[2] G. Cybenko, "Fast approximation of dominant harmonics," *SIAM J. Sci. Stat. Comput.*, vol. 5, pp. 317-331, June 1984.
[3] P. Stoica and A. Nehorai, "The poles of symmetric linear prediction models lie on the unit circle: A new proof and computational aspects," Rep. 8606, Center for Syst. Sci., Yale Univ., New Haven, CT, July 1986.
[4] ——, "The poles of symmetric linear prediction models lie on the unit circle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1344-1346, Oct. 1986.
[5] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification, Lecture Notes in Control and Information Sciences*. Berlin: Springer-Verlag, 1983.
[6] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.
[7] P. Stoica and A. Nehorai, "On stability and root location of linear prediction models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 582-584, Apr. 1987.
[8] J. Makhoul, "Linear prediction—A tutorial review," *Proc. IEEE*, vol. 63, pp. 561-580, 1975.
[9] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, CA, May 1975.
[10] P. Stoica, B. Friedlander, and T. Söderström "Asymptotic bias of the high-order autoregressive estimates of sinusoidal frequencies," *Circ. Syst. Signal Processing*, vol. 6, pp. 287-298, 1987.
[11] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 347-366, 1973.
[12] P. Delsarte and Y. V. Genin, "The split Levinson algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 470-478, June 1986.
[13] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, S35(A), 1975.
[14] F. K. Soong and B.-H. Juang, "Line spectrum pair (LSP) and speech data compression," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, San Diego, CA, Mar. 1984, pp. 1.10.1-1.10.4.
[15] G. Cybenko, "Computing Pisarenko frequency estimates," in *Proc. 1984 Conf. Inform. Syst. Sci.*, Princeton Univ., Mar. 1984.

# Maximum Likelihood Estimation of Linear Signal Parameters for Poisson Processes

MICHAEL UNSER AND MURRAY EDEN

*Abstract*—**The estimation of linear signal parameters is studied under the hypothesis of independently Poisson distributed measurements. A simple iterative maximum likelihood estimator (MLE) is derived and is optimized for rapid convergence. It is shown to be statistically optimal in the sense of providing unbiased and minimum variance estimates. Experimental conditions are identified where MLE results in significant performance improvement when compared to conventional linear least squares (LLS).**

## I. Introduction

Within the armamentarium of instruments for biomedical research there are a number whose data are acquired by counting independent events that fall into particular intervals of the measured variables. For example, flow cytometers count cells within bins or channels corresponding to particular intervals of DNA mass per cell [1], PET scanners count coincidences of gamma rays emitted from a positron annihilation corresponding to photo-emission in crystal pairs occurring within a short time interval [2], energy dispersive X-ray spectrometers or electron energy loss spectrometers count X-ray photons and electrons respectively in different energy intervals [3]–[4]. The resulting histograms of count versus channel number are then analyzed in order to estimate averages, standard deviations, and other statistics of the experiment. In essence, the analysis is equivalent to making a model of the underlying process and estimating its parameters.

Two factors need to be taken into account. First, the channel counts per unit time, denoted by $\{x_k\}$, where $k = 1, \cdots, N$ is the channel index, are generally assumed to be independent Poisson random variables with parameters $\{\mu_k, k = 1, \cdots, N\}$. Second, the likelihood that a count will be added to any particular channel depends on physical or instrumental factors which determine the choice of an appropriate signal representation. A wide range of problems is approached by assuming that the signal, defined as $\{\mu_k = E\{x_k\}, k = 1, \cdots, N\}$, may be linearly decomposed as

$$\mu_k = \sum_{i=1}^{M} a_i t_k^{\langle i \rangle}, \qquad (k = 1, \cdots, N) \tag{1}$$

