

(4)

FILE COPY

AD-A 198 867

HIGH RESOLUTION RADAR IMAGING

Semi-Annual Progress Report No. 4,
O.N.R. Contract N00014-86-K-0370
Period: 1 December 1987—31 May 1988

**ELECTRONIC SYSTEMS
AND SIGNALS
RESEARCH LABORATORY**

Department of Electrical Engineering
Campus Box 1127
One Brookings Drive
Washington University

Principal Investigator: Donald L. Snyder

DTIC
ELECTE
AUG 24 1988
S
A
D

This document has been approved
for public release and sale in
distribution is unlimited.

014

HIGH RESOLUTION RADAR IMAGING

Semi-Annual Progress Report

Office of Naval Research Contract Number N00014-86-K-0370

Period Covered: 1 December 1987 - 30 May 1988

Principal Investigator:

Donald L. Snyder
Director, Electronic Systems and Signals Research Laboratory
Washington University
Campus Box 1127
One Brookings Drive
St. Louis, Missouri 63130

Scientific Program Director:
Dr. Rabinder Madan
Office of Naval Research
Code 1114SE
800 North Quincy Street
Arlington, Virginia 22217-5000

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<i>NA</i>
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	



DISTRIBUTION

	<u>copies</u>
Mr. John W. Michalski Office of Naval Research Resident Representative Federal Building, Room 286 536 South Clark Street Chicago, Illinois 60605-1588	1
Dr. Rabinder Madan Office of Naval Research Code 1114SE 800 North Quincy Street Arlington, Virginia 22217-5000	1
Director Naval Research Laboratory Attn: Code 2627 Washington, DC 20375	1
Defense Technical Information Center Building 5 Cameron Station Alexandria, Virginia 22314	12
Mr. Harper J. Whitehouse Naval Ocean Systems Center Code 7402 San Diego, California 92152	1
Dr. James Fienup Environmental Research Institute of Michigan P. O. Box 5615 Ann Arbor, Michigan 48107	1
Dr. Kenneth Senne M. I. T. Lincoln Laboratory Lexington, Massachusetts 02173	1
Dr. Richard E. Blahut I. B. M. Owego, New York	1

Table of Contents

1. Introduction	1
2. Summary of Work Accomplished	6
2.1. Estimation-Theory Approach to Imaging	6
2.1.1. Inclusion of a Specular Component	6
2.1.2. Performance Evaluation	6
2.1.3. Forming Images Subject to a Constraint on Input Signal-to-Noise Ratio ..	7
2.1.4. Conditions for Uniqueness of Target Images	7
2.1.5. Signal Selection for Target Imaging	7
2.2. Chirp-Rate Modulation Approach to Imaging	7
3. References	7
4. Appendices	9
4.1. Appendix 1. Imaging Specular Targets	9
4.2. Appendix 2. Performance Analysis, Forming Images under a Power Constraint	13
4.3. Appendix 3. Performance Analysis, Signal Design	25
4.4. Personnel	33

1. Introduction

This semi-annual progress report contains a summary of work accomplished on O. N. R. Contract N00014-86-K-0370, *High Resolution Radar Imaging*, during the period from 1 December 1987 to 31 May 1988.

The goal of this project is to formulate and investigate new approaches for forming images of radar targets from spotlight-mode, delay-doppler measurements. These measurements could be acquired with a high-resolution radar-imaging system operating with an optical- or radio-frequency carrier. Two approaches are under study. The first is motivated by an image-reconstruction method used in radionuclide imaging called the *confidence-weighted algorithm*; here, we refer to this approach as the *chirp-rate modulation approach*. The second approach is based on more fundamental principles which starts with a mathematical model that accurately describes the physics of an imaging radar-system and then uses statistical-estimation theory with this model to derive methods for producing target images; we refer to this as the *estimation-theory approach*. — (RFA)

Figs. 1 to 4 summarize a broad context in which forming radar images is important and the way in which the methods we are developing may be useful in this context. Fig. 1 depicts a

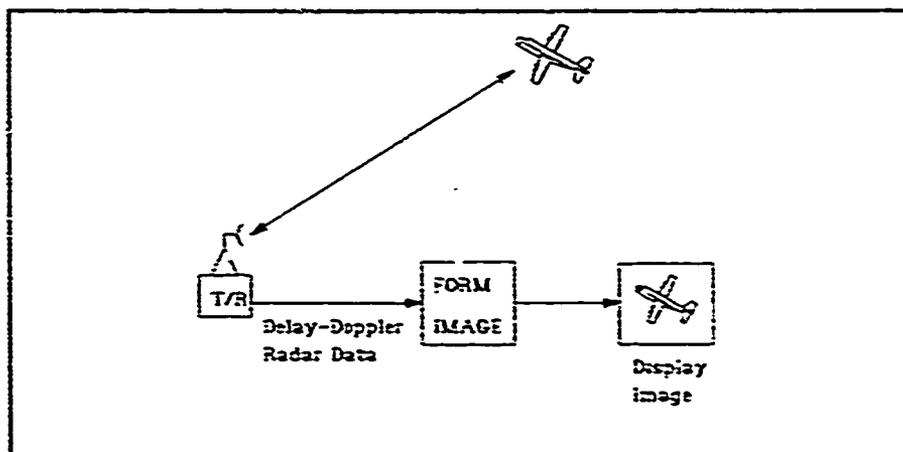


Figure 1. Radar Imaging System

A radar transmitter/receiver illuminates the target with a series of pulses and observes echo-data containing delay-doppler information. These data are processed to form an image of the target.

radar imaging system. Shown is a single target, but multi-target and clutter environments can be accommodated as well. Images are conventionally formed by processing data acquired with a series of target illuminations using chirped-sinusoidal or stepped-frequency waveforms. The usual two-dimensional Fourier transform processing, described by Wehner [1], is based on a deterministic model in which there is no randomness in the target's reflectivity and no noise in the radar data, an idealization which is often not met in practice. Another form of processing, described by Mensa [2], is based on tomographic principles and is also based on a deterministic model for the problem. The chirp-rate modulation approach which we are developing is based on tomographic principles; it extends the approach described by Mensa through the use of a variable chirp rate, as first described by Bernfeld [3] and Feig and Grunbaum [4], so that only small angular rotations of the target are necessary for forming the image. Our tomographic method is based upon developments in radionuclide imaging and differs from that of Bernfeld, Feig and Grunbaum in that we can accommodate practical ambiguity functions that do not yield the ideal line integrals required with the usual tomographic approaches they employ. The more fundamental estimation-theory approach which we are developing uses a target model that accounts for diffuse, random reflectivity and for noise. The image formation process is derived mathematically using statistical-estimation theory. The new processing we obtain is more complex and computationally demanding than either the Fourier-transform or tomographic approaches, but improved images can result because it more accurately accommodates the physics of radar imaging problem. At the present stage of our investigations, we are not overly concerned about the computational load of the approach because the equations to be evaluated are quite amenable to evaluation on massively parallel computing architectures.

We are beginning to evaluate the performance of the estimation-theory approach in comparison to conventional Fourier transform processing. Only an extremely preliminary result is available now, but we include it to indicate what we hope the gains with our method may be as we complete a more complex evaluation. Shown in Fig. 2 are the average results obtained from 3000 realizations with a computer simulation in which a single point-target of unit scattering strength (i.e., random reflectivity with power $\sigma^2 = 1$) is illuminated with a sinusoidal pulse.

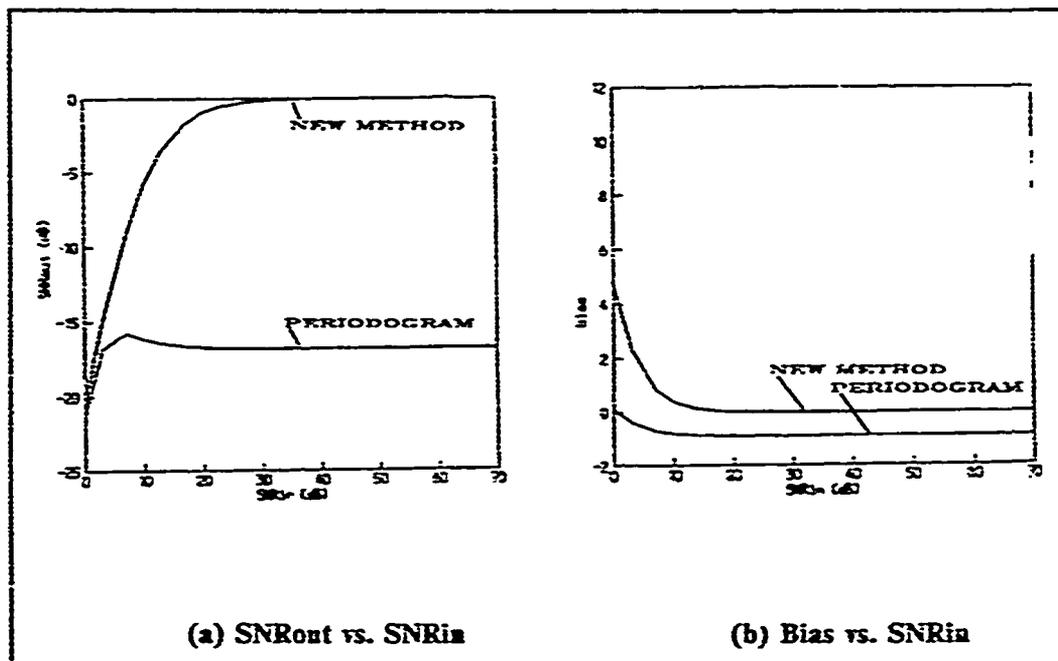


Figure 2. Performance Comparison for a Point Target

Shown are the output signal-to-noise ratio and bias for estimating the scattering function of a point target with a periodogram and with the estimation-theory method.

Shown are the bias, defined by

$$\text{BIAS} = E[\hat{\sigma}^2] - \sigma^2,$$

and output signal-to-noise ratio, defined by,

$$\text{SNR}_{\text{out}} = \frac{\sigma^2}{\sqrt{\text{MSE}}},$$

where MSE is the mean-square error defined by

$$\text{MSE} = E[(\hat{\sigma}^2)^2] - E^2[\hat{\sigma}^2].$$

The input signal-to-noise ratio is defined by

$$\text{SNR}_{\text{in}} = \frac{\sigma^2}{N_0} = \frac{1}{N_0}.$$

where N_0 is the power density of additive receiver noise. The scattering strength of the point target was estimated using a Fourier transform method (periodogram) and our estimation-theory method. For input signal-to-noise ratios above about 5 dB, the periodogram estimate is strongly biased and has a low output signal-to-noise ratio. In this same regime, the estimation-theory method yields an unbiased estimate with about a 15 dB greater output signal-to-noise ratio. While this is an extremely elementary example, it does give us some optimism that the new approach will result in improved images compared to the Fourier transform approach when applied in more realistic situations.

