Tools for multivariable spectral and coherence analysis

AUTHOR: Tryphon Georgiou
AFFILIATION: University of Minnesota

REPORT: February 2007

Summary

This report discusses recent research on high resolution spectral and coherence analysis and outlines directions we are actively pursuing. The focus is on developing quantitative tools for spectral analysis and system modeling, extend current resolution techniques to the case of multidimensional signals, and to fill an apparent need for robust, efficient, and high resolution tools for use in sensor networks and arrays, data mining, and spectral analysis. Signal analysis is often the hidden technology behind a wide range of applications. In particular, techniques that we have recently developed have have been applied to non-invasive temperature sensing via ultrasound (intended to facilitate computer guided tumor ablation and therapy) in collaboration with E. Ebbini, and tested for the purpose of vibration tracking and isolation in collaboration with D. Herrick. Such application areas underscore the need for the proposed research and for further improvements of the relevant tools.

A fundamental issue discussed in this report is the question of how to quantify uncertainty in spectral analysis and how to assess/improve resolution. The need for robustness and high resolution in signal analysis is self evident. Yet, at the present time, traditional assessment of such properties remains largely ad-hoc. The development of suitable metrics for quantifying distances between statistics and power spectra is the initial focus. Section 1.1 presents a model for the uncertainty in estimating covariance statistics, and outlines steps for improving the reliability and consistency of such statistical estimates. Section 1.2 presents certain (new) metrics for quantifying distances between power spectral densities. The approach mimics the development of the Fischer information metric and Kullback-Leibler distance of Information Geometry. It gives rise to a Riemannian geometry for spectral density functions rooted in prediction theory. The subsequent sections suggest improvements and generalizations of techniques in spectral analysis that are being studied. In particular, Section 1.3 seeks to transport the efficiency of periodogram-like methods and of the time-tested contribution of Blackman and Tukey, to multi-variable and multi-dimensional signal analysis. Section 1.4 seeks to explore convex optimization for the analysis of *multivariable & multidimensional* power spectra. Section 1.5 discusses the classical concept of coherence, and possible generalizations to an angular distance that reflects causal dependence as well. Such causal dependence is central in control applications as suggested by certain feed-forward control problems that one encounters when seeking noise suppression and disturbance rejection. The merit of the new approach relies on the metric concepts introduced, and in suggesting the possible integration of such metrics for developing robust high resolution analysis tools.

Our goal is enhance technological advances which, themselves, have a direct impact on human lives. Recent collaborative work with E. Ebbini and A.N. Amini on noninvasive temperature sensing via ultrasound (intended to facilitate computer controlled tumor ablation and therapy) identified as the single most important obstacle the task of translating the ultrasound echo into the map of a temperature field inside the tissue. Resolution and efficiency of the signal processing algorithms are the key. The steps discussed herein address a number of shortcomings in earlier techniques of signal analysis. The tools that we develop have the potential to impact a range of application areas involving distributed sensor arrays, identification, modeling and data mining.

Background

Our recent research has led to a range of tools for high resolution analysis of signals. The analysis helps identify a wide range of underlying physical causes and dynamical dependencies. On the application side, we have explored the relevance of our techniques in two main areas. First, we conducted case studies in measuring the temperature of (artificial) tissue in a noninvasive manner by analysing ultrasound echo. The purpose of such "non-invasive sensing" is to provide reliable measurement of tissue temperature for computer guided tumor ablation and therapy [48]. The results have been encouraging and substantially better than earlier state-of-the art. Second, we explored the use of such high resolution techniques in synthetic apperture radar (SAR) with similar results [51].

The most recent development, bearing on our present research plan, is a breakthrough in analyzing data from distributed sensor arrays (e.g., see [17]). Interestingly, this advancement shares the same basic framework with our techniques for high resolution spectral analysis. Identifying power distributions consistent with given measurements is treated as an inverse problem. The family of such distributions is suitably parametrized, and the size of the family represents a measure of uncertainty. The data/measurements may represent radar/sonar echo, a speech recording, etc., and may be sampled non-uniformly with gaps on the record. Our computational theory allows solving general such multivariable and multidimensional moment problems. Besides the relevance of these techniques in analysis, they also impact on the design and optimal distribution of sensors.

Our joint work with C. Byrnes and A. Lindquist [5], where foundations of our framework were first laid out, received the G.S. Axelby outstanding paper award from the IEEE Control Systems Society in 1993, and a U.S. patent [63] which was based on this and subsequent work. We wish to mention that earlier joint work with M.C. Smith on metrics for robust control, supported by NSF, received the G.S. Axelby outstanding paper award twice, in 1992, and again in 1999 (for a linear and a nonlinear theory, respectively). The most recent work can be broadly classified into the following categories with some overlaps. It has been supported by the AFOSR, the NSF, and the Vincentine Hermes-Luh endowment.

High resolution spectral analysis and applications

Publications: [48, 52, 49, 19, 32, 13, 33, 16, 22, 23, 24, 50, 42, 26, 27, 62, 53]

Our framework was initiated in [28, 30, 31]. It was influenced by collaborative work with C. Byrnes and A. Lindquist in [5, 6] and led to a U.S. patent [63] on tunable high-resolution spectral estimators. Publication [49] explores basic tradeoffs between resolution and robustness of such estimators, and outlines how to tune these for optimal performance. In [52, 53, 62, 51] we explain advantages of the new techniques and insight in antenna arrays, SAR, and in multi-rate signal processing. In [48, 50] we demonstrated the use of our new methods in non-invasive ultrasound temperature sensing for computer controlled tumor ablation. In [22, 23, 24] we pointed out that a key step in spectral analysis, the step of estimating statistics, may not provide data which are consistent with underlying dynamics. In this case, resolution can be dramatically improved if care is taken to adjust the statistics so as to conform with known underlying dynamics. Publications [42, 33, 19, 32, 27] deal with assessing the level of spectral uncertainty, and then presenting canonical decompositions for use in spectral analysis problems.

Multi-variable & multi-dimensional moments

Publications: [17, 19, 32, 16, 20, 22, 23, 26]

In these publications we solve the general multi-variable and multi-dimensional moment problem. Data for modeling, identification, spectral analysis, etc. often specify moment constraints on a power density function —possibly multivariable (matrix-valued) and multidimensional (spatio-temporal). The development in [17] gives a way to determine and parametrize all consistent distributions. Publications [32, 19]

develop further a very important "boundary" case of singular data sets. Publications [20, 22, 23] are mostly on a static but multivariable version of such problems and corresponding numerical issues.

Analytic interpolation with degree constraint: Publications: [7, 32, 16, 17, 26]

The problem of analytic interpolation with degree constraint was introduced in the early 1980's (Georgiou's 1983 Ph.D. thesis). Important contributions by Chris Byrnes and Anders Lindquist re-kindled interest in the problem, and in recent work [7] a rather complete theory for the (scalar version) of the problem was finally completed. Interest in this problem stems from applications in control and signal processing. The theory in [32, 16, 17, 26] is relevant in addressing the multivariable version of the problem (which is essential for multivariable control applications). Work on the multivariable problem is still in progress along the lines of [17] and will appear shortly [61]. This work entails a complete parametrization of all solutions to a multivariable Nehari type of interpolation problem, which have a Macmillan degree $\leq$ to the generic degree prescribed by the problem data.

1 High resolution tools

The goal has been to develop theory and techniques for robust & high resolution spectral analysis and system identification. The approach seeks a quantitative assessment of uncertainty as well as suitable models/spectra that are optimal with respect to suitable criteria. The data are typically statistics of multivariable time-series. The main tools consist of suitable distance measures and convex optimization. We finally focus on certain disturbance isolation experiments and issues on how to utilize information from distributed sensor arrays.

There are several analogies and a certain compatibility of our framework with that of modern robust control. In particular, certain of the results that we have already obtained echo analogous constructions in robust control (e.g. cf. [5, 30]). Then, also, the main focus in our study is on metric uncertainty and how this depends on the data and the underlying (physical) assumptions about their origin. This "systems viewpoint" has very much been underlying our recent work on high resolution spectral analysis which was funded by a previous NSF grant.

We present our basic framework in the context of spectral analysis of time-series. In particular, we underscore the relevance of our approach on several basic questions in signal analysis and identification.

1.1 Distance measures between statistics and related approximation problems

Consider a zero-mean, stationary random process u_k and let $R_k := E\{u_\ell u^*_{\ell+k}\},\ k = 0, 1, \dots$ denote its autocorrelation function. Most modern nonlinear spectral analysis techniques rely on estimates of the autocorrelation matrix

$$\mathbf{T} : \ = \ E\{x_k x_k^*\} = \begin{bmatrix} R_0 & R_1 & \dots & R_\ell \\ R_{-1} & R_0 & \dots & R_{\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{-\ell} & R_{-(\ell-1)} & \dots & R_0 \end{bmatrix} \tag{1}$$

where $x_k := \begin{bmatrix} u'_k & u'_{k+1} & \dots & u'_{k+\ell} \end{bmatrix}'$, and prime (resp. star) denotes transposition (resp. complex conjugation and transposition). When u_k is a scalar-process $\mathbf{T}$ has a Toeplitz structure, and when u_k is a (column-)vectorial process $\mathbf{T}$ has a block-Toeplitz structure. Estimates of such statistical quantities are typically obtained via sample averaging, e.g. using

$$\hat{\mathbf{T}} := \frac{1}{N} \sum_{k=0}^{N-1} x_k x_k^*, \tag{2}$$

and relying on a finite observation record $\{u_0, u_1, \ldots, u_{N+\ell-1}\}$. Not surprisingly, $\hat{\mathbf{T}}$ fails to be (block) Toeplitz—a fact which adversely affects all subsequent analyses.

