[10] S. Weiss, R. W. Stewart, and W. Liu, "A broadband adaptive beam-former in subbands with scaled aperture," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Monterey, CA, Nov. 2002, pp. 1298–1302.

[11] M. M. Goodwin and G. W. Elko, "Constant beamwidth beamforming," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, Apr. 1993, vol. 1, pp. 169–172.

[12] J. H. Doles, III and F. D. Benedict, "Broad-band array design using the asymptotic theory of unequally spaced arrays," *IEEE Trans. Antennas Propag.*, vol. 36, pp. 27–33, Jan. 1988.

[13] S. C. Chan and K. S. Pun, "On the design of digital broadband beam-former for uniform circular array with frequency invariant characteristics," in *Proc. IEEE Int. Symp. Circuits Systems*, Phoenix, AZ, May 2002, vol. 1, pp. 693–696.

[14] S. C. Chan and H. H. Chen, "Theory and design of uniform concentric circular arrays with frequency invariant characteristics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 4, pp. 805–808.

[15] S. C. Chan and H. H. Chen, "Theory and design of uniform concentric spherical arrays with frequency invariant characteristics," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Toulouse, France, May 2007, vol. IV, pp. 1057–1060.

[16] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Amer.*, vol. 97, no. 2, pp. 1023–1034, Feb. 1995.

[17] S. Haykin and J. Kesler, "Relation between the radiation pattern of an array and the two-dimensional discrete Fourier transform," *IEEE Trans. Antennas Propag.*, vol. AP-23, no. 3, pp. 419–420, May 1975.

[18] T. Sekiguchi and Y. Karasawa, "Wideband beamspace adaptive array utilizing FIR fan filters for multibeam forming," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 277–284, Jan. 2000.

On the Deterministic CRB for DOA Estimation in Unknown Noise Fields Using Sparse Sensor Arrays

Martin Kleinsteuber and Abd-Krim Seghouane, *Member, IEEE*

Abstract—The Cramér–Rao bound (CRB) plays an important role in direction of arrival (DOA) estimation because it is always used as a benchmark for comparison of the different proposed estimation algorithms. In this correspondence, using well-known techniques of global analysis and differential geometry, four necessary conditions for the maximum of the log-likelihood function are derived, two of which seem to be new. The CRB is derived for the general class of sensor arrays composed of multiple arbitrary widely separated subarrays in a concise way via a coordinate free form of the Fisher Information. The result derived in [1] is confirmed.

Index Terms—Cramér–Rao bound (CRB), differential geometry, direction of arrival (DOA) estimation, maximum likelihood.

I. INTRODUCTION

The maximum likelihood technique is a widely used tool for directions of arrival (DOA) estimation. Many log-likelihood functions and estimation algorithms have been proposed in the literature depending on the structure of the noise covariance matrix which make them sensitive to the assumed noise model. In most practical situations, the noise model is unknown and to effectively handle unknown noise environments several methods have been proposed. The most recent one consists of spacing the array geometry in certain ways. In this correspondence, the general case of sensor arrays composed of multiple arbitrary widely separated subarrays [1] is considered. In such arrays, intersubarray spacings are substantially larger than the signal wavelength and the noise covariance matrix of the whole array is block-diagonal.

The classical way for deriving the maximum likelihood estimate of the DOA is by setting the derivative of the log-likelihood function with respect to the DOA parameters to zero and solving the formed equation set. Note, that two different types of data models are used in applications for DOA estimation. The so-called conditional model, where the signal is supposed to be nonrandom, and the unconditional model, where the signal is assumed to be random [2]. Since the results derived in this correspondence are extensions of previous results derived in [1], we exclusively focus on the first case—the conditional model and the corresponding likelihood function.

To assess the performance of these derived maximum likelihood estimators, the Cramér–Rao bound (CRB) plays an important role because it is always used as a benchmark for comparison. The derivation of closed-form expressions for the CRB for the general unknown noise model have been approached in [3]–[5] and obtained for the uniform and nonuniform white noise case in [6] and [7]. An extension of the

work provided in [6] was used in [8] to derive a closed-form expression for the CRB in the most general case of an arbitrary unknown noise field.

In this correspondence, we consider the general class of sensor arrays composed of multiple arbitrary widely separated subarrays [1]. Using well-known techniques of global analysis and differential geometry, the derivative and the Hessian form of the log-likelihood function are computed. The latter one is used to derive a coordinate free form of the Fisher information. In contrast to earlier approaches, this allows us to directly compute the CRB of linear transformations for the DOA. Choosing a standard basis yields the results obtained in [1].