Shown in Fig. 3 is a target identification system that uses radar images to aid in the

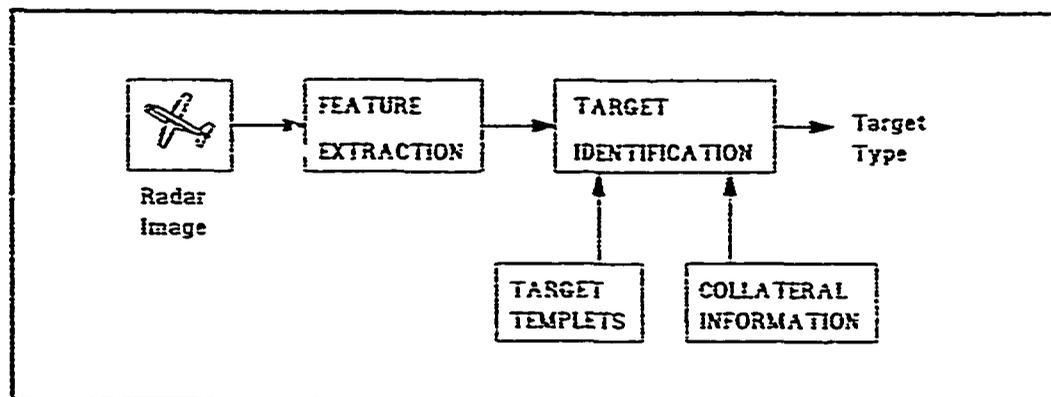


Figure 3. A Target Identification System

Shown is a system in which targets are identified on the basis of features extracted from radar images of the target, target templates, and other collateral information.

identification process. Such a system commonly starts with an image formed by the Fourier-transform method. Features such as edges, regions of similar texture, and shapes are extracted. These are then used with a catalog of possible targets and other available information, such as target-track data, surveillance data, and target environment, to identify the target type. Since the identification process will be aided by having improved target images from which to extract features, we expect that more reliable identification will result when images produced with our estimation-theory method are used. Note that in this system, two separate and independent steps are involved, that of image formation followed by that of feature extraction for target identification. One may expect that combining these two steps in a unified manner should result in further improvements in identification.

Shown in Fig. 4 is a potentially improved target identification system in which images are

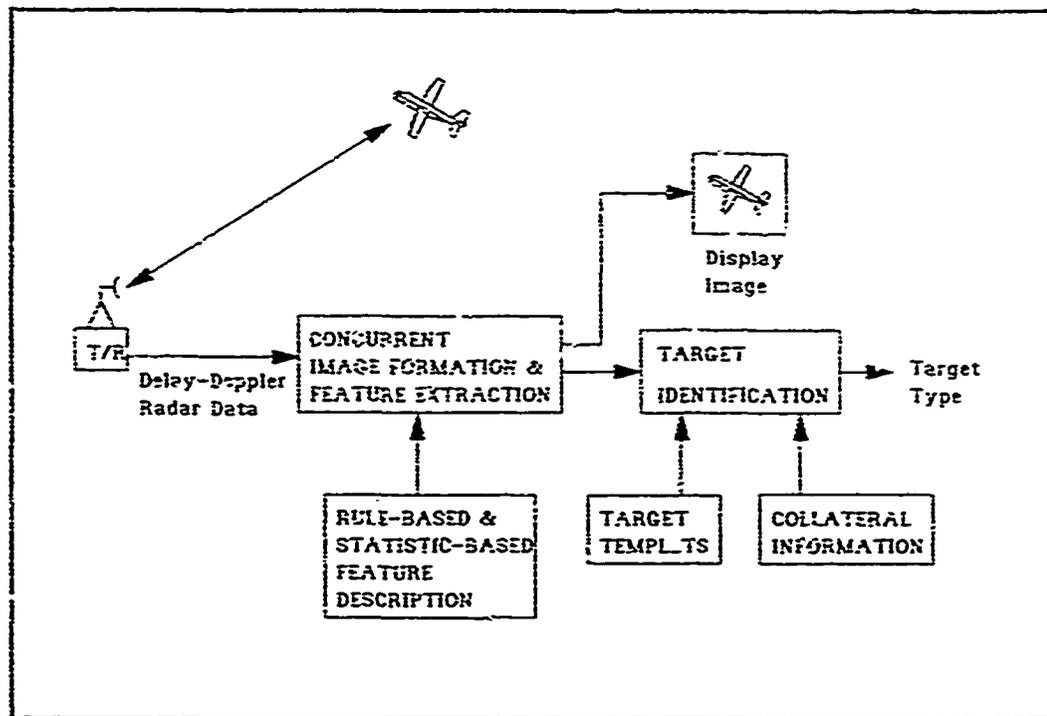


Figure 4. A Unified Target Imaging and Identification System

Shown is a system in which targets are imaged and identified in a coordinated manner in an extended estimation-theory method in which rule- and statistic-based constraints are recognized.

formed and features extracted in a combined, unified manner. This is accomplished by incorporating constraints during the image formation and extraction process. Constraints may be rule-based, such as to enforce curvature rates on edges and closure of boundaries, or statistic-based, such as roughness of texture and prior statistical knowledge of the target. The estimation-theory approach we are developing is ideally suited for incorporating rule and statistic based constraints in a single, coordinated process that should result in improved target identification.

At the present time, research under this contract is focused on two methods to improve target images. Such improvements would be useful in the broad context of target identification, particularly for the separated identification system of Fig. 3. In our laboratory, a strong research effort on image estimation subject to rule- and statistic-based constraints has been

initiated [5], with a view towards implementations on highly parallel computer architectures. We envision that these developments will be important for a unified target imaging and identification system in the form of Fig. 4.

2. Summary of Work Accomplished

2.1. Estimation-Theory Approach to Imaging

Progress during this reporting period has been made on: a, extending the estimation-theory approach to include a specular or glint component in the radar-echo data in addition to the diffuse component presently in the model; b, analyzing the performance of the estimation-theory approach both analytically through Cramér-Rao bounds and with simple computer-simulations; c, extending the estimation-theory approach to accommodate constraints on the received signal-to-noise ratio; d, identifying analytical conditions for the uniqueness of images produced with the estimation-theory approach; and e, addressing how radar signals should be selected to form good images. Each of these areas is described briefly below and more completely in the appendices.

2.1.1. Inclusion of a Specular Component

The estimation-theory approach to imaging which we have developed to date is based upon targets having a diffuse reflectivity with no specular or glint component; the reflectivity is modeled as a zero-mean, complex-valued Gaussian random process [6]. We are now developing an analogous approach for targets that are described by a collection of purely specular reflections with no diffuse component. The model we are using and a status report for this effort are contained in Appendix 1, by K. Krause.

2.1.2. Performance Evaluation

The estimation-theory approach to imaging yields an iterative algorithm for producing target images [6]. As a result, the performance of the resulting images is difficult to predict analytically. For this reason, we have recently developed a Cramér-Rao lower bound on the mean square-error in estimating the target's scattering function. This is discussed in Appendices 2 and 3. Another approach we are beginning to use for studying the performance

and for making comparisons to alternative image formation strategies is computer simulations. These are very demanding computationally. Our first results are, therefore, for very simple special cases. These are discussed in Appendix 2.

2.1.3. Forming Images Subject to a Constraint on Input Signal-to-Noise Ratio

We have extended the estimation-theory approach to include a specified input signal-to-noise ratio. The result, obtained by P. Moulin, is discussed in Appendix 2.

2.1.4. Conditions for Uniqueness of Target Images

Conditions for the uniqueness of target images formed with the estimation-theory method have been identified by J. O'Sullivan in terms of the Fisher information-matrix and the Cramér-Rao bound. This is discussed in Appendix 3.

2.1.5. Signal Selection for Target Imaging

An effort has been initiated to establish a method for determining good radar signals to transmit for delay-doppler imaging. Preliminary results, obtained by J. O'Sullivan, are discussed in Appendix 3.

2.2. Chirp-Rate Modulation Approach to Imaging

Work is continuing on the chirp-rate modulation approach to imaging. Using maximum-likelihood estimation, the processing to be used on radar-echo signals has been identified when a linearity constraint is placed on the processing. This *confidence-weighted* processing consists of a bank of bandpass matched filters followed by additional filtering specified by solving a set of normal equations. These equations are valid for a variety of modulation formats, and we are presently specializing them for the stepped-frequency waveform so that we may compare the processing to conventional two-dimensional Fourier transform processing.

3. References

1. D. R. Wehner, *High Resolution Radar*, Artech House, Dedham, MA, 1987.
2. D. L. Measa, *High Resolution Radar Imaging*, Artech House, Dedham, MA, 1984.

3. M. Bernfeld, "CHIRP Doppler Radar," *Proceedings of the IEEE*, Vol. 72, pp. 540-541, April 1984.
4. E. Feig and F. A. Grunbaum, "Tomographic Methods in Range-Doppler Radar," *Inverse Problems*, Vol. 2, pp. 185-195, 1986.
5. M. I. Miller, "Mapping Rule-Based Regular Grammars to Connection Architectures Via Markov Random Fields: Applications to Image Reconstruction," presented at the AMS-IMS-SIAM Joint Summer Research Conference on *Spatial Statistics and Imaging*, (A. Poggio, University of Washington, Chairman), Bowdoin College, June 18-24, 1988.
6. D. L. Snyder, J. A. O'Sullivan, and M. I. Miller. "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar Targets," *IEEE Transaction on Information Theory*, to appear 1988.

4. Appendices

4.1. Appendix 1. Imaging Specular Targets

Maximum-Likelihood Approach to Specular Imaging

Kenneth E. Krause

K. E. Krause
June 9, 1988

STATUS REPORT:

MAXIMUM LIKELIHOOD APPROACH TO SPECULAR TARGET IMAGING

The following describes progress in statistical model formulation and imaging approach for the Maximum Likelihood Estimation based imaging of delay and doppler spread specular targets. That is, targets composed of multiple reflectors, each of which is characterized by a specular reflection process. According to Wehner (ref. 1) the reflectors on such targets may be thought of as having the following properties: reasonably constant reflection amplitude with respect to small aspect change, rapidly varying phase with respect to aspect, and weak (for this effort negligible) relationship of echoing area to frequency.

