A DOA ESTIMATOR BASED ON LINEAR PREDICTION AND TOTAL LEAST SQUARES

Amauri Lopes - Ivanil S. Bonatti - Pedro L. D. Peres - Ricardo F. Colares - Carlos A. Alves

FEEC-UNICAMP - 13083-970 - P.O.Box 6101 - Campinas - SP - Brazil
amauri@decom.fee.unicamp.br

ABSTRACT

We propose an estimator for the direction of arrival (DOA) of plane waves incident on a linear equally spaced array of sensors. The estimator uses the forward - backward linear prediction filter optimized by means of the total least squares criterion subject to constraints. The DOA angles are estimated by the zeros of the optimum filter. The proposed procedure performs better than the maximum likelihood methods present in the literature.

1. INTRODUCTION

The problem of estimating the direction of arrival (DOA) of plane waves incident on a linear equally spaced array of sensors is considered in this paper. The DOA problem and the array signal processing are present in mobile communication, radar, sonar, seismology, radio astronomy and industry applications [1].

The best methods for DOA estimation present in the literature are those derived from the maximum likelihood estimator (MLE), like the MODE [2] and the MODEX [3].

We propose a DOA estimator, named FBCTLS, derived from the forward-backward linear prediction (FBLP) filter [4], optimized using the constrained total least squares criterion (CTLS) [5]. Some of the zeros of the optimized FBLP filter lead to the estimation of the DOA angles. The zeros that best estimate the angles are selected by means of a maximum likelihood cost criterion.

The proposed estimator is compared with the MODE and MODEX and we show that the FBCTLS presents advantages.

2. SIGNAL AND NOISE MODEL

The problem of estimating the direction of M narrowband plane waves impinging on a uniform linear array of N sensors ($N > M$), can be reduced to the problem of estimating the frequencies $\boldsymbol{\omega} \in \mathcal{R}^{M\times 1}$ in the following model [2]:

$$\mathbf{y}_k = \mathbf{A}\mathbf{x}_k + \mathbf{n}_k \ ; \ k = 1, ..., K \tag{1}$$

where

$\mathbf{y}_k \in \mathcal{C}^{N\times 1}$ is the kth noisy data vector (snapshot);

K is the number of data vectors collected from the array through the time;

$\mathbf{A} = [\mathbf{a}_1 \ ... \ \mathbf{a}_M]$;

$\mathbf{a}_i = [1 \ e^{j\omega_i} ... \ e^{j(N-1)\omega_i}]^T; \ i = 1, \ ... \ , M$;

$\mathbf{x}_k \in \mathcal{C}^{M\times 1}$ is the signal vector ;

$\mathbf{x}_k = [c_k(1)e^{j\phi_k(1)} \ ... \ c_k(M)e^{j\phi_k(M)}]^T$;

$c_k(i)$ and $\phi_k(i)$ are the signal amplitudes and phases, respectively;

$\mathbf{n}_k \in \mathcal{C}^{N\times 1}$ is the noise vector;

$(.)^T$ denotes the transpose.

The signal and the noise are independent zero mean complex stationary Gaussian random processes with the following second-order moments:

$$\begin{array}{ll} E\{\mathbf{x}_k\mathbf{x}_l^H\} = \mathbf{C}\delta_{k,l} & ; \ E\{\mathbf{x}_k\mathbf{x}_l^T\} = \mathbf{0} \\ E\{\mathbf{n}_k\mathbf{n}_l^H\} = \sigma^2\mathbf{I}\delta_{k,l} & ; \ E\{\mathbf{n}_k\mathbf{n}_l^T\} = \mathbf{0} \end{array} \tag{2}$$

where $E\{\}$ is the expectation, $\mathbf{C}$ is the unknown signal covariance matrix, $\delta_{k,l}$ is the Kronecker delta operator, σ^2 is the unknown noise power, $\mathbf{I}$ is the identity matrix and $(.)^H$ denotes the conjugate transpose.

All the concepts and equations to be presented in the sections 3 and 4 and in the Appendix A refer to the kth snapshot. Then we omitted the subscript "k", like that in the symbol $\mathbf{y}_k$, in order to simplify the notation in those sections and Appendix.