where the $t_k^{\langle i \rangle}$'s are known characteristic (or basis) functions of the channel number $(k)$, and where $M$ is usually much smaller than $N$. These basis functions are usually directly related to some elementary events of interest. This type of representation uses the principle of superposition. The unknown parameters or expansion coefficients $\{a_i, i = 1, \cdots, M\}$, which are the weights of those functions, usually provide measures of the occurrence of the events of interest. For example, in X-ray spectrometry, a particular nonvolatile element in a sample of fixed dimension will induce X-ray emission of photons with a characteristic energy spectrum. The expected value for the number of photons generated per unit time within any particular energy interval (channel) is linearly related to the number of atoms of that element. Thus, for a mixed sample of several species, the measured spectrum is the linear superposition of a set of characteristic functions (the individual reference spectra, and the background radiation) weighted by the relative concentrations of the elements. In practice, reference spectra are either recorded as a whole and corrected for the average effect of background or decomposed as a sum of a few characteristic lines which are usually well represented by Gaussian curves [3], [4].

As in the example of X-ray analysis, physical or instrumental factors may also include a background, or "dark current," unrelated to the events of interest. Such a slowly varying continuum is usually well represented by a lower order polynomial function [5]. In some other instances, the signal is known to be band-limited, due to the physics of the process that is being monitored, or as a result of the convolution with the characteristic response of the detector which often has the effect of a low-pass filter. In such a case, a representation with a limited number of Fourier coefficients may be appropriate.

It is customary for investigators to estimate the parameters of such models using least squares techniques [3], [6]. Linear least squares (LLS) is widely used because it is familiar to virtually all biological scientists and it is simple to do. LLS is optimal for additive white Gaussian noise, but is not the most appropriate for nonstationary Poisson noise, where the variance is equal to the expected value of the underlying signal. In this correspondence, we investigate the use of maximum likelihood estimation (MLE) which, in the present case, is shown to be optimal in the sense of providing minimum mean square error estimates of the signal parameters. An important practical issue is to explore whether there are ranges of experimental interest for which improvement of performance due to MLE is sufficient to justify added complexity when compared to conventional LLS.

Manuscript received January 9, 1987; revised December 14, 1987.
The authors are with the Biomedical Engineering and Instrumentation Branch, Division of Research Services, National Institutes of Health, Bethesda, MD 20892.
IEEE Log Number 8820430.

0096-3518/88/0600-0942$01.00 © 1988 IEEE

## II. Parameter Estimation

The problem stated in the Introduction is equivalent to estimating the mean vector of a Poisson process characterized by the linear model

$$E\{\boldsymbol{x}\} = \boldsymbol{\mu} = \boldsymbol{T}\boldsymbol{a} \tag{2}$$

where $\boldsymbol{\mu} = [\mu_1 \cdots \mu_N]^T$ is the mean vector of the measurement vector $\boldsymbol{x} = [x_1 \cdots x_N]^T$, $\boldsymbol{T} = [t_1 \cdots t_M]$ an $N \times M$ matrix, and $\boldsymbol{a} = [a_1 \cdots a_M]^T$ the vector of expansion coefficients. An estimate $\hat{\boldsymbol{a}}$ of the parameter vector $\boldsymbol{a}$ is computed on the basis of an experimental instance of $\boldsymbol{x}$ using the known transformation matrix $\boldsymbol{T}$ whose column vectors are obtained from the characteristic functions in (1).

### A. Properties of the Linear Poisson Model

The sequence of Poisson parameters $\{\mu_k\}$ represents the mean of the Poisson process but also the variance of the individual measurements. For independently distributed Poisson variables, the log-likelihood function of $\boldsymbol{x}$ is

$$L(\boldsymbol{x}|\boldsymbol{\mu}) = \log\left(\sum_{k=1}^{N} \frac{e^{-\mu_k} \cdot \mu_k^{x_k}}{x_k!}\right) = \sum_{k=1}^{N} \left(-\mu_k + x_k \log(\mu_k) - \log(x_k!)\right) \tag{3}$$

and its covariance matrix is the diagonal matrix: diag $\{\boldsymbol{\mu}\}$.