The rest of this correspondence is organized as follows. Some basics in differential geometry are provided in Section II. Necessary conditions for the existence of the maximum likelihood are derived in Section III. In Section IV, the general closed-form expression for the CRB is derived and the relation with the particular case of [1] is discussed. A conclusion is given in Section V.

II. PRELIMINARIES ON DIFFERENTIAL GEOMETRY

We recall some basic facts and definitions on global analysis (cf. [9] and [10]). Let M be a smooth manifold of dimension n. A *curve* through $x \in M$ is a smooth map

$$\gamma : I \longrightarrow M$$

where $I \subset \mathbb{R}$ is an open interval containing 0 and $\gamma(0) = x$. Let U be a neighborhood of x and let $\phi : U \longrightarrow \mathbb{R}^n$ be a chart. Then

$$\phi \circ \gamma : I \longrightarrow \phi(U) \subset \mathbb{R}^n$$

is differentiable. Two curves γ_1 and γ_2 through $x \in M$ are said to be *equivalent* if $(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0)$ holds for some and therefore any chart ϕ. This defines an equivalence relation on the set of all curves through x. A *tangent vector* at x is then an equivalence class $\xi := [\gamma]$ of a curve γ and the *tangent space* $T_x M$ is the set of all tangent vectors. It can be shown to be an n-dimensional real vector space.

A trivial example of a manifold is an open subset U of $\mathbb{R}^n$ together with the identity mapping as the chart. In this case, the tangent space at any point of U can be identified with $\mathbb{R}^n$.

Now let M, N be manifolds and let $f : M \longrightarrow N$ be smooth. If γ is a curve through $x \in M$, then $f \circ \gamma$ is a curve through $f(x) \in N$ and equivalent curves through x are mapped to equivalent curves through $f(x)$. We can therefore define the *derivative* of f at $x \in M$ as the linear map

$$Df(x) : T_x M \longrightarrow T_{f(x)} N$$

given by $Df(x)[\gamma] = [f \circ \gamma]$ for all tangent vectors $[\gamma] \in T_x M$. If $f : M \longrightarrow \mathbb{R}$ is a smooth real valued function, we identify $T_y(\mathbb{R}) = \mathbb{R}$ for all $y \in \mathbb{R}$ and define a *critical point* of f as a point $x \in M$ such that $Df(x)\xi = 0$ for all $\xi \in T_x M$. The *Hessian* of f at a critical point x then is the symmetric bilinear form

$$\mathcal{H}_f(x) : T_x M \times T_x M \longrightarrow \mathbb{R}$$
$$(\xi_1, \xi_2) \longmapsto \frac{1}{2}(\mathcal{H}_f(x)(\xi_1 + \xi_2, \xi_1 + \xi_2)$$
$$- \mathcal{H}_f(x)(\xi_1, \xi_1) - \mathcal{H}_f(x)(\xi_2, \xi_2)) \tag{1}$$

where $\mathcal{H}_f(x)([\gamma], [\gamma]) := (f \circ \gamma)''(0)$. It can be shown that this definition is independent of the choice of the representative γ only if $\gamma(0)$

is a critical point of f. The Hessian is therefore only well defined at critical points of f. A critical point is *nondegenerate* if its Hessian is nondegenerate. If x is a local maximum (minimum), then $\mathcal{H}_f(x)$ is negative (positive) semidefinite. On the other hand, if $\mathcal{H}_f(x)$ is negative (positive) definite, then x is a local maximum (minimum).

III. LOG-LIKELIHOOD FUNCTION

Let an array of n sensors having unknown gains and phases receive signals from m ($m < n$) narrowband far-field sources with unknown DOAs $\{\theta_1, \ldots, \theta_m\}$. The $n \times 1$ array snapshot vectors can be modeled as [1]

$$y(t) = \Gamma(\lambda)A(\theta)x(t) + v(t) \quad t = 1, \ldots, N \tag{2}$$

where $\theta = [\theta_1, \ldots, \theta_m]^\top$ is the $m \times 1$ vector of signal DOAs, $A(\theta) = [a(\theta_1), \ldots, a(\theta_m)]$ is the $n \times m$ source direction matrix, $a(\theta)$ is the $n \times 1$ steering vector, $x(t) = [x_1(t), \ldots, x_m(t)]^\top$ is the $m \times 1$ vector of the source waveforms, $v(t) = [v_1(t), \ldots, v_m(t)]^\top$ is the $n \times 1$ vector of sensor noise, $\Gamma(\gamma)$ is a diagonal matrix containing the unknown complex-valued sensor responses, i.e., $\Gamma(\lambda) = \text{diag}\{\lambda_1, \ldots, \lambda_n\}, (\cdot)^\top$ denotes transpose and N is the number of statistically independent snapshots. In this case, the array model can be rewritten as [1], [11]

$$Y = \Gamma(\lambda)A(\theta)X + V = \Gamma A X + V \tag{3}$$

where $Y = [y(1), \ldots, y(N)], X = [x(1), \ldots, x(N)], V = [v(1), \ldots, v(N)]$ are the $n \times N$ array data matrix, the $m \times N$ source waveform matrix, and the $n \times N$ sensor noise matrix, respectively.