Target Model The target is assumed to be composed of an array, in the delay-doppler plane, of specular reflectors as described above. The generic return from each reflector, with attributes sufficient to capture the aspects of Wehner's properties, is postulated. It is

$$s_r(t, \theta_{nm}, B_{nm}) = \sqrt{2E_t} B_{nm} f(t - \tau_n) \cos [2\pi(f_0 + f_m)t - 2\pi f_0 \tau_n + \phi(t - \tau_n) + \theta_{nm}]$$

where t is the continuous time variable, τ_n is the scatterer delay coordinate, f_m is the scatterer doppler coordinate, E_t is the transmitted signal energy, B_{nm} is the deterministic amplitude of the reflection coefficient, f_0 is the radar carrier frequency, $f(t)$ is the amplitude of the modulation's complex envelope, $\phi(t)$ is the phase of the carrier modulation's complex envelope, and θ_{nm} is the random phase of the reflection coefficient. Phase θ_{nm} is assumed to arise from the following probability density given in ref. 2:

$$p(\theta_{nm}, \Lambda_{nm}) = \frac{\exp[\Lambda_{nm} \cos(\theta_{nm})]}{2\pi I_0(\Lambda_{nm})} \quad -\pi < \theta_{nm} < \pi$$

Λ_{nm} is a constant which can be adjusted to model the a priori knowledge of the randomness of the phase variations. A value 0 implies a uniform distribution

or nothing known about the phase and a value of infinity implies the phase is known exactly.

Imaging Approach The received signal available from which to form images is a sum of returns from elements like those described above located at an array of locations (τ_n, f_m) in the delay-doppler plane. This signal is corrupted by additive white gaussian noise. The received signal may thus be expressed in terms of the generic returns from individual scattering elements as:

$$s_{r2D}(t; \Theta, B) = \sum_{n=0}^{N-1} \sum_{m=-M}^M s_r(t; \theta_{nm}, B_{nm}) + w(t) \quad 0 \ll T$$

N is the number of discrete delay coordinates where scatterers exist and M is the number of discrete doppler coordinates on either side of 0 at which scatterers exist. $w(t)$ is the additive white gaussian noise.

The desired image is the maximum likelihood estimate of the magnitude of B_{nm} at each point in the target array.

Technique 1 By making some assumptions on the relation between T and the doppler resolution and by restricting the class of allowed complex envelopes somewhat, a likelihood ratio can be derived which has the form of a product of ref. 2's likelihood ratios for the detection of a known signal with an unwanted phase. Thus, the 2D imaging problem may reduce in level of complexity to that of estimating the reflection magnitude of a single generic target. To date, the form of the solution for case $\Lambda_{nm}=0$ and Λ_{nm} approaching infinity have been investigated. A relatively simple expression results for the latter case (corresponding to phase known exactly). For $\Lambda_{nm}=0$ (phase uniformly distributed), an expression containing modified Bessel functions of the first kind and orders 0 and 1 results which would have to be iteratively matched with a linear function of their arguments to converge on a solution. The case for Λ_{nm} arbitrary yields even more complicated expressions which would have to be solved using computer iterations.

Technique 2 Currently, the EM-algorithm of ref. 3 is being investigated in quest of a more tractable iterative solution to the problem. The first and most obvious form of the complete data space model looked at in this regard appears to have a degeneracy which practically would cause the accumulation of time samples beyond the first 2 to not be meaningfully used in the estimation procedure. This will be rechecked. The next step, currently in work, is to look

for a new complete data space model which will provide meaningful utilization of all data with improved tractability over Technique 1.

REFERENCES

1. D. R. Wehner, High Resolution Radar (Artech House, 1987)
2. H. L. Van Trees, Detection, Estimation, and Modulation Theory-Part I (Wiley, 1968)
3. A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *J. of the Royal Statistical Society, Vol. B, 39*, pp. 1-37, 1977.

4.2. Appendix 2. Performance Analysis, Forming Images under a Power Constraint

Performance Analysis for Maximum-Likelihood Spectrum

Pierre Moulin

Donald L. Snyder

Joseph A. O'Sullivan

This appendix contains a preprint of a paper in review for presentation at the 1988 Allerton Conference to be held at the University of Illinois, Urbana, IL.

Performance Analysis for Maximum Likelihood Spectrum Estimation of Periodic Processes from Noisy Data *

P. Moulin
D. L. Snyder
J. A. O'Sullivan

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
Saint Louis, MO 63130

1. Introduction

A promising approach to maximum-likelihood estimation of Toeplitz constrained covariance matrices has been proposed recently [1]. Several developments can be considered. First, this method also applies to the dual problem of spectrum estimation. Another issue of interest is that the statistical model can account for the presence of additive noise corrupting the observations and for linear transformations of the process whose covariance or spectrum is sought. These considerations motivated a new approach to high-resolution delay-doppler radar imaging, where a major goal is to produce estimates of the target's scattering function [2]. In the special case where the transmitted signal is a constant and there is only one delay bin, this reduces to a spectrum estimation problem. This paper describes some recent results obtained for this simplified problem.

Our model is presented in Section 2. A maximum-likelihood formulation of the problem is given in Section 3, and it is shown how the equations can be solved. Section 4 is the main section. The statistical performance of the estimator is studied, and it is shown how it compares with more traditional methods such as the periodogram. The results are encouraging, and future directions for research are proposed. One such direction consists of using a priori information on the signal. When this information is in the form of a constraint on the signal power, a new estimator can be derived by maximizing the likelihood subject to that constraint. Theoretical issues of existence and uniqueness of the solution are discussed in the last section.

2. Model

We now present two models for the spectrum estimation problem. They fundamentally differ in that the first one accounts for noise corrupting the observations.

2.1. Model 1

This follows from the model presented in [2], when the transmitted signal is a constant and only one delay bin is considered. Let r be an N -vector sample of a real Gaussian process corrupted by an additive white Gaussian noise :

$$r = b + w, \quad (2.1)$$

where

w is an additive white noise with covariance matrix $N_0 I_N$, uncorrelated with b , where I_N is the $N \times N$ identity matrix.

* This work was supported by contract number N0001-85-K-0370 from the Office of Naval Research.

b is made of samples of a Gaussian periodic process with period $P \geq N$.

Both b and w are zero-mean. The periodicity assumption guarantees that the likelihood function is bounded above; therefore, there exists a maximum-likelihood estimator [1]. The assumption also implies that the covariance matrix K_P for a full period of the process is a Hermitian symmetric, circulant, Toeplitz matrix. Its eigenvectors are columns of the $P \times P$ DFT matrix W_P , so

$$K_P = W_P^\dagger A_P W_P, \quad (2.2)$$

where superscript \dagger denotes the Hermitian-transpose operator on matrices, and A_P is a $P \times P$ real diagonal matrix whose diagonal entries are nonnegative and are spectrum samples.

Assume now that we are only interested in estimating M of the P spectrum samples ($1 \leq M \leq P$), the other spectrum samples being zero. This assumption is introduced to deal with the bandlimited spectra encountered in radar applications [2]. We define the diagonal matrix of spectral parameters Σ by

$$A_P = M_M \Sigma M_M, \quad (2.3)$$

where

$$M_M = \begin{bmatrix} I_{(M+1)/2} & 0 & 0 \\ 0 & 0 & I_{(M-1)/2} \end{bmatrix}$$

and $I_{(M+1)/2}$ and $I_{(M-1)/2}$ are respectively the $(M+1)/2 \times (M+1)/2$ and the $(M-1)/2 \times (M-1)/2$ identity matrices. For simplicity, we consider M to be odd. Only the first $(M+1)/2$ and the last $(M-1)/2$ diagonal entries of A_P are nonzero, and these define the diagonal entries $\sigma^2(i)$ of the $M \times M$ diagonal matrix Σ . The $N \times N$ covariance matrix K_s of b is the upper left corner of K_P . From (2.2) and (2.3),

$$K_s = \Gamma^\dagger \Sigma \Gamma, \quad (2.4)$$

where we define the $M \times N$ matrix to be

$$\Gamma = M_M W_P \{I_N \ 0\}^\dagger. \quad (2.5)$$

The covariance matrix for r is then given by

$$K_r = \Gamma^\dagger \Sigma \Gamma + N_0 I_N. \quad (2.6)$$

2.2. Model 2

In most spectrum estimation problems, it is assumed that the observations are samples of the process of interest. This is also the model assumed in [1]. This is a special case of (2.1) in which

$$r = b. \quad (2.7)$$

From the definitions in (2.3) and (2.5) for Σ and Γ , the covariance matrix for r is now given by

$$K_r = \Gamma^\dagger \Sigma \Gamma. \quad (2.8)$$

3. Spectrum Estimators

In our simulations in the next section, we compare the performance of three estimators for the spectrum of the process b given r from equation (2.1). These three estimators are presented in this section. There are two maximum-likelihood estimators derived from models 1 and 2, and denoted by ML1 and ML0, respectively. The third estimator is the periodogram.

