QUADRATIC–INVERSE EXPANSION OF THE RIHACZEK DISTRIBUTION

David J. Thomson

Queen's University, Kingston Ontario,
djt@mast.queensu.ca (613) 533-2426

ABSTRACT

This paper describes some unexpected relationships between multitaper estimates of the spectrum and time–frequency distributions. In particular, there is an orthogonal decomposition of a localized Rihaczek distribution in terms of quadratic–inverse expansions. The frequency marginal of this estimate is the standard multitaper estimate and thus much more accurate than conventional forms that have the periodogram as the frequency marginal. The time marginal is a smoothed version of the instantaneous power. The first three terms of the quadratic–inverse expansion are approximately the multitaper estimate and the quadratic–inverse estimates of the time and frequency derivatives of the spectrum.

1. INTRODUCTION

During the last 25 years non-parametric spectrum estimates have been evolving in two parallel branches, windowed Fourier transforms and time–frequency methods. These two branches differ in outlook with, for example, the "windowed Fourier transformers" largely concerned with statistical properties while the "time-frequency" group is concerned with properties of a given signal and reconstruction. There are, naturally, many common ideas and techniques but, to a large extent, proponents of the different branches "drink in different bars." The purpose of this paper is to demonstrate some relations between multitaper estimates and time–frequency distributions and specifically, to write the Rihaczek distribution in terms of quadratic–inverse expansions. This approach gives bounds on achievable accuracy.

I assume a finite sample, $x(t)$, $t = 0, 1, \ldots, N-1$ with Fourier transform

$$y(f) = \sum_{t=0}^{N-1} x(t) e^{-i2\pi ft} \tag{1}$$

The process $x(t)$ is assumed to have a spectral representation valid for harmonizable processes (and hence all the ones encountered in practice)

$$x(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\xi t} dX(\xi) \tag{2}$$

For a zero–mean stationary process one has $\mathbf{E}\{dX(f)\} = 0$ and the spectrum is *defined* by

$$S(f)df = \mathbf{E}\{|dX(f)|^2\} \, . \tag{3}$$

Combining (1) and (2), one has

$$y(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Phi(f-\xi) \frac{\sin N\pi(f-\xi)}{\sin \pi(f-\xi)} dX(\xi) \, , \tag{4}$$

Thanks to the Canada Research Chairs program for funding.

the fundamental equation of spectrum estimation. Here $\Phi(f) = \exp\{i\pi f(N-1)\}$ is a phase factor reflecting the choice of origin at 0 instead of the center of the observation interval.

2. NOTATION AND BACKGROUND

As above, I assume W is the analysis bandwidth width with $0 < W < \frac{1}{2}$, and that N is the number of data samples. The Slepian sequences are solutions of the symmetric Toeplitz matrix eigenvalue problem

$$\lambda_k \, v_n^{(k)} = \sum_{m=0}^{N-1} \frac{\sin 2\pi W(n-m)}{\pi(n-m)} \, v_m^{(k)} \tag{5}$$

standardized to be real, positive at the center of the range for even sequences, and to have positive central slope for odd sequences with, as usual

$$\sum_{n=0}^{N-1} v_n^{(j)} \, v_n^{(k)} = \delta_{j,k} \tag{6}$$

and are ordered by their eigenvalues

$$1 > \lambda_0 > \lambda_1 > \cdots > \lambda_{N-1} > 0 \tag{7}$$

with the first $K \approx 2NW$ eigenvalues approximately 1. There are two useful definitions of the Slepian functions. Slepian, [1] defined the discrete prolate spheroidal wave functions as

$$U_k(N, W; f) = \epsilon_k \sum_{n=0}^{N-1} v_n^{(k)} e^{i2\pi f(n - \frac{N-1}{2})} \tag{8}$$

where ϵ_k is 1 for k even and i for k odd. They satisfy the integral equation

$$\lambda_k U_k(f) = \int_{-W}^{W} \frac{\sin N\pi(f-f')}{\sin \pi(f-f')} U_k(f') df' \tag{9}$$

and the orthogonality conditions are

$$\int_{-W}^{W} U_j(f) U_k(f) df = \lambda_j \delta_{j,k} \tag{10}$$

and

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} U_j(f) U_k(f) df = \delta_{j,k} \, . \tag{11}$$

