance in many statistical signal-processing problems, for example in determining convergence behavior of LMS algorithms for adapting the weights of linear antenna or sensor arrays or for adapting the coefficients of the FIR digital filters common in noise-cancellation and channel-equalization [1] applications. Most commonly the element values of some key Toeplitz autocorrelation matrix $\mathbf{R}$ are sampled from the central part of the autocorrelation function of some random process, $\mathbf{R}_{mn} = R(m-n)$. If the Toeplitz autocorrelation matrix $\mathbf{R}$ is not naturally related to an autocorrelation function in this way, such an autocorrelation function can in fact be constructed [2], so this development applies to all Toeplitz autocorrelation matrices.

It is well-known that as matrix $\mathbf{R}$ becomes larger and larger, so that it includes more and more autocorrelation-function samples, the maximum and minimum eigenvalues of $\mathbf{R}$ asymptotically approach the maximum and minimum of the spectral density $S(f)$ that is the Fourier transform of autocorrelation function $R(k)$. For well over two decades, the usual argument for this [3] has centered around the asymptotic approximation of $\mathbf{R}$ by a cyclic matrix and the cyclic-matrix property that its eigenvalues are the (unitary) DFT coefficients of its first column (or row). Many find this approach awkward however, and it offers little intuition for the behavior of the asymptotic-approximation error as spectral characteristics and matrix order are varied. The alternative view of the extreme eigenvalues presented here allows signal-processing engineers to use their understanding of FIR-filter design both to understand the classic result and to gain insight into the approximation issues so important in practice.

Section II comprises the simple derivation of the relevant mathematical underpinnings and a brief discussion of the relation of the present approach to earlier results, particularly to the family of eigenfilter methods [4] for FIR-filter design. The key insights accorded by the present approach are subsequently developed in Section III through examples.

## II. FIR Filters versus Eigenvalues: the Mathematics

Suppose $N \times N$ complex (or real) matrix $\mathbf{R}$ is Toeplitz, Hermitian, and nonnegative definite, with $\mathbf{R}_{mn} = R(m-n)$ for some sequence $R(k)$ having nonnegative real discrete-time Fourier transform $S(f)$. Further suppose that complex (or real) random process $z(n)$ with stationary autocorrelation function $R(k)$ and spectral density $S(f)$ is used as a test input to a filter with complex (or real) unit-sample response $h(n)$. The average power in filter output $y(n) = z(n) \star h(n)$ is

$$E\left[|y(n)|^2\right] = \sum_j \sum_k h(k)^* E\left[z(n-j) z(n-k)^*\right] h(j).$$

If $h(n)$ is the response of an $N$-tap FIR filter, it can be represented by some $N$ consecutive samples in $N$-vector $\mathbf{h}$. Since the expectation is just $R(k-j)$, this filter output power $P_o(S)$, which results from input spectrum $S(f)$, can be written in terms of autocorrelation matrix $\mathbf{R}$ as quadratic form $P_o(S) = \mathbf{h}^\dagger \mathbf{R} \mathbf{h}$, where $(\cdot)^\dagger$ denotes the Hermitian (conjugate) transpose.

A standard result in linear algebra states that the maximum and minimum eigenvalues of $\mathbf{R}$ are* the maximum and minimum values of $\{\mathbf{h}^\dagger \mathbf{R} \mathbf{h} : \|\mathbf{h}\|^2 = 1\}$. We can easily derive a conveniently normalized version of this key result while relating it to the randomly driven FIR filter.

The author is with the Naval Research Laboratory, Washington DC 20375 (email: jeffc@alum.mit.edu). This work was performed while the author was with Michigan Technological University and was supported by National Science Foundation grant MIP-9409686.

*This is sometimes called Rayleigh's principle [5], and $\mathbf{h}^\dagger \mathbf{R} \mathbf{h} / \|\mathbf{h}\|^2$ is sometimes called Rayleigh's ratio. Rayleigh's principle can also be derived using a Lagrange multiplier [6].