Another example originates in recent trends in high resolution spectral analysis. These focus on the dependence of statistics on known dynamics dictating the distribution u_k. This dependence generalizes the "Toeplitz structure" to more general (linear, algebraic, etc.) contstraints on the statistics and on the power density of u_k. For instance, when data is collected by an array of sensors which is not "linear and equispaced," the geometry dictates a transcendental dependence between the transmission delays of the individual sensor outputs. A fairly general and theoretically advantageous starting point is to assume that the statistics consist of the state-covariance of a known linear system (which for simplicity can be thought to be finite-dimensional). The structure of state-covariances is not arbitrary and has been described in [15, 28, 31]. Not surprisingly, when sample averaging is used, very much like in $\hat{\mathbf{T}}$, sample state-covariances fail to have the correct structure. This adverse affects accuracy and resolution.

Thus, a key problem is to estimate second-order statistics a dynamical process in a way *consistent* with the underlying dynamics. In the case of a time-series u_k, the "dynamical" dependence between variables u_k, u_{k+1} is trivially only a time-delay while, in the more general case, nontrivial dynamics dictate the evolution of the entries of x_k and hence, the relevant statistics.

In principle, a sample covariance can be thought of as a random variable and a maximum likelihood value can be sought that is consistent with any dynamical constraints. However, the computation of such a maximum likelihood solution is quite complicated in general (even for the most rudimentary case when u_k is Gaussian and $\mathbf{T}$ is required to be Toeplitz). A commonly used method which allows enforcing structure and positivity in a consistent manner, is the celebrated Burg's algorithm [39]. This relies on a clever forward/backward scheme to obtain values for the "reflection coefficients" which ensure positivity. Unfortunately, the algorithm applies only to ensuring a Toeplitz structure for the autocorrelation *scalar* time-series and based on *one contiguous observation record*. In spite of how significant the problem is, there has been no generalization of Burg's algorithm (although this is possible for block-Toeplitz structures using matrix "orthogonal polynomials," though very tedious, see discussion in [32].) In practice, the fact that e.g., $\hat{\mathbf{T}}$ is not Toeplitz is often overlooked and/or ad-hoc corrections are being made. The PI has studied and advocated alternatives for some time now [22]. Our current plan is to explore a new, simple, and quite versatile approach. This is based on postulating a *model for the sampling errors*, and then to minimizing the variance of such errors in a way so as to achieve consistency [29].

Because the time-index is not essential we suppress it and set $x = x_k$ (for any particular fixed value k). We postulate the following *model for the statistical errors in the estimation process*:

$$\hat{\mathbf{T}} = E\{(x+v)(x+v)^*\} \tag{3}$$

where v is a zero-mean, stationary noise process, possibly correlated with x. Note that $\hat{\mathbf{T}}$, as defined earlier, is a finite sum of products of random variables (hence, a random variable itself). Thus, equation (3) provides a model, postulating that $\hat{\mathbf{T}}$ can be accurately represented as *the* covariance of a *perturbed* random vector. Assuming such a model, it is natural to *seek the least variance perturbation which achieves consistency*. More specifically, if we denote by

$$\mathbf{M} \;=\; E\{\begin{bmatrix} x \\ v \end{bmatrix} \begin{bmatrix} x^* & v^* \end{bmatrix}\} = \begin{bmatrix} \mathbf{T} & \mathbf{S} \\ \mathbf{S}^* & \mathbf{Q} \end{bmatrix} \tag{4}$$

the joint covariance matrix of x_k and v_k, then we seek a noise component of least variance so that

$$\hat{\mathbf{T}} \;=\; \begin{bmatrix} I & I \end{bmatrix} \mathbf{M} \begin{bmatrix} I \\ I \end{bmatrix} = \mathbf{T} + \mathbf{S} + \mathbf{S}^* + \mathbf{Q}$$

where I denotes the identity matrix of compatible dimension. Thus, we have studied a "*minimal-variance correction*" as a distance measure between a (sample) covariance matrix $\hat{\mathbf{T}}$ and a convex class of matrices,

such as Toeplitz matrices or other structured matrices of compatible size. It is important to note that, given $\hat{\mathbf{T}}$ and a (convex) class of non-negative matrices $\mathcal{T}$ of compatible size, the optimization problem

$$d_M(\hat{\mathbf{T}}, \mathcal{T}) \quad := \quad \min\{\text{trace}\,(\mathbf{Q}) \mid (5) \text{ holds}, \mathbf{M} \geq 0, \mathbf{T} \in \mathcal{T}\} \tag{5}$$

is convex. If we assume that the postulated estimation noise v_k is uncorrelated with the random vector x_k, the least variance perturbation is again obtained by a semi-definite program:

$$d_Q(\hat{\mathbf{T}}, \mathcal{T}) \quad := \quad \min\{\text{trace}\,(\mathbf{Q}) \mid \hat{\mathbf{T}} = \mathbf{T} + \mathbf{Q}, \mathbf{Q} \geq 0, \mathbf{T} \in \mathcal{T}\} \tag{6}$$

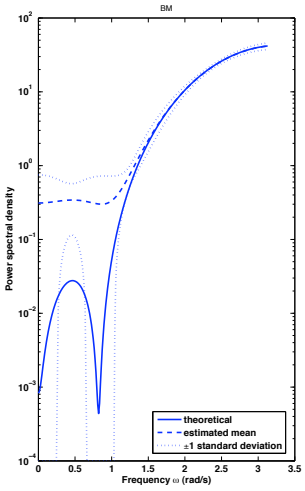
An alternative model can be based on a symmetric allocation of "corrections." To this end we postulate that $\mathbf{T}$ and $\hat{\mathbf{T}}$ are covariances of vectors x and $\hat{x}$, respectively, and that the two are noisy measurement of yet a third vector z. Hence, $x + v = z = \hat{x} + \hat{v}$. If (4) and an analogous definition of $\hat{\mathbf{M}}$ as the covariance of $\begin{bmatrix} \hat{x}' & \hat{v}' \end{bmatrix}'$ hold, then symmetrized distances can be defined as follows:

$$\delta_1(\mathbf{T}, \hat{\mathbf{T}}) \quad := \quad \min\{\text{trace}\,(\mathbf{Q} + \hat{\mathbf{Q}}) \mid \hat{\mathbf{T}} + \mathbf{Q} = \mathbf{T} + \hat{\mathbf{Q}}, \mathbf{Q} \geq 0, \hat{\mathbf{Q}} \geq 0\} \tag{7}$$

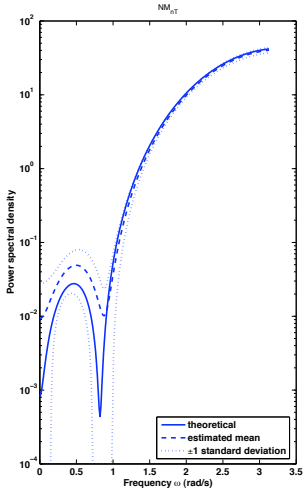
$$\delta_2(\mathbf{T}, \hat{\mathbf{T}}) \quad := \quad \min\{\text{trace}\,(\mathbf{Q} + \hat{\mathbf{Q}}) \mid \mathbf{M} \geq 0, \hat{\mathbf{M}} \geq 0, \text{ and } \begin{bmatrix} I & I \end{bmatrix} (\mathbf{M} - \hat{\mathbf{M}}) \begin{bmatrix} I \\ I \end{bmatrix} = 0\}. \tag{8}$$

The expression in (7) is a special case of (8) where the matrices $\mathbf{M}, \hat{\mathbf{M}}$ are taken to be block diagonal. Interestingly, $\delta_1(\mathbf{T}, \hat{\mathbf{T}})$ is a metric (shown by the PI in [29]). It was kindly suggested to us by P. Parrilo, it can also be written in the form $\|\mathbf{T} - \hat{\mathbf{T}}\|_1$, where $\|M\|_1 = \text{trace}\,((MM^*)^{1/2})$ is the "trace norm." The important fact is that such distances between a given $\hat{\mathbf{T}}$ and suitable convex classes $\mathcal{T}$, can be readily and efficiently computed using standard semidefinite programming software.

The estimation of moving average power spectra is a notoriously difficult task [57, 58]. Yet, following our rationale, we may choose $\mathcal{T}$ as being banded Toeplitz matrices corresponding to moving-average (MA) models of any given order. This family is convex, efficiently characterized by linear matrix inequalities via the Kalman-Yakubovic-Popov lemma [19, 57], and $\min_{\mathbf{Q}} d_Q(\hat{\mathbf{T}}, \mathcal{T})$ is readily computable. In recent work (jointly with P. Stoica, Lin Du, and Jian Li), we sought to compare on a typical example the accuracy of prior *state-of-the art* [57] with our approach based on $\mathbf{T} \equiv \hat{\mathbf{T}} - \arg\min_{\mathbf{Q}} d_Q(\hat{\mathbf{T}}, \mathcal{T})$. The plots on the right compare the theoretical spectrum with the mean, and the mean ± 1 standard deviation, for spectra obtained by following [57] and then using $\mathbf{T}$ as above. The improvement is dramatic.



Method in [57]



Based on $\min_{\mathbf{Q}} d_Q(\hat{\mathbf{T}}, \mathcal{T})$

Spectral analysis of vector-valued processes is relevant in wide range of applications. In sensor technology (e.g., in polarimetric synthetic aperture radar) a collection of correlated echoes at different wavelength/polarization/etc. encode attributes of a scattering field. In system identification, spectral analysis of the vectorial process which consists of inputs and outputs of a dynamical system provides models for a given system, possibly operating in closed loop. An early investigation of such a framework for system identification in the context of robust control providing confidence intervals in the "gap metric," is outlined in [21]. The framework that we presented needs to be integrated with the theory in [5, 6, 17, 15, 28, 30, 31] in order to provide a toolbox for high resolution (multivariable) spectral analysis and system identification. Particular questions/tasks are as follows:

Problem 1 Develop an optimal approximation theory in d_M, d_Q, δ_i as well as, "weighted" counterparts.