3. FORWARD-BACKWARD LINEAR PREDICTION

The linear prediction leads to an estimator for the frequencies $\boldsymbol{\omega}$ [6]. To present the estimator, consider a forward linear prediction error filter of order L, $M \leq L < N$, and coefficients $\mathbf{b}^f = [-1 \ b_1^f \ ... \ b_L^f]^T$, processing $\mathbf{y}$ (for the kth snapshot) and producing the prediction error [7]:

$$e^f(n) = -y(n) + \sum_{i=1}^{L} b_i^f y(n-i);\; n = L+1,\; ...\;, N \quad (3)$$

Defining $\mathbf{e}^f = \left[e^f(L+1)\ldots e^f(N)\right]^T$ and

$$\mathbf{Y}^f = \begin{bmatrix} y(L+1) & \cdots & y(1) \\ \vdots & \ddots & \vdots \\ y(N) & \cdots & y(N-L) \end{bmatrix} \quad (4)$$

the equation (3) can be written as $\mathbf{Y}^f\,\mathbf{b}^f = \mathbf{e}^f$.

Consider now a backward linear prediction error filter of order L and coefficients $\mathbf{b}^b = [-1\; b_1^b\; ..\; b_L^b]^T$, processing $\mathbf{y}$ and producing the prediction error :

$$e^b(n) = -y(n) + \sum_{i=1}^{L} b_i^b y(\,n+i);\;\; n = 1,\; ...\;, N-L \quad (5)$$

Defining $\mathbf{e}^b = \left[\; e^b(1) \;\;\cdots\;\; e^b(N-L)\;\right]^T$ and $\mathbf{Y}^b = \mathbf{Y}^f\mathbf{J}$, where $\mathbf{J}$ is a permutation matrix (its anti-diagonal is composed by "ones", whereas the others elements are "zeros"), the equation (5) can be written as $\mathbf{Y}^b\,\mathbf{b}^b = \mathbf{e}^b$.

For stationary signals and in the limit $N \to \infty$, it can be shown that $\left(\mathbf{Y}^f\right)^H\mathbf{Y}^f = \left(\left(\mathbf{Y}^b\right)^H\mathbf{Y}^b\right)^*$, implying that $\mathbf{b}_{ot}^f = \left(\mathbf{b}_{ot}^b\right)^*$, where $(.)^*$ denotes the complex conjugate [7].

This result motivated the definition of the forward-backward linear prediction, where $\mathbf{b}^f = \left(\mathbf{b}^b\right)^*$is used even for finite N. Defining $\mathbf{b} = \mathbf{b}^f = \left(\mathbf{b}^b\right)^*$, $\mathbf{Y} = \left[\left(\mathbf{Y}^f\right)^T\;\left(\mathbf{Y}^b\right)^H\right]^T$ and $\mathbf{e} = \left[\left(\mathbf{e}^f\right)^T\;\left(\mathbf{e}^b\right)^H\right]^T$, the formulation of the FBLP problem can now be expressed as $\mathbf{Yb} = \mathbf{e}$. Then the optimum vector $\mathbf{b}_{ot}$ is obtained by minimizing $\|\mathbf{e}\|_2$, where $\|.\|_2$ denotes the 2-norm.

Consider the polynomial $P(z) = -z^L + b_1 z^{L-1} + ... + b_L$, obtained after the FBLP filter has been optimized. In the absence of noise, $P(z)$ has M zeros on the unit circle at the positions $\{\exp(jw_i)\;;\; i = 1,\; ...\;, M\}$ (the signal zeros), whereas the remaining $(L-M)$ zeros (the noise zeros) are situated inside the unit circle [6], [7]. Therefore, the frequencies $\boldsymbol{\omega}$ can be estimated searching for the signal zeros on the unit circle [6].

The presence of the noise causes the zeros to fluctuate around their noise free positions. But for signal-to noise ratios (SNR) not excessively low, the frequencies $\boldsymbol{\omega}$ can be estimated searching for the M zeros that are closest to the unit circle [6].