3.1. ML1 Estimator

This estimator is derived from model 1. Following the formulation in [2], the equation for the received signal can be rewritten as

$$r = \Gamma^\dagger c + w, \quad (3.1)$$

where c is a zero-mean Gaussian random vector with diagonal covariance Σ . The likelihood function for K_r is

$$L(r, K_r) = -\frac{1}{2} \ln \det K_r - \frac{1}{2} r^T K_r^{-1} r. \quad (3.2)$$

From (2.6) and (3.2), the likelihood function for Σ is

$$L(r, \Sigma) = -\frac{1}{2} \ln \det (\Gamma^T \Sigma \Gamma + N_0 I_N) - \frac{1}{2} r^T (\Gamma^T \Sigma \Gamma + N_0 I_N)^{-1} r \quad (3.3)$$

Maximizing the likelihood with respect to Σ yields the necessary trace condition which the estimate $\hat{\Sigma}$ must satisfy [1,2]:

$$\text{Tr} \{ \Gamma (\Gamma^T \hat{\Sigma} \Gamma + N_0 I_N)^{-1} (r r^T - \Gamma^T \hat{\Sigma} \Gamma - N_0 I_N) (\Gamma^T \hat{\Sigma} \Gamma + N_0 I_N)^{-1} \Gamma^T \delta \hat{\Sigma} \} = 0. \quad (3.4)$$

for all $M \times M$ diagonal matrices $\delta \hat{\Sigma}$. This trace condition is a nonlinear equation in $\hat{\Sigma}$. Generally it cannot be solved directly in closed-form, so some numerical search procedure must be implemented. An elegant solution is the expectation-maximization (EM) algorithm used in [1,2]. An initial estimate $\hat{\Sigma}^{(0)}$ is selected. At step $k+1$ ($k = 0, 1, \dots$) the estimate is updated according to

$$\hat{\Sigma}^{(k+1)} = \text{argmax}_{\Sigma} Q(\Sigma | \hat{\Sigma}^{(k)}) \quad (3.5)$$

where

$$Q(\Sigma | \hat{\Sigma}^{(k)}) = -\frac{1}{2} \sum_{i=0}^{M-1} \ln \sigma^2(i) - \frac{1}{2} \sum_{i=0}^{M-1} E_i \{ |c(i)|^2 | r, \hat{\Sigma}^{(k)} \} / \sigma^2(i). \quad (3.6)$$

and

$$E_i \{ |c(i)|^2 | r, \hat{\Sigma}^{(k)} \} = \{ \hat{\Sigma}^{(k)} - \hat{\Sigma}^{(k)} \Gamma (\Gamma^T \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1} \Gamma^T \hat{\Sigma}^{(k)} + \hat{\Sigma}^{(k)} \Gamma (\Gamma^T \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1} r r^T (\Gamma^T \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1} \Gamma^T \hat{\Sigma}^{(k)} \} (i, i). \quad (3.7)$$

This algorithm produces a sequence of estimates

$$\hat{\sigma}^2(i)^{k+1} = E_i \{ |c(i)|^2 | r, \hat{\Sigma}^{(k)} \} \quad (3.8)$$

having increasing likelihood. It can be shown that the stable points of this algorithm satisfy the necessary trace condition for a maximizer [2]. The issue of uniqueness is addressed in a companion paper [3].

Special case: $N = M = P = 1$, $N_0 = 0$

A closed-form expression for $\hat{\Sigma}$ can be derived in this special case:

$$\hat{\sigma}^2(i) = \max(0, r^2 - N_0). \quad (3.9)$$

3.2. MLO Estimator

When model 2 holds, the equation for the received signal can be rewritten as

$$r = \Gamma^T c. \quad (3.10)$$

where c is a zero-mean Gaussian random vector with diagonal covariance Σ . From (2.8) and (3.2), the likelihood function for Σ is

$$L(r, \Sigma) = -\frac{1}{2} \ln \det (\Gamma^T \Sigma \Gamma) - \frac{1}{2} r^T (\Gamma^T \Sigma \Gamma)^{-1} r. \quad (3.11)$$

As for model 1, a trace condition for a maximizer of the likelihood can be derived. The EM algorithm can also be used to solve this nonlinear equation in $\hat{\Sigma}$. The sequence of estimates is still given by (3.5), where now

$$E_i \{ |c(i)|^2 | r, \hat{\Sigma}^{(k)} \} = \{ \hat{\Sigma}^{(k)} - \hat{\Sigma}^{(k)} \Gamma (\Gamma^T \hat{\Sigma}^{(k)} \Gamma)^{-1} \Gamma^T \hat{\Sigma}^{(k)} + \hat{\Sigma}^{(k)} \Gamma (\Gamma^T \hat{\Sigma}^{(k)} \Gamma)^{-1} r r^T (\Gamma^T \hat{\Sigma}^{(k)} \Gamma)^{-1} \Gamma^T \hat{\Sigma}^{(k)} \} (i, i)$$

Clearly MLO and ML1 are equivalent for noise-free problems ($N_0 = 0$).

Special case: $N = M$

The problem for which the number of observations (N) is equal to the number of parameters to be estimated (M) is of some practical interest. It also turns out that the trace condition can be solved in closed-form in this instance. The matrix $\hat{\Gamma}$ is then invertible. Equation (3.10) indicates that there is a one-to-one mapping between r and c , and the MLO estimator is simply

$$\hat{\sigma}^2(i) = |(\Gamma^{-1}r)\hat{c}_i|^2 \quad (3.12)$$

where Γ^{-1} denotes $(\Gamma^{-1})^T$.

3.3. Periodogram

The periodogram estimate of the P -spectrum of the periodic process is simply the magnitude-squared Fourier transform of the observations [4],

$$\hat{A}_P = (P/N) W_P \begin{bmatrix} r \\ 0 \end{bmatrix} r^T \Theta^T W_P^T \quad (3.13)$$

Assume we are looking only at the first M spectrum samples. Then, from (2.3), (2.5) and (3.13), $\hat{\Sigma}$ is given by

$$\hat{\sigma}^2(i) = (P/N) |(\Gamma r)\hat{c}_i|^2 \quad (3.14)$$

Special case : $N = M = P$

When $N = M = P$, the matrix Γ is equal to W_P and periodogram and MLO estimates are the same. In this case, a full period of the process is estimated.

4. Performance Analysis

In this section, we apply the three estimators presented above to model 1 and study their statistical performances. The bias and variance are evaluated for each estimator, where

$$\text{Bias}[\hat{\Sigma}] = E[\hat{\Sigma}] - \Sigma \quad (4.1)$$

and

$$\text{Var}[\hat{\Sigma}] = E[\hat{\Sigma}^2] - (E[\hat{\Sigma}])^2 \quad (4.2)$$

As we shall see in Section 4.3, the performance strongly depends upon the input signal to noise ratio defined by

$$\text{SNR}_{in} = E_0 / N_0 \quad (4.3)$$

where E_0 is the average power of the process, defined by

$$E_0 = K_0(0) = (1/P) \text{Tr} [\Sigma] \quad (4.4)$$

From (4.1) and (4.2), we derive the diagonal mean-squared error (MSE) matrix

$$\text{MSE}[\hat{\Sigma}] = E\{(\hat{\Sigma} - \Sigma)^2\} = \text{Var}[\hat{\Sigma}] + (\text{Bias}[\hat{\Sigma}])^2 \quad (4.5)$$

The output signal to noise ratio matrix is defined as follow :

$$\text{SNR}_{out}[\hat{\Sigma}] = \Sigma (\text{MSE}[\hat{\Sigma}])^{-1} \quad (4.6)$$

In the following section, we evaluate the bias and mean-squared error for the estimator derived in Section 3. Whenever closed-form expressions for the MLO estimates cannot be derived, computer simulations are performed. Typically 3000 realizations are generated for each process. For a given estimator, (4.1) and (4.6) are then estimated from the 3000 estimates.

4.1. Performance Analysis

Closed form expressions for the bias and mean-squared error are derived for MLO and the periodogram when they exist. Simulations were carried out to compare the performance of the estimators for various levels of input SNR. The performance is then compared to the Cramer-Rao

Lower bound for the variance of unbiased estimators. Much effort was made for the special case $M = N$. This provides insight into the problem since the MLO equations can be solved in closed form. The choice of P is free, so long as $P \geq N$ [2].

4.2. Closed-form Expressions for Estimator Performance

(a) ML1

As indicated in Section 3.2, no closed-form expression for the estimator is available, so the evaluation of bias and variance is obtained by computer simulation.

(b) MLO

Closed-form expressions for MLO can be derived when $M = N$. The results are presented below.

Bias

Taking the expectation of (3.12) and using (2.6), we get

$$E[\hat{\sigma}^2(i)] = \sigma^2(i) + N_0(\Gamma\Gamma^T)^{-1}(i,i) \quad (4.7)$$

so we have

$$\text{Bias}[\hat{\sigma}^2(i)] = N_0(\Gamma\Gamma^T)^{-1}(i,i). \quad (4.8)$$

The bias is due to the noise corrupting the observations and is proportional to N_0 . The sensitivity of the bias to the noise is determined by the diagonal entries of the matrix $(\Gamma\Gamma^T)^{-1}$.

Mean-Squared Error

Combining (3.1) and (3.12), we can write

$$\hat{\sigma}^2(i) = |(c + \Gamma^{-1}w)(i)|^2. \quad (4.9)$$

Taking the expectation of (4.9) squared, we obtain

$$\begin{aligned} E\{[\hat{\sigma}^2(i)]^2\} &= \sigma^4(i) (2 + \delta_{00}) + N_0 \sigma^2(i) \left(4(\Gamma\Gamma^T)^{-1}(i,i) + 2 \sum_{j=0}^{M-1} \text{Re}[\Gamma^{-1}(i,j)]^2 \right) \\ &\quad + N_0^2 \left(2(\Gamma\Gamma^T)^{-1}(i,i)^2 + \sum_{j=0}^{M-1} |\Gamma^{-1}(i,j)|^4 \right). \end{aligned} \quad (4.10a)$$

After some algebraic manipulations, this expression can be lower-bounded by

$$E\{[\hat{\sigma}^2(i)]^2\} \geq 2 \left[\sigma^4(i) + N_0(\Gamma\Gamma^T)^{-1}(i,i) \right]^2 = 2 \left(E[\hat{\sigma}^2(i)] \right)^2. \quad (4.10b)$$

From (4.7) and (4.10), (4.5) becomes

$$\text{MSE} [\hat{\sigma}^2(i)] = \left(E\{[\hat{\sigma}^2(i)]\} \right)^2 + N_0 (\Gamma\Gamma^T)^{-1}(i,i)^2, \quad (4.11a)$$

and

$$\text{MSE} [\hat{\sigma}^2(i)] \geq \sigma^4(i) + 2 \sigma^2(i) N_0 (\Gamma\Gamma^T)^{-1}(i,i) + 2 (N_0 (\Gamma\Gamma^T)^{-1}(i,i))^2. \quad (4.11b)$$

(c) Periodogram

Bias

Combining (3.1) and (3.14), we can write

$$\hat{\sigma}^2(i) = (P/N) |(\Gamma^T c + \Gamma w)(i)|^2. \quad (4.12)$$

Taking the expectation of (4.12), we get

$$E[\hat{\sigma}^2(i)] = (P/N) (\Gamma^T \Sigma \Gamma^T + N_0 \Gamma^T)(i,i). \quad (4.13)$$

and

$$\text{Bias}[\hat{\sigma}^2(i)] = (P/N) (\Gamma^T \Sigma \Gamma^T) - \Sigma)(i, i) + (P/N) N_0 (\Gamma^T)(i, i). \quad (4.14)$$

The bias is made of two terms. The second one is due to the noise and is proportional to N_0 . The sensitivity of the bias to the noise is determined by the diagonal entries of the matrix Γ^T . The first one is independent of N_0 . Even for noise-free observations the periodogram is a biased estimator of Σ unless Γ^T is the identity matrix. This would be the case only for $N = M = P$ (observation of a full period of the process) or $N/M \rightarrow \infty$.