This question entails studying diverse convex families of covariances functions (banded, moving-average, state-covariances and those with short-correlation structure as in [32]). Further, it is not known how "close" δ_2 is to being a metric, and it is not known when the minimizer, in the case δ_2 or δ_M are used, is unique. The effect of weighted distance measures when seeking to minimize trace $(\mathbf{Q} + W\hat{\mathbf{Q}})$ instead of trace $(\mathbf{Q} + \hat{\mathbf{Q}})$ penalize differently the variability of elements in $\hat{\mathbf{T}}$. This extra degree of freedom has obvious practical significance. We also plan to study approximation problems with low rank or sparse correlation structures (studied in recent years by D. Donoho, L. El Ghaoui, and others). A most challenging question is to investigate how such approximation schemes perform when a covariance matrix is to be estimated from very few "dyads" in (2). Such problems are central in, e.g., bio-informatics where x is a rather long vector, although there, any constraints on the elements of covariances may not be apparent.

1.2 Distances between power spectra and entropy functionals

Despite the centrality of spectral analysis in a wide range of disciplines, no agreement exists as to what an appropriate distance measure between spectral density functions is. Some of the key contenders have been Bregman distances, the Kullback-Leibler-von Neumann distance, the Itakura-Saito distance, and finally Battacharrya and Mahalanobis-type variants. Certain of these distances have a definite relevance when used to discriminate between two probability density functions. Yet none seems to have a physically meaningful interpretation when applied to power spectra. We begin with a new distance measure between power spectral densities and in fact, a (pseudo-) metric, which has a clear interpretation rooted in prediction theory. This is based on [34, 35].

Our starting point is to consider the degradation of the variance of the prediction error when a predictor is based on the wrong choice among two alternatives. More specifically, let f_1, f_2 represent spectral densities of discrete-time zero-mean random processes $u_{f_i}(k)$ ($i \in \{1,2\}$ and $k \in \mathbb{Z}$), and let $p_{f_i}(\ell)$ ($\ell \in \{1,2,3,\ldots\}$) be values for the coefficients that minimize the prediction error variance

$$\mathcal{E}\{|u_{f_i}(0) - \sum_{\ell=1}^{\infty} p(\ell) u_{f_i}(-\ell)|^2\}.$$

The optimal set of coefficients depends on the power spectral density function of the process, a fact which is acknowledged by the subscript in the notation $p_{f_i}(\ell)$. It is reasonable to consider as a distance between f_1 and f_2, the degradation of predictive error variance when the coefficients $p(\ell)$ are selected assuming one of the two, and then used to predict a random process with the other spectral density function. The ratio of the "degraded" predictive error variance over the optimal error variance

$$\rho(f_1, f_2) := \frac{\mathcal{E}\{|u_{f_1}(0) - \sum_{\ell=1}^{\infty} p_{f_2}(\ell) u_{f_1}(-\ell)|^2\}}{\mathcal{E}\{|u_{f_1}(0) - \sum_{\ell=1}^{\infty} p_{f_1}(\ell) u_{f_1}(-\ell)|^2\}}$$

equals the *ratio of the arithmetic over the geometric means* of the fraction of the two densities, namely

$$\rho(f_1, f_2) = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} \frac{d\theta}{2\pi} \right) / \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\frac{f_1(\theta)}{f_2(\theta)}) \frac{d\theta}{2\pi} \right),$$

see [34, 35]. Then, since $\rho(f_1, f_2) \geq 1$, either $\delta(f_1, f_2) := \log \rho(f_1, f_2)$ or $\gamma(f_1, f_2) := \rho(f_1, f_2) - 1$ represent measures of dissimilarity between the "shapes" of f_1 and f_2 and, can be viewed, as analogous to "divergences" of Information Theory (such as the Kullback-Leibler relative entropy). By considering the incremental degradation between a nominal power spectral density f and a perturbations $f + \Delta$ (e.g., $\gamma(f, f + \Delta)$), the quadratic term defines the Riemannian metric

$$g_f(\Delta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2 \frac{d\theta}{2\pi} - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} \frac{d\theta}{2\pi} \right)^2 \tag{9}$$

on density functions. It is a pleasant surprise that, geodesic paths f_τ ($\tau \in [0,1]$) connecting spectral densities f_0, f_1 can be explicitly computed [34]. Interestingly, these turn out to be logarithmic intervals (also referred to as exponential families), $f_\tau(\theta) = f_0^{1-\tau}(\theta) f_1^{\tau}(\theta)$ for $\tau \in [0,1]$, between the two extreme points. Furthermore, the length along such geodesics can be *explicitly* computed in terms of end points

$$d_g(f_1, f_2) := \int_0^1 \sqrt{\delta(f_\tau, f_{\tau+d\tau})} = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\log \frac{f_1(\theta)}{f_2(\theta)}\right)^2 \frac{d\theta}{2\pi} - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{f_1(\theta)}{f_2(\theta)} \frac{d\theta}{2\pi}\right)^2}. \qquad (10)$$

This is a "standard-deviation-like" measure of the difference $\log(f_1) - \log(f_2)$. It is a *pseudo-metric* in that it does not account for *constant* multiplicative factors.

It is rather interesting to point out that $f \mapsto \log(f)$ maps power spectral densities onto a Euclidean space where quadratic norms such as (10) have a clear interpretation. In fact, with respect to the Riemannian metric (9), the space has zero curvature since geodesics are "logarithmic" straight lines. From this vantage point one may also consider alternative norms such as $\|\log(\frac{f_1}{f_2})\|_2$, etc. though without yet a natural interpretation.

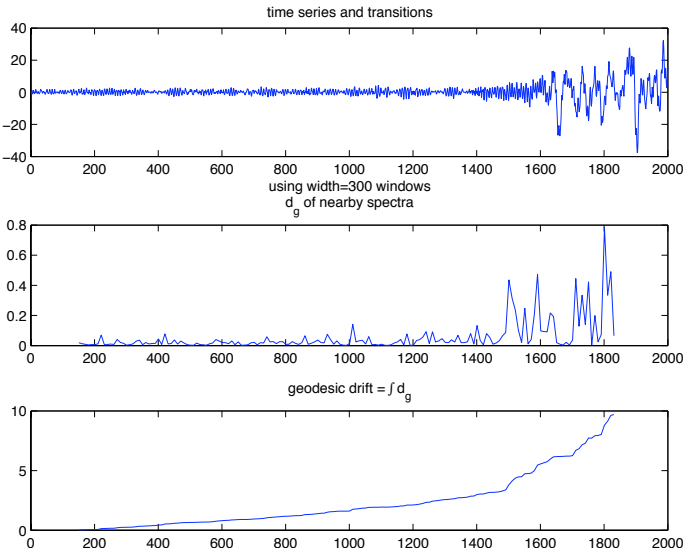
It is interesting to compare the differential structure on power densities induced by (9) with the corresponding structure of "Information Geometry." In Information Geometry $f(\theta)$ represents a probability density on $[-\pi, \pi]$ and the natural Riemannian metric is the *Fisher information metric* is (cf. [1, page 28]) which can be expressed as

$$g_{\text{Fisher},f}(\Delta) \quad = \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Delta(\theta)^2}{f(\theta)} \frac{d\theta}{2\pi} \qquad (11)$$

(with $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\frac{d\theta}{2\pi} = 1$ and $\frac{1}{2\pi}\int_{-\pi}^{\pi} \Delta(\theta)\frac{d\theta}{2\pi} = 0$ since both f, $f+\Delta$ need to be probability densities). Direct comparison reveals that the powers of $f(\theta)$ in (9) and (11) are different. Thus, it is curious and worth underscoring that in either differential structure, geodesics and geodesic lengths can be computed. For completeness we note that Information Geometry is a vast subject, originating in the work of Rao, Amari, Cencov and others, with a large following directed towards analogous geometric interpretations in Quantum theory. The starting point of Information Geometry may be considered, in a way analogous to our development, to be the degradation of coding efficiency when the wrong choice between two probability distributions f_1 and f_2 is assumed. This degradation is precisely the Kullback-Leibler distance between the two, which can gives rise to the Riemannian metric $g_{\text{Fisher},f}(\Delta)$, in a way analogous to our construction of $g_f(\Delta)$. Our plan regarding distance measures on power densities is summarized as follows:

Problem 2 *Study (9) and the corresponding differential structure on power spectra —explore analogues and connections with (11) and Information Geometry. Determine other natural distance measures between power densities (e.g., based on smoothing and other type of filtering). Utilize such distance measures in solving spectral-inverse problems and in quantifying distance between solutions.*

The way that such metrics are to be utilized in practice is examplified by a case study in trying to identify the drift in the spectral make-up of a time-series. The data were collected in a vibration experiment. The analysis was carried out in collaboration with Dr. D. Herrick. The data was processed by sliding a window 100[ms] and evaluating the distance between the spectra of nearby windows. Spectral distances and the geodesic drift, shown on the right, reveal a very fast transition/change in spectral content beginning at 1500[ms], and provide a reliable indicator of the on-set of strong disturbance.



We are also interested in generating metrics for the case where deterministic components are present. Such components are "invisible" in the above predictive framework. The idea to employ the degradation of performance with regard to specific tasks extends easily to a variety of contexts, and should be useful for that purpose as well. An alternative paradigm can be built on smoothing problems which we take up next but with a different goal in mind, namely, to provide a critique and an alternative to the maximum entropy principle.

The maximum entropy principle, as it is often invoked in time-series analysis ([39, 9, 40, 45]), suggests the selection of a power spectrum which is consistent with autocorrelation data and corresponds to a random process least predictable from past observations. While this is a reasonable dictum when one is interested in prediction, it is often used regardless of the specific intent for the sought spectrum. The point we wish to raise becomes apparent when considering the relevance of another dictum, equally pertinent, albeit based on smoothing instead of prediction.