Mercer's theorem expresses $\mathbf{R} = \sum_i \lambda_i \phi_i \phi_i^\dagger$ in terms of the eigenvalues $\{\lambda_i\}$ (real and nonnegative because $\mathbf{R}$ is Hermitian and nonnegative definite) and an orthonormal set $\{\phi_i\}$ of corresponding eigenvectors of $\mathbf{R}$. If the latter set is used as a basis in which to express the filter's unit-sample response vector $\mathbf{h} = \sum_j \alpha_j \phi_j$, the filter output power becomes $P_o(S) = \mathbf{h}^\dagger \mathbf{R} \mathbf{h} = \sum_i \lambda_i (\mathbf{h}^\dagger \phi_i)(\phi_i^\dagger \mathbf{h}) = \sum_i \lambda_i |\alpha_i|^2$. But a white input spectrum $W(f) = a > 0$ results in filter output power $P_o(W) = \mathbf{h}^\dagger (a\mathbf{I})\mathbf{h} = a\|\mathbf{h}\|^2$, where $a$ is just the total input power—call it $P_i(W)$—so $P_o(W)/P_i(W) = \|\mathbf{h}\|^2 = \sum_j |\alpha_j|^2$. A normalized output power can then be written

$$\frac{P_o(S)}{P_o(W)/P_i(W)} = \sum_i \lambda_i \frac{|\alpha_i|^2}{\sum_j |\alpha_j|^2},$$

an arbitrary convex combination of the set of eigenvalues $\{\lambda_i\}$, a combination whose possible values for arbitrary $\mathbf{h} \neq 0$ are therefore precisely those in the closed interval $[\lambda_{\min}, \lambda_{\max}]$, the convex hull of that set of eigenvalues. It follows that

$$\lambda_{\min} = \min \left\{ \frac{P_o(S)}{P_o(W)/P_i(W)} : \mathbf{h} \neq 0 \right\} \qquad (1)$$

$$\lambda_{\max} = \max \left\{ \frac{P_o(S)}{P_o(W)/P_i(W)} : \mathbf{h} \neq 0 \right\}. \qquad (2)$$

The extreme eigenvalues of autocorrelation matrix $\mathbf{R}$ can now be seen to be just the extrema, as the filter response is varied, of the ratio of the test-input power passed by the filter to the fraction of the power that would be passed for a white reference signal. The usual frequency-domain power relationships now lead to

$$\lambda_{\min} = \min \left\{ \frac{\int |H(f)|^2 S(f)\, df}{\int |H(f)|^2\, df} \right\} \qquad (3)$$

$$\lambda_{\max} = \max \left\{ \frac{\int |H(f)|^2 S(f)\, df}{\int |H(f)|^2\, df} \right\}, \qquad (4)$$

where the integrals are over any unit interval of normalized frequency $f$, and where the extremization is over FIR-filter responses $H(f)$ that are not identically zero and with the number of FIR filter taps equal to the autocorrelation-matrix order. It is the interpretation of this simple form that proves to be of such value in understanding the extreme eigenvalues.

Before exploring that interpretation, it is interesting to compare the present approach with earlier results. Makhoul [6] summarized two early uses of test inputs and extremization of the resulting FIR-filter output power: Robinson [7] designed an FIR filter to maximize a signal-to-noise ratio for an optimal detector, and Gueguen [8,9] designed a linear predictor with coefficient vector $\mathbf{h}$ by minimizing error power $\mathbf{h}^\dagger \mathbf{R} \mathbf{h}$ subject to $\|\mathbf{h}\|^2 = 1$. The latter foreshadowed the eigenfilter method of FIR filter design introduced by Vaidyanathan and Nguyen [4] and later generalized by Nguyen [10], Pei and Shyu [11, 12], and others. In many such methods, however, the kernel $\mathbf{A}$ is not Toeplitz and

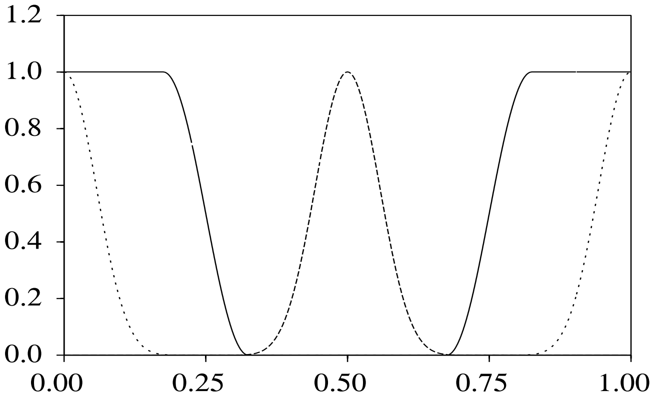


Fig. 1. The solid curve (linear scale versus normalized frequency $f$) is an example spectral density $S(f)$ corresponding to an autocorrelation function from which an $11 \times 11$ Toeplitz autocorrelation matrix $\mathbf{R}$ is sampled. The associated random signal drives an 11-tap FIR filter $H(f)$. The dotted and dashed curves are the $|H(f)|^2$ that respectively maximize and minimize $\int |H(f)|^2 S(f)\, df \div \int |H(f)|^2\, df$ and fix this ratio at the maximum and minimum eigenvalues of $\mathbf{R}$.

the formulation in terms of some test-input process $z(n)$ is not applicable.