Mean-Squared Error

Taking the expectation of (4.12) squared, we obtain

$$\begin{aligned} E\{[\hat{\sigma}^2(i)]^2\} &= (P^2/N^2) \left[2 \left[\sum_{j=0}^{M-1} \sigma^2(j) |(\Gamma^T)(i, j)|^2 \right]^2 + 4 N_0 \sum_{j=0}^{M-1} \sigma^2(j) |(\Gamma^T)(i, j)|^2 (\Gamma^T)(i, i) \right. \\ &\quad \left. + 2 N_0^2 (\Gamma^T)(i, i)^2 \right. \\ &\quad \left. + |\sigma^2(0) (\Gamma^T)(i, 0)|^2 + \sigma^2(M/2) |(\Gamma^T)(i, M/2)|^2 + N_0 \sum_{j=0}^{M-1} |(\Gamma^T)(i, j)|^2 \right]. \end{aligned} \quad (4.15a)$$

This expression is lower-bounded by

$$2 \langle P/N \rangle^2 [(\Gamma^T \Sigma \Gamma^T + N_0 \Gamma^T)(i, i)]^2 = 2 (E[\hat{\sigma}^2(i)])^2. \quad (4.15b)$$

From (4.13) and (4.15), (4.5) becomes

$$\text{MSE}[\hat{\sigma}^2(i)] = (E[\hat{\sigma}^2(i)])^2 + [(P/N) (\Gamma^T \Sigma \Gamma^T - (N/P) \Sigma + N_0 \Gamma^T)(i, i)]^2, \quad (4.16a)$$

and

$$\begin{aligned} \text{MSE}[\hat{\sigma}^2(i)] &\geq (P^2/N^2) \{ [(\Gamma^T \Sigma \Gamma^T)(i, i)]^2 + (\Gamma^T \Sigma \Gamma^T - (N/P) \Sigma)(i, i)^2 \\ &\quad + 2 N_0 (\Gamma^T)(i, i) (2 \Gamma^T \Sigma \Gamma^T - (N/P) \Sigma)(i, i) \\ &\quad + 2 [N_0 (\Gamma^T)(i, i)]^2 \}. \end{aligned} \quad (4.16b)$$

4.3. Simulation results

Process 1

Here we consider a lowpass process of period $P = 10$. There are $M = 5$ nonzero spectrum samples

$$\sigma^2(i) = 1 \quad i = 0, \dots, 4.$$

The number of observations is $N = M = 5$, and the noise variance N_0 ranges from 0 to 1. Figures 1 and 2 give a plot of the bias and SNR_{est} for the estimators of $\sigma^2(2)$ in function of SNR_{in} , according to the definitions (4.1), (4.3), and (4.6). In the absence of additive noise ($SNR_{in} \rightarrow \infty$), ML1 and ML0 are the same. Both are unbiased estimators. The periodogram, however, is biased, and its MSE is also larger than the MSE for the ML estimators. When N_0 increases from 0, the performance of the estimators is roughly constant so long as SNR_{in} remains above some threshold. For larger N_0 , all three estimators exhibit a strong degradation in performance. Comparing the thresholds for ML0 and ML1 we see the tremendous improvements brought by taking the noise into account in the model. Typically, for a same SNR_{est} , ML1 will have the same performance as ML0 operating in a 20 dB noisier environment.

We also notice that the threshold for the periodogram is located at a lower SNR_{in} than for the ML estimators. In Sections 4.2(b) and 4.2(c), we indicated how the sensitivity of the performance to the noise can be determined for ML0 and the periodogram when $N = M$. It turns out that for the process considered here, with a uniform spectrum, the periodogram has a lower sensitivity than ML0 and ML1. This is thought to be due to the smooth spectrum used in the simulation.

Process 2

It has been conjectured that the periodogram does not perform well for peaky spectra [5]. This motivated our study of a sharply peaked spectra. The process has period $P = 10$, and is made of a single spectrum component

$$\sigma^2(0) = 1.$$

There is just $N = M = 1$ observation. Bias and SNR_{est} for the estimators of $\sigma^2(0)$ are plotted as a function of SNR_{in} in figures 3 and 4. In the absence of additive noise, the periodogram is very strongly biased, and its MSE is large. Furthermore, in high-noise environment the periodogram is no longer more robust than the ML estimators. Clearly, the periodogram is outperformed. It should also be noticed that for this process, the improvement of ML1 over ML0 is quite reduced.

Computational Considerations

The convergence rate of the EM algorithm depends on several parameters. The computation time for each iteration is of order MN^2 . The number of iterations required for convergence of the algorithm grows as M and N increase. For ML1, more iterations are required as N_0 increases, especially in the threshold region and beyond. Typical figures are: for process 1 with $N_0 = 0.1$, 30 iterations are required before the spectrum estimates are stable; when $N_0 = 1$, 300 iterations must be performed. Our algorithm is implemented on a Masscomp model 5500. Running the program on 3000 realizations in the latter case is typically completed in 6 CPU hours.

4.4. Cramer-Rao Bounds

In this section we study how the performance of the estimators considered so far relates to the Cramer-Rao bound on the variance of any unbiased estimator. The Cramer-Rao bound on the variance of any unbiased (UB) estimator of $\sigma^2(i)$ for our model has been found to be [3]

$$UB-CR[\hat{\sigma}^2(i)] = (\sigma^2(i) + N_0(\Gamma\Gamma^T)^{-1}(i,i))^2. \quad (4.17)$$

From (4.5) and (4.17), the MSE for an unbiased estimator reaching the Cramer-Rao bound is given by

$$MSE[\hat{\sigma}^2(i)] = (\sigma^2(i) + N_0(\Gamma\Gamma^T)^{-1}(i,i))^2. \quad (4.18)$$

The Cramer-Rao bound on the variance of a biased (B) estimator of $\sigma^2(i)$ is given by

$$B-CR[\hat{\sigma}^2(i)] = UB-CR[\hat{\sigma}^2(i)] (\partial E[\hat{\sigma}^2(i)] / \partial \sigma^2(i))^2. \quad (4.19)$$

From (4.5), (4.17) and (4.19), the MSE for a biased estimator reaching the Cramer-Rao bound is given by

$$MSE[\hat{\sigma}^2(i)] = (\sigma^2(i) + N_0(\Gamma\Gamma^T)^{-1}(i,i))^2 (\partial E[\hat{\sigma}^2(i)] / \partial \sigma^2(i))^2 + (Bias[\hat{\sigma}^2(i)])^2. \quad (4.20)$$

From the analytical expressions given for $E[\hat{\sigma}^2(i)]$ in Section 4.2, we can now calculate the gradient of $E[\hat{\sigma}^2(i)]$ for ML0 and the periodogram. Then the minimum MSE for a biased estimator having the same bias as ML0 and the periodogram is derived, and a comparison with the actual MSE is made. No closed-form expression has been found for ML1.

ML0

From (4.7),

$$\partial E[\hat{\sigma}^2(i)] / \partial \sigma^2(i) = 1. \quad (4.21)$$

From (4.8), (4.20) and (4.21), the MSE is lower-bounded by

$$MSE_{cr}[\hat{\sigma}^2(i)] = \sigma^4(i) + 2\sigma^2(i)N_0(\Gamma\Gamma^T)^{-1}(i,i) + 2[N_0(\Gamma\Gamma^T)^{-1}(i,i)]^2. \quad (4.22)$$

Periodogram

From (4.13),

$$\partial E[\hat{\sigma}^2(i)] / \partial \sigma^2(i) = (P/N)(\Gamma\Gamma^T)(i,i). \quad (4.23)$$

Combining (4.14), (4.20) and (4.23), the MSE is lower-bounded by

$$\begin{aligned} \text{MSE}_{\sigma}[\hat{\sigma}^2(i)] &= (P^2/N^2) \left([(\Gamma^T \Sigma \Gamma)(i,i)]^2 + (\Gamma^T \Sigma \Gamma^T - (N/P) \Sigma)(i,i)^2 \right) \\ &\quad + 2 N_0 (\Gamma^T)(i,i) (2 \Gamma^T \Sigma \Gamma^T - (N/P) \Sigma)(i,i) \\ &\quad + 2 [N_0 (\Gamma^T)(i,i)]^2 \end{aligned} \quad (4.24)$$

Comparison of MSE's with Cramer-Rao bounds

The Cramer-Rao bounds (4.22) and (4.24) on the MSE are the same as the bounds (4.11b) and (4.16b) derived algebraically from the exact expressions (4.11a) and (4.16a). Figure 5 shows how the lower bounds compare with the exact expressions for process 1. The actual MSE's are 3-4 dB above their respective bounds.

4.5. Discussion

The results derived above suggest additional comments on a comparison between periodogram and ML estimators. Typically each component of the gradient of $E[\hat{\sigma}^2(i)]$ given in (4.23) is smaller than unity for the processes we consider, and the Cramer-Rao bound on the variance of the periodogram-like biased estimator is much smaller than the Cramer-Rao bound on the variance of unbiased estimators. When the variance dominates the MSE, the periodogram offers then a good MSE performance. This was the case for process 1. For a more peaky spectrum such as the one chosen for process 2, the bias dominates the MSE and the periodogram is outperformed by the ML estimators.

5. Constrained maximum-likelihood estimation

5.1. Description of the problem

An examination of figures 1-4 suggests that ML suffers in certain situations. When the input SNR_{in} is low, the estimates are biased and their variance is large. Although the maximum-likelihood estimator is asymptotically unbiased and efficient, these properties are not guaranteed in small samples. For the problems considered in Section 4, the number of samples is only equal to the number of parameters to estimate. This limitation can be alleviated if a priori knowledge, such as SNR_{in} , is available. Since N_0 is known, such a constraint on the signal-to-noise ratio can be translated into a constraint on the signal power that must be satisfied by the maximum-likelihood estimates. In the following we show how this constraint can be incorporated into the EM algorithm. The constrained estimates exist, and are unique.