The variance $\mathcal{E}\{|u(0)-\sum_{\ell=1}^{\infty} p_f(\ell)u(-\ell)|^2\}$ of the optimal one-step-ahead (linear) predictor $\hat{u}(0|\text{past}) := \sum_{\ell=1}^{\infty} p_f(\ell)u(-\ell)$ is the *geometric mean* (see [38, page 183], [65, Chapter 6]) of f, i.e.,

$$\mathcal{E}\{|u(0) - \hat{u}(0|\text{past})|^2\} = m_{0,f} := \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(f(\theta)\right) d\theta\right).$$

This is the content of the celebrated Kolmogorov-Szegö theorem. The entropy rate [39] is then defined as the negative integral of the logarithm of f (i.e., as $-\int \log(f(\theta))d\theta$). The notation $m_{0,f}$, taken from [3, page 23] for the *geometric mean*, is sought to contrast with the expression for the variance of the error

$$\mathcal{E}\{|u(0) - \hat{u}(0|\text{past} + \text{future})|^2\} = m_{-1,f} := \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)^{-1} d\theta\right)^{-1}$$

for the optimal smoothing filter $\hat{u}(0|\text{past} + \text{future}) := \sum_{\ell\neq 0}^{\infty} q_f(\ell)u(-\ell)$. This expression represents the *harmonic mean* of f [33]. Applications abound where records need to be interpolated, or where the indexing of data collected via a sensor array represents spatial-ordering and not time-ordering. In all such applications there is no natural "time-arrow" and, hence, it is imperative that Burg's maximum entropy principle is re-evaluated.

Thus, in the context of time series analysis, both Burg's entropy $= \int \log(f(\theta)^{-1})d\theta$ and the smoothing analog $\int f(\theta)^{-1}d\theta$ ([33]) relate to the level of unpredictability in these two different situations. Burg's entropy has been also used as a regularizing functional in inverse problems (see [17]). But the latter functional can be used equally well for similar modeling purposes. For instance, we have shown in [33] that extremal spectra with respect to the second choice give rise to all-pole Markovian models very much like Burg's maximum entropy AR-models, but with one important difference. The poles in these models appear with fractional powers. Such fractional powers are often encountered in processes with long "memory." This motivates:

Problem 3 *Study the relevance of fractional dynamics in modeling long-range correlations.*

There is an apparent dichotomy depending on whether we consider a one-sided or a two-sided past. Stationary time-series are said to be deterministic in the Kolmogorov sense if $\log(f) \notin L_1$. When we consider determinism with respect to a two-sided past, then the corresponding condition weakens to $f^{-1} \notin L_1$, because it is only then that the smoothing error is zero. This dichotomy raises similar questions for spatial processes and fields. This is especially pertinent when *space-time data* are collected via sensor arrays. The type of questions we will address are exemplified by the following two problems:

Problem 4 *Given (partial correlation) moments of a two-dimensional distribution, determine the distribution corresponding to a random process least predictable at a particular range of coordinates from values over a certain other range (the latter is thought of as the "past" or, as the "available sensor readings").*

Problem 5 *If we postulate that a spatial process consists of a noise component superimposed on top of a sought deterministic component, how can we identify the power spectral content of the latter from partially known statistics?*

Problem 4 suggests a possible contact with the theory Markov random fields and of reciprocal processes (e.g., see [46]), and alludes to a suitable generalization of the maximum entropy paradigm. Problem 5, when specialized to ordinary scalar time-series, has also a long history (see [32]). It underlies many of the most widely used high resolution methods of signal analysis. In the current formulation Problem 5 calls for mutli-dimensional generalizations as well as an investigation of what kind of decompositions we can expect when we use different notions of determinism.

From a mathematical and computational standpoint, entropy functionals can be thought of as natural barriers of convex sets (i.e., of probability simplices, or of cones of power densities). They thus can be used to construct solutions to ill-posed inverse problems. Problems 4 and 5 in particular can be addressed as moment problems. The history of such a viewpoint can be looked at in a recent publication [17]. Besides the Shannon-von Neuman, Burg, and Kullback-Leibler-Linblad-Leib functionals, discussed in [17] there is a plethora of alternatives, such as the Rényi and Tsallis entropies, and other generalized means (see [34], and also Ferrante *etal.* [11]) As part of the current project we intend to address:

Problem 6 *Study alternative notions of entropy focusing on closed-form formulae for their extrema and their relevance in spectral analysis and moment problems.*

Our ultimate goal is to acquire convenient tools for incorporating uncertainty into correlation measurements. The practical significance of such an undertaking is to integrate statistical uncertainty with the inherent uncertainty of inverse problems. Such an assessment of uncertainty is clearly relevant in prediction, control, and modeling in general.

We note in passing that matricial entropy functionals in [17], have led to a parametrization of all (generically) minimal degree solutions to robust multivariable (1-block) control problems [61], generalizing the work in [5, 14, 7].

1.3 Generalized statistics & generalized periodogram

Our next step is to explain what is meant by "generalized statistics" and why is this concept important. Consider a stationary random process u_k as before, with zero mean and power spectral density $f_u(\theta)$. The autocorrelation samples $R_k = E\{u_\ell u_{\ell+k}\}$ of u_k are the Fourier coefficients of $f_u(\theta)$, i.e.,

$$R_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j\theta} f_u(\theta) d\theta, \text{ for } k = 0, \pm 1, \pm 2, \dots .$$

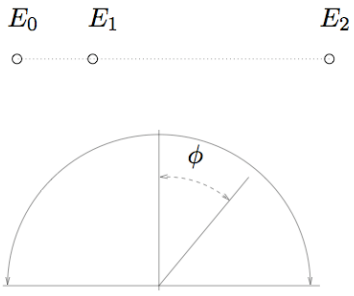
Occasionally, u_k is not directly observable in which case one may not be able to estimate autocorrelation samples. For instance, if $x_k = ax_{k-1} + u_k - bu_{k-1}$ is a first-order system ($-1 < a < 1$) and if only x_k is available, then it is natural to estimate statistics of x_k instead. These statistics represent moments of $f_u(\theta)$ with respect to kernel functions which differ from the usual $e^{-jk\theta}$. For example, the variance of x_k,

$$\mathcal{E}\{x_k^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\frac{e^{j\theta}-b}{e^{j\theta}-a}|^2 f_u(\theta) d\theta$$

is a moment of $f_u(\theta)$ with respect to the kernel function $|(e^{j\theta}-b)/(e^{j\theta}-a)|^2$. Such filtering may be part of a measuring apparatus, but it may also be introduced to improve S/R and resolution as in e.g, [49, 30].

Thus, in general, it is customary to refer to any moments of $f_u(\theta)$ as *generalized statistics* of the underlying random process. Not all such moments originate in ordinary time-filtering, and not all correspond to rational kernel functions. In fact, a most challenging and very common situation arises when the indexing in $\{u_k\}$ refers to space and not time.

Take for instance an array of sensors with three elements, linearly spaced at distances 1 and $\sqrt{2}$ wavelengths from one another, and assume that (monochromatic) planar waves, originating from afar, impinge upon the array. This is exemplified in the figure on the right. Assuming that the sensors are sensitive to disturbances originating over one side of the array, with sensitivity independent of direction, the signal at the ℓth sensor is typically represented as a superposition



Non-equispaced sensor array

$$u_\ell(t) = \int_0^\pi A(\theta) e^{j(\omega t - p x_\ell \cos(\theta) + \phi(\theta))} d\theta,$$

of waves arising from all spatial directions $\theta \in [0, \pi]$, where ω is the angular time-frequency (as opposed to "spatial"), x_ℓ the distance between the ℓth and the 0th sensor, p the wavenumber, and $A(\theta)$ the amplitude and $\phi(\theta)$ a random phase of the θ-component. Typically, the phase $\phi(\theta)$ for various values of θ are uncorrelated. The term $p x_\ell \cos(\theta)$ in the exponent accounts for the phase difference between reception at different sensors. For simplicity we assume that $p = 1$ in appropriate units. Correlating the sensor outputs we obtain

$$R_k = E\{u_{\ell_1} \bar{u}_{\ell_2}\} := \int_0^\pi e^{-jk\cos(\theta)} f(\theta) d\theta \tag{12}$$

where $f(\theta) = |A(\theta)|^2$ now represents *power density*, and $k = \ell_1 - \ell_2$ with $\ell_1 \geq \ell_2$ belonging to $\{0, 1, \sqrt{2}+1\}$ (k is kept as a "non-integer" index in R_k from mnemonic purposes). Thus, $k \in \mathcal{I} := \{0, 1, \sqrt{2}, \sqrt{2}+1\}$. The only significance of our selection of distances between sensors, that gave rise to this rather unusual indexing set $\mathcal{I}$, is to underscore that there is no algebraic dependence between the kernel functions

$$1,\ e^{-j\cos(\theta)},\ e^{-j\sqrt{2}\cos(\theta)},\ e^{-j(\sqrt{2}+1)\cos(\theta)}.$$

Even more challenging situations arise when (i) the kernel functions represent Green's functions or transfer functions in a general spatial domain [10], e.g., in case sensors are scattered in a random pattern in $\mathbb{R}^3$, and (ii) when statistics are obtained from observations non-equispaced in time (also, random sampling).

Let us revisit the situation of ordinary autocorrelation samples, and let $\mathbf{T}$ be a corresponding Toeplitz matrix as in (1). Then $\mathbf{T} = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(e^{j\theta}) f_u(\theta) G(e^{j\theta})^* d\theta$, where

$$G(e^{j\theta}) = \begin{bmatrix} 1 & e^{j\theta} & \dots & e^{j\ell\theta} \end{bmatrix}'$$

(and as before, "$'$" denotes transpose while "$*$" denotes complex conjugate transpose). The (column) $G(e^{j\theta})$ is referred to as a "Fourier vector," and as θ varies, it defines a curve in a complex space which in the array signal processing literature is known as the "array manifold". In this section we assume $\mathbf{T}$ known and do not address issues of approximating $\mathbf{T}$ from sample statistics that we raised earlier.

