Fig. 5. Output SINR's of the arrays versus LDE with two external interferences and with $\mu$ as a parameter ($M = 10$ and $\gamma_i = 20$ dB).

In this note, we proposed a signal subspace method for a beam-steered adaptive array and analyzed its performance analytically when steering errors exist and when they do not. From analytical and simulation results, we showed that the signal subspace method is very robust to steering error due to look-direction error or random steering error.

V. Conclusion

In this note, we proposed a signal subspace method for a beam-steered adaptive array and analyzed its performance analytically when steering errors exist and when they do not. From analytical and simulation results, we showed that the signal subspace method is very robust to steering errors which are caused by look-direction mismatch or random errors due to array imperfections. Especially, in the absence of interferences, the output SNR of the array with proposed method is always $M$ (the number of array elements) times the input SNR of the array irrespective of the presence of steering errors. The proposed method works as far as the input covariance matrix can be decomposed into signal and noise subspaces.

A Class of Statistical and Spectral Distance Measures Based on Bose–Einstein Statistics

Luc Knockaert

Abstract—A class of statistical distance measures and their spectral counterparts are presented. They have strong physical foundations since they are based on the combinatorial law leading to Bose–Einstein statistics in statistical physics. It is shown that these distance measures are very closely related to the recently introduced Jensen-Shannon divergence measure. The Kullback-Leibler number is found to be a limit case of this class.

I. Introduction

The Kullback-Leibler information number [1] or cross-entropy [2] is a well-known tool in statistical analysis. Many theorems show that it is a useful measure to discriminate between probability distributions [3], [4]. In addition, for Gaussian processes it can be shown [4] that the KL information number is closely related to the Itakura–Saito spectral distance measure [5]; more precisely, Itakura–Saito distance is equal to $4\pi$ times the asymptotic KL number.

It is known [6] that the concept of entropy and cross-entropy is directly related to the multinomial law. This is easily demonstrated: in the multinomial law, which describes the combinatorial problem of distributing $N$ items (balls) over $M$ levels (boxes) and where each level has a given a priori probability of occupation, the likelihood $P_M$ of a particular realization $\{n_i\}$ is

$$ P_M(n_1, \ldots, n_M | q_1, \ldots, q_M) = N! \prod_{i=1}^{M} q_i^{n_i}/n_i! $$

where the $q_i$ are the priors and of course $\sum_{i=1}^{M} q_i = 1$. Letting $p_i = n_i/N$, taking logarithms, and invoking Stirling’s formula we readily obtain the log-likelihood formula in the limit for $N \to \infty$:

$$ \lim_{N \to \infty} \frac{1}{N} \ln P_M = \sum_{i=1}^{M} p_i \ln \frac{q_i}{p_i} = -KL(p, q) $$

References

where $KL(p, q)$ is the KL number for the discrete distributions $p_i$ and $q_i$. Hence, it is seen that the concept of minimum cross-entropy can be traced back to maximum likelihood. This is explained in great detail in Frieden [6], [7].

We are now in a position to ask the most important question of all. What is the physical rationale for using the multinomial law? The multinomial law is used as a starting point in statistical physics and can be traced back to maximum likelihood. This is explained in detail in Frieden [6], [7].

The question is, what is the probability of a particular realization $\{n_i\}$ given the degeneracies $\{\omega_i\}$?

The solution to this problem is [8]

$$P_b(n_1, \ldots, n_M|\omega_1, \ldots, \omega_M) = \frac{1}{C} \prod_{i=1}^M \frac{(n_i + \omega_i - 1)!}{n_i!(\omega_i - 1)!},$$

where the normalizing factor $C$ is given by

$$C = \frac{(N + \sum \omega_i - 1)!}{N!(\sum \omega_i - 1)!}.$$  

On the other hand, we need to know the conditional probability of the realization $\{n_i\}$ given the prior probabilities $\{q_i\}$ of each level. To find the answer consider first the simple two-level ($M = 2$) case with prior probabilities $q_1 = 1/3, q_2 = 2/3$. Obviously the second level has twice the prior probability of the first level, but one sees that this can also be realized by assigning a degeneracy factor 2 to the second level and a degeneracy factor 1 to the first. Hence, generally speaking, the prior probabilities can be dealt with by assigning a degeneracy factor

$$\omega_i = Kq_i, \quad i = 1, \ldots, M$$

to level $i$, where $K$ is the least common multiple of the denominators of the $q_i$, which are assumed to be rational numbers.