Since the number of parameters ($M$) in (2) is generally much smaller than the number of measurements ($N$), the estimation of the expansion coefficients for a given $\boldsymbol{x}$ generally results in substantial noise reduction. More specifically, the noise contribution before data reduction is measured by

$$E\{\|\boldsymbol{x} - \boldsymbol{\mu}\|^2\} = \operatorname{tr}(\operatorname{diag}\{\boldsymbol{\mu}\}) = \sum_{i=1}^{N} \mu_k. \tag{4}$$

When $\boldsymbol{x}$ is used to compute an estimate $\hat{\boldsymbol{a}}$ of the signal parameter in (2), the residual noise after processing is measured by the mean square error between the estimated and true mean vectors:

$$E\{\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2\} = E\{\|\boldsymbol{T}(\hat{\boldsymbol{a}} - \boldsymbol{a})\|^2\} = \operatorname{tr}\{\boldsymbol{T}\boldsymbol{C}_{\hat{\boldsymbol{a}}}\boldsymbol{T}^T\} \tag{5}$$

which is a function of the estimator's covariance matrix: $\boldsymbol{C}_{\hat{\boldsymbol{a}}} = E\{(\hat{\boldsymbol{a}} - \boldsymbol{a})(\hat{\boldsymbol{a}} - \boldsymbol{a})^T\}$. This criterion provides a good performance indicator. It also indicates that noise reduction is maximized when the variances of expansion coefficients are minimized.

An upper bound for the performance of any estimator may be obtained by using the Cramer–Rao inequality [7]. By substituting (2) in (3) and taking the partial derivatives with respect to $\boldsymbol{a}$, the Fisher information matrix is computed as

$$\begin{aligned}\boldsymbol{J} &= E\{(L(\boldsymbol{x}|\boldsymbol{a})/\partial\boldsymbol{a})(L(\boldsymbol{x}|\boldsymbol{a})/\partial\boldsymbol{a})^T\} \\ &= E\{\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\} \\ &= \boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}.\end{aligned} \tag{6}$$

It follows that the covariance of any unbiased estimator of $\boldsymbol{a}$ is bounded by

$$\boldsymbol{C}_{\hat{\boldsymbol{a}}} \geq \boldsymbol{J}^{-1} = (\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T})^{-1} \tag{7}$$

### B. Linear Least Squares (LLS)

The simplest estimation technique is LLS where the parameter vector is determined by minimizing the quadratic error between the model (2) and the measurement vector $\boldsymbol{x}$, and leads to the well-known solution

$$\hat{\boldsymbol{a}}_{\mathrm{LS}} = (\boldsymbol{T}^T\boldsymbol{T})^{-1}\boldsymbol{T}^T \cdot \boldsymbol{x} = \boldsymbol{T}^{\dagger} \cdot \boldsymbol{x}. \tag{8}$$

$\boldsymbol{T}^{\dagger}$ is the so-called pseudoinverse of $\boldsymbol{T}$ and is equal to $\boldsymbol{T}^T$ when the $t_i$'s are orthogonal. It is straightforward to show that the LLS estimator of $\boldsymbol{a}$ is unbiased and that its covariance matrix is given by

$$\boldsymbol{C}_{\mathrm{LS}} = E\{(\boldsymbol{a} - \hat{\boldsymbol{a}}_{\mathrm{LS}})(\boldsymbol{a} - \hat{\boldsymbol{a}}_{\mathrm{LS}})^T\} = \boldsymbol{T}^{\dagger} \operatorname{diag}\{\boldsymbol{\mu}\}(\boldsymbol{T}^{\dagger})^T. \tag{9}$$

By using the fact that $\hat{\boldsymbol{a}}$ is linearly related to $\boldsymbol{x}$ and that the Poisson distribution can be frequently approximated by a Gaussian law, it follows that the distribution of $\hat{\boldsymbol{a}}$ is approximately multivariate Gaussian: $\hat{\boldsymbol{a}}_{\mathrm{LS}} \sim N(\boldsymbol{a}, \boldsymbol{C}_{\mathrm{LS}})$.