There is a rather rich theory on how much $\mathbf{T}$ is telling us about the power spectum, and how to reconstruct representative spectra (maximum-entropy, etc.) which are consistent with the partial sequence of autocorrelation statistics. This goes back to the theory of the trigonometric moment problem and of orthogonal polynomials [37], and forms the basis of the so-called "modern nonlinear spectral analysis methods" [39]. Yet, a more common way to reconstruct $f_u(\theta)$, based on $\mathbf{T}$, is the time-tested periodogram/correlogram

$$\hat{f}(\theta) \ := \ \frac{1}{\ell+1} G(e^{j\theta})^* \mathbf{T} G(e^{j\theta}) \ = \ \dots + \frac{\ell}{\ell+1} R_1 e^{j\theta} + R_0 + \frac{\ell}{\ell+1} R_1 e^{j\theta} + \dots . \tag{13}$$

This is an approximation of $f_u(\theta)$ — see [59]. Another equally direct way is due to Capon:

$$\hat{f}(\theta) := \frac{1}{\ell+1} \left(G(e^{j\theta})^* \mathbf{T}^{-1} G(e^{j\theta}) \right)^{-1}. \tag{14}$$

In either case, weighted versions of the autocorrelation coefficients can be used instead, in order to trade-off resolution with robustness. I.e., using various windowing functions w_k (Hamming, Kaiser, etc.) one may replace R_k with $R_k w_k$ in the above. These ideas are classical, were extensively studied decades ago, and remain the workhorse of signal analysis applications to this day. Yet, it is a striking fact that a multivariable version of such successful tools has largely been absent (i.e., a periodogram-like method for inherently multivariable processes).

A further fact is that the corresponding issues when $G(e^{j\theta})$ is not an ordinary Fourier vector, have not been studied with the exception of the somewhat ad-hoc beamspace techniques. The recent work in [17, 16, 62] has attempted to address such issues on a firm theoretical basis. For instance, returning to the example of the non-equispaced antenna array and R_k for $k \in \mathcal{I}$ in (12), it is important to determine whether estimated values for the moments are consistent with the geometry of the array, and if so to characterize all consistent power spectra. In the present situation we package R_k's in (12) into a matrix and set $\tau = \cos(\theta)$). The nonnegativity of

$$\mathbf{R} := \int_{-1}^{1} \begin{bmatrix} 1 \\ e^{-j\tau} \\ e^{-j\sqrt{2}\tau} \end{bmatrix} \frac{f(\cos^{-1}(\tau))}{\sqrt{1-\tau^2}} \begin{bmatrix} 1 & e^{j\tau} & e^{j\sqrt{2}\tau} \end{bmatrix} d\tau \;=\; \begin{bmatrix} R_0 & R_1 & R_{\sqrt{2}+1} \\ \bar{R}_1 & R_0 & R_{\sqrt{2}} \\ \bar{R}_{\sqrt{2}+1} & \bar{R}_{\sqrt{2}} & R_0 \end{bmatrix}, \tag{15}$$

is only a necessary condition. The fact that this condition is *not* sufficient (see e.g., [15, page 786]) motivated our recent work and led to an approach documented in [17, 16].

We now highlight some of the important findings and open questions. First, as alluded to earlier, we generalize the form of the "Fourier vector" $G(e^{j\theta})$, replacing it with the transfer function of a linear, time-invariant, discrete-time (input-to-state) dynamical system

$$x_k = Ax_{k-1} + Bu_k, \text{ with } k \in \mathbb{Z},$$

x_k being the state-vector and A, B matrices in $\mathbb{R}^{n\times n}$ and $\mathbb{R}^{n\times m}$, respectively[1]. The input-to-state transfer function $G(e^{j\theta}) = (I - e^{j\theta}A)^{-1}B$ could be matricial and the random process u_k vectorial (when $m > 1$). If u_k is white noise (with covariance matrix $Q \geq 0$), then it is well known that the state covariance

$$\mathbf{R} := E\{x_k x_k'\}$$

satisfies the Lyapunov equation $\mathbf{R} - A\mathbf{R}A' = BQB'$. The case where u_k is not white was dealt only recently ([30, 31]). The correspondence between $\mathbf{R}, A, B$ and input power spectra f_u is detailed in [30, 31, 32]. Briefly, a state covariance for the above system satisfies

$$\operatorname{rank} \begin{bmatrix} \mathbf{R} - A\mathbf{R}A^* & B \\ B^* & 0 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix} \tag{16}$$

where 0 is the zero matrix of appropriate dimension. An alternative characterization amounts to the solvability of

$$\mathbf{R} - A\mathbf{R}A' = BH' + HB'$$

for a matrix H which is of the same size as B. Conversely, provided $\mathbf{R}$ satisfies either of the above two equivalent conditions, and provided it is non-negative definite, there exists a power spectrum for a candidate input that gives rise to such state-statistics (this was shown in [30]). The parametrization of all consistent power spectra and related computational issues has been the subject of [30, 31, 32]. The relevant realization theory for matricial power spectral densities amounts to analytic interpolation with positive-real matricial functions and thus, echoes the usual tools and constructions in H_∞-control theory.

The motivation for considering state-covariances of linear systems, was to develop a theory for high resolution spectral analysis following [5, 6, 28]. Our joint work with C. Byrnes and A. Lindquist led to a

[1] This can only approximate the case (12), to which we will return shortly.

U.S. patent [63]. The main idea in [5] arose from the simple observation that the autocorrelation samples of a time-series correspond to interpolation conditions for a positive-real function related to the power spectrum, at the origin. In some detail,

$$F(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1+ze^{j\theta}}{1-ze^{j\theta}} f_u(\theta)d\theta = R_0 + 2R_1 z + 2R_2 z^2 \dots$$

is a positive-real function, and the R_k's relate to the value of $F(z)$ and its derivatives at the origin. This generalizes to statistics of the state or output of any dynamical system. E.g., if $x_k = ax_{k-1} + u_k$ as before, and if $-1 < a < 1$, then

$$E\{x_k^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|e^{j\theta}-a|^2} f(\theta)d\theta = \frac{1}{1-a^2}F(a)$$

from which we readily obtain an interpolation constraint on $F(z)$ at $z = a$. In general, superior resolution is achieved by selecting data-dependent interpolation constraints at points proximal to the unit-disc sector of a targeted frequency band [6, 63]. The filter may reflect sensor dynamics, but it can also be virtual, focusing on the frequency range of interest. Given interpolation constraints for the power spectrum, a whole range of tools of the nonlinear methods [39] extends to this framework (encompassing so-called beamspace techniques in antenna arrays). The design of input-to-state filters and relevant tradeoffs between robustness and resolution have been addressed in [49], and will be part of our continuing research and development of such algorithms (available at: [36]).

We now highlight the case where the "Fourier vector" is replaced by a Green's/transfer function $G(e^{j\theta})$ with no apparent shift structure. Turning once more to the non-equispaced antenna array we introduced earlier, we seek a power density function $f(\theta)$ consistent with the statistics which is closest to a "prior" $f_{\text{prior}}(\theta)$ in the sense of, say, a Kullback-Leibler distance

$$\mathbb{S}(f||f_{\text{prior}}) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(f_{\text{prior}}\log(f_{\text{prior}}) - f_{\text{prior}}\log(f)\right)d\theta.$$

The minimizing solution can be written in closed form $f(\theta) = f_{\text{prior}}(\theta)/\text{Re}\{\lambda_o G(e^{j\theta})\}$ where λ_o denotes a (row) vector of Lagrange multipliers for the minimization problem. These multipliers can be easily computed so that $f(\theta)$ abides by the given statistics, **provided** of course that the statistics are consistent with the structure of $G(e^{j\theta})$ (which underscores the importance of our earlier considerations in approximating sample covariances). A homotopy method was proposed in [16, 17] leading to a differential equation for $\lambda(\tau)$ in a homotopy variable τ. If the statistics are consistent with the structure of $G(e^{j\theta})$, then $\lambda(\tau) \to \lambda_o$ as $\tau \to 1$, otherwise $\lambda(\tau)$ escapes to ∞. The rôle of f_{prior} is to introduce prior information, but can also be used to parametrize all solutions to the moment problem, since choices of f_{prior} lead to the complete set of f's such that

$$\mathbf{R} = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(e^{j\theta})f(\theta)G(e^{j\theta})^* d\theta. \tag{17}$$

We would like to emphasize that the theory in [17] applies to the case of matricial power density functions f_u (e.g., spectral density functions of multi-variable processes), as well as to cases where the support is multi-dimensional (e.g., space-time distributions, or $\theta \in \mathbb{R}^\ell$ with $\ell > 1$ in general) in which case the integrals are interpreted accordingly. A challenge at present is to improve the computational efficiency of obtaining density (possibly, matrix) functions, for either case where G has or has not a shift structure (i.e., in the case where G is the input-to-state transfer function of a linear filter, or simply a multivariable array manifold with no apparent structure, in general). Thus, we plan to address the following specific issues.

Problem 7 *Study the following "periodogram-like" analog for matricial moment problems*

$$f_{\text{perio}}(\theta) := \frac{1}{\ell+1} G(e^{j\theta})^* \mathbf{R} G(e^{j\theta}),$$

ℓ being the size of $\mathbf{R}$. More specifically, given an arbitrary (smooth matrix-valued) function $G(e^{j\theta})$ and a matricial moment $\mathbf{R}$ of compatible size, determine how far is $f_{\text{perio}}(\theta)$, as a matrix-valued density function, from being compatible with (17).