In this correspondence, we consider the case of sparse arrays composed of q arbitrary subarrays whose intersubarray displacements are substantially larger than the signal wavelength. As a result, sensor noises can be assumed to be statistically independent between different subarrays. This leads to a noise covariance matrix, say Q, that has a block form. The size of each block, say n_i, corresponds to the numbers of sensors in the corresponding subarray ($n = \sum_{i=1}^q n_i$). In other words, $Q \in \mathcal{Q}$ with

$$\mathcal{Q} := \left\{ \begin{bmatrix} Q_1 & & \\ & \ddots & \\ & & Q_q \end{bmatrix} | Q_i \in \mathbb{C}^{n_i \times n_i}, Q_i > 0 \right\}$$
$$=: \text{bdiag}(Q_1, \ldots, Q_q) = E\{v(t)v(t)^\dagger\} \tag{4}$$

where we write shortly $Q_i > 0$ for the positive definite noise covariance matrix of the ith subarray Q_i, bdiag$\{\}$ denotes the block-diagonal matrix operator, $(\cdot)^\dagger$ denotes conjugate transpose, and $E\{\}$ is the statistical expectation. Note, that $\mathcal{Q}$ is open in the set of Hermitian block-diagonal matrices of appropriate blocksize and therefore a manifold whose tangentspace at each point can be identified with

$$T_Q \mathcal{Q} = \{\text{bdiag}(H_1, \ldots, H_q), H_i \in \mathbb{C}^{n_i \times n_i}, H_i^\dagger = H_i\}. \tag{5}$$

Taking into account the special structure of the $(n \times m)$ source direction matrix, it varies over the set

$$\mathcal{A} := \left\{ \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_m \\ \vdots & & \vdots \\ z_1^{n-1} & \cdots & z_m^{n-1} \end{bmatrix} | z_i \in \mathbb{C}, |z_i| = 1 \right\} \tag{6}$$

which is diffeomorphic to the m-torus via the obvious mapping

$$
\Phi : (z_1, \ldots, z_m) \mapsto \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_m \\ \vdots & & \vdots \\ z_1^{n-1} & \cdots & z_m^{n-1} \end{bmatrix}
$$

and hence a smooth and compact manifold. Note that the set of structure matrix variation of the source direction matrix (6) also covers the n sensors nonuniform linear array. Indeed, by definition, an n-element nonuniform linear array can be viewed as a subarray of the n_α-element uniform linear array. For the corresponding n_α-element uniform linear array, the source direction matrix is of Vandermonde structure [12].

In the following, the tangent space is derived according to Section II. To this end, let $A = \Phi(a_1, \ldots, a_m) \in \mathcal{A}$ and let γ be a curve through A given by

$$
\gamma : I \to \mathcal{A}, \quad t \mapsto \Phi(a_1 \exp(\mathbf{i} t \theta_1), \ldots, a_m \exp(\mathbf{i} t \theta_m)) \quad (7)
$$

where $\theta := (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ and $\mathbf{i} := \sqrt{-1}$. Differentiating with respect to t and setting $t = 0$ yields the tangent space

$$
T_A \mathcal{A} = \{ \mathbf{i} A \odot \mathbf{n} \theta^\top, \theta \in \mathbb{R}^m \} \quad (8)
$$

where $\odot$ denotes the matrix Hadamard product and the vector

$$
\mathbf{n} := (0, 1, \ldots, n-1)^\top \in \mathbb{R}^n.
$$

Similarly, the normalized diagonal matrix Γ that contains the unknown sensor responses varies over

$$
\mathcal{T} := \{ \mathrm{diag}(z_1, \ldots, z_n) \mid z_i \in \mathbb{C}, |z_i| = 1 \} \quad (9)
$$

which is diffeomorphic to the n-torus with tangent space

$$
T_\Gamma \mathcal{T} = \{ \mathbf{i} \Gamma D \mid D \in \mathbb{R}^{n \times n} \text{ is diagonal} \} \quad (10)
$$

at $\Gamma \in \mathcal{T}$. Let the array data matrix $Y \in \mathbb{C}^{n \times N}$ be given. The conditional log-likelihood function (LL-function) is given by [11]

$$
f : \mathcal{Q} \times \mathcal{T} \times \mathcal{A} \times \mathbb{C}^{m \times N} \to \mathbb{R}
$$
$$
(Q, \Gamma, A, X) \mapsto -N \log \det Q - \mathrm{tr}[(Y - \Gamma A X)^\dagger Q^{-1} (Y - \Gamma A X)]. \quad (11)
$$

For convenience, we further shortly write

$$
G := Y - \Gamma A X.
$$

The derivatives of f with respect to Q will be denoted by $D_Q f$ and similar the notation $D_\Gamma f$, $D_A f$ and $D_X f$ is used. In what follows, $\Re(z)$ represents the real part of z.