4. THE FBLP AND THE CONSTRAINED TOTAL LEAST SQUARES CRITERION

The least squares criterion is usually employed to optimize the forward-backward filter. This criterion assumes that just one of the columns of $\mathbf{Y}$ has errors or noise, whereas, in fact, all the columns are affected by the noise. The total least squares (TLS) criterion [8], [9] should be employed in this situation and its application to the minimization of the energy of the FBLP error $\mathbf{e}$ leads to following problem:

$$\min_{(\Delta\mathbf{Y},\mathbf{b})} \;\|\Delta\mathbf{Y}\|_F \quad \text{subject to} \;\; (\mathbf{Y}+\Delta\mathbf{Y})\,\mathbf{b} = \mathbf{0} \quad (6)$$

where $\|.\|_F$ denotes the Frobenius norm, and $\Delta\mathbf{Y}$ is a matrix composed of independent, identically distributed, zero mean white random variables representing perturbations to the matrix $\mathbf{Y}$ [8], [9].

Another aspect suggesting the use of the TLS criterion is the Toeplitz and Hankel structures of the matrix $\mathbf{Y}$. They are not taken into account in the least squares optimization process. Nevertheless, the constrained total least squares (CTLS) criterion is able to take into account those structures [5], [9].

In the appendix A it is demonstrated that the application of the CTLS criterion to the minimization of energy of the FBLP error leads to following problem:

$$\min_{(\mathbf{b},\Delta\mathbf{y}^{fb})} \left\{\left(\Delta\mathbf{y}^{fb}\right)^H \mathbf{P}\;\Delta\mathbf{y}^{fb}\right\} \quad \text{subject to} \quad \mathbf{Yb} + \mathbf{B}^H\Delta\mathbf{y}^{fb} = \mathbf{0} \quad (7)$$

As $\mathbf{B}$ has rank $2(N-L)$, the solution for (7) is [5]

$$\mathbf{b}_{ot} = \arg\min_{\mathbf{b}} \left\{\mathbf{b}^H\mathbf{Y}^H\left(\mathbf{B}^H\mathbf{P}^{-1}\mathbf{B}\right)^{-1}\mathbf{Yb}\right\} \quad (8)$$

5. THE FBCTLS METHOD

The problem (8) refers to the kth snapshot. When all the K snapshots are considered, it is desirable to have the same $\mathbf{b}_{ot}$ solution for all of them, since the signal and noise are stationary and the DOA angles are the same for all the snapshots. Then the equations (20) and (26) in the appendix A, show that both the matrices $\mathbf{P}$ and $\mathbf{B}$ will also be the same for all the snapshots.

One possible strategy is to minimize the sum of the terms inside the brackets in the equation (8) for $k = 1, ..., K$, leading to

$$\mathbf{b}_{ot} = \arg\min_{\mathbf{b}} \left\{\mathbf{b}^H \sum_{k=1}^{K}\left[\mathbf{Y}_k^H\left(\mathbf{B}^H\mathbf{P}^{-1}\mathbf{B}\right)^{-1}\mathbf{Y}_k\right]\mathbf{b}\right\} \quad (9)$$

It is shown in the appendix B that the problem (9) can be written as

$$\mathbf{b}_{ots} = \arg\min_{\mathbf{b}} \left\{ \mathbf{b}^H \mathbf{D}^H \mathbf{D} \mathbf{b} \right\} \text{ subject to } b_0 = -1 \tag{10}$$

where $\mathbf{D}$ is a matrix with dimensions $4M(N-L) \times (L+1)$. This is a fourth order minimization problem with respect to $\mathbf{b}$.

It will now be presented an iterative algorithm to solve the problem (10). In the first step it is imposed $\mathbf{B} = \mathbf{I}$ and the corresponding matrix $\mathbf{D}$ is calculated. Therefore, the minimization problem is reduced to a second order one which can be solved using the QR decomposition [2]. This procedure leads to a first value to $\mathbf{b}$, which is used to update the $\mathbf{B}$ and $\mathbf{D}$ matrices. In the second step, a new value for $\mathbf{b}$ is calculated using the updated matrix $\mathbf{D}$, and so on for the next steps. In general, three such steps are sufficient for an adequate convergence [2], [3].