Makhoul's paper finally abandoned the power-ratio approach and mathematically explored the eigenstructure of symmetric Toeplitz matrices. Farhang-Boroujeny [13] independently discovered the connection between extreme FIR-filter power gains and extreme eigenvalues, mentioning it in passing before focusing on the addition of linear orthogonality constraints to the extremization to obtain the complete eigenstructure. He later [14] used these ideas to design FIR filter banks for normalization of the LMS adaptive-filter algorithm.

Venkatraman and Coleman [15] bounded output powers of the form $\mathbf{h}^\dagger \mathbf{R} \mathbf{h}$ in response to test inputs in constrained FIR-filter optimization. The eigenfilter work cited earlier steered away from test-input formulations, but Coleman [16, 17] steered it back, developing embedded FIR filter design as a generalized eigenproblem.

In all but [13], which had a different focus, the goal was the FIR filter and the eigenvalues were tools. The next section shows through spectral visualization and example that it can be turned around: filter design is the tool by which, at least in principle, extreme eigenvalue behavior can be understood.

## III. Interpretation

The basic idea in (3) and (4) is most easily understood through an example. Suppose we are interested in the extreme eigenvalues of the $11 \times 11$ autocorrelation matrix $\mathbf{R}$ associated with the 30% raised-cosine spectral density (typical of an adaptive-equalization application) $S(f)$ depicted by the solid (flat-topped) curve of Fig. 1. Imagine a 11-tap FIR filter $H(f)$ driven by a signal of that spectral density, so that the filter output power is $\int |H(f)|^2 S(f)\, df$. According to (3) and (4), the ratio of this power to the area under $|H(f)|^2$ varies between $\lambda_{\min}$ and $\lambda_{\max}$ as $|H(f)|^2$ is varied.

How should $|H(f)|^2$ be chosen to find these extremes?

The (dotted) Fig. 1 curve that peaks at the origin is the $|H(f)|^2$ that maximizes the ratio

$$\frac{\int |H(f)|^2 S(f)\, df}{\int |H(f)|^2\, df}$$

and therefore sets it equal to $\lambda_{\max}$. And the (dashed) curve that peaks at $f = 0.5$ is the $|H(f)|^2$ that minimizes the ratio and therefore sets it equal to $\lambda_{\min}$. In one case, the curve is concentrated under the highest part of the input spectrum $S(f)$. In the other case, it is concentrated in the part of the spectrum where $S(f)$ maintains its lowest value (here zero). Several general properties of the extreme eigenvalues can be easily inferred by thinking about examples of this type.

*A. A line spectrum can cause singularity:*

Suppose $S(f)$ is a line spectrum with spectral lines at precisely $M$ distinct frequencies per unit of normalized frequency. An FIR filter with more than $M$ taps has at least $M$ zeros, and a choice of $|H(f)|^2$ is possible that zeros each spectral line. Therefore, the minimum eigenvalue $\lambda_{\min} = 0$ for autocorrelation matrices larger than $M \times M$. For matrices $M \times M$ or smaller, a choice of $|H(f)|^2$ that zeros the output power is impossible, so $\lambda_{\min} > 0$.

*B. The asymptotic limit:*

As the size of the autocorrelation matrix $\mathbf{R}$, and hence the number of filter taps, is increased, it will be possible for $|H(f)|^2$ to be made narrower and narrower. If the spectral extremes of $S(f)$ occur at isolated frequencies (rather than over intervals as in the example of Fig. 1), these extremizing choices of $|H(f)|^2$ will asymptotically approach impulses, so that $\lambda_{\min}$ will approach the minimum of $S(f)$, and if $S(f)$ is continuous, $\lambda_{\max}$ will approach the maximum of $S(f)$. (For the density $S(f)$ of Fig. 1, the extreme eigenvalues asymptotically approach zero and unity.) The argument is more involved if $S(f)$ has extended minima or maxima, but the outcome is the same. If an impulse is present in the density, $\lambda_{\max}$ approaches infinity, because if a constant area is maintained for an impulse-aligned $|H(f)|^2$ pulse in order to keep the normalization constant, the height of that $|H(f)|^2$ pulse, picked out by the impulse in the integral, will increase without bound as the pulse width decreases.†

*C. How the limit is approached:*

The tails of $|H(f)|^2$ are never precisely zero except at isolated frequencies. To the extent that $|H(f)|^2$ is not an impulse and has significant tails, it de-emphasizes the extreme of $S(f)$ in the integral and picks up a less extreme contribution from the rest of $S(f)$. This means that for finite matrix order, $\min\{S(f)\} \le \lambda_{\min} \le \lambda_{\max} \le \max\{S(f)\}$, or to say it another way, the asymptotic limits are approached from the inside. If the density contains no impulses, these inequalities become equalities only for a white $S(f)$.