5.2. Equations

The equations for ML presented in Section 3.1 can now be modified as follows. At each step of the EM algorithm, we maximize $Q(\Sigma | \hat{\Sigma}^{(k)})$ defined in (5.6), subject to the power constraint

$$\sum_{i=0}^{M-1} \sigma^2(i) = P E_0 = S, \quad (5.1)$$

where E_0 is the constraint on the signal power. The solution also maximizes

$$Q(\Sigma | \hat{\Sigma}^{(k)}) + \lambda \left(\sum_{i=0}^{M-1} \sigma^2(i) - S \right), \quad (5.2)$$

where λ is a Lagrange multiplier. Taking the gradient of (5.2) with respect to Σ , we obtain a quadratic equation for each spectral component

$$2 \lambda \sigma^4(i) - \sigma^2(i) + C_i = 0, \quad (5.3)$$

where

$$C_i = E[|c(i)|^2 | r, \hat{\Sigma}^{(k)}]$$

is calculated according to (3.7). The solution to (5.3) is

$$\begin{aligned} \sigma^2(i) &= (1 + I_i \sqrt{1 - 8C_i \lambda}) / 4\lambda & : \lambda \neq 0 \\ &= C_i & : \lambda = 0. \end{aligned} \quad (5.4)$$

where I_i is either +1 or -1. The equation for λ is

$$4S\lambda - M = \sum_{i=0}^{M-1} I_i \sqrt{1 - 8C_i \lambda}. \quad (5.5)$$

In general this nonlinear equation in λ cannot be solved in closed-form. Furthermore, an ambiguity subsists about the choice of the signs I_i . The latter problem is solved by application of the following theorem:

Theorem

Assume that $C_0 > C_i, i = 1, \dots, M-1$. Then

(1)

$$\begin{aligned} I_i &= -1 & : i = 1, \dots, M-1 \\ I_0 &= +1 & : S < 2C_0 [M - \sum_{i=1}^{M-1} I_i \sqrt{1 - (C_i/C_0)}] \\ &= -1 & : \text{else} \end{aligned}$$

(2) λ is the largest nonzero solution of

$$(4S\lambda - M + \sum_{i=1}^{M-1} I_i \sqrt{1 - (C_i/C_0)})^2 = 1 - 8C_0\lambda, \text{ for } S \neq \sum_{i=1}^{M-1} C_i, \text{ and} \quad (5.6a)$$

$$\lambda = 0, \text{ for } S = \sum_{i=1}^{M-1} C_i. \quad (5.6b)$$

λ is upper-bounded by $1/8C_0$, and the equation (5.6a) can be solved numerically for λ . Note that the particular case (5.6b) is also the solution to the unconstrained maximization problem. Next $\hat{\sigma}^2(i)^{j^{i+1}}$ is calculated from (5.4). The whole procedure is repeated at each maximization step of the EM algorithm.

5.3. Discussion

An algorithm has been derived to produce maximum-likelihood estimates subject to the power constraint (5.1). Currently we are preparing a computer implementation of this algorithm, and the performances of the constrained estimator will be studied. As indicated in Section 5.1, noticeable improvements in high-noise situations are expected.

Conclusions

In this paper, we have described our approach to spectrum estimation from noisy data, based upon a statistical model for the observations. From this model, a maximum-likelihood estimator is derived, and its bias and MSE are computed and compared with two methods that do not take the additive noise into account. The new estimator can perform significantly better than its competitors.

A priori knowledge of the signal power can be introduced as a constraint to improve performance. Our current research activities include evaluation of the performance of the new estimator. We expect significant improvements for low SNR_m . Among future directions for research, let us mention another issue of practical interest. The information about SNR_m might have the form of an inequality rather than an equality constraint. We believe that the method described in Section 5 can easily be modified to incorporate such a constraint.

References

[1] M. L. Miller, D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proc. IEEE*, Vol. 75, No. 7, July 1987.

- [2] D. L. Snyder, J. A. O'Sullivan, M. I. Miller, "The Use Of Maximum-Likelihood Estimation For Forming Images Of Diffuse Radar-Targets From Delay-Doppler Data," *IEEE Trans. on Information Theory*, to appear in 1988.
- [3] J. A. O'Sullivan, P. Moulin, D. L. Snyder, "Cramer-Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging," submitted to 1988 Allerton Conference.
- [4] W. Davenport, W. Root, *Random Signals and Noise*. McGraw Hill, 1958.
- [5] D. Fuhrmann, personal communication.

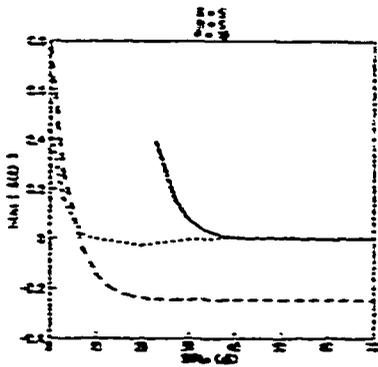


Figure 1. $Bias(\sigma^2(2))$ for Process 1
 MLO (solid line), ML1 (dotted line),
 periodogram (dashed line)

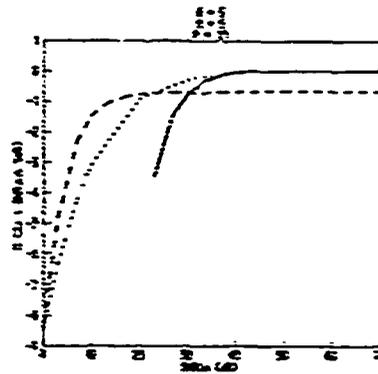


Figure 2. $SNR_{eff}(\sigma^2(2))$ for Process 1
 MLO (solid line), ML1 (dotted line),
 periodogram (dashed line)

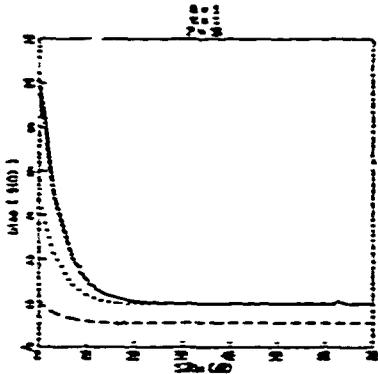


Figure 3. $Bias(\sigma^2(0))$ for Process 2
 MLO (solid line), ML1 (dotted line),
 periodogram (dashed line)

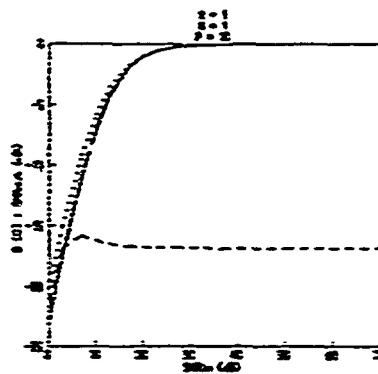


Figure 4. $SNR_{eff}(\sigma^2(0))$ for Process 2
 MLO (solid line), ML1 (dotted line),
 periodogram (dashed line)

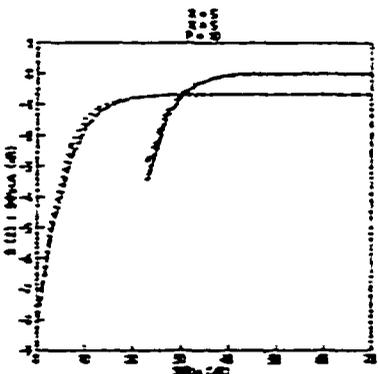


Figure 5. Comparison of $SNR_{eff}(\sigma^2(2))$ and
 Cramer-Rao bounds for Process 1
 MLO (solid line), its CR bound (short-
 dashed line), periodogram (long-dashed
 line), its CR bound (dotted line)

4.3. Appendix 3. Performance Analysis, Signal Design

Cramér-Rao Bounds for Constrained Spectrum Estimation with
Applications to a Problem in Radar Imaging

Joseph A. O'Sullivan

Pierre Moulin

Donald L. Snyder

This appendix contains a preprint of a paper in review for presentation at the 1988 Allerton Conference to be held at the University of Illinois, Urbana, IL.

Cramer Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging

Joseph A. O'Sullivan
Pierre Moulin
Donald L. Snyder

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
St. Louis, MO 63130

1. Introduction

This paper discusses some recent theoretical results which we have derived for maximum likelihood spectrum estimation problems. The practical problem which has motivated much of our study is a delay-doppler radar imaging problem where the spectrum to be estimated corresponds to an image of the target. For this problem it is the scattering function, which is a delay dependent power spectrum, that is of interest. For related problems, the parameters of interest are the lag covariances. These problems are not always equivalent and one goal of this paper is to point out some of the issues which are unique to the spectrum estimation formulation of the problem.

The results are presented in a rather general manner and are then specialized to problems of interest. First, problems of interest are discussed. This section also introduces some of the notation used throughout the remainder of the paper. The third section derives the Cramer-Rao lower bound for the variance of unbiased estimates of the spectrum samples. This bound is used in each of the other sections of the paper. The equivalence of nonuniqueness of maximum likelihood estimates for the spectrum samples and infinite magnitude of the Cramer-Rao lower bound is then demonstrated. The fourth section addresses a problem more specific to the radar imaging problem. The task is to choose a transmitted signal such that the spectrum estimates have minimum variance. This is accomplished by minimizing the Cramer-Rao lower bound with respect to the transmitted signal. The final section discusses the implications of these results and suggests further work.

2. Problem Definition

There are three problems which are of interest. The first two are special cases of the third, but some of the results to be presented are applicable to only one of the special cases.

Let W_P be a $P \times P$ DFT matrix. The k,m entry of W_P is $(1/\sqrt{P}) e^{-j2\pi km/P}$. In each of the three problems, r is a linear combination of a signal vector and additive white Gaussian noise. The goal of each problem is to find the maximum likelihood estimate for the power spectrum of the desired signal given one observation vector r .