It can be shown [2] that the solution to the second order minimization problem at each step is given by $\mathbf{b}_{ots} = \left[-1\ \boldsymbol{\eta}^T\right]^T$ where $\boldsymbol{\eta} = \mathbf{R}_0^{-1}\mathbf{Q}_1^H\mathbf{D}_1$; $\mathbf{D} = [\mathbf{D}_1\ \mathbf{D}_2]$; $\mathbf{D}_1$ = first column of $\mathbf{D}$; $\mathbf{D}_2 \in \mathcal{C}^{4M(N-L)\times L}$; $\mathbf{D}_2 = \mathbf{QR}$ (QR decomposition with $\mathbf{Q} \in \mathcal{C}^{4M(N-L)\times 4M(N-L)}$ and $\mathbf{R} \in \mathcal{C}^{4M(N-L)\times L}$) and

$$\mathbf{Q} = \left[\ \mathbf{Q}_1 \quad \mathbf{Q}_2\ \right];\ \mathbf{Q}_1 \in \mathcal{C}^{4M(N-L)\times L} \tag{11}$$

$$\mathbf{R} = \left[\begin{array}{c} \mathbf{R}_0 \\ \mathbf{0} \end{array} \right];\ \mathbf{R}_0 \in \mathcal{C}^{L\times L} \tag{12}$$

Once the vector $\mathbf{b}_{ot}$ has been calculated, the polynomial $P(z)$ is formed and the corresponding zeros are calculated. Then the M signal zeros are estimated searching for those M zeros that are closest to the unit circle.

The above procedure was applied to an example with the following parameters: number of plane waves $M = 2$, signal correlation matrix $\mathbf{C} = \mathbf{I}$ (uncorrelated signals), number of sensors $N = 10$, frequencies to be estimated $\omega_1 = 0.5455$ and $\omega_2 = 0.8131$, order of the FBLP filter $L = 6$, number of snapshots $K = 100$. One hundred different experiments were produced with the above specifications, but with different realizations for the random signal and noise. Referring to the expression (2), the signal-to-noise ratio (SNR) is defined as SNR= $trace\,(\mathbf{C})\,/M\sigma^2$.

The Figure 1 presents the superposition of the zeros of the polynomial $P(z)$ for the 100 experiments and SNR= 15 dB. There are two signal zeros close to the ideal positions defined by the desired frequencies on the unit circle, whereas the noise zeros are distributed inside the unit circle. Therefore, the desired frequencies can be estimated searching for the two zeros that are closest to the unit circle.

However, the variance of the zeros increases as the SNR decreases and for SNR $<$ 5dB there are noise zeros closer to unit circle than the signal zeros. Therefore, the estimation strategy based on searching for the two zeros that are closest to the unit circle leads to a poor performance.

We adopted another criterion for the selection of the signal zeros to overcome this problem. We select the M signal zeros as those that minimize the maximum likelihood cost function for the DOA problem, given by [3]

$$J(\omega) = \left[\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H - \mathbf{I}\right]\left(\sum_{k=1}^{K}\mathbf{y}_k\mathbf{y}_k^H(k)\right) \tag{13}$$

where $\mathbf{A}$ is formed by the arguments of the zeros to be tested. Then, for each group of M zeros among the L zeros obtained at each experiment, we form the corresponding $\mathbf{A}$ matrix and calculate $J(\omega)$.

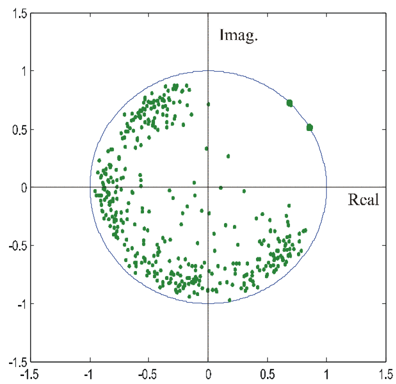

Fig. 1 - Zeros in the z-plane: SNR=15 dB.

The Figure 2 shows the results achieved with the proposed procedure applied to the above example with $L = 7$. The root mean square error is calculated for both frequencies after 100 different experiments and for various values of the SNR. The Figure 2 also presents the performance of the methods MODE [2] and MODEX [3].