Computationally, I define the Slepian functions as

$$V_k(f) = \sum_{n=0}^{N-1} v_n^{(k)} e^{-i2\pi nf} \tag{12}$$

The V's (or $V_k(N,W;f)$ in explicit notation) are thus complex functions of frequency so, by elementary properties of the Fourier transform, $V_k^*(f) = V_k(-f)$. Comparing (12) and (8) , one has

$$U_k(f) = \epsilon_k e^{-i2\pi f(\frac{N-1}{2})} V_k(-f) \; . \tag{13}$$

The U_k's have the advantage of being real–valued whereas the V_k's are easily computed by standard FFTs.

A useful identity similar to that noted in [2] is

$$e^{-i2\pi ft} = \sum_{j=0}^{N-1} v_t^{(j)} V_j(f) \tag{14}$$

valid for $|f| \leq W$ and $0 \leq n \leq N-1$.

2.1. Multitaper estimates

In [3] I introduced the idea of multitaper spectrum estimates The basic idea was to regard the fundamental equation (4) as an *integral equation*, as opposed to a convolution, and attempt to solve it. This was done in a frequency band $(f-W, f+W)$ where W is a preselected bandwidth. Originally, the process was assumed to be stationary and the spectrum approximately flat or "locally white'' over $(f-W, f+W)$. This estimate was shown to be more efficient than other nonparametric spectrum estimates [3, 4] and approximately maximum–likelihood [5].

The locally white condition was eliminated in [6] with the introduction of quadratic inverse methods for stationary processes. Quadratic inverse theory was extended to nonstationary methods in [7] but, again, with the locally white condition. A major idea in multitaper analysis was to find an approximate solution of the fundamental integral equation (4) in a series expansion of Slepian functions. This, in common with many wavelet methods, gives an overcomplete basis. If one denotes the starting time of the block by t_o and the center frequency by f_o, this results in the eigencoefficients

$$y_k(t_o, f_o) = \sum_{t=0}^{N-1} x(t_o+t)\, v_t^{(k)} e^{-i2\pi f_o t} \tag{15}$$

This gives the approximate solutions in the band $(f_o - W, f_o + W) \times [t_o, t_o + N - 1]$

$$\hat{x}(t_o+t, f_o) = \sum_{k=0}^{K-1} y_k(t_o, f_o)\, v_k^{(t)} \tag{16}$$

and

$$\widehat{dZ}(t_o, f_o+f) = \sum_{k=0}^{K-1} y_k(t_o, f_o)\, \mathcal{V}_k(f) df \tag{17}$$

where $K \approx 2NW$ is the dimension of the space. In most of these estimates one averages over both t_o and f_o but, in the remainder of this paper we will drop explicit dependence on t_o and f_o in the y_k's and concentrate on the local properties. We also denote the vector of eigencoefficients by $\mathbf{Y}(f)$,

$$\mathbf{Y}(f) = [y_0, y_1, \ldots, y_{K-1}]^T \tag{18}$$

so the simplest multitaper estimate of the spectrum is

$$\hat{S}(f) = \frac{1}{K} ||\mathbf{Y}(f)||^2 = \frac{1}{K} \sum_{k=0}^{K-1} |y_k|^2 \tag{19}$$

Note that the "raw" eigencoefficients are rarely used directly with the adaptively–weighted [3] form being most common and the more difficult coherent sideband subtraction [8] used less in general, but ubitiquously for reshaping around line components.