*D. Bounds and estimates:*

The extremizing choices of $|H(f)|^2$ can be determined by inspection only in special cases, and actually calculating the extremizing choices amounts to calculating the eigenvectors associated with the extreme eigenvalues. And it is of course the outcome of the eigenvalue calculation that we wish to predict. But since the minimizing (or maximizing) choice of $|H(f)|^2$ yields $\lambda_{\min}$ (or $\lambda_{\max}$), every other choice of $|H(f)|^2$ yields an upper (or lower) bound for $\lambda_{\min}$ (or $\lambda_{\max}$), so that if FIR response $H(f)$ is not identically zero,

$$\begin{aligned} \min\{S(f)\} &\le \lambda_{\min} \\ &\le \frac{\int |H(f)|^2 S(f)\, df}{\int |H(f)|^2\, df} \\ &\le \lambda_{\max} \\ &\le \max\{S(f)\}. \end{aligned}$$

Response $H(f)$ can be chosen to aim at tight bounds on either $\lambda_{\min}$ or $\lambda_{\max}$. Even bounds based on crudely estimated values for the ratio can be quite useful in getting an intuitive feel for $\lambda_{\min}$ and $\lambda_{\max}$.

As an example, again consider the spectral density $S(f)$ of Fig. 1. Intuitively, 11 taps should certainly be enough to permit nearly all of the energy in $|H(f)|^2$ to be placed under the flat top of the raised-cosine spectrum, so it is reasonable to estimate that $\lambda_{\max}$ will be very slightly less than unity. Indeed, the actual $\lambda_{\max} \approx 0.999995$. Estimating $\lambda_{\min}$ is trickier, but even there some feel for the sidelobe levels that can be obtained in simple equiripple designs of FIR filters is enough to get a sense of the order of magnitude of $\lambda_{\min}$ and its behavior as parameters are varied.

For example, most of the energy of $|H(f)|^2$ can be placed in the $S(f) = 0$ region of the spectrum by designing $H(f)$ with linear phase, a one-point unity-gain passband at $f = 0.5$, and an equiripple stopband extending over the $S(f) \neq 0$ region. A stopband level of $-44.5$ dB results. Ignoring many factors, each on the order of unity, in the ratio of integrals and keeping only the key sidelobe level, we might guess that $10^{-4.45} \approx 3.6 \times 10^{-5}$ would give an order-of-magnitude estimate of an upper bound on $\lambda_{\min}$. Indeed, a numerical integral reveals the bound associated with this equiripple $H(f)$ to be $\approx 7.2 \times 10^{-5}$, so the ignored factors amounted cumulatively to a factor of two. Here the actual $\lambda_{\min} \approx 5.2 \times 10^{-6}$, so this simple choice of a not-quite-minimizing $H(f)$ has roughly indicated the magnitude of $\lambda_{\min}$ rather well.

Clearly the purpose of this approach is less to provide precise numerical estimates than to offer the system designer a feel for the behavior of $\lambda_{\min}$ as a function of matrix order and the various spectral parameters such as the raised-cosine excess bandwidth of this example. Neither here requires more than some intuition for the amount of

†It appears that a bit of generalization is possible by replacing "is continuous" with "contains no more than a finite number of discontinuities" and by replacing "minimum" and "maximum" with "infimum" and "supremum" respectively. Restriction to a finite number of impulses may be technically necessary as well. In most applications, however, the density is continuous or continuous with a few added impulses, and such refinements are not needed.

stopband suppression that might be obtained with different numbers of taps and different transition bandwidths in ordinary equiripple design. Based on this notion, an engineer experienced in FIR filter design is likely to recognize instantly, for example, that an order of magnitude drop in $\lambda_{\min}$ would require only a small increase in matrix size,[‡] that a change in excess raised-cosine bandwidth from 30% to 25% would drop $\lambda_{\min}$ modestly,[§] but that a halving of the horizontal scale of the entire raised cosine, corresponding to a doubling of sample rate, would be expected to lower $\lambda_{\min}$ by many orders of magnitude.[¶]