The FBCTLS presents better performance than the classical MODE for SNR values lesser than 5 dB and is competitive with the MODEX.

In order to compare the computational complexity of three methods, observe that the MODE and the MODEX solve problems similar to (10). But the MODEX and the FBCTLS use $L > M$, whereas the MODE uses $L = M$. So the MODE is the most economical in this part of the algorithm. Also, MODE and MODEX demand similar computational effort to obtain their optimum **b** vector. The MODE is also the most economical when searching for the signal zeros because it uses $L = M$, that is, just the signal zeros. In the other hand, the MODEX and the FBCTLS use the maximum likelihood cost function in (13) to select the signal zeros. Therefore, the MODE is the most economical in terms of computational complexity.

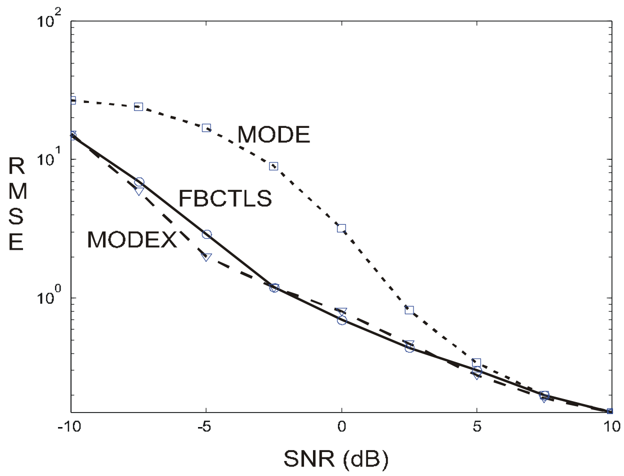

Fig. 2 - Root mean square error as a function of the SNR for the FBCTLS, MODE and MODEX.

Although MODEX and FBCTLS are similar to each other in the above calculations, the MODEX search for the signal zeros in a ensemble composed by the sum of the $L > M$ zeros and the M MODE zeros. So the MODEX is less economical than the FBCTLS as it demands additional efforts to execute the MODE and to search for the signal zeros.

The conclusion is that the FBCTLS and the MODEX present similar performance. They perform better than the MODE at a cost of additional computational efforts. But the FBCTLS is more efficient than the MODEX in terms of computational effort.

6. CONCLUSIONS

We proposed a new DOA method, FBCTLS, based on the forward-backward linear prediction and the constrained total least squares criterion.

The proposed procedure leads to the same minimization problem as those that are present in the most popular maximum likelihood methods. The solution was achieved by means of the same iterative algorithm employed in the MODE and the MODEX methods.

The comparison study showed that the FBCTLS and the MODEX perform similarly and that both are better than the MODE. However, both demand additional computational effort when compared to the MODE. But the FBCTLS demand less computational effort than the MODEX.

7. REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[2] J. Li, P. Stoica, and Z. Liu, "Comparative study of IQML and MODE direction-of-arrival estimators," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 149–160, Jan. 1998.

[3] A.B. Gershman and P. Stoica, "New MODE-based techniques for direction finding with an improved threshold performance," *Signal Processing*, , no. 76, pp. 221–235, 1999.

[4] S.M. Kay, *Modern Spectral Estimation - Theory and Application*, Prentice Hall Signal Processing Series, Englewood Cliffs, NJ, 1988.

[5] T.J Abatzoglou, J.M. Mendel, and G.A. Harada, "The constrained total least squares technique and its application to harmonic superresolution," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1070–1087, May 1991.

[6] D.W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. of the IEEE*, vol. 70, pp. 975–989, Sept. 1982.

[7] S. Haykin, *Adaptive Filter Theory*, Prentice Hall Information and System Sciences Series, Englewood Cliffs, NJ, 2nd edition, 1991.

[8] G.H. Golub and C.F. Van Loan, "An analysis of the total least squares problem," *SIAM Journal on Numerical Analysis*, vol. 17, pp. 883–893, 1980.

[9] R.P. Lemos, *Mínimos-quadrados totais e máxima verossimilhança em estimação de freqüências*, Ph.D. thesis, School of Electrical and Computer Engineering - UNICAMP, Campinas-SP-Brazil, June 1997.