### C. Maximum Likelihood Estimation (MLE)

The maximum likelihood estimator $\hat{\boldsymbol{a}}_{\mathrm{ML}}$ is defined as the parameter vector that maximizes the joint probability density function of the measurement vector. By using the constraint that $\boldsymbol{\mu} = \boldsymbol{T}\boldsymbol{a}$, $\hat{\boldsymbol{a}}_{\mathrm{ML}}$ is determined by taking the partial derivative of (3) with respect to $\boldsymbol{a}$, and solving

$$\begin{aligned}\partial L(\boldsymbol{x}|\boldsymbol{a})/\partial\boldsymbol{a} &= \boldsymbol{T}^T \cdot \partial L(\boldsymbol{x}|\boldsymbol{\mu})/\partial\boldsymbol{\mu} \\ &= \boldsymbol{T}^T \cdot \left[\left(1 - \frac{x_1}{\mu_1}\right) \cdots \left(1 - \frac{x_N}{\mu_N}\right)\right]^T = 0.\end{aligned} \tag{10}$$

A numerical solution can be found by using a standard steepest descent or gradient method. An iteration of this algorithm is described by

$$\boldsymbol{a}_{(k+1)} = \boldsymbol{a}_{(k)} + \alpha_k \cdot \boldsymbol{T}^T\boldsymbol{e}_{(k)} \tag{11a}$$

or, equivalently, by

$$\boldsymbol{\mu}_{(k+1)} = \boldsymbol{\mu}_{(k)} + \alpha_k \cdot (\boldsymbol{T}\boldsymbol{T}^T) \cdot \boldsymbol{e}_{(k)} \tag{11b}$$

where $\boldsymbol{e}_{(k)}$ is the gradient of log-likelihood function with respect to $\boldsymbol{\mu}$ evaluated as $\boldsymbol{\mu}_{(k)}$ given by

$$\boldsymbol{e}_{(k)}^T = \left[\frac{x_1 - \mu_{(k),1}}{\mu_{(k),1}} \cdots \frac{x_N - \mu_{(k),N}}{\mu_{(k),N}}\right], \tag{12}$$

and which measures the relative error between $\boldsymbol{x}$ and the current mean estimate $\boldsymbol{\mu}_{(k)}$.

The rate of convergence is improved significantly by selecting, at each iteration, the step size $\alpha_k$ that provides the greatest increase in the likelihood function. This is achieved by choosing $\alpha_k$ such that the gradient at iteration $(k + 1)$ is orthogonal to the gradient at iteration $(k)$ [8]. Using (11b) and (12), the error at iteration $(k + 1)$ can be expressed as

$$\begin{aligned}\boldsymbol{e}_{(k+1)} &= \operatorname{diag}\{\boldsymbol{\mu}_{(k+1)}\}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{(k)} - \alpha_k(\boldsymbol{T}\boldsymbol{T}^T)\boldsymbol{e}_{(k)}) \\ &\cong (1 - \alpha_k \operatorname{diag}\{\boldsymbol{\mu}_{(k)}\}^{-1}(\boldsymbol{T}\boldsymbol{T}^T))\boldsymbol{e}_{(k)}\end{aligned}$$

where the simplified expression on the right-hand side is obtained by assuming that $\boldsymbol{\mu}_{(k)}$ is reasonably close to $\boldsymbol{\mu}_{(k+1)}$ and that (diag $\{\boldsymbol{\mu}_{(k+1)}\}^{-1}$ diag $\{\boldsymbol{\mu}_{(k)}\}$) is therefore almost equal to the $N \times N$ identity matrix. Hence, the scalar product between two successive evaluations of $\partial L(\boldsymbol{x}|\mu)/\partial\boldsymbol{a}$ is approximately equal to

$$\begin{aligned}\boldsymbol{e}_{(k)}^T\boldsymbol{T} \cdot \boldsymbol{T}^T\boldsymbol{e}_{(k+1)} &\approx \boldsymbol{e}_{(k)}^T(\boldsymbol{T}\boldsymbol{T}^T) \\ &\cdot (1 - \alpha_k \operatorname{diag}\{\boldsymbol{\mu}_{(k)}\}^{-1}(\boldsymbol{T}\boldsymbol{T}^T))\boldsymbol{e}_{(k)}.\end{aligned}$$

The optimal step is found by setting this expression to zero and is given by

$$\alpha_k = \frac{\|\boldsymbol{T}^T\boldsymbol{e}_{(k)}\|^2}{(\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{e}_{(k)})^T \operatorname{diag}\{\boldsymbol{\mu}_{(k)}\}^{-1}(\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{e}_{(k)})}. \tag{13}$$