Lemma 1: The partial derivatives of the LL-function are given by

$$
D_Q f : T_Q \mathcal{Q} \to \mathbb{R}, \quad H \mapsto \mathrm{tr}[Q^{-1} G G^\dagger Q^{-1} H] - N \mathrm{tr}[Q^{-1} H] \quad (12)
$$

$$
D_\Gamma f : T_\Gamma \mathcal{T} \to \mathbb{R}, \quad \xi \mapsto 2\Re \mathrm{tr}[G^\dagger Q^{-1} \xi A X] \quad (13)
$$
$$
D_A f : T_A \mathcal{A} \to \mathbb{R}, \quad \psi \mapsto 2\Re \mathrm{tr}[G^\dagger Q^{-1} \Gamma \psi X] \quad (14)
$$
$$
D_X f : \mathbb{C}^{m \times N} \to \mathbb{R}, \quad S \mapsto 2\Re \mathrm{tr}[G^\dagger Q^{-1} \Gamma A S]. \quad (15)
$$

Proof: Equations (13)–(15) follow straightforwardly by the product rule since the second term of the LL-function is the squared norm of G with respect to the real inner product $\Re \mathrm{tr}[G_1^\dagger Q^{-1} G_2]$ with $G_1, G_2 \in \mathbb{C}^{n \times N}$. Since they all are derived in a very similar way, we restrict ourself to deduce (14). Let γ be given as in (7) with $\psi := \dot{\gamma}(0) \in T_A \mathcal{A}$. Then

$$
\begin{aligned}
D_A f(\xi) &= \frac{\mathrm{d}}{\mathrm{d}t} \big|_{t=0} f(Q, \Gamma, \gamma(t), X) \\
&= -\Re \mathrm{tr}[(-\Gamma \dot{\gamma}(0) X)^\dagger Q^{-1} (Y - \Gamma \gamma(0) X)] \\
&\quad - \Re \mathrm{tr}[(Y - \Gamma \gamma(0) X)^\dagger Q^{-1} (-\Gamma \dot{\gamma}(0) X)] \\
&= 2\Re \mathrm{tr}[G^\dagger Q^{-1} \Gamma \psi X]. \quad (16)
\end{aligned}
$$

For (12), note that $\log \det Q = \mathrm{tr} \log Q$, implying

$$
D_Q(\log \det Q)(H) = \mathrm{tr}[D_Q(\log Q)(H)] = \mathrm{tr}[Q^{-1} H]
$$

and differentiating $QQ^{-1} = I$ on both sides yields

$$
D_Q Q(H) \cdot Q^{-1} + Q \cdot D_Q(Q^{-1})(H) = 0
$$

and hence $D_Q(Q^{-1})(H) = -Q^{-1} H Q^{-1}$. ∎

From the above Lemma, we immediately have the following theorem, where part 1) and 4) have already been derived in a different way in [8]. These results have been used in [1] to derive an algorithm that iteratively estimates the DOA.

Theorem 1: Let p denote the orthogonal projection from the set of Hermitian $n \times n$-matrices onto $T_Q \mathcal{Q}$ with respect to the inner product $\mathrm{tr}[Q_1 Q_2]$. Necessary conditions for a critical point $(Q_0, \Gamma_0, A_0, X_0)$ of the LL-function are:
1) $p(G_0 G_0^\dagger) = N Q_0$;
2) the diagonal entries of $\Gamma_0 A_0 X_0 G_0^\dagger Q_0^{-1}$ are real;
3) the vector $(A_0^\top \odot X_0 G_0^\dagger Q_0^{-1} \Gamma_0) \cdot \mathbf{n}$ has real entries;
4) $A_0^\dagger \Gamma_0^\dagger Q_0^{-1} \Gamma_0 A_0 X_0 = A_0^\dagger \Gamma_0^\dagger Q_0^{-1} Y$, which simplifies if and only if A has full rank into

$$
X_0 = \left(A_0^\dagger \Gamma_0^\dagger Q_0^{-1} \Gamma_0 A_0 \right)^{-1} A_0^\dagger \Gamma_0^\dagger Q_0^{-1} Y.
$$

Note, that since A is a Vandermonde matrix, it has full rank if and only if the entries z_i are pairwise distinct.