Hence, the Bose–Einstein likelihood function is

$$P_b(n_1, \ldots, n_M|\omega_1, \ldots, \omega_M) = \frac{N! (K - 1)!}{(N + K - 1)!} \cdot \prod_{i=1}^M \frac{(n_i + Kq_i - 1)!}{n_i!(Kq_i - 1)!}.$$  

Taking the negative logarithm of (13) and Stirling’s approximation with $n_i = Np_i, K = N\alpha$, we obtain the Bose–Einstein equivalent of the KL number, which we call BE number

$$BE_\alpha(p, q) = \lim_{N \to \infty} - \frac{1}{N} \ln P_b$$

$$= \sum_{i=1}^M \left[ p_i \ln p_i + \alpha q_i \ln q_i - (p_i + \alpha q_i) \ln \left( \frac{p_i + \alpha q_i}{1 + \alpha} \right) \right].$$

The strict convexity [14] of the function $x \ln x$ over the interval $[0, 1]$ guarantees that the BE numbers are nonnegative. The continuous version of (14) is

$$BE_\alpha(p, q) = \int \left[ p \ln p + \alpha q \ln q - (p + \alpha q) \ln \left( \frac{p + \alpha q}{1 + \alpha} \right) \right] dp.$$  

Note that

$$\alpha = \lim_{N \to \infty} K$$

which can be any positive real number.
In terms of the Shannon entropy

\[ H(p) = -\int p \ln p \, d\mu \quad (17) \]

the BE numbers can be neatly expressed as

\[ BE_\alpha(p, q) = (1 + \alpha) \left[ H \left( \frac{p + \alpha q}{1 + \alpha} \right) - \frac{\alpha}{1 + \alpha} H(p) \right]. \quad (18) \]

This is very closely related to the Jensen–Shannon divergence [9]

\[ JS_\alpha(p, q) = H(\pi, p + \pi_2 q) - \pi_1 H(p) - \pi_2 H(q), \quad (19) \]

where \( \pi_1, \pi_2 \geq 0, \pi_1 + \pi_2 = 1 \) are the weights of the two probability distributions \( p \) and \( q \), respectively. If one takes \( \pi_1 = 1/(1 + \alpha) \), it is clear that

\[ BE_\alpha(p, q) = (1 + \alpha)JS_\alpha(p, q). \quad (20) \]

For \( \alpha = 1 \) we obtain a symmetric distance measure which is called the \( L \)-divergence in [9]

\[ L(p, q) = BE_1(p, q) = 2H\left( \frac{p + q}{2} \right) - H(p) - H(q). \quad (21) \]

It is straightforward to prove the following properties of the BE numbers:

1) The BE numbers are Csiszar \( f \)-divergences with

\[ f(x) = ax \ln x - (1 + ax) \ln \left( \frac{1 + ax}{1 + \alpha} \right). \]

The \( f \)-divergence constitutes an important family of divergence measures introduced by Csiszar [15]. Many of the properties of \( f \)-divergence can be found in [4, 16].

2) \( BE_\alpha(p, q) = 0 \) if \( p = q \) almost everywhere

3) \( \frac{\partial}{\partial \alpha} BE_\alpha(p, q) = KL \left( q, \frac{p + \alpha q}{1 + \alpha} \right) \geq 0. \quad \alpha \to -\infty \)

4) \( \lim_{\alpha \to -\infty} BE_\alpha(p, q) = KL(p, q). \)

5) \( BE_{\alpha}(p, q) = \frac{1}{\alpha} BE_\alpha(q, p). \)

6) \( BE_\alpha(p, q) = \alpha KL(q, p), \) for \( \alpha \) sufficiently small.