Direct comparison with (13) is very suggestive. Given the computational simplicity of the above formula, it is imperative to understand *when* it can be used. Very much as in the case of the classical periodogram, it is only an approximation. Thus, we plan to study in what sense it is an approximation and to develop quantitative answers to this question. Further, a choice of weights can dramatically enhance robustness of such a matrix-valued generalization of the periodogram. This echoes the way weights reduce variability of spectral estimates in the standard Blackman-Tukey techniques—yet, now, for multivariable spectral and arbitrary dynamics in A. A natural way that we may introduce weights is by using the algebraic structure of Schur products: given two matrices $\mathbf{R}, W$ of same size, the Schur product $\mathbf{R} \bullet W$ is defined as another matrix of the same size formed via term-wise multiplication of the entries of $\mathbf{R}$ and W, i.e., the (i,k)-entry of $\mathbf{R} \bullet W$ is $R_{i,k} \cdot W_{i,k}$. The Schur product is also known as Haddamard product and represents a commutative operation. An important fact is that if $\mathbf{R} > 0$ and $W > 0$, in the positive-definite sense, then so is their Schur product. Preliminary research suggests that the following generalization of the periodogram is especially versatile and useful.

Problem 8 *(Generalization of Blackman-Tukey) Let W be a positive definite matrix with the same size as $\mathbf{R}$ and let*

$$f_{\text{perio},W}(\theta) := G(e^{j\theta})^* (\mathbf{R} \bullet W) G(e^{j\theta}).$$

Determine the relationship between $f_{\text{perio},W}(\theta)$, the "weight function" $G(e^{j\theta})^* W G(e^{j\theta})$, and the class of matrix functions $f(\theta)$ consistent with (17).

Throughout the superscript * denotes the complex conjugate transpose. It is easy to see that $f_{\text{perio},W}(\theta)$ generalizes the classical periodogram when $\mathbf{R}$ is a Toeplitz matrix (and u_k scalar). In this case, $f_{\text{perio},W}(\theta)$ is a scalar function since G is the usual Fourier vector, and it is also the convolution of a generic function f which is consistent with (17) with $G(e^{j\theta})^* W G(e^{j\theta})$. Broadly speaking, our goal is to generalize the Blackman-Tukey techniques to the multivariable case by explaining how the choice of W affects the variability of the estimates, and how those relate to the underlying power spectra.

Similar set of issues will be taken up for a multivariable (also multi-dimensional) analog of (14). It is interesting to point out (see [28]) that when u_k is scalar, the density function in (14), as well as its analog when G, T_n are replaced by an input-to-state filter and the corresponding state-covariance $\mathbf{R}$, represent power spectral envelops. I.e., $\hat{f}(\theta)$ represents the maximal energy of a periodic signal at frequency θ, for each θ, which is compatible with the given statistics. No corresponding interpretation exists at the moment for the case where u_k is vectorial. Yet, experience suggests that an analogous interpretation ought to be true. The following summarizes pertinent issues we wish to pursue.

Problem 9 *For a given $W > 0$ of same size as $\mathbf{R}$, provide an interpretation of*

$$f_{\text{capon},W}(\theta) := \frac{1}{n+1} \left(G(e^{j\theta})^* (\mathbf{R}^{-1} \bullet W) G(e^{j\theta}) \right)^{-1}.$$

as it pertains to spectral density functions $f(\theta)$ consistent with (17).

This generalizes the Capon spectral envelopes [59], hence the subscript. The introduction of weights incurs minimal computation cost and it appears to have similar benefits as in classical periodogram techniques. The benefits, significance, and interpretation of such weights will be a subject for investigation. We will also study the sensitivity of the techniques on the structure of $\mathbf{R}$ and the benefits of the approximation theory we outlined earlier in correcting inconsistencies in sample covariances.

1.4 Correlation range and convex analysis

The observation that singularities in a covariance matrix reveal deterministic linear dependences between observed quantities, forms the basis of a wide range of techniques, from Gauss' least squares, to principle component analysis (PCA, GPCA), to modern subspace methods in time-series analysis. This observation suggests that a decomposition of covariance data into "signal + noise," in accordance with a suitable postulate, leads to identification of such deterministic dependences.

We first discuss the implications of the observation in time-series analysis. Here, one may seek a white-noise component of maximal variance which is consistent with estimated statistics. For instance, if u_k is a scalar random process as before, the minimal eigenvalue $\lambda_{\min}(\mathbf{T})$ of $\mathbf{T}$ represents the maximal power of white noise which is consistent with this autocorrelation data. Furthermore, $T_n - \lambda_{\min}(T_n)I$ is singular and corresponds to a deterministic random process made up of at most n-complex sinusoidal components. This fact (albeit in a different language) was already known to Carathéodory and Fejér in the early part of the 20th century. It was recognized by Pisarenko in the 1960's for its relevance in signal analysis and forms the basis of certain widely used high resolution methods for spectral analysis known as MUSIC (MUltiple SIgnal Classification) and ESPRIT (EStimation of Parameters by Rotational Invariant Techniques) —see [15, 28, 59].

A theory for a multivariable Carathéodory-Fejér-Pisarenko decomposition was presented recently by the P.I. in [32]. In [32] we have shown that after we account for white noise of maximal power, the remaining variance cannot be accounted for by pure sinusoids (it is considerably more complicated). Then, since the "white noise" hypothesis is often suspect anyway, and since in sensor arrays the hypothesis of mutual couplings and local scatterering effects suggests the presence of short range correlation noise (e.g., the analog of say, $MA(1)$ or $MA(2)$ in time-series), [32, 19] develop canonical decompositions accounting for noise with such short-range correlations. These problems are formulated as semi-definite programs and efficiently solved with existing software [4]. In this direction we plan to address the following.

Problem 10 *Extend the theory and techniques in [32] to a general transfer function/array manifold $G(\theta)$ with no apparent shift structure (i.e., not necessarily one in the form $(I - e^{j\theta}A)^{-1}B$). More specifically, given a covariance matrix $\mathbf{R}$ which originates by correlating outputs from a spatially distributed sensor array, develop efficient numerical techniques for decomposing $\mathbf{R}$ in accordance with the hypothesis that the data contain a strong short-range-correlation noise component.*

Thus, we plan to develop tools that are suitable to deal with structured noise statistics (typified by mutual couplings and interference in sensor arrays) as well as with the system theoretic maxim that a maximal set of dependences is to be sought. To this end, recent developments in imposing rank constraints in such additive decomposition of covariances [54, 47] will be studied (cf. Section 1.1 as well).

1.5 Multivariable identification and causal coherence

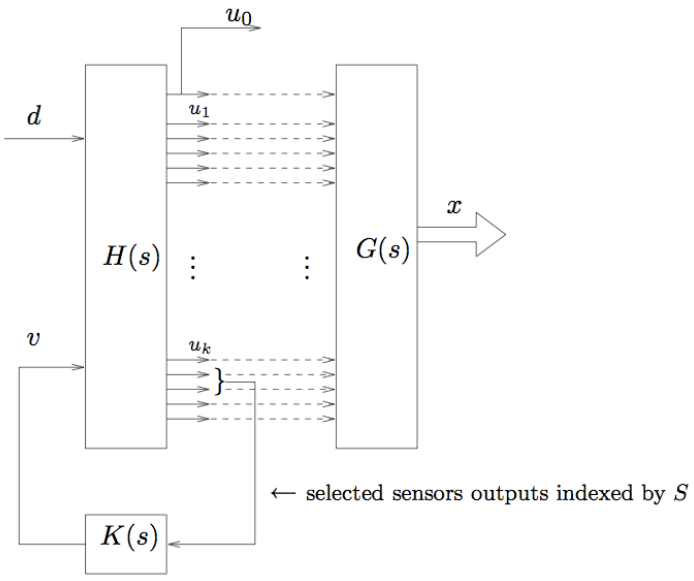
Questions regarding coherence and relevance of signals to one another are certainly not new. Yet, in a variety of applications, the size of the database and the purpose of seeking such signal affinities, demand a fresh and effective way to deal with such questions. The basic issue is to identify the dynamical dependance and quantify coherence between collections of signals. What we are aiming at is a way to select, among the readings available via an array of sensors, a most relevant signal (or collection of such) for use in feedback control. Coherence alone is not sufficient to designate a signal as relevant, since the phase is of crucial importance for control. For instance, a delayed version of a system output is highly correlated to the actual (without delay) output, yet it may be useless for disturbance rejection purposes, whereas another less coherent noisy signal may be preferable if it provides information on the output at an earlier time.

The setting we have selected begins with data collected via distributed sensor arrays. In a typical situation we may assume that (column) vectors of observations

$$u(t) = \begin{bmatrix} u_0(t) & u_1(t) & \dots & u_n(t) \end{bmatrix}'$$

where $t \in \{0, 1 \ldots, N\}$ (herein discrete-time and equi-spaced for simplicity) are available. The number n of entries is typically very large and the entries $u_k(t)$ themselves $(k = 0, 1, \ldots, n)$ could be vectorial in general. We often suppress the time dependence for notational convenience.

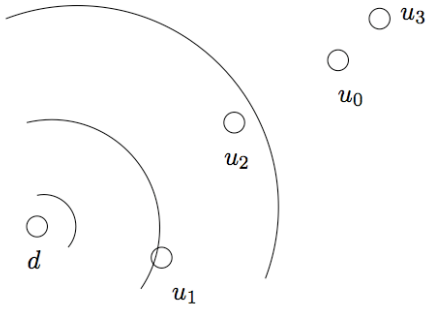
The entries u_k $(k = 0, 1, \ldots, N)$ represent measurements collected by an array of sensors which are distributed over a spatial domain at known spatial coordinates. The spatial medium is assumed non-dispersive and allows the flow of vibrations from unkown disturbance locations to some or all of the sensors. *The location where u_0 is being measured is special.* It may represent reading at the location of an instrument (laser gun, tracking device) which we seek to isolate from vibrations through feedback control. Thus, u_0 depends on a control actuation signal v and a disturbance d. A schematic on the right, represents the dynamical dependence between measurables u_k's and d, v via an unknown (linear) dynamical system with a $(n + 1) \times 2$-transfer function matrix H. The pathway from the control signal v to u_0 can be direct, with a transfer function of 1, if we chose to neglect actuator dynamics.



