ON THE CONCENTRATED STOCHASTIC LIKELIHOOD FUNCTION IN ARRAY SIGNAL PROCESSING

Petre Stoica and Arve Nehorai

Abstract. The stochastic likelihood function [(STO)LF] associated with the narrowband signal processing problem can be concentrated with respect to the signal covariance matrix elements and the noise power. Although this is a known fact, no clear-cut derivation of the concentrated (STO)LF appears to be available in the literature. In this short paper we provide a simple, complete proof of the concentrated (STO)LF formula.

1. Introduction and preliminaries

Consider a situation in which $n$ narrowband signals $\{x_k(t)\}_{k=1}^n$ impinge on an array comprising $m$, ($m > n$) sensors. Let $a(\hat{\theta})$ denote the response of the array to a unit-amplitude signal generated by a source with location parameter(s) $\hat{\theta}$ (depending on the application under consideration, $\hat{\theta}$ may include the source’s azimuth, elevation, and range, or it may be simply the direction-of-arrival corresponding to the source in question). The array output $y(t)$, observed at time instant $t$, can then be described by the following equation [1]–[4], [7]–[9]:

$$y(t) = Ax(t) + e(t),$$

where $x(t) = [x_1(t) \cdots x_n(t)]^T$ is the signal vector ($n \times 1$) as measured at some reference point, $e(t)$ is an additive noise vector ($m \times 1$), and

$$A = [a(\theta_1) \cdots a(\theta_n)] (m \times n).$$

It is assumed that the array is calibrated so that the functional form of $a(\cdot)$ is known, and also that the array is unambiguous so that rank ($A$) = $n$. The location
parameter vector

\[ \theta = \begin{bmatrix} \theta_1 & \vdots & \theta_n \end{bmatrix} \]  

(3)

is to be estimated from \( N \) snapshots \( \{y(t), \ldots, y(N)\} \) of the array output.

Under the assumption that the snapshots are independent and identically distributed complex Gaussian random variables with zero mean, the log-likelihood function of the samples \( \{y(t)\}_{t=1}^N \) is given by (see, e.g., [1]–[4], [8]):

\[ LF = -N \left[ m \ln(\pi) + \ln|R| + \text{tr}(R^{-1}\hat{R}) \right] , \]  

where \( R \) is the theoretical covariance matrix of \( y(t) \),

\[ R = Ey(t)y^*(t) \]  

(5)

and \( \hat{R} \) is the sample covariance matrix

\[ \hat{R} = \frac{1}{N} \sum_{t=1}^{N} y(t)y^*(t) . \]  

(6)

Hereafter, the superscript * denotes the conjugate transpose, and the notation \(|R|\) and \(\text{tr}(R)\) denote, respectively, the determinant and the trace of \( R \).

Under the additional assumption that the signals are stationary stochastic sequences with covariance matrix \( P = Ex(t)x^*(t) \), the noise is spatially white and has the same power in all sensors \( Ee(t)e^*(t) = \sigma I \), and the signals and the noise are uncorrelated, the covariance matrix \( R \) has the expression:

\[ R = APA^* + \sigma I . \]  

(7)

The likelihood function (4), (7) is called the (STO)LF because it corresponds to the assumption that the signals \( \{x_k(t)\} \) are stochastic \([1, 2, 4, 8]\). This is to be contrasted to the assumption that \( \{x_k(t)\} \) in (1) are deterministic signals, which leads to the so-called deterministic LF \([7, 9]\).

Maximization of (4), (7) with respect to \( P, \sigma, \) and \( \theta \) yields the (stochastic) maximum likelihood estimates (MLE) of these parameters, which are known to be asymptotically statistically efficient \([1, 8]\). This maximization problem is significantly complicated by the presence of \( P \) and \( \sigma \) in (4), (7). As most commonly \( \{\theta_k\} \) are the parameters of major interest, one should try to concentrate out the nuisance parameters \( P \) and \( \sigma \) from (4). When doing so, one may constrain the matrix \( P \) to be nonnegative definite \([2]\). However, this complicates the concentration of (4) with respect to \( P \) and \( \sigma \) significantly, and the gain obtained by imposing the nonnegative definiteness condition on \( P \) is unclear. By assuming that \( P \) is a general Hermitian matrix, Böhme \([1]\) was apparently the first to report the following concentrated (STO)LF result.
Theorem 1.1. For given $\theta$, the Hermitian matrix $P$ and the positive scalar $\sigma$ that maximize (4) are given by

$$\hat{P}(\theta) = A^+ \hat{R} A^* + \hat{\sigma}(\theta)(A^* A)^{-1}$$

(8)

$$\hat{\sigma}(\theta) = \text{tr}(\Pi^l \tilde{R})/(m - n)$$

(9)

where

$$A^+ = (A^* A)^{-1} A^*$$

(10)

$$\Pi^l = I - AA^+.$$  

(11)