We now show that the minimization of the \( L \) number under constraints (4) yields the Bose–Einstein statistics. The minimum condition is

\[ \ln p - \ln (p + q) + \Psi = 0, \]

where \( \Psi = \Phi - 1 + \ln 2 \) and \( \Phi \) is defined in (6). The solution is simply

\[ p = \frac{q}{e^{\frac{q}{\alpha}} - 1}, \quad (22) \]

and we recognize the celebrated Bose–Einstein term.

As a last result we show that the BE numbers for zero-mean Gaussian distributions with respective covariances \( R \) and \( S \) are approximately Chernoff distances. We approximate the \( \ln (p + \alpha q) \) term by a normal i.e.,

\[ p + \alpha q = N(0, D), \quad (24) \]

In order to have equality of moments up to order 2 we should have:

\[ D = \int \frac{p + \alpha q}{1 + \alpha} \, d\mu = \frac{R + \alpha S}{1 + \alpha}. \quad (25) \]

This leads to the distance measure between covariances

\[ d_\alpha(R, S) = BE_\alpha(p, q) \]

\[ = \frac{1 + \alpha}{2} \ln \left[ \frac{R + \alpha S}{2} \right] - \frac{1}{2} \ln |R| - \frac{\alpha}{2} \ln |S|. \quad (26) \]

This is directly related to the Chernoff distance \( C_\alpha(R, S) \) between two zero-mean Gaussians with respective covariances \( R \) and \( S \) [11]:

\[ d_\alpha(R, S) = (1 + \alpha) C_\alpha(R, S), \quad \text{with } \tau = \frac{\alpha}{1 + \alpha}. \quad (27) \]

Finally, applying Whittle’s result [12], [17], yields a distance between spectra:

\[ d_s[\tau(\theta), s(\theta)] = \lim_{N \to \infty} 4\pi \int_{-\pi}^{\pi} \frac{1}{1 - \tau - i\epsilon} \ln \left( \frac{1 - \tau \tau(\theta) + \tau s(\theta)}{r(\theta)^{-1}s(\theta)^{-1}} \right) d\theta \]

where \( \tau \) is the asymptotic Chernoff distance [11]. For \( \alpha = 1 \) or \( \tau = \frac{\alpha}{1 + \alpha} \) we obtain the Bhattacharyya distance [10].

**ACKNOWLEDGMENT**

The author would like to thank Dr. J. Van Calck for some stimulating discussions, and the anonymous reviewers for their helpful comments.

**REFERENCES**


Analysis of Linear Prediction by Matrix Approximation

Alex C. Kot, Donald W. Tufts, and Richard J. Vaccaro

Abstract—We present a new method of analysis of the effects of data perturbations on the estimates of signal parameters which are obtained by the method of linear prediction. The variances of signal-parameter estimates to signals which are linear combinations of complex-valued exponential functions.