Feedback using a sub-collection of a sensor array

Our goal is to seek in a systematic manner a sub-collection of u_k-signals (ideally a "small" sub-collection) which most accurately captures the path of the disturbance through the medium and is maximally relevant in controlling u_0.

A simple schematic showing the relative location of sensors $0, 1, 2, 3$ along with the location where the disturbance enters into a planar medium is shown on the right. Large arcs are drawn to suggest wave propagation from the source of the disturbance towards the sensors. The dynamical dependence is characterized by a suitable transfer function $H_{u_k,d}$ in each case. Based on relative distances and locations, it is intuitively obvious that although u_0 and u_3 are highly correlated, u_3 is not very useful for control purposes since the disturbance impinges upon u_0 first, and then upon u_3. The ratio $H_{u_0,d}(e^{j\theta})/H_{u_3,d}(e^{j\theta})$ would manifest a "phase-lead" (i.e., "time-advance") reflecting the inverse of the time-delay between locations 0 and 3 along the path of the disturbance. Similarly, it is intuitively obvious that u_2 may be of significant value since its distance to the point of entry of the disturbance is less than that of u_1.



Vibrational source and array elements

In a static context, where no dynamic dependence is taken into account, classical coherence techniques are based on analysis of the covariance matrix $\mathbf{R} = E\{uu'\}$. Assuming all variables have zero mean, one typically computes and compares the Hotelling canonical correlations

$$c_{0,k} := \frac{\mathcal{E}\{u_0 u_k\}}{\sqrt{\mathcal{E}\{u_0^2\}\mathcal{E}\{u_k^2\}}} =: \cos(\theta_{0,k})$$

which represent cosines of the angle between the random variables as elements in an underlying Hilbert space. Similarly the angle ("minimal angle") between the subspace spanned by a collection u_k $(k \in S$

a set of indices) and u_0 is the arcsine of $\sqrt{1 - R_{00}^{-1} R_{0,S} R_{S,S}^{-1} R_{0,S}' R_{00}^{-1/2}}$, where $R_{00} = \mathcal{E}\{u_0^2\}$, R_{0S} is the vector of covariances between u_0 and those indexed by elements of S, and similarly R_{SS} is the covariance matrix of the sub-collection of the u_k's indexed by S. In this context, we are interested in the following:

Problem 11 *Identify a sub-collection of m random variable u_k so that u_0 is closest to their span, for any given m.*

When n is large, the search for such a "closest subspace" could be daunting. To simplify the task, one may use the Hadamard ratio

$$0 \leq \frac{\det(\mathbf{R})}{\prod_{\ell \in S} R_{\ell,\ell}}$$

which quantifies the linear dependence between elements indexed by S. Multivariable generalizations of these are quite natural. Indeed, the Hadamard product of, say a collection u_ℓ with $\ell \in S = S_1 \cup S_2$ ($S_1 \cap S_2 = \emptyset$) factors into the Hadamard product corresponding to each sub-collection of indices S_1, S_2 times an angular distance between the respective subspaces spanned by the random variables in the two sub-collections (e.g., see [56, Equation (40)]). Using such a tool it is possible to guide the search for relevancy among large numbers of sensor outputs u_k. Another useful tool is the algebra of the Schur complement [8]. For instance, u_k's which are maximally coherent with u_0 are necessarily near the principal components of the Schur complement $\mathbf{R}/R_{00} := R_{1:n,1:n} - R_{1:n,0} R_{00}^{-1} R_{0,1:n}$ (using a "Matlab-inspired" notation for ranges of indices) which can guide the relevant search. We plan to develop efficient tools for addressing the above problem. We also intend to study the relevance of low rank approximations of $\mathbf{R}$, or perhaps even better, of sparse approximations for $\mathbf{R}^{-1}$ (since the inverse covariance characterizes conditional dependence).

In a dynamic context, where the time-history is taken into account, we are interested in the coherence between random processes. This can be formalized via angular distances in the spectral domain [2]. In other words, a coherence function can be calculated at various frequencies as the sine of the angle between respective spectral components. An average coherence can be conveniently defined as a mean value (arithmetic, geometric, etc.) across a frequency range. Our goal is to address the following:

Problem 12 *Given a collection of time-series data $u_0(t), \ldots, u_n(t)$, ($t = 0, 1, \ldots, N$) as before, or possibly, given partial auto-covariance statistics for the vectorial $u := [u_0, \ldots, u_n]'$, determine a sub-collection of m variables (u_k with k in a suitable indexing subset $S \subset \{1, \ldots, n\}$), so that the mean value of the coherence function over a given frequency range is maximal.*

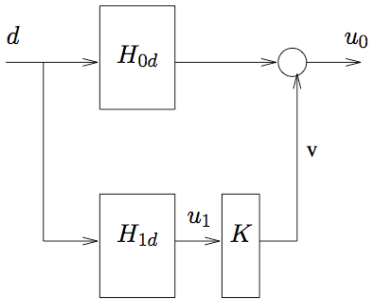
The particular way we quantify coherence is important. The above follow classical guidelines. However, if the sub-collection of random processes is to be used for feedback control, we need to base our selection on non-traditional metrics as we explain next.

Consider two random processes $u_0(t), u_1(t)$. If there is a dynamic dependence between the two, as in the case where both are outputs of a linear dynamical system with a scalar "disturbance" input d, the joint spectral density function is of rank 1 across frequencies because

$$f(\theta) = \begin{bmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{bmatrix} = \begin{bmatrix} H_{0d}(e^{j\theta}) \\ H_{1d}(e^{j\theta}) \end{bmatrix} f_{dd}(\theta) \begin{bmatrix} H_{0d}(e^{j\theta})^* & H_{1d}(e^{j\theta})^* \end{bmatrix},$$

where f_{dd} is the spectral density of the disturbance and H_{id} represents the transfer function from d to u_i. Given $f(\theta)$ we can readily identify the fraction $H_{0d}(e^{j\theta})/H_{1d}(e^{j\theta}) = f_{00}(\theta)/f_{01}(\theta)$.

Now consider that u_1 is to be used for mediating the effect of the disturbance d to the node u_0, via a suitably designed controller K as shown in the schematic on the right. In this, for simplicity, we assume that the control signal v affects u_0 directly. Our control authority in mediating the effect of disturbances depends heavily on the relationship between the particular transfer functions $H_{id}(s)$ (e.g., for $i \in \{0, 1\}$). Indeed, the quality of disturbance attenuation is readily quantified as the solution to the following standard (Nehari) problem in H_∞-control:



Disturbance rejection "wiring" diagram

$$\inf_{K \in H_\infty} \|H_{0d}(z) - K(z)H_{1d}(z)\| = \|\Gamma_{H_{0d}(z)/H_{1d}(z)}\|$$

taking H_{1d} to be inner. Here $\Gamma_{H(z)}$ denotes the Hankel operator with symbol $H(z)$ as is customary in H_∞-control. The above is a standard model-matching problem. We also note that it is quite easy and natural to introduce weights so as to incorporate prior information about the spectrum of d.

The coherence function between u_0 and u_1 is given by

$$\gamma(\theta) = \frac{|f_{01}|}{\sqrt{f_{00}f_{11}}}$$

and is identically equal to 1 under the above assumption of "noise free" u_0, u_1. In general, the rank of $f(\theta)$ will be 2 and the coherence < 1. Yet, an appropriate quantity which quantifies our "control authority" at the node u_0 when we know u_1, is the norm

$$\|\Gamma_{f_{00}/f_{01}}\|$$

of the Hankel operator with symbol $f_{01}(\theta)/f_{00}(\theta) = H_{0d}(e^{j\theta})/H_{1d}(e^{j\theta})$ as discussed earlier. Returning to the situation of many signals $u_0(t), \ldots, u_n(t)$ $(t = 0, 1, \ldots, N)$, we can similarly argue that the quantity which captures our authority in controlling u_0 based on measurements of $u_\ell(t)$, $\ell \in S$ is

$$\inf_{K_\ell \in H_\infty, \ell \in S} \|H_{0d} - \sum_{\ell \in S} H_{0\ell}K_\ell\|_\infty. \tag{18}$$

Estimates of H_{0d}/H_{id} can be computed directly from the data and their statistics. For example, we can estimate covariances $R_{ij}(\ell) = E\{u_i(t)u_j(t+\ell)'\}$ and then utilize either the "central" solution to respective moment problems given in [31] or multivariable periodograms described in Section 1.3, in order to estimate the required ratios. High resolution estimates of such ratios can be obtained using (generalized) statistics via an "input-to-state" filter which aggregates all measurements as shown in the schematic of page 13. The input-to-state filter is denoted by $G(s)$. This brings us to one final key problem we wish to address:

Problem 13 *With the notation and context given above, determine a sub-collection of m variables u_k, so that they are maximally relevant in negating the effect of a disturbance d on u_0, in the sense of minimizing (18).*

References

[1] S. Amari, **Differential-geometrical methods in statistics**, Lecture notes in statistics, Springer-Verlag, Berlin, 1985.

[2] J.S. Bendat and A.G. Piersol, **Engineering Applications of Correlation and Spectral Analysis**, Wiley-Interscience, Second Edition, 1993.

[3] E.F. Beckenbach and R. Bellman, **Inequalities**, Springer-Verlag, Berlin-Heidelberg, 198 pages,1965.

[4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, vol. 15, Studies in Applied Mathematics, SIAM, Philadelphia, PS, June 1994.

[5] C. I. Byrnes, T.T. Georgiou, and A. Lindquist, "A generalized entropy criterion for rational Nevanlinna-Pick interpolation: A convex optimization approach to certain problems in systems and control," *IEEE Trans. on Automatic Control*, **45(6):** 822-839, June 2001.

[6] C. I. Byrnes, T.T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. on Signal Proc.*, **48(11):** 3189-3206, November 2000.

[7] C.I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in H^∞ with a complexity constraint," *Trans. of the American Math. Society*, (electronically published on December 9, 2004) *Trans. Amer. Math. Soc.* **358(3)**: 965-987, 2006.