As both vectors $(\boldsymbol{T}^T\boldsymbol{e}_{(k)})$ and $(\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{e}_{(k)})$ have to be evaluated successively in order to compute the update in (11a, b), the determination of $\alpha_k$ is relatively straightforward. A cycle of this algorithm is more or less equivalent in complexity to an LLS in the simplest case of orthogonal basis functions. In such a situation, the correction term $\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{e}_{(k)}$ may be interpreted as the projection of the error on the subspace defined by $t_1, \cdots, t_M$. The least squares estimate given by (6) provides an initial guess $\boldsymbol{a}_{(0)}$ which is usually very close to the correct ML solution. Under such circumstances, it has been verified experimentally that the algorithm converges rapidly to the correct solution with a number of iterations approximately equal to $M$ (the dimensionality of $\boldsymbol{a}$).

It is demonstrated in the Appendix that the MLE provides the best possible estimator of $\boldsymbol{a}$ in the sense that it is unbiased and has minimum variance; this property is usually referred to as *efficiency*. It is also shown that its covariance matrix corresponds to the lower Cramer-Rao bound given by (7):

$$\boldsymbol{C}_{\mathrm{ML}} = E\left\{(\boldsymbol{a} - \hat{\boldsymbol{a}}_{\mathrm{LS}})(\boldsymbol{a} - \hat{\boldsymbol{a}}_{\mathrm{LS}})^T\right\} = \left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right)^{-1} \tag{14}$$

As for the LLS, the joint distribution of $\hat{\boldsymbol{a}}_{\mathrm{ML}}$ is approximately multivariate Gaussian. This assertion is further supported by the well-known asymptotic properties of maximum likelihood estimators [7].

## III. Simulation and Discussion

For the comparison of LLS and MLE, some general observations can be made on the basis of a simple two component signal model:

$$\mu_k = a_1 + a_2 t_k \tag{15}$$

where $a_1$ represents a constant background on which has been superposed some waveform represented by a unique function $\{t_k, k = 1, \cdots, N\}$. Typical examples of such signals are given in Fig. 1. The loss of performance of LLS when compared to MLE may be measured from the ratios of the variances of the estimated expansion coefficients

$$\rho_i = \operatorname{Var}\{\hat{a}_{i\,\mathrm{ML}}\}/\operatorname{Var}\{\hat{a}_{i\,\mathrm{LS}}\} \leq 1; \qquad i = 1, \cdots, M < N, \tag{16}$$

or by a more global criterion which is the ratio of their expected mean square errors for the estimation of $\boldsymbol{\mu}$, and which is evaluated as

$$\rho = \operatorname{tr}\{\boldsymbol{T}\boldsymbol{C}_{\mathrm{ML}}\boldsymbol{T}^T\}/\operatorname{tr}\{\boldsymbol{T}\boldsymbol{C}_{\mathrm{LS}}\boldsymbol{T}^T\} \leq 1. \tag{17}$$

An important observation is to be made: for Poisson signals, multiplying the mean $\boldsymbol{\mu}$ by a factor $\alpha$ (which is usually achieved by increasing the acquisition time in the same proportion), decreases the signal-to-noise ratio by $\sqrt{\alpha}$ while the relative performance criteria (16) and (17) remain unchanged. Consequently, for the two component model (15), $\rho_1$, $\rho_2$, and $\rho$ depend only on the ratio $a_2/a_1$ (which is invariant to any scaling factor $\alpha$) and may as well be expressed as a function of $\max\{\mu_k\}/\min\{\mu_k\}$ which, we feel, provides a more useful indicator.

The results of the computation of $\rho$ as a function $\max\{\mu_k\}/\min\{\mu_k\}$ for the cosine, straight line, and quadratic reference signals are shown in Fig. 2. For these particular models, it can be seen that the use of MLE may result in a 20–30 percent greater noise reduction than LLS when $\max\{\mu_k\}/\min\{\mu_k\} > 100$. When the sequence $\{t_k\}$ was chosen to be zero mean, and therefore orthogonal to the background component, the improvement of the performance of MLE was essentially due to a reduction of the variance of $a_2$. In this particular case, both methods performed equally well for the estimation of $a_1$ ($\rho_1 = 1$).