I. INTRODUCTION

A new backward error analysis is presented here based on matrix approximation for linear prediction methods, or equivalently Prony methods. There have been different methods proposed by other authors in the past few years to get a precise, statistical characterization of the performance of linear prediction methods, for parameter estimation, system identification and high-resolution spectral estimation. One approach to analyze the statistical performance of linear prediction methods for parameter estimation in [1] is based on Wilkinson's approach [2] and the statistical results in [3] to derive a first-order approximation to the first and second moments of the perturbed signal eigenvalues and eigenvectors. However, a more tractable asymptotic derivation for the second moment can be found in [4]. A similar approach can be found in [5], wherein the derivation of the first derivative of singular values and vectors with respect to perturbations in the data is used to calculate the perturbation of the linear prediction coefficients. Some asymptotic perturbation analysis in frequency estimation based on large data point assumption can be found in [6]. More recent perturbation analysis in linear prediction methods can be found in [7]-[9]. However, all of the previous analytical approaches have been forward error analyses based on the assumption of small perturbations on the data. The new approach taken in this performance study is a backward error analysis. That is, we consider linear prediction as a minimization problem, and approximate this problem by keeping only first-order perturbation terms. Then, an exact solution to the approximate minimization problem is obtained and analyzed. However, this new approach yields much simpler analytical results, especially for multicomponent signals. In our analysis, we partition the perturbation into signal and orthogonal subspace and consider only the perturbation to the signal. The calculated variance expression is essentially the same as in [8], wherein a result due to Wedin [10] was used as a major ingredient in their analysis. However, our work is simpler and the analysis is self-contained. In addition, the analysis yields insights for design and tractable results even for the case of multicomponent signals, and for both Toeplitz and Hankel data matrices. Computer simulations are provided to verify our analysis.

II. PERTURBATION ANALYSIS

In this section, we assume the measured data is slightly perturbed. We first analyze the prediction coefficient perturbation using matrix approximation for linear prediction (LP) with singular value decomposition (SVD). Then we use the first-order relationship between the perturbations in prediction coefficients, and root locations. The perturbations in root locations can be expressed as radius and angle perturbations.

A. Problem Formulation

We begin by establishing some notation. Given a noisy data sequence \( \{y_i = y_i + n_i\} \), where \( y_i = \sum_{k=0}^{N} A_k e^{j\omega_i k} \), \( A_i \) is the amplitude, \( \omega_i \) is the frequency of the signal to be estimated, \( n_i \) is the \( k \)th noise component, \( p \) is the number of complex sinusoids. The \((N - L) \times L\) matrix \( \hat{Y} \) and the \((N - L) \times 1\) vector \( \hat{y} \) are formed

\[
\hat{Y} = \begin{bmatrix}
    \hat{y}_L & \hat{y}_{L-1} & \cdots & \hat{y}_1 \\
    \hat{y}_{L+1} & \hat{y}_L & \cdots & \hat{y}_2 \\
    \vdots & \vdots & \ddots & \vdots \\
    \hat{y}_{N-1} & \cdots & \cdots & \hat{y}_N
\end{bmatrix} \quad \hat{y} = \begin{bmatrix}
    \hat{y}_{L+1} \\
    \hat{y}_{L+2} \\
    \vdots \\
    \hat{y}_N
\end{bmatrix}
\]

where \( L \) is the length of the LP vector which is chosen to be larger than the number of exponential which describes the signal, and \( (\cdot) \) will denote the noisy data.

We let the matrix \( \hat{Y} \) have a singular value decomposition \( \hat{Y} = \hat{U} \hat{S} \hat{V}_Y^H \) which is partitioned as follows:

\[
\hat{Y} = \begin{bmatrix} \hat{U} & \hat{O} \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \overline{\Sigma} \end{bmatrix} \begin{bmatrix} \hat{V}_Y^H \\ 0 \end{bmatrix}
\]

\[
= \hat{U} \Sigma \hat{V}_Y^H + \hat{O} \overline{\Sigma} \tilde{V}_o \tilde{V}_o^H
\]

where \( m = N - L \) and the superscript \( ^* \) denotes the conjugate transpose.

The partitioning of the SVD to yield a rank \( p \) matrix \( \hat{Y} \), can be done on the basis of \textit{a priori} knowledge that the data contains \( p \) exponential components. Alternatively, a value of \( p \) can be chosen by determining the number of "dominant" singular values. A linear prediction procedure which uses the SVD of \( \hat{Y} \) is described in [11]. We also denote \( \hat{Y} = Y + N \), where

\[
\text{Signal matrix } Y' = [\hat{y}^T, Y] = [\hat{y}_N + 1, \hat{U}, \Sigma, V_Y^H]
\]

\[
\text{Noise matrix } N' = [\hat{y}^T, N]
\]