## IV. Summary

If an $N$-tap FIR filter is driven by a random process, the spectral density of that process determines its autocorrelation function, which in turn determines the $N \times N$ Toeplitz autocorrelation matrix of the random vector whose elements are linearly combined in the filter. Or given an $N \times N$ Toeplitz autocorrelation matrix at the outset, a power spectrum and filter such that these relationships hold can be found. If the filter frequency response is chosen to maximize (minimize) the ratio of the filter output power to that which would result with a unit-power white input, that ratio is the maximum (minimum) eigenvalue of the autocorrelation matrix. This fact leads immediately to: (1) a pure line spectrum with fewer distinct lines than the matrix order leads to a singular autocorrelation matrix; (2) the spectral minimum (maximum) is a lower (upper) bound on the minimum (maximum) eigenvalue; and (3) those bounds are approached asymptotically with increasing matrix order (a classic result). It is often relatively simple to intuitively visualize the nature of a filter response that will come close to extremizing the power ratio. This often leads directly to an estimate of the extreme eigenvalues whose variation with matrix order and key spectral parameters are intuitively apparent.

## References


[1] Edward A. Lee and David G. Messerschmitt. *Digital Communication.* Kluwer Academic Publishers, Boston, 1988. Section 9.1.3.

[2] Boaz Porat. *Digital Processing of Random Signals.* Prentice-Hall, 1994. Chap. 6, page 171.

[3] Robert Gray. The asymptotic eigenvalue distribution of toeplitz matrices. *IEEE Transactions on Information Theory*, 18:725–730, April 1972.

[4] P . P . Vaidyanathan and Truong Q. Nguyen. Eigenfilters: A new approach to least-squares FIR filter design and applications including Nyquist filters. *IEEE Transactions on Circuits and Systems*, CAS-34(1):11–23, January 1987.

[5] B. Nobel and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall, 1977.

[6] John Makhoul. On the eigenvectors of symmetric toeplitz matrices. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(4):868–872, August 1981.

[7] E. Robinson. *Statistical Communication and Detection.* Griffin, London, 1967.

[8] C. Gueguen. The modified linear prediction: A factor analysis approach to speech analysis. Technical Report Rep. UCLA-ENG 7540, Dept. Syst. Sci., Univ. California, Los Angeles, 1975.

[9] G. Carayannis and C. Gueguen. The factorial linear modelling: A karhunen-loeve approach to speech analysis. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP '76)*, pages 489–492, April 1976.

[10] Truong Q. Nguyen. The design of arbitrary FIR digital filters using the eigenfilter method. *IEEE Transactions on Signal Processing*, 41(3):1128–1139, March 1993.

[11] S.-C. Pei and J.-J. Shyu. Complex eigenfilter design of arbitrary complex coefficient FIR digital filters. *IEEE Trans. on Circuits and Systems II*, CAS-40(1):32–39, January 1993.

[12] S.-C. Pei and J.-J. Shyu. A unified approach to the design of quadrantally symmetric linear-phase two-dimensional FIR digital filters by eigenfilter approach. *IEEE Trans. on Circuits and Systems II*, CAS-42(10):2886–2890, October 1994.

[13] B. Farhang-Boroujeny. On eigenvalues distribution of correlation matrices. In *Proc. IEEE Int'l Symposium on Circuits and Systems*, pages 1251–1254, June 1991.

[14] B. Farhang-Boroujeny and S. Gazor. Selection of orthonormal transforms for improving the performance of the transform domain normalised LMS algorithm. *IEE Proceedings, Part F*, 139(5):327–335, October 1992.

[15] Rajesh R. Venkatraman and Jeffrey O. Coleman. A test-waveform view of the quadratic design of FIR filters. In *Proc. the 7th Int'l Conf. on Signal Processing Applications and Technology (ICSPAT '96)*, October 1996.

[16] J. O. Coleman. Embedded FIR generalized eigenfilters using test inputs. In *Proc. 1998 IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP '98)*, May 1998.

[17] J. O. Coleman. Quadratic FIR-filter design as a generalized eigenproblem. In *Proc. 1998 Eureopean Signal Processing Conf. (EUSIPCO '98)*, September 1998. accepted.


‡ From $11 \times 11$ to $13 \times 13$ in fact.

§ It roughly halves it.

¶ Over six orders of magnitude, actually.