Proof: We will drop the index "0" during the proof. At a critical point (Q, Γ, A, X), all partial derivatives have to vanish.
1) For $D_Q f \equiv 0$, this means that

$$
\mathrm{tr}[(Q^{-1} G G^\dagger Q^{-1} - N Q^{-1}) H] = 0
$$

for all $H \in T_Q \mathcal{Q}$, implying

$$
p(Q^{-1} G G^\dagger Q^{-1} - N Q^{-1}) = 0.
$$

Now taking into account the block structure of Q^{-1}, this is equivalent to $p(GG^\dagger) - NQ = 0$ and 1) is shown.
2) Setting $D_\Gamma f \equiv 0$ and using the special structure of the tangent space elements (10), one has

$$
\Re \mathrm{tr}[\mathbf{i} A X G^\dagger Q^{-1} D] = 0
$$

for all real diagonal $(n \times n)$-matrices D.

3) Note, that for three matrices of appropriate size the identity

$$\mathrm{tr}(A \odot B)C^\top = \mathrm{tr}(A \odot C)B^\top$$

holds (cf. [9]). Hence, $D_A f \equiv 0$ yields

$$\Re\mathrm{tr}[\mathbf{i}XG^\dagger Q^{-1}\Gamma(A \odot \mathbf{n}\theta^\top)] = 0$$

for all $\theta \in \mathbb{R}^m$, which is equivalent to

$$\Re\mathrm{tr}[\mathbf{i}(A^\top \odot XG^\dagger Q^{-1}\Gamma)\mathbf{n}\theta^\top] = 0$$

for all $\theta \in \mathbb{R}^m$.

4) For $D_X f \equiv 0$, we have equivalently

$$A^\dagger\Gamma^\dagger Q^{-1}G = 0 \Leftrightarrow A^\dagger\Gamma^\dagger Q^{-1}Y = A^\dagger\Gamma^\dagger Q^{-1}\Gamma AX.$$

Now let A have full rank and let $x \in \mathbb{C}^m \setminus \{0\}$. Then, $y := \Gamma Ax \neq 0$ and by the positive definiteness of Q^{-1}, we obtain $x^\dagger A^\dagger\Gamma^\dagger Q^{-1}\Gamma Ax > 0$. Therefore, $A^\dagger\Gamma^\dagger Q^{-1}\Gamma A$ is positive definite and hence invertible. If, on the other hand, A does not have full rank, there exists $x \in \mathbb{C}^m \setminus \{0\}$ such that $\Gamma Ax = 0$ and in this case, $A^\dagger\Gamma^\dagger Q^{-1}\Gamma A$ has eigenvalue 0 and is not invertible

. ∎

IV. CRAMÉR–RAO BOUND

To derive the CRB, the Hessian at the critical point $p_0 = (Q_0, \Gamma_0, A_0, X_0)$ has to be computed. We shortly denote

$$D_{QQ}f(H_1, H_2) = D_Q(D_Q f(p_0)(H_1))(H_2)$$

and similar $D_{QA}f(H, \psi) = D_Q(D_A f(p_0)(\psi))(H)$, and so on. Note that the Hessian is symmetric, i.e., $D_{QA}f(H, \psi) = D_{AQ}f(\psi, H)$, etc. Again, the index 0 for indicating the critical point is dropped in the following. From (12)–(15), we derive

$$\begin{aligned}D_{QQ}f(H_1, H_2) = &N\mathrm{tr}[Q^{-1}H_2Q^{-1}H_1] \\ &- \mathrm{tr}[Q^{-1}H_2Q^{-1}GG^\dagger Q^{-1}H_1] \\ &- \mathrm{tr}[Q^{-1}GG^\dagger Q^{-1}H_2Q^{-1}H_1]. \end{aligned} \quad (17)$$

With $\xi_i = \mathbf{i}\Gamma D_i$ as in (10), $i = 1, 2$, one obtains

$$\begin{aligned}D_{\Gamma\Gamma}f(\xi_1, \xi_2) &= D_{\Gamma\Gamma}f(D_1, D_2) \\ &= -2\Re\mathrm{tr}[(\Gamma D_2 AX)^\dagger Q^{-1}\Gamma D_1 AX] \\ &- 2\Re\mathrm{tr}[G^\dagger Q^{-1}\Gamma D_2 D_1 AX]. \end{aligned} \quad (18)$$