[8] R.W. Cottle, "Manifestations of the Schur Complement," *Linear Algebra and its Appl.*, **8:** 189-211, 1974.

[9] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Probability,* **19(4)**: 2032-2066, 1991.

[10] B.W. Dickinson, "Two dimensional Markov spectrum estimates need not exist," *IEEE Trans. on Information Theory*, **IT-26**: 120-121, 1980.

[11] A. Ferrante, M. Pavon, F. Ramponi "Hellinger vs. Kullback-Leibler multivariable spectrum approximation," (http://arxiv.org/abs/math..OC/0702212), February 2007.

[12] B.R. Frieden, **Science from Fisher information: a unification**, Cambridge University Press, 2004.

[13] T.T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. on Information Theory*, **49(11)**, November 2003.

[14] T.T. Georgiou, "The interpolation problem with a degree constraint," *IEEE Trans. on Automatic Control*, **44(3)**: 631-635, March 1999.

[15] T.T. Georgiou, "Signal Estimation via Selective Harmonic Amplification: MUSIC redux," *IEEE Trans. on Signal Processing*, **48(3)**: 780-790, March 2000.

[16] T.T. Georgiou, "Solution of the general moment problem via a one-parameter imbedding," *IEEE Trans. on Automatic Control*, **50(6)**: 811-826, June 2005.

[17] T.T. Georgiou, "Relative Entropy and the multi-variable multi-dimensional Moment Problem," *IEEE Trans. on Information Theory*, **52(3):** 1052 - 1066, March 2006.

[18] T.T. Georgiou and A. Lindquist, "Remarks on control design with degree constraint," *IEEE Trans. on Automatic Control,* **51(7):** 1150-1156, July 2006.

[19] T.T. Georgiou, "Decomposition of Toeplitz matrices via convex optimization," *IEEE Signal Processing Letters,* **13(9):** 537- 540, September 2006.

[20] T.T. Georgiou, "A differential equation for the LMI feasibility problem," submitted to *IEEE Trans. on Automatic Control, under revision.*

[21] T.T. Georgiou, C. Shankwitz and M.C. Smith, "Identification of linear systems: a stochastic approach based on the graph," Proceedings of the 1992 American Control Conference, Chicago, June 1992, pp. 307-312.

[22] T.T. Georgiou, " Toeplitz covariance matrices and the von Neumann relative entropy," in Control and Modeling of Complex Systems: Cybernetics in the 21st Century: Festschrift volume for Professor H. Kimura; K. Hashimoto, Y. Oishi, and Y. Yamamoto, Eds. Boston, MA: Birkhauser, 2003.

[23] T.T. Georgiou, " Structured covariances and related approximation questions," in Lecture Notes in Control and Information Sciences, Volume 286, 2003, pp. 135-140, Springer Verlag, 2003.

[24] T.T. Georgiou, " The mixing of state covariances," in Lecture Notes in Control and Information Sciences, Volume 289, 2003, pp. 207-212, Springer Verlag, 2003.

[25] T.T. Georgiou and A. Lindquist, "Two Alternative Views on Control Design with Degree Constraint," Proc. 44th IEEE Conference on Decision and Control, December 2005, pp. 3645-3650.

[26] T.T. Georgiou, "Relative Entropy and Moment Problems," Proc. 44th IEEE Conference on Decision and Control, December 2005, pp. 4397 - 4403.

[27] T.T. Georgiou, "Singular decomposition of state covariances," Proceedings of the 2006 American Control Conference, 6 pp., June 2006.

[28] T.T. Georgiou, " Spectral Estimation via Selective Harmonic Amplification, *IEEE Trans. on Automatic Contr.*," January 2001, **46(1)**: 29-42.

[29] T.T. Georgiou, "Distances between time-series and their autocorrelation statistics, (http://arxiv.org/abs/math.OC0701181); January 2007.

[30] T.T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. on Automatic Control*, **47(7):** 1056-1066, July 2002.

[31] T.T. Georgiou, "Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization," *IEEE Trans. on Automatic Control*, **47(11):** 1811-1823, November 2002.

[32] T.T. Georgiou, "The Caratheéodory-Fejér-Pisarenko decomposition and its multivariable counterpart," (http://arxiv.org/abs/math.PR/0509225); *IEEE Trans. on Automatic Control*, to appear in February 2007.

[33] T.T. Georgiou, "The maximum entropy ansatz in the absence of a time-arrow: fractional pole models," preprint at: (http://arxiv.org/abs/math/0601648/), submitted to the *IEEE Trans. on Information Theory*, January 2006.

[34] T.T. Georgiou, "Distances between power spectral densities," (http://arxiv.org/abs/math/0607026), *IEEE Trans. on Signal Processing*, accepted, to appear in 2007.

[35] T.T. Georgiou, "An intrinsic metric for power spectral density functions," preprint at: (http://arxiv.org/abs/math/0608486), *IEEE Signal Processing Letters*, accepted, to appear, August 2007.

[36] http://www.ece.umn.edu/users/georgiou/files/reports.html

[37] Ya. L. Geronimus, **Orthogonal Polynomials**, English translation from Russian by Consultants Bureau, New York, 570 pages, 1961.

[38] U. Grenander and G. Szegö, **Toeplitz Forms and their Applications**, Chelsea, 1958.

[39] S. Haykin, **Nonlinear Methods of Spectral Analysis,** Springer-Verlag, New York, 247 pages, 1979.

[40] E.T. Jaynes, "On the rationale of maximum entropy methods," *Proc. IEEE*, **70**: 939-952, 1982.

[41] R.E. Kalman, "Identification of noisy systems," *Russian Math. Surveys*, **40(4)**: 25-42, 1985.

[42] J. Karlsson and T.T. Georgiou, "Signal analysis, moment problems & uncertainty measures," Proc. 44th IEEE Conference on Decision and Control, December 2005, pp. 5710 - 5715.

[43] M. Kumar, D.P. Garg, and R.A. Zachery, "A generalized approach for inconsistency detection in data fusion from multiple sensors," Proc. of the 2006 American Control Conf. pages 2078-2083, June 2003.

[44] S.W. Lang and J.H. McClellan, "Multidimensional MEM spectral estimation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-30(6)**: 880-887, 1982.

[45] R.D. Levine and M. Tribus (editors), **The Maximum Entropy Formalism**, MIT Press, Cambridge, 1979.

[46] B.C. Levy, R. Frezza, and A.J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," IEEE Trans. Automatic Control, 35:1013–1023, September 1990.

[47] M. Mesbahi and G.P. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," *IEEE Trans. on Aut. Control*, **42(2)**: 239-243, 1997.

[48] A. Nasiri Amini, E. Ebbini, and T.T. Georgiou, "Noninvasive tissue temperature estimation using high-resolution spectral analysis techniques," *IEEE Trans. on Biomedical Engineering*, **52(2)**: 221-228, February 2005.

[49] A. Nasiri Amini and T.T. Georgiou, "Tunable Spectral Estimators Based on State-Covariance Subspace Analysis," *IEEE Trans. on Signal Processing,* **54(7):** 2662-2671, July 2006.

[50] A. Nasiri Amini, E. Ebbini, and T. Georgiou, "Noninvasive tissue temperature estimation via state-covariance spectral estimation," *Proc. 10th IEEE DSP workshop,* Taos, NM, August 2004.

[51] A. Nasiri Amini and T.T. Georgiou, "SAR imaging via state-covariance subspace methods," submitted to *IEEE Trans. on Aerospace and Electronics Systems*, under revision.

[52] A. Nasiri Amini and T.T. Georgiou, "Avoiding Ambiguity in Beamspace Processing," *IEEE Signal Processing Letters,* **12(5)**: 372 - 375, May 2005.

[53] A. Nasiri Amini, S. Takyar, and T.T. Georgiou, "A Homotopy Approach for Multirate Spectrum Estimation," Proceedings of the 2006 IEEE International Conf. on Acoustics, Speech and Signal Processing, vol. 3, pages 532-535, May 2006.

[54] P. Parillo, **Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization**, Ph.D. thesis, Caltech, 2000.

[55] O. Reiersol, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, **9**: 1-24, 1950.

[56] L.L. Scharf and C.T. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Trans. on Signal Processing*, **48(3)**: 824-831, March 2000.

[57] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Trans. on Signal Processing,* **48**: 1999-2012, July 2000.

[58] B. Dumitrescu, I. Tabus, and P. Stoica, "On the parametrization of positive real sequences and MA parameter estimation," *IEEE Trans. on Signal Processing,* **49**: 2630-2639, November 2001.

[59] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 2005.

[60] S. Takyar, A. Nasiri Amini, and T.T. Georgiou, "Sensitivity shaping with degree constraint via semidefinite programming," Proceedings of the 2006 American Control Conference, 4 pp., June 2006.

[61] S. Takyar, Ph.D. thesis in preparation, University of Minnesota.

[62] S. Takyar, A. Nasiri Amini, and T.T. Georgiou, "Spectral analysis from multi-rate observations: Itakura-Saito based approximation of power spectra and Capon-like spectral envelopes," Proceedings of the Fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM-2006), Waltham, Massachusetts, July 2006

[63] United States. Patent 6,400,310, "A tunable high-resolution spectral estimator," (joint co-inventors with C.I. Byrnes and A. Lindquist), June 4, 2002.

[64] H.L. Van Trees, **Optimum Array Processing: part IV of Detection, Estimation and Modulation Theory**, Wiley-Interscience, 2002.

[65] S.R.S. Varadhan, **Probability Theory**, AMS, 2000.

[66] S. Varigonda, T.T. Georgiou, and P. Daoutidis, "Numerical solution of the optimal periodic control problem using differential flatness," *IEEE Trans. on Automatic Control*, **49(2)**:271 - 275, February 2004.

[67] P. Whittle, **Prediction and Regulation**, The English Universities Press, 1963.