Problem one:

$$r = b + w, \quad (2.1)$$

where b is an N -vector sample from a zero mean Gaussian process with Toeplitz covariance matrix K_1 , and w is white Gaussian noise with covariance $N\sigma^2 I_N$. The matrix K_1 is constrained to be of the form

$$K_1 = \begin{bmatrix} J_N & 0 \\ 0 & 0 \end{bmatrix} W_P^H \Sigma W_P \begin{bmatrix} I_N \\ 0 \end{bmatrix}, \quad (2.2)$$

where Σ is a $P \times P$ real diagonal matrix with nonnegative entries, I_N is an $N \times N$ identity matrix, and \dagger denotes complex conjugate transpose. This is equivalent to assuming that b is Γ_1^\dagger times c , where c is a zero mean Gaussian vector with covariance Σ and

$$\Gamma_1 = W_P \begin{bmatrix} I_N \\ 0 \end{bmatrix}. \quad (2.3)$$

Problem two:

$$r = S^\dagger b + w, \quad (2.4)$$

where w is as before, b is an M length vector,

$$b = [I_M \ 0] W_P^\dagger c, \quad (2.5)$$

but now there is an $N \times M$ matrix S^\dagger multiplying b . The vector c is zero mean Gaussian distributed with real diagonal covariance given by Σ . Under these assumptions, the covariance of r is $K_2 + N \sigma^2 I_N$, where

$$K_2 = S^\dagger [I_M \ 0] W_P^\dagger \Sigma W_P \begin{bmatrix} I_M \\ 0 \end{bmatrix} S = \Gamma_2^\dagger \Sigma \Gamma_2, \quad (2.6)$$

where

$$\Gamma_2 = W_P \begin{bmatrix} I_M \\ 0 \end{bmatrix} S. \quad (2.7)$$

This problem reduces to the first when $N = M$ and $S = I_N$.

Problem three:

$$r = S^\dagger b + w \quad (2.8)$$

where w is as before, S^\dagger is an N by $M I_R$ signal matrix, and now b is a sample from a process whose covariance matrix is $M I_R \times M I_R$ block diagonal with each $M \times M$ block being Toeplitz. Separate b into I_R subvectors each of length M and label these subvectors sequentially from 0 to $I_R - 1$. The covariance of $b(k)$ is given by

$$K_3(k) = [I_M \ 0] W_L^\dagger \Sigma(k) W_L \begin{bmatrix} I_M \\ 0 \end{bmatrix} \quad (2.9)$$

where $K_3(k)$ is the k^{th} diagonal block of the covariance of b . If the $L \times L$ matrix $\Sigma(k)$ is placed in the k^{th} diagonal block of a $L I_R \times L I_R$ block diagonal matrix, Σ , then we can write

$$K_3 = \Gamma_3^\dagger \Sigma \Gamma_3, \quad (2.10)$$

where the matrix Γ_3 is defined as the product of a block diagonal matrix and S ; the block diagonal matrix has I_R blocks each of which equals $W_L [I_M \ 0]^\dagger$. For this problem, let $P = L I_R$. This problem reduces to the second when there is only one block on the diagonal ($I_R = 1$). Since each of the other problems is obtained by simplifying this problem, we can focus our attention on this problem.

We follow the analysis in [1,2]. Let k denote either 1, 2, or 3. Problem k may be restated as follows. Assume the observation vector r is given,

$$r = \Gamma_k^\dagger c + w, \quad (2.11)$$

where c is a Gaussian random variable with covariance Σ , and w is white Gaussian noise with intensity $N \sigma^2$. Find the maximum likelihood estimate for Σ . Under the assumptions, r is a Gaussian distributed random variable with covariance

$$K_r = \Gamma_k^\dagger \Sigma \Gamma_k + N \sigma^2 I_N. \quad (2.12)$$

The loglikelihood function is (neglecting terms not involving Σ)

$$L(r, \Sigma) = -\frac{1}{2} \ln \det(\Gamma_k^\dagger \Sigma \Gamma_k + N \sigma^2 I_N) - \frac{1}{2} r^\dagger (\Gamma_k^\dagger \Sigma \Gamma_k + N \sigma^2 I_N)^{-1} r. \quad (2.13)$$

Taking the partial derivative with respect to Σ and setting it to zero yields the trace condition which a

solution must satisfy [1,2]. A necessary condition for Σ to be the maximum likelihood estimate for the spectrum for problem k is that

$$\text{Tr}[\Gamma_k(\Gamma_k^H \Sigma_k + N\sigma^2)^{-1}(r^k - \Gamma_k^H \Sigma_k - N\sigma^2)(\Gamma_k^H \Sigma_k + N\sigma^2)^{-1} \Gamma_k \delta \Sigma_k] = 0 \quad (2.14)$$

for all diagonal matrices $\delta \Sigma$ of the same size as Σ . Typically this equation cannot be solved for Σ explicitly and some iterative technique for determining the maximum likelihood estimate must be used. The algorithm used in [1,2] is the FM algorithm. One important question which this paper addresses is the uniqueness of maximum likelihood solutions. A result derived in the fourth section states that a sufficient condition for the uniqueness of maximum likelihood estimates is that the Cramer-Rao lower bound be infinite.

3. Cramer-Rao Lower Bound

This section contains a derivation of the Cramer-Rao lower bound for unbiased estimates of the covariance matrices for the problems from Section 2. It is important to remember that, for biased estimates, the lower bound for the variance of the estimates is proportional to this bound, so these calculations are relevant for that case as well. In fact, the maximum likelihood estimation procedure currently implemented in our laboratory generates biased estimates for the spectrum samples under certain conditions [7].

Following Van Trees [2, p. 79], the variance of any unbiased maximum likelihood estimate of a nonrandom parameter is greater than or equal to the diagonal elements of the inverse of the Fisher information matrix. The Fisher information matrix has entries equal to the negatives of the expected value of second the derivatives of $L(r, \Sigma)$ with respect to the parameters (in this case Σ).

Lemma 1: For problem k, the m,n element of the Fisher information matrix equals the magnitude squared of m,n element of the matrix

$$\Gamma_k(\Gamma_k^H \Sigma_k + N\sigma^2)^{-1} \Gamma_k^H \quad (3.1)$$

Proof: In the appendix of [1] an expression for the second derivative of $L(r, \Sigma)$ is obtained. Taking the negative of the expected value yields the above expression. \square

The above result is very important for the simulations in order to determine the performance of the estimation scheme used. Some of the simulations performed in our Laboratory are reported in [4] and the performance is compared to this bound. In addition, the algorithm proposed in [1] is compared to other algorithms for estimating spectrum samples and it is shown that in certain cases this algorithm outperforms the others.

One of the goals of calculating the Cramer-Rao bound is to use it to determine the well posedness of estimation problems.

Lemma 2: For problem one the rank of the Fisher information matrix is less than or equal to $2N-1$. Furthermore, if $P > 2N-1$ then there is no unique Σ positive definite which maximizes the loglikelihood.

Proof: The proof involves writing the Fisher information matrix as the product of three matrices one of which is $2N-1 \times 2N-1$. The rank of this matrix is thus shown to be less than or equal to $2N-1$. The third matrix is the Hermitian transpose of the first. The first matrix is P by $2N-1$ and has k,m element given by

$$\frac{1}{P} e^{-j2\pi k(m-N+1)/P} \quad (3.2)$$

where k denotes the row and ranges from 0 to $P-1$ and m ranges from 0 to $2N-2$. Let the α, β element of $(\Gamma_k^H \Sigma_k + N\sigma^2)^{-1}$ be denoted by $f_{\alpha\beta}$. Then the m,n element of the second matrix is

$$\sum_{m=0}^{2N-2} \sum_{n=0}^{2N-2} f_{\alpha+\alpha-N+1, \beta-N+1} f_{\alpha\beta}^* \quad (3.3)$$

where m and n range from 0 to $2N-2$ and * denotes complex conjugate. Multiplying these three matrices completes the first part of the proof. For the second part of the proof, it is easily shown that the addition of a real diagonal matrix with entries all of the same magnitude with alternating signs to any matrix does not change the value of the loglikelihood if $P > N-1$. Thus no positive definite Σ yields a unique value for the loglikelihood. By the same arguments, no Σ with one diagonal entry equal to 0 and the rest positive yields a unique value of $L(r, \Sigma)$. If two elements are zero, however, the argument fails and other tests must be

used. \square

Lemma 2 gives a precise statement about the uniqueness of maximum likelihood estimates for the problems under consideration. In particular, it states precisely the number of parameters which can be estimated uniquely for a problem of type one. We believe the bound is tight. That is, for a problem of type one, we believe that a unique maximum likelihood estimate exists whenever $P \leq 2N-1$ and that the Fisher information matrix is of rank P for this case. The basis for this belief is that the first matrix computed in the proof has full row rank for $P \leq 2N-1$. For cases of interest, the second matrix constructed is of full rank. If this second matrix is of full rank, the inequality in the first sentence of the lemma statement can be changed to equality.

The proof of the last lemma did not require that the elements of the Γ matrix be entries from \mathcal{W}_P . It only required a certain stationarity property.

Lemma 3: Let the i,m element of Γ_k be $g_{i,m}$ and suppose that $g_{i,m}g_{i,m}^* = h_{i,m-n}$ (the product depends only on the row number and the difference $m-n$). Then the rank of the Fisher information matrix is at most $2N-1$ and if $P > 2N-1$ there is no unique Σ positive definite which maximizes the loglikelihood.

Proof: Modify the proof of the last lemma by changing the first matrix to have k,m element $h_{k,m-N+1}$. The proof of the first part is then identical to that proof. To show that there is no unique Σ , we construct a matrix such that if this matrix is added to Σ then the value of the loglikelihood does not change. This diagonal perturbational matrix is formed as follows. Let γ_i denote the i^{th} row of Γ_k . Then we have

$$\Gamma_k^H \Sigma \Gamma_k = \sum_{i=0}^{P-1} (2) h_i \gamma_i^H \gamma_i. \quad (3.4)$$

The potential elements of this matrix are completely determined by the elements of the outer products of γ_i with itself for i from 0 to $P-1$. Form P row vectors with the elements of these outer products by taking the Kronecker products of γ_i with γ_i^* . Take these P vectors, transpose them, and put them in the columns of a $N^2 \times P$ matrix. Denote this matrix G^T . This matrix has rank at most $2N-1$ over the complex numbers due to the stationarity assumption on elements of Γ_k stated in the lemma. The $2N-1$ linearly independent rows have i^{th} element $h_{i,m-N+1}$. One of these linearly independent rows has all real elements while the others are in general complex and appear in complex conjugate pairs (the real row has entries $h_{i,0} = |g_{i,m}|^2$). Thus this matrix has column rank at most $2N-1$ when taken over the real numbers. Any real vector in the kernel of this matrix may be placed along the diagonal of a matrix and then added to Σ without changing the value of $L(r, \Sigma)$ because the diagonal matrices formed in this way have the property that they are mapped to the zero matrix when premultiplied by Γ_k^H and postmultiplied by Γ_k . \square

Again the bound is tight because in general the matrix formed of Kronecker products of the rows of Γ_k will have rank $2N-1$ over the reals. Using the matrix G defined in the proof of Lemma 3 a much more general result may be obtained.