For $\psi_i = \mathbf{i}A \odot \mathbf{n}x_i^\top$ as in (8), $i = 1, 2$,

$$\begin{aligned}&D_{AA}f(\psi_1, \psi_2) \\ &= D_{AA}f(\theta_1, \theta_2) \\ &= -2\Re\mathrm{tr}\left[\left(\Gamma\left(A \odot \mathbf{n}\theta_2^\top\right)X\right)^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_1^\top\right)X\right] \\ &- 2\Re\mathrm{tr}\left[G^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_2^\top \odot \mathbf{n}\theta_1^\top\right)X\right] \end{aligned} \quad (19)$$

holds and

$$D_{XX}f(S_1, S_2) = -2\Re\mathrm{tr}[(\Gamma AS_2)^\dagger Q^{-1}\Gamma AS_1]. \quad (20)$$

Moreover

$$D_{Q\Gamma}f(H, D) = -2\Re\mathrm{tr}[G^\dagger Q^{-1}HQ^{-1}\mathbf{i}\Gamma DAX], \quad (21)$$

$$D_{QA}f(H, \theta) = -2\Re\mathrm{tr}[G^\dagger Q^{-1}HQ^{-1}\mathbf{i}\Gamma(A \odot \mathbf{n}\theta^\top)X] \quad (22)$$

$$D_{QX}f(H, S) = -2\Re\mathrm{tr}[G^\dagger Q^{-1}HQ^{-1}\Gamma AS] \quad (23)$$

$$\begin{aligned}D_{\Gamma A}f(D, \theta) = &-2\Re\mathrm{tr}[(\Gamma DAX)^\dagger Q^{-1}\Gamma(A \odot \mathbf{n}\theta^\top)X] \\ &- 2\Re\mathrm{tr}[G^\dagger Q^{-1}\Gamma D(A \odot \mathbf{n}\theta^\top)X] \end{aligned} \quad (24)$$

$$\begin{aligned}D_{\Gamma X}f(D, S) = &2\Re\mathrm{tr}[\mathbf{i}(\Gamma DAX)^\dagger Q^{-1}\Gamma AS] \\ &+ 2\Re\mathrm{tr}[\mathbf{i}G^\dagger Q^{-1}\Gamma DAS] \end{aligned} \quad (25)$$

and, finally

$$\begin{aligned}D_{AX}f(\theta, S) = &2\Re\mathrm{tr}[\mathbf{i}(\Gamma(A \odot \mathbf{n}\theta^\top)X)^\dagger Q^{-1}\Gamma AS] \\ &+ 2\Re\mathrm{tr}[\mathbf{i}G^\dagger Q^{-1}\Gamma(A \odot \mathbf{n}\theta^\top)S]. \end{aligned} \quad (26)$$

In order to derive the Fisher information matrix, we have a look at the expectation value $\mathbf{E}[\cdot]$ of the above terms (17)–(27). Using the fact that at the maximum $\mathbf{E}[G] = 0$ and $\mathbf{E}[G^\dagger G] = NQ$ immediately yields

$$\begin{aligned}&\mathbf{E}[D_{QQ}f(H_1, H_2)] \\ &\quad = -N\mathrm{tr}[Q^{-1}H_2Q^{-1}H_1] \\ &\mathbf{E}[D_{\Gamma\Gamma}f(D_1, D_2)] \\ &\quad = -2\Re\mathrm{tr}[(\Gamma D_2 AX)^\dagger Q^{-1}\Gamma D_1 AX] \\ &\mathbf{E}[D_{AA}f(\theta_1, \theta_2)] \\ &\quad = -2\Re\mathrm{tr}\left[\left(\Gamma\left(A \odot \mathbf{n}\theta_2^\top\right)X\right)^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_1^\top\right)X\right] \\ &\mathbf{E}[D_{XX}f(S_1, S_2)] \\ &\quad = -2\Re\mathrm{tr}[(\Gamma AS_2)^\dagger Q^{-1}\Gamma AS_1] \\ &\mathbf{E}[D_{Q\Gamma}f(H, D)] = 0 \\ &\mathbf{E}[D_{QA}f(H, \theta)] = 0 \\ &\mathbf{E}[D_{QX}f(H, S)] = 0 \\ &\mathbf{E}[D_{\Gamma A}f(D, \theta)] \\ &\quad = -2\Re\mathrm{tr}[(\Gamma DAX)^\dagger Q^{-1}\Gamma(A \odot \mathbf{n}\theta^\top)X] \\ &\mathbf{E}[D_{\Gamma X}f(D, S)] \\ &\quad = 2\Re\mathrm{tr}[\mathbf{i}(\Gamma DAX)^\dagger Q^{-1}\Gamma AS] \\ &\mathbf{E}[D_{AX}f(\theta, S)] \\ &\quad = 2\Re\mathrm{tr}[\mathbf{i}(\Gamma(A \odot \mathbf{n}\theta^\top)X)^\dagger Q^{-1}\Gamma AS]. \end{aligned} \quad (27)$$

Gathering the derived results yields the following theorem.