The Gaussian peak model provides a simple example where the relative improvement of the MLE can be made arbitrarily large by decreasing the width of the peak. This is illustrated by Fig. 3 which shows the relative performance criterion computed for various peak widths. It appears that the loss of performance of the LLS increases dramatically as the peak gets narrower. The superiority of the MLE approach is due to the fact that it incorporates the property that, for Poisson measurements, the variance is equal to the mean value. In contrast, LLS applies identical weights to all measurements and therefore provides a fit that is dominated statistically by high signal values. This results in a greater estimation error which is further amplified when the proportion of high values is comparatively small, as is the case when the peak width is narrow. As for the previous example, MLE clearly outperforms LLS when $\max\{\mu_k\}/\min\{\mu_k\} > 10$.

In conclusion, the loss of performance of the LLS over the MLE is far from negligible under the following conditions: i) when the ratio between the maximum and minimum signal level is large, say greater than 10; and ii) when the overall proportion of large signal values is relatively small, which is usually the case when one is estimating peak heights. These are examples for which the usual hypothesis of stationary Gaussian noise is strongly contradicted and where the use of the MLE is to be preferred. These results should be useful for background and peak estimation in low dose scanning

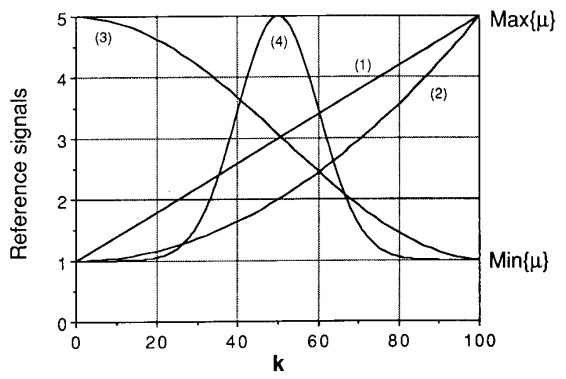


Fig. 1. Examples of reference signals used for simulation. (1) Straight line: $t_k = k + c$; (2) quadratic: $t_k = k^2 + c$; (3) cosine: $t_k = \cos(2\pi k/N)$; (4) Gaussian peak: $t_k = \exp\{-(k - \Delta)^2/2\sigma^2\}$ where $\Delta = N/2$, $\sigma = Nw$, and $k = 0, \cdots, 99$ ($N = 100$).

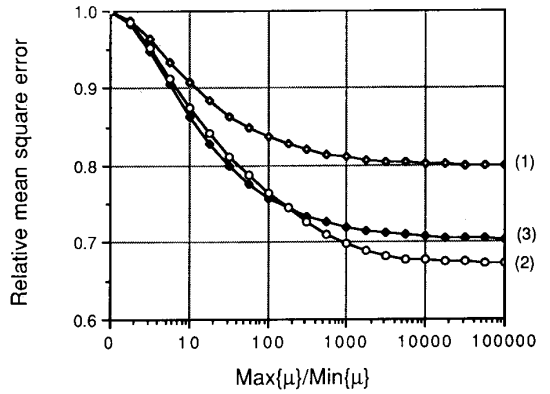


Fig. 2. Ratio of the mean square estimation errors for MLE and LLS (criterion $\rho$) as a function of $\mathrm{Max}\{\mu\}/\mathrm{Min}\{\mu\}$ for signal models (1), (2), and (3) in Fig. 1 and $N = 100$.