Lemma 4: Let G be the $P \times N^2$ matrix whose i^{th} row is the Kronecker product of γ_i with γ_i^* . Assume that K_r is positive definite. Then the Fisher information matrix is singular if and only if the rank of G is less than P . If a positive definite Σ yields a unique maximum of the loglikelihood function then G has rank P .

Proof: The Fisher information matrix may be written as

$$G(K_r^{-1} \oplus K_r^{-1})G^T. \quad (3.5)$$

If K_r is nonsingular then the rank of this matrix equals the rank of G and it is singular if and only if the rank of G is less than P . If the rank of G is less than P then pick a real vector s such that $s^T G = 0$ (such an s exists because the Fisher information matrix has real entries). If these vector elements are placed along the diagonal of a matrix and added to Σ the matrix K_r remains unchanged and thus the loglikelihood is unchanged. On the other hand, if the rank of G equals P , then there is no such vector s and any change of the diagonal entries of Σ yields a different K_r . \square

4. Signal Selection

This section presents some new thinking on the problem of signal selection for radar imaging systems. The approach is somewhat general and it can be changed to fit several problems of interest.

Most approaches to radar imaging are based on deterministic techniques. Typically, the reflections from the target are assumed to be deterministic and the radar returns are processed using a two dimensional transform technique. The transform used (or in general the linear processing scheme used) may be modified for the problem at hand, but the overall processing scheme is fixed. The signal selection problem may be thought of in a couple of different ways. One way is to select the signal which optimizes the radar return for the processing scheme used. For example, stepped frequency waveforms described in [5] may be thought of as optimizing the signal assuming that two dimensional Fourier transforms are used to process the radar return. In other words, this transmitted signal attempts to make the return signals look like the two dimensional Fourier transform of the target image. A second type of signal selection is more general. In this case, the signal is selected to achieve pulse compression so that by combining the returned signal properly the result of processing achieves desired range and crossrange resolution [5].

Our approach for radar imaging differs from conventional approaches by starting from a model which accurately accounts for the true nature of radar reflections. The literature on the physics of radar reflections [3,6] states that the reflections are random not deterministic. We start with a model which takes into account this randomness (the discrete version of this model is given by problem three above). The processing to obtain the "best" image is then derived [1]. This processing is valid for any transmitted signal as long as the model is valid. Thus, this processing could be used on the radar returns from any system presently implemented if the target falls into the category covered by the model. What we aim to do in this section of the paper is to introduce a procedure for selecting the transmitted signal optimally. The signal is selected to maximize a measure of the achievable performance. The potential use of this procedure in practise and open questions which we are exploring are addressed shortly.

The measure which is used to determine the relative merit of a transmitted signal is the performance which is achievable. The performance is measured in terms of the Fisher information matrix.

Definition: Let S denote the set of admissible signal matrices. Any signal matrix $S \in S$ defines a matrix Γ_3 from (2.10) and a $P \times N^2$ matrix G as defined in Lemma 4. Let Ω denote the set of possible target images for Σ . Let $\|\cdot\|_P$ be a given norm on $P \times P$ matrices. A signal matrix $S \in S$ is said to be optimal if S achieves the supremum

$$\sup_{S \in S} \inf_{\Sigma \in \Omega} \|G(K_r^{-1} \oplus K_r^{-1})G^T\|_P. \quad (4.1)$$

The three objects which must be given in order to attempt to calculate the optimal S are S , Ω , and $\|\cdot\|_P$. In order to simplify the derivation, we assume that the set of possible target images, Ω , has only one member, Σ . This restriction will be discussed later. For the norm, the trace operator is chosen. Since all of the matrices under consideration are hermitian symmetric and positive semidefinite, the trace is always greater than or equal to zero and is zero only when all of the eigenvalues are zero (and hence the matrix is the zero matrix because it is similar to a zero matrix). The trace is additive so the triangle inequality holds and a scalar times a matrix scales the trace. Thus for the set of positive semidefinite hermitian symmetric matrices, the trace is a valid norm.

There are several possibilities for the set of signal matrices. For both intuitive and analytical reasons it is important to have a constraint on the admissible signal matrices. Intuitively, since the noise intensity does not depend on the signal magnitude, one would expect that increasing the transmitted energy would decrease the effect of the noise. This is the case analytically also. Parameterize the signal matrix as $S = \sqrt{E_t} \hat{S}$, where E_t may be thought of as the transmitted energy. Obviously, Γ_3 may then be written as $\sqrt{E_t} \hat{\Gamma}_3$, where $\hat{\Gamma}_3$ is obtained from Γ_3 by substituting \hat{S} for S . Similarly, from (2.10) and (2.11)

$$K_3 = E_t \hat{\Gamma}_3^T \Sigma \hat{\Gamma}_3 = E_t \hat{K}_3, \quad (4.2)$$

and

$$K_r = E_t \hat{K}_3 + N_0 I_N \quad (4.3)$$

From $\hat{\Gamma}_3$ form the matrix \hat{G} the same way G was formed from Γ_3 . Then the Fisher information matrix (3.5) may be rewritten as

$$E_t^2 \hat{G}^T \{ (E_t \hat{K}_3 + N_0 I_N)^{-1} \oplus (E_t \hat{K}_3 + N_0 I_N)^{-1} \} \hat{G}^T = \hat{G}^T \{ (\hat{K}_3 + (N_0/E_t) I_N)^{-1} \oplus (\hat{K}_3 + (N_0/E_t) I_N)^{-1} \} \hat{G}^T \quad (4.4)$$

Letting E_t get large decreases the eigenvalues of $\hat{K}_3 + (N_0/E_t) I_N$, increasing the eigenvalues of its inverse.

This increases the eigenvalues of the Fisher information matrix and hence increases its trace (the chosen norm). Thus if the signal matrix were unconstrained in size, the optimal signal would be infinite.

There are many ways to constrain the allowed magnitude of S . Suppose that $M: S \rightarrow R^+$ is a measure of how large $S \in S$ is. Then we may reformulate the problem as maximize the norm of the Fisher information matrix subject to the constraint that $M(S) = E_t$. For example, if the energy is a quadratic function of S , M may be selected to be the trace of $S^T S$. At this point the measure is not too important as long as it constrains the largest value any element of S may take. This is not to say that the computational results will not depend on M , because they will.

Using any standard text on optimization theory (for example, see [7]), a necessary condition for the solution to this problem is that there exist a λ and an $S \in S$ such that

$$\frac{\partial \text{Tr}\{G(K_r^{-1} \otimes K_r^{-1})G^T\}}{\partial S} + \lambda \frac{\partial M(S)}{\partial S} = 0, \quad (4.5)$$

and

$$M(S) = E_t. \quad (4.6)$$

The partial derivative of the trace of the Fisher information matrix with respect to S in the direction δS is

$$\text{Tr}\{K_r^{-1} S^T V^T V S K_r^{-1} (\delta S^T V^T V S + S^T V^T V \delta S)\}, \quad (4.7)$$

where V is the matrix from (2.9), (2.10), and the accompanying discussion such that $\Gamma_3 = VS$. After selecting a particular M , the previous equations are solved for the signal matrix.

This approach to signal selection has the disadvantage that the equations to be solved are highly nonlinear. It should be noticed that the signal matrix enters the partial derivative trace result both directly and through the matrix K_r^{-1} . Even though these equations may be difficult to solve, in practice they are usually computed offline and thus would not affect the realizability of the proposed system. Any constraints on the signal matrix are included in S and M so any signal computed by the above scheme should produce an implementable signal matrix.

As mentioned previously, the choice of M plays a large part in the complexity of the calculations. A good choice of M could yield a calculation for the signal matrix in closed form.

The results derived so far pertain to the special case where $\Omega = \Sigma$. In an implementation, the actual Σ is unknown. One possible way to take this into account is to transmit multiple pulses. For the first pulse, transmit the signal which would be ideal for some assumed generic target (perhaps a uniform spectrum). Calculate an image using the algorithm in [1]. For pulse k , transmit the pulse which is ideal for the image which is the output of the $k-1$ imaging stage. The convergence of this scheme must be explored.

5. Conclusions

This paper has two major results on spectrum estimation problems where the observed data is the sum of white noise and a linear combination of a realization of a portion of a periodic process. The results of each part are derived from the Cramer-Rao lower bound for the variance of spectrum estimates. The first set of results deals with uniqueness of maximum likelihood estimates of the spectrum samples. In particular, if the Fisher information matrix is singular then there is no positive definite spectrum which yields a unique maximum of the likelihood function. The second set of results proposes a new method for selecting radar signals based on the Fisher information matrix. Transmitted signals are selected to maximize the performance achievable. It is attempted to minimize the variance of the estimates of the spectrum samples. Thus the images produced may be relied upon with more confidence.

We are still in the process of implementing the algorithm proposed in [1] to produce delay doppler images using maximum likelihood estimation techniques. Some preliminary experiments have been performed on one dimensional versions of the imaging problem and the results appear promising [8]. When going to the full two dimensional problem, the issue of choosing a good transmitted signal arises. It is hoped that the method proposed here will yield better images than those presently used.

6. References

1. D. L. Snyder, J. A. O'Sullivan, and M. I. Miller, "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar-Targets from Delay-Doppler Data," *IEEE Transactions on Information Theory*, to appear 1988.
2. M. I. Miller and D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proc. IEEE*, Vol. 75 No. 7, July 1987.
3. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III*, Wiley and Sons, New York, 1971.
4. P. Morin, D. L. Snyder, and J. A. O'Sullivan, "Performance Analysis for Maximum Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," submitted to 26th Allerton Conference.
5. P. R. Wehner, *High Resolution Radar*, Artech House, Norwood, MA, 1987.
6. J. Shapiro, B. A. Capron, and R. C. Haney, "Imaging and Target Detection with a Heterodyne-Reception Optical Radar," *Applied Optics*, Vol. 20, No. 19, pp. 3292-3313, October 1981.
7. D. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison Wesley, Reading, MA, 1973.

4.4. Personnel

The following individuals contributed to this research project during the reporting period.

Faculty:

Joseph A. O'Sullivan

Donald L. Snyder

Graduate Students:

Kenneth E. Krause

Pierre Moulin

J. Trent Wohlschlaeger