3. STATIONARY QUADRATIC-INVERSE FUNCTIONS

Denote by $B_l(N,W,K;\xi)$, hereinafter written $B_l(\xi)$, the real eigenfunctions of the degenerate integral equation

$$g_l B_l(\xi) = 2W \int_{-W}^{W} \left| \frac{\sin N\pi(f-\eta)}{\sin \pi(f-\eta)} \right|^2 B_l(\eta) d\eta \tag{20}$$

ordered by the eigenvalues $g_0 \geq g_1 \geq \ldots \geq g_{2K-1} \geq 0$ and normalized so that

$$\frac{1}{2W} \int_{-W}^{W} B_l(f) B_m(f) df = \delta_{l,m}. \tag{21}$$

Note that the kernel of (20) is the *square* of the kernel in (9) and the normalization factor $2W$ in (20) is included to standardize the trace at $(2NW)^2$. The elements of the $K \times K$ complex matrix $\mathbf{B}^{(l)}$ are given by

$$B_{jk}^{(l)} = \frac{1}{\sqrt{\lambda_j \lambda_k}} \int_{-W}^{W} V_j(\xi)\, V_k^\star(\xi) B_l(\xi) d\xi \; . \tag{22}$$

Rewriting this equation in terms of the real–valued Slepian functions, U_k, (13) gives

$$B_{jk}^{(l)} = \frac{\epsilon_j^* \epsilon_k}{\sqrt{\lambda_j \lambda_k}} \int_{-W}^{W} U_j(\xi) U_k(\xi) B_l(\xi) d\xi \; . \tag{23}$$

so the $\mathbf{B}^{(l)}$'s are real symmetric for l even and imaginary skew-symmetric for l odd.

These matrices have the remarkable property

$$tr\{\mathbf{B}^{(l)} \mathbf{B}^{(m)}\} = g_l \delta_{l,m} \tag{24}$$

where tr denotes trace. This trace-orthogonality is a general feature of product kernels and may be verified by using (22) to expand both matrices in (24), identify both sums as the kernel and the results follow. (Proofs may be simplified by truncating the kernels to their $K \times K$ projection equivalents.)

In stationary quadratic–inverse theory the spectrum is expanded within the inner bandwidth, $|\xi| < W$, as

$$\hat{S}_{sqi}(f_o + \xi) = \sum_{l=0} \hat{b}_l(f_o) B_l(\xi) \tag{25}$$

where the coefficients are estimated by the quadratic forms

$$\hat{b}_l(f_o) = \frac{1}{g_l} \; \mathbf{X}^\dagger \mathbf{B}^{(l)} \mathbf{X} \; . \tag{26}$$

Under the locally–white conditions the variance of $\hat{b}_p(f)$ is $S^2(f)/g_p$ so, because $g_l \asymp 2NW - p/2$, the variance of $\hat{b}_0$ is the same as that of the usual multitaper spectrum estimate while that of the "first derivative " term, $\hat{b}_1(f)$ is only slightly larger.

4. NONSTATIONARY QUADRATIC-INVERSE FUNCTIONS

The basis functions for the non-stationary quadratic inverse theory are solutions of the algebraic eigenvalue equation

$$\alpha_l A_l(n) = N \sum_{m=0}^{N-1} \left[\frac{\sin 2\pi W(n-m)}{\pi(n-m)} \right]^2 A_l(m) \qquad (27)$$

standardized to be orthonormal

$$\sum_{t=0}^{N-1} A_l(t) A_m(t) = \delta_{l,m} \ . \qquad (28)$$

The leading factor of N in (27) standardizes the trace to $(2NW)^2$. Again, the kernel of (27) is the square of that in (5). The basis matrices

$$A^{(l)}_{jk} = \sqrt{\lambda_j \lambda_k} \sum_{n=0}^{N-1} v_n^{(j)} v_n^{(k)} A_l(n) \qquad (29)$$

are real, symmetric, and trace-orthogonal,

$$tr\{\mathbf{A}^{(l)} \mathbf{A}^{(m)}\} = \alpha_l \delta_{l,m}. \qquad (30)$$

Again, there are $4NW$ such functions corresponding to significantly non-zero eigenvalues so the time resolution is limited to essentially $1/4W$, half the Nyquist spacing.