Theorem 2: The bilinear form corresponding to the Fisher information is given by

$$\begin{aligned}&\mathbf{F}((H_1, D_1, \theta_1, S_1), (H_2, D_2, \theta_2, S_2)) \\ &= N\mathrm{tr}[Q^{-1}H_2Q^{-1}H_1] \\ &\quad + 2\Re\mathrm{tr}[(\Gamma D_2 AX)^\dagger Q^{-1}\Gamma D_1 AX] \\ &\quad + 2\Re\mathrm{tr}\left[\left(\Gamma\left(A \odot \mathbf{n}\theta_2^\top\right)X\right)^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_1^\top\right)X\right] \\ &\quad + 2\Re\mathrm{tr}[(\Gamma AS_2)^\dagger Q^{-1}\Gamma AS_1] \\ &\quad + 2\Re\mathrm{tr}\left[(\Gamma D_1 AX)^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_2^\top\right)X\right] \\ &\quad + 2\Re\mathrm{tr}\left[(\Gamma D_2 AX)^\dagger Q^{-1}\Gamma\left(A \odot \mathbf{n}\theta_1^\top\right)X\right] \\ &\quad - 2\Re\mathrm{tr}[\mathbf{i}(\Gamma D_1 AX)^\dagger Q^{-1}\Gamma AS_2] \\ &\quad - 2\Re\mathrm{tr}[\mathbf{i}(\Gamma D_2 AX)^\dagger Q^{-1}\Gamma AS_1] \\ &\quad - 2\Re\mathrm{tr}\left[\mathbf{i}\left(\Gamma\left(A \odot \mathbf{n}\theta_1^\top\right)X\right)^\dagger Q^{-1}\Gamma AS_2\right] \end{aligned}$$

$$- 2\Re\mathrm{tr}\left[\mathbf{i}\left(\Gamma\left(A \odot \mathbf{n}\theta_2^\top\right)X\right)^\dagger Q^{-1}\Gamma AS_1\right]. \qquad (28)$$

Clearly, a matrix representation of $\mathbf{F}$ depends on the choice of a basis $\mathcal{B}$ of the tangent space at the maximum, which is given by $T_Q\mathcal{Q} \times T_\Gamma\mathcal{T} \times T_A\mathcal{A} \times \mathbb{C}^{m \times N}$. Let $\mathcal{B}_H, \mathcal{B}_D, \mathcal{B}_\theta$, and $\mathcal{B}_S$ be basis of $T_Q\mathcal{Q}, T_\Gamma\mathcal{T}, T_A\mathcal{A}$, and $\mathbb{C}^{m \times N}$, respectively. If $\mathcal{B}$ is chosen to be

$$\mathcal{B} := (\mathcal{B}_\theta, \mathcal{B}_D, \mathcal{B}_S, \mathcal{B}_H) \qquad (29)$$

then the matrix representation of $\mathbf{F}$ takes the form

$$F_\mathcal{B} = \begin{bmatrix} F_{\theta\theta} & F_{\theta D} & F_{\theta S} & 0 \\ F_{\theta D}^\top & F_{DD} & F_{DS} & 0 \\ F_{\theta S}^\top & F_{DS}^\top & F_{SS} & 0 \\ 0 & 0 & 0 & F_{HH} \end{bmatrix}. \qquad (30)$$

The matrices F_{ij} depend on the choice of the basis $\mathcal{B}_i$ and $\mathcal{B}_j$ for $i, j \in \{\theta, D, S, H\}$. Using the partitioned matrix inversion formula, the $(m \times m)$ CRB-matrix for θ is given by

$$C_\theta = \left(F_{\theta\theta} - [F_{\theta D} \quad F_{\theta S}]\begin{bmatrix} F_{DD} & F_{DS} \\ F_{DS}^\top & F_{SS} \end{bmatrix}\begin{bmatrix} F_{\theta D}^\top \\ F_{\theta S}^\top \end{bmatrix}\right)^{-1}. \qquad (31)$$