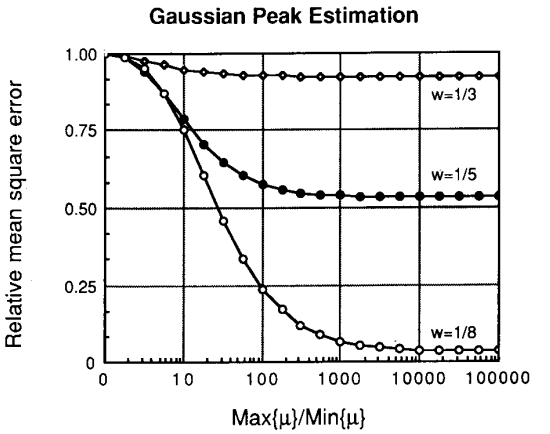


Fig. 3. Ratio of the mean square estimation errors for MLE and LLS (criterion $\rho$) as a function of $\mathrm{Max}\{\mu\}/\mathrm{Min}\{\mu\}$ and the peak width for reference signal (4) in Fig. 1 and $N = 100$. The relative width $w$ is defined as the ratio of the standard deviation of the peak reported to the size of the estimation region.

transmission electron microscopy (STEM), electron energy loss spectrometry (EELS), or energy dispersive X-ray spectrometry where the hypothesis of independent counting statistics is generally well accepted.

## Appendix
## Efficiency of Poisson ML Estimator

The efficiency of the MLE is demonstrated indirectly by using the following theorem.

*Theorem [7, p. 185]:* If an estimator exists such that equality is satisfied in the Cramer–Rao inequality, it can be determined as a solution of the maximum likelihood equation.

We will show that such an estimator exists. This implies that the covariance matrix of the MLE is given by the inverse of the Fisher information matrix.

Let us consider the linear minimum mean square (LMMS) estimator where the samples are weighted by the inverse of their standard deviation:

$$\hat{\boldsymbol{a}}_{\text{LMMS}} = \left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{x} = \boldsymbol{R}\boldsymbol{x}. \quad \text{(A.1)}$$

Note that this formula cannot be applied in a practical situation because it requires the true mean $\boldsymbol{\mu}$ to be known. The LMMS estimator is unbiased and its covariance matrix is computed as

$$\begin{aligned}\boldsymbol{C}_{\text{LMMS}} &= E\left\{\boldsymbol{R}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{R}^T\right\} = \boldsymbol{R}\operatorname{diag}\{\boldsymbol{\mu}\}\boldsymbol{R}^T \\ &= \left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right)^{-1}\left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right) \\ &\quad \cdot \left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right)^{-1} \\ &= \left(\boldsymbol{T}^T \operatorname{diag}\{\boldsymbol{\mu}\}^{-1}\boldsymbol{T}\right)^{-1}. \quad \text{(A.2)}\end{aligned}$$

This expression corresponds to the lower limit specified by the Cramer–Rao bound (7) and therefore proves the efficiency of the MLE.

## References


[1] W. H. Schuette, S. E. Shackney, M. A. MacCollum, and C. A. Smith, "A count-dependent filter for smoothing flow cytometric histograms," *Cytometry*, vol. 5, pp. 487–493, 1984.

[2] R. A. Brooks, M. R. Keller, C. M. O'Connor, and W. T. Sheridan, "Progress toward quantitative computed tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-27, pp. 1121–1127, 1980.

[3] J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*. New York and London: Plenum, 1982.

[4] D. C. Joy, A. D. Romig, and J. I. Goldstein, *Principles of Analytical Electron Microscopy*. New York: Plenum, 1986.

[5] R. J. Gehrke and R. C. Davies, "Spectrum filtering techniques for energy dispersive X-ray analysis of oxides and silicates with electron microbream excitation," *Anal. Chem.*, vol. 48, no. 1, pp. 1537–1546, 1975.

[6] H. Nullens, P. Van Espen, and F. Adams, "Linear and non-linear fitting in energy-dispersive X-ray fluorescence," *X-Ray Spectrometry*, vol. 8, no. 3, pp. 104–109, 1979.

[7] H. W. Sorenson, *Parameter Estimation: Principles and Problems*. New York and Basel: Marcel Dekker, 1980.

[8] C. N. Dorny, *A Vector Space Approach to Models and Optimization*. New York: Wiley, 1975.