Parallelling the stationary case, in the nonstationary case the A's are used to estimate the evolution of power within a block, that is

$$\widehat{S}_{nsqi}(t_o + t, f) = \sum_{l=0} \hat{a}_l(t_o, f_o)\ A_l(t) \qquad (31)$$

again with the coefficients estimated by the quadratic forms

$$\hat{a}_l(t_o, f_o) = \frac{1}{\alpha_l}\ \mathbf{X}^\dagger \mathbf{A}^{(\mathbf{l})} \mathbf{X} \qquad (32)$$

As in the stationary case, the eigenvalues $\alpha_p \asymp 2NW - p/2$. Because the Slepian sequences and functions closely approximate the continuous–time prolate spheroidal wave functions, and these are their own Fourier transforms, $v^{(k)}_{\frac{N}{2}(1+x)} \sim V_k(Wx)$ for $-1 \leq x \leq 1$ and, similarly, $A_l \sim B_l$ so

$$\mathbf{B}^{(l)}_{j,k} \sim \epsilon_j^* \epsilon_k \mathbf{A}^{(l)}_{j,k} \ . \qquad (33)$$

This demonstrates an important difference between the stationary and non-stationary cases; the $\mathbf{A}^{(l)}$'s are symmetric, the $\mathbf{B}^{(l)}$'s Hermitian. In particular $tr\{\mathbf{A^{(1)}B^{(1)}}\} = 0$ so one can form uncorrelated estimates of the spectrum, it's frequency derivative $b_1(f) \sim \frac{\partial}{\partial f} S$, and it's time derivative $a_1(f) \sim \frac{\partial}{\partial t} S$. The problem has been how to combine the stationary and nonstationary forms in general.

5. THE RIHACZEK DISTRIBUTION

The Rihaczek distribution, [9], is one of numerous interrelated time–frequency distributions, see *e.g.* [10, 11, 12]. It is generally defined as

$$\mathcal{R}(t, f) = x(t) y^*(f) e^{-i2\pi ft} \qquad (34)$$

where $y(t)$ is the Fourier transform defined in (1) and superscript "*" indicates complex conjugate. This distribution was apparently introduced in [13] and appeared in [14] but Rihaczek described the physical motivation for using it in engineering applications. The marginal distributions are $|x(t)|^2$ and $|y(f)|^2$. One notes that the frequency marginal is, except for a factor of $1/N$, just the periodogram and consequently horribly biased [8].

6. LOCAL RIHACZEK ESTIMATES

In the remainder of this paper we will restrict our attention to the area $[0 \leq t \leq N-1] \times [-W \leq f \leq W)$, work with the sequence $z(t)$ and its Fourier transform $Z(f)$ defined by

$$z(t) = \sum_{k=0}^{K-1} y_k\, v_k^{(t)} \qquad (35)$$

and

$$Z(f) = \sum_{k=0}^{K-1} y_k\, V_k(f) \qquad (36)$$

so the *local* Rihaczek distribution becomes

$$\mathcal{R}_L(t, f) = \frac{1}{N} \sum_{j,k=0}^{K-1} y_j y_k^*\, v_j^{(t)}\, V_k^\star(f) e^{-i2\pi ft} \qquad (37)$$

where the normalization factor $1/N$ has been included to standardize the frequency marginal to have the same units as a spectrum. We now expand $\mathcal{R}_L(t, f)$ in a product form of the quadratic–inverse functions,

$$\mathcal{R}_L(t, f) = \sum_{l,m=0} \mathcal{R}_{l,m} A_l(t)\ B_m(f) \qquad (38)$$