Now denote by e_i, f_i, g_i the standard basis vectors of $\mathbb{R}^m, \mathbb{R}^n, \mathbb{R}^N$, respectively, i.e., having ith entry 1 and zeros elsewhere. By choosing the basis

$$\mathcal{B}_\theta = \left(\mathbf{i}A_0 \odot \mathbf{n}e_i^\top, i = 1, \dots, m\right)$$
$$\mathcal{B}_D = \left(\mathbf{i}\Gamma_0 f_i f_i^\top, i = 1, \dots, n\right)$$
$$\mathcal{B}_S = \left(e_i g_j^\top, \mathbf{i}e_i g_j^\top, i = 1, \dots, m, j = 1, \dots, N\right) \qquad (32)$$

equation (31) is equivalent to (89) in [1]. More generally, if $T \in \mathbb{R}^{m \times m}$ is a change of coordinates in $T_A\mathcal{A}$, i.e.,

$$\tilde{\mathcal{B}}_\theta := \mathbf{i}A_0 \odot \mathbf{n}\left((Te_1)^\top, \dots, (Te_m)^\top\right) \qquad (33)$$

then the CRB-matrix with respect to this new basis is given by

$$\tilde{C}_\theta = T^{-1}C_\theta T^{-\top}. \qquad (34)$$

Example: We illustrate the above by means of a simple example. Let the source direction matrix $A_0 = \begin{bmatrix} 1 & 1 \\ a_1 & a_2 \end{bmatrix}$ be given, where $a_1 = e^{\mathbf{i}\theta_1}$ and $a_2 = e^{\mathbf{i}\theta_2}$ and θ_1, θ_2 are the directions of arrival. Differentiating A_0 with respect to θ_i yields $(d/d\theta_i)A_0 = \mathbf{i}A_0 \odot \begin{bmatrix} 0 \\ 1 \end{bmatrix} e_i^\top \theta_i$, which corresponds to the canonical basis $\mathcal{B}_\theta$ in (32) that finally gives the CRB C_θ for (θ_1, θ_2) as in [1]. Assume now that we are interested in $\tilde{\theta}_1, \tilde{\theta}_2$ such that

$$T\begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \qquad (35)$$

for some invertible 2×2 matrix T. Then

$$\frac{d}{d\tilde{\theta}_i}A_0 = \mathbf{i}A_0 \odot \begin{bmatrix} 0 \\ 1 \end{bmatrix}(Te_i)^\top \tilde{\theta}_i. \qquad (36)$$

Hence, choosing a basis

$$\mathcal{B}_{\tilde{\theta}} = \left(\mathbf{i}A_0 \odot \begin{bmatrix} 0 \\ 1 \end{bmatrix}(Te_i)^\top, \quad i = 1, 2\right) \qquad (37)$$

in (29) leads to the CRB $C_{\tilde{\theta}}$ for $(\tilde{\theta}_1, \tilde{\theta}_2)$ for which $C_{\tilde{\theta}} = T^{-1}C_\theta T^{-\top}$ holds, without explicitly computing T^{-1}.

V. CONCLUSION

Using well-known techniques of global analysis and differential geometry, the determination of the derivatives of the maximum likelihood function is easy and concise.

The Fisher information has been derived in terms of a coordinate free bilinear form. Different choices of basis in the tangent space at the maximum of the log likelihood function lead to different Fisher information matrices and hence to different CRBs. The connections between these CRBs have been explained in (34).

One of the benefits of the proposed approach is, that in order to derive the CRB for $\tilde{\theta}$, satisfying $T\tilde{\theta} = \theta$, the matrix T^{-1} does not have to be computed explicitly.

REFERENCES

[1] S. A. Vorobyov, A. B. Gershman, and K. M. Wong, "Maximum likelihood direction of arrival estimation in unknown noise fields using sparse sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 34–43, Jan. 2005.

[2] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[3] L. L. Scharf and L. T. McWhorter, "Geometry of the Cramér–Rao bound," *Signal Process.*, vol. 31, pp. 301–3011, 1993.

[4] H. Ye and R. DeGroat, "Maximum likelihood DOA estimation and asymptotic Cramér–Rao bounds for additive unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 938–949, Apr. 1995.

[5] A. J. Weiss and B. Friedlander, "On the Cramér-Rao bound for direction finding of correlated sources," *IEEE Trans. Signal Process.*, vol. 41, pp. 495–499, 1993.

[6] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.

[7] M. Pesavento and A. B. Gershman, "Maximum likelihood direction of arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, pp. 1310–1324, 2001.

[8] A. B. Gershman, P. Stoica, M. Pesavento, and E. G. Larsson, "Stochastic Cramér-Rao bound for direction estimation in unknown noise fields," *Proc. IEEE—Radar Sonar Navigat.*, vol. 149, pp. 2–8, 2002.

[9] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. London, U.K.: Springer, 1994.

[10] M. Spivak, *Calculus on Manifolds.*. New York: W. A. Benjamin, 1965.

[11] D. Astléy, A. L. Swindlehurst, and B. Ottersten, "Spatial signature estimation for uniform linear arrays with unknown receiver gain and phases," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2128–2138, Aug. 1999.

[12] Y. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays—Part II: Partially augmentable arrays," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1492–1507, Jun. 2003.