Furthermore, the concentrated negative (STO)LF (obtained by inserting (8) and (9) in (4) with changed sign) is (to within an additive constant)

$$F(\theta) = \ln A^+ \hat{P}(\theta) A^* + \hat{\sigma}(\theta) I$$

(12)

The minimizing argument $\hat{\theta}$ of (12) is the (STO)MLE of the location parameter vector $\theta$. The ML estimates of $P$ and $\sigma$, if desired, can be obtained by inserting $\hat{\theta}$ in (8) and (9).

The cited reference [1], however, does not contain any proof of the above theorem. Jaffee [4], assuming that $\sigma$ is known, considered the problem of concentrating (4) only with respect to $P$ (assumed to be a general Hermitian matrix). The result derived in [4] may be used to obtain a proof of the previous theorem, but that would require several steps; and considerable care, as several terms depending on $\sigma$ have been dropped in the concentrated (w.r.t. $P$) (STO)LF of [4] (recall that $\sigma$ was assumed to be known there); however, all those terms should be taken into consideration when concentrating the likelihood function with respect to $\sigma$.

Theorem 1.1 is an important result for the recent research in array signal processing. Our goal here is to provide a simple, complete proof of this theorem. Our approach is purely algebraic throughout. This is to be contrasted to the previous (partial) analyses of the (STO)LF problem, which used differential calculus (see, e.g., [4]).

2. Derivation of the concentrated (STO)LF result

Maximization of (4) with respect to $P$, $\sigma$, and $\theta$ is equivalent to the minimization of the function

$$L(\theta, P, \sigma) = \ln |R| + \text{tr}(R^{-1} \tilde{R}).$$

(13)

By using the fact that for two arbitrary matrices $C$ and $D$, the following determinantal equality holds $|CD + I| = |DC + I|$ (see [6]), we can write

$$|R| = \sigma^n |APA^*/\sigma + I| = \sigma^n |PA^* A/\sigma + I|$$

$$= \sigma^{n-1} |A^* A| |P + \sigma (A^* A)^{-1}|.$$  

(14)
Next, make use of the matrix inversion lemma [6], to obtain

\[ R^{-1} = \frac{1}{\sigma} \left( I + APA^*/\sigma \right)^{-1} = \frac{1}{\sigma} \left[ I - A(\sigma I + PA^*)^{-1}PA^* \right] . \]  

(15)

The second term in (15), when inserted in (13), yields an expression that can be processed to a more convenient form as follows:

\[
\begin{align*}
\text{tr} \left[ A(PA^*A + \sigma I)^{-1}PA^* \hat{R} \right] &= \text{tr} \left[ (PA^*A + \sigma I)^{-1}(PA^* \hat{R}A) - (A^*A)^{-1}A^* \hat{R}A + (A^*A)^{-1}A^* \hat{R}A \right] \\
&= \text{tr} \left[ (A^*A)^{-1}A^* \hat{R}A \right] \\
&\quad + \text{tr} \left[ (PA^*A + \sigma I)^{-1} \left[ PA^* \hat{R}A - (PA^*A + \sigma I)(A^*A)^{-1}A^* \hat{R}A \right] \right] \\
&= \text{tr}(A^* \hat{R}A) - \sigma \text{tr} \left[ \left[ P + \sigma (A^*A)^{-1} \right]^{-1}A^T \hat{R}A^T \hat{R}A^{1*} \right] .
\end{align*}
\]  

(16)

From (15) and (16), we get

\[ \text{tr}(R^{-1} \hat{R}) = \frac{1}{\sigma} \text{tr}(\hat{R}) - \frac{1}{\sigma} \text{tr}(A^* \hat{R}A) + \text{tr}(S^{-1}A^* \hat{R}A^{1*}) , \]  

(17)

where

\[ S = P + \sigma (A^*A)^{-1} . \]  

(18)