where the A's and B's are the basis functions defined in (27) and (20), respectively, and $\mathcal{R}_{l,m}$ depends implicitly on t_o and f_o. Now replace the $\exp\{-i2\pi ft\}$ in (37) with the expansion (14), multiply by $A_l(t)\ B_m(f)$ and integrate and sum over the region. Identifying the sum in t with (29) and the integral in f with (22), one obtains

$$\mathcal{R}_{l,m} = \frac{1}{2NW} \sum_{j,k,p=0}^{K-1} y_j y_k^* \mathbf{A}^{(l)}_{j,p} \mathbf{B}^{(m)}_{p,k} \qquad (39)$$

or

$$\mathcal{R}_{l,m} = \frac{1}{2NW} \mathbf{X}^\dagger \mathbf{A^{(l)} B^{(m)} X} \ , \qquad (40)$$

an equation of the same general form as the stationary and nonstationary quadratic–inverse estimates.

The inverse relation may be found by rewriting (39) as

$$\frac{1}{N} \sum_{j,k=0}^{K-1} y_j y_k^*\, v_j^{(t)}\, V_k^\star(f) = e^{i2\pi ft} \sum_{l,m=0} \mathcal{R}_{l,m} A_l(t)\ B_m(f) \qquad (41)$$

and, as before, using (14). Now multiply by the two Slepian functions, integrate, and sum to get

$$y_j y_k^* = \sum_{l,m=0} \mathcal{R}_{l,m} (\mathbf{A^{(l)} B^{(m)}})_{j,k} \qquad (42)$$

so, as with other time–frequency distributions, the local signal can be recovered. Equations (40) and (42) jointly imply

$$tr\{\mathbf{A^{(l)} B^{(m)} A^{(p)} B^{(q)}}\} = 2NW \delta_{l,p} \delta_{m,q} \qquad (43)$$

similar in ways to the two–dimensional orthogonality of the "induced basis" discussed in [15, 14] in connection with expansions of the ambiguity function on an orthogonal basis.

7. DISCUSSION

The preceding derivation shows that the local Rihaczek distribution can be obtained from the multitaper eigencoefficients and *vice–versa*. The global distribution can be obtained from local averages and this method has the advantage of essentially eliminating cross–terms. As presented above, the theory is essentially deterministic. Defining the eigencoefficient covariance, or *ecco*, matrix as in [8]

$$\mathbf{C}_{j,k}(f) = \mathbf{E}\{y_j y_k^*\} \tag{44}$$

all the expectations carry through the various linear operators so the second–order terms are basically identical to the deterministic expressions.

The interpretation of the various terms is of interest. First, $\mathbf{A}^{(0)}$ and $\mathbf{B}^{(0)}$ are approximately identity matrices and, consequently, $\mathcal{R}_{0,0}$ is approximately the multitaper spectrum. Next, the coefficient of $\mathbf{A}^{(0)}\mathbf{B}^{(1)}$ is approximately $\mathbf{B}^{(1)}$, so $\mathcal{R}_{0,1}$ is, roughly, the *frequency derivative* of the spectrum. Similarly,$\mathcal{R}_{1,0}$ is roughly the *time derivative* of the spectrum. This derivation illuminates several points:

- In standard quadratic inverse theory, the normalization factor is the reciprocal of the eigenvalue, α_l or g_m and the quadratic–inverse estimates are minimum–variance and unbiased, [6]. This implies that the local Rihaczek distribution is biased with the bias worse on the higher–order terms.
- Beginning with the $K \approx 2NW$ basic eigencoefficients, the dimensionality of the space, one cannot have uncorrelated estimates of the $2K$ coefficients implied by either set of eigencoefficients from a single realization.
- Worse, the covariance matrix (44) is $K \times K$ and so contains K^2 elements. The basis set $\mathbf{A}^{(l)}\mathbf{B}^{(m))}$ contains $4K^2$ elements that all cannot be linearly independent. This implies that one cannot tell a stationary process with rapidly varying spectrum from a nonstationary process. Note that even the K^2 elements of $\mathbf{C}$ can not be arbitrary because, being a covariance matrix, $\mathbf{C}$ must be non-negative definite.
- The form of the estimate implies a bound on the "derivative" estimates. All the coefficients, (40) are quadratic forms so, taking the stationary case as an example, if the average power $b_0(f)$ from (26) is fixed, then $b_1(f)$ is bounded proportionally to the largest eigenvalue of the matrix pencil $\{\mathbf{B^{(0)}}, \mathbf{B^{(1)}}\}$ and similarly with the other terms.
- While explicit variance terms have not yet been derived for the general nonstationary case, one expects that the derivation in [6, 8] will extend. Thus one expects that the degrees–of–freedom will be reduced by approximately a factor of $1 + c_{0,1}(\frac{\partial}{\partial f} \ln S(t,f))^2 + c_{1,0}(\frac{\partial}{\partial t} \ln S(t,f))^2 + \ldots$.
- Both sets of eigenvalues, the g_m's and α_m's are assymptotically of the form $2NW - m/2$ plus a small exponential "tail" and the m^{th} basis function, A_m or B_m has m zeroes in its principal domain. Thus the resolution is effectively limited by the finest zero spacing, that is $2w/4NW = 1/2N$ in frequency or $1/4W$ in time, that is the frequency resolution is limited to one–half the Rayleigh resolution and the time resolution to one–half the Nyquist interval. Moreover, the terms approaching these limits have much larger variances than the low order terms.

Finally, and most important, the derivation unifies time–frequency distributions with quadratic–inverse methods and, moreover, points the direction for a unified nonstationary, non locally–white theory.

8. REFERENCES

[1] David Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: the discrete case," *Bell System Tech. J.*, vol. 57, pp. 1371–1429, 1978.

[2] B. R. Frieden, "Evaluation, design and extrapolation methods for optical signals based on use of the prolate functions," *Prog. Optics*, vol. X, pp. 312–407, 1969.

[3] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.

[4] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2941–2946, 1992.

[5] P. Stoica and T. Sundlin, "On nonparametric spectral estimation," *Circuits Systems Signal Process.*, vol. 18, pp. 169–181, 1999.

[6] D. J. Thomson, "Quadratic-inverse spectrum estimates: applications to paleoclimatology," *Phil. Trans. R. Soc. Lond. A*, vol. 332, pp. 539–597, 1990.

[7] D. J. Thomson, "Non-stationary fluctuations in "stationary" time series," *Proceedings of the SPIE*, vol. 2027, pp. 236–244, 1993.

[8] David J. Thomson, "Multitaper analysis of nonstationary and nonlinear time series data," in *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young, Eds., pp. 317–394. Cambridge Univ. Press, 2001.

[9] A. W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Information Theory*, vol. IT-14, pp. 369–374, 1968.

[10] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[11] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution — a tool for time–frequency signal analysis part iii: Relations with other time–frequency transformations," *Phillips J. Res.*, vol. 35, pp. 372–389, 1980.

[12] L. L. Scharf, B. Friedlander, and D. J. Thomson, "Covariant estimators of time-frequency descriptors for nonstationary random processes," in *Proc. of the Thirty-Second Asilomar Conf. on Signals, Systems, and Computers*, Madison, WI, 1998, pp. 808–811, Omnipress, IEEE Cat. No. 98CH36284.

[13] H. Margenau and R. N. Hill, "Correlation between measurements in quantum mechanics," *Prog. Theoretical Physics*, vol. 26, pp. 722–738, 1961.

[14] C. A. Stutt, "Some results on real–part/imaginary–part and magnitude–phase relations in ambiguity functions," *IEEE Transactions on Information Theory*, vol. IT-10, pp. 321–327, 1964.

[15] S. M. Sussman, "Least–square synthesis of radar ambiguity functions," *IRE Trans. on Information Theory*, vol. IT–8, pp. 246–254, 1962.