# On the Performance Analysis of the MVDR Beamformer in the Presence of Correlated Interference

MICHAEL D. ZOLTOWSKI

***Abstract*—Expressions are developed which describe the output of an adaptive array based on the minimum variance distortionless response (MVDR) beamformer when multiple correlated interfering signals are present. Such signals may arise in certain multipath scenarios or in the case of "smart" jamming. Previous analyses considered only the case of a single correlated interferer. Three scenarios of interest are examined: very high SNR, moderately high SNR, and very low SNR. The analysis is based on simple linear algebra concepts and generalizes a number of previously reported results proved for special cases. In addition, a modification of MVDR beamforming based on Total Least Squares is introduced in the final section as a means for mitigating the sensitivity of the MVDR beamformer to uncertainties in element positions, pointing angle, etc.**

Manuscript received October 22, 1986; revised January 9, 1988.

The author is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.

IEEE Log Number 8820835.

## I. Introduction

In [1], the signal cancellation and interference rejection properties of the minimum variance distortionless response (MVDR) beamformer were studied in detail for the case of a single correlated interferer (in addition to the desired source) and additive noise for various levels of signal-to-noise (SNR) ratio. Although it is pointed out that the results derived in [1] can be extended for the case of more than one interferer, a general analysis based on the approach taken there requires calculations that are quite tedious. Here, a different approach is taken which facilitates the performance analysis of the MVDR beamformer for an arbitrary number of correlated/uncorrelated interferers. Interfering signals which are correlated with the desired source may arise in certain multipath scenarios or in the case of "smart" jamming [1]–[3]. Three scenarios are examined: very high SNR, moderately high SNR, and very low SNR. The approach taken here, based on simple linear algebra, produces results that agree with and extend upon those presented in [1]. In addition, a modification of MVDR beamforming based on Total Least Squares is introduced in the final section to mitigate the sensitivity of the MVDR beamformer to uncertainties in element positions, pointing angle, etc.

## II. Array Signal Model and the Correlation Matrix

Consider an array of $m$ antennas receiving signals from $d$ sources of emission at directions $\theta_k$, $k = 1, \cdots, d$ with respect to the array. The angles of arrival are enumerated such that $\theta_1$ is the angle of the desired source. We will assume that the desired source is narrow-band and that narrow-band filtering about the center frequency of the desired source, $f_0$, occurs at the front end of the receivers such that the $d-1$ interfering signals are narrow-band and co-located in frequency with the desired signal at the beamformer input. We will also assume additive, "spatially white" noise. Let $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^T$, where $x_i(t)$ is the signal received at the $i$th array element. Given the above-assumed scenario, $\boldsymbol{x}(t)$ can be expressed in the following fashion:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t). \quad (1)$$

$\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_d)]$ where $\boldsymbol{A}_{ik} = a_i(\theta_k)$ is the response of the $i$th array element relative to that of the first when a single signal arrives at $\theta_k$, $\boldsymbol{s}(t) = [s_1(t), \cdots, s_d(t)]^T$ where $s_k(t)$ is the signal associated with the $k$th source as received at the first array element, and $\boldsymbol{n}(t) = [n_1(t), n_2(t), \cdots, n_m(t)]^T$ where $n_i(t)$ is the additive noise present at the $i$th receiver. Note that we here assume the $d$ columns of $\boldsymbol{A}$, $\boldsymbol{a}(\theta_k)$, $k = 1, \cdots, d$, to be linearly independent.

With the assumptions made previously, it follows that the correlation matrix of the array element outputs can be written in the following form:

$$\boldsymbol{R} = E\left[\boldsymbol{x}(t)\,\boldsymbol{x}^H(t)\right] = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma_n^2\boldsymbol{I} \quad (2)$$

where $H$ denotes conjugate (Hermitian) transpose. $\boldsymbol{S}$ is the $d \times d$ source covariance matrix associated with the $d$ signal sources: $\boldsymbol{S} = E[\boldsymbol{s}(t)\,\boldsymbol{s}^H(t)]$. The elements of $\boldsymbol{S}$ are $\boldsymbol{S}_{ki} = E[s_k(t)\,s_i^*(t)] = \rho_{ki}\sigma_k\sigma_i$, $k, i = 1, \cdots, d$, where $\rho_{ki}$ is the complex correlation

0096-3518/88/0600-0945$01.00 © 1988 IEEE