The insertion of (14) and (17) into (13) yields:

\[ L(\theta, P, \sigma) = (m - n) \ln \sigma + \ln |A^*A| + \ln |S| \\
+ \text{tr}(S^{-1}A^* \hat{R}A^{1*}) + \frac{1}{\sigma} \text{tr}(\Gamma A^T \hat{R}) . \]  

(19)

The only terms in (19) that depend on \( P \) are the third and fourth ones. To minimize (19) with respect to \( P \) (for given \( \theta \) and \( \sigma \)) we make use of the following result. For an arbitrarily given positive definite \((n \times n)\)-matrix \( \Gamma \), the inequality below holds for any positive definite \((n \times n)\) \( S \):

\[ \ln |S| + \text{tr}(S^{-1} \Gamma) \geq n \ln |\Gamma| . \]  

(20)

Furthermore, the equality in (20) is achieved for \( S = \Gamma \) (which is obvious). To prove (20), we use the following inequality shown in [3]:

\[ \text{tr}(S^{-1} \Gamma) \geq n|S^{-1} \Gamma|^\frac{1}{2} \]  

(21)

to write

\[
\begin{align*}
\ln |S| + \text{tr}(S^{-1} \Gamma) &\geq \ln |S| + n|S^{-1} \Gamma|^\frac{1}{2} \\
&= n \left[ \ln |S|^{\frac{1}{2}} - \ln |\Gamma|^{\frac{1}{2}} + \frac{1}{n} \ln |\Gamma| + |S^{-1} \Gamma|^\frac{1}{2} \right] \\
&= \ln |\Gamma| + n(\alpha - \ln \alpha) ,
\end{align*}
\]  

(22)

where \( \alpha \) is the positive root of

\[ \alpha = n(\alpha - \ln \alpha) \]  

from (22).

The minimization of (19) yields

\[ \hat{P}(\theta, \sigma) = \frac{1}{\sigma} \left( \frac{S^{-1}A^* \hat{R}A^{1*}}{\text{trace}(S^{-1}A^* \hat{R}A^{1*})} \right) , \]

which is the result we set out to prove for \( \hat{P}(\theta, \sigma) \), the estimate of \( P \) from the observation \( A^T \hat{R}A \).

Finally, we have obtained the result fairly generally stated in the Appendix B.

where \( \alpha = |S^{-1} \Gamma|^\frac{1}{2} > 0 \). As \( \ln \alpha \leq \alpha - 1 \) for any \( \alpha > 0 \), (20) follows at once from (22). By using (20), the minimization of (19) with respect to \( P \) is immediate. The minimizing \( \hat{P} \) is
\[
\hat{P}(\theta, \sigma) = A^\dagger \hat{R} A^* = \sigma (A^* A)^{-1}
\]
(23)

(which proves (8)), and the function (19) with \( P \) concentrated out is given by
\[
L(\theta, \hat{P}(\theta, \sigma), \sigma) = (m - n) \ln \sigma + \frac{1}{\sigma} \operatorname{tr}(\Pi^\dagger \hat{R}) + \ln |A^* A| + \ln |A^\dagger \hat{R} A^*| + n.
\]
(24)

Minimization of (24) with respect to \( \sigma \) yields:
\[
\check{\sigma}(\theta) = \operatorname{tr}(\Pi^\dagger \hat{R}) / (m - n)
\]
(25)

(which is (9)). Finally, by inserting (25) in (24) we obtain (to within an additive constant):
\[
F(\theta) = \ln \left\{ \check{\sigma}(\theta)^{m-n} |(A^* A)A^\dagger \hat{R} A^*| \right\}
\]
\[
= \ln \left\{ \check{\sigma}(\theta)^{m-n} |(A^* A) \left[ \hat{P}(\theta) + \check{\sigma}(\theta) (A^* A)^{-1} \right]| \right\}
\]
\[
= \ln \left\{ \check{\sigma}(\theta)^{m} |(A^* A) \hat{P}(\theta) / \check{\sigma}(\theta) + I| \right\}
\]
\[
= \ln |A \hat{P}(\theta) A^* + \check{\sigma}(\theta) I|,
\]
(26)

which is the sought expression (12). Note that \( \hat{P}(\theta) \) above is a short notation for \( \hat{P}(\theta, \check{\sigma}(\theta)) \) and that the second equality in (26) follows from (23). With this observation, the proof of Theorem 1 is complete.

Finally, we note that the result of Theorem 1.1 can be readily extended to a fairly general class of unknown (parameterized) noise covariance matrices, see [5, Appendix D] for details.

References