[10] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 5, pp. 1081–1089, Oct. 1986.

[11] R.P. Lemos and A. Lopes, "A unifying framework to total least squares and approximate maximum likelihood," in *Bi-Annual Intern. Telecom. Symposium - ITS'96*, Acapulco-Mexico, Oct. 1996, pp. 153–157.

8. APPENDICES

A. SOLUTION TO THE CTLS PROBLEM

The results presented in [5] permits to show that the application of the CTLS criterion to the minimization of the energy of the FBLP error leads to following problem:

$$\min_{(\Delta\mathbf{Y},\mathbf{b})} \|\Delta\mathbf{Y}\|_F \quad \text{subject to} \tag{14}$$

$$(\mathbf{Y}+\Delta\mathbf{Y})\,\mathbf{b}=\mathbf{0} \tag{15}$$

where

$$\Delta\mathbf{Y}=\begin{bmatrix} \Delta\mathbf{Y}^f \\ \left(\Delta\mathbf{Y}^b\right)^* \end{bmatrix} \tag{16}$$

$$\Delta\mathbf{Y}^f=\left[\mathbf{F}_1^f\Delta\mathbf{y}....\mathbf{F}_{L+1}^f\Delta\mathbf{y}\right] \tag{17}$$

$$\Delta\mathbf{Y}^b=\left[\mathbf{F}_1^b\Delta\mathbf{y}....\mathbf{F}_{L+1}^b\Delta\mathbf{y}\right] \tag{18}$$

$$\Delta\mathbf{y}=\left[\Delta y(1)...\Delta y(N)\right]^T \tag{19}$$

$\Delta\mathbf{y}$ is a vector composed of independent, identically distributed, zero mean white random variables representing perturbations to the vector $\mathbf{y}$. The $\mathbf{F}_i^f \in \mathcal{R}^{(N-L)\times N}$ and the $\mathbf{F}_i^b \in \mathcal{R}^{(N-L)\times N}$ are matrices composed by "ones" and "zeros". Each row has $(N-1)$ "zeros" and just one "one", whose position is chosen to assure that $\mathbf{Y}^f=\left[\mathbf{F}_1^f\mathbf{y}\ ....\ \mathbf{F}_{L+1}^f\mathbf{y}\right]$ and $\mathbf{Y}^b=\left[\mathbf{F}_1^b\mathbf{y}\ ...\ \mathbf{F}_{L+1}^b\mathbf{y}\right]$. That is, the $\mathbf{F}_i$ matrices describe the structures of the $\mathbf{Y}^f$ and $\mathbf{Y}^b$ matrices. Therefore, the equations (17) and (18) show that $\Delta\mathbf{Y}^f$ and $\Delta\mathbf{Y}^b$ are matrices composed by the perturbation variables and present the same structure as $\mathbf{Y}^f$ and $\mathbf{Y}^b$, respectively.

Referring to $\|\Delta\mathbf{Y}\|_F$ in the problem (14) and using the expressions (16), (17), (18) and (19), it is possible to write

$$\|\Delta\mathbf{Y}\|_F^2=\left(\Delta\mathbf{y}^{fb}\right)^H\ \mathbf{P}\ \Delta\mathbf{y}^{fb} \tag{20}$$

where $\mathbf{P}=diag\left[p_1\ ...\ p_N\ \ p_1\ ...\ p_N\right]_{2N\times 2N}$ and the real numbers $p_i,(i=1,\ ...,N)$ are weighting factors that take into account how many times $\Delta y(i)$ and $\Delta y(i)^*$ appear in the $\Delta\mathbf{Y}$ matrix.

Referring to the equation (15), it is possible to verify that

$$\begin{aligned}\Delta\mathbf{Y}^f\mathbf{b} &= \left[-F_1^f+\sum_{i=2}^{L+1}F_i^f b_{i-1}\right]\Delta\mathbf{y} \qquad (21)\\ \left(\Delta\mathbf{Y}^b\right)^*\mathbf{b} &= \left[-F_1^b+\sum_{i=2}^{L+1}F_i^b b_{i-1}\right]\Delta\mathbf{y}^*\end{aligned}$$

Let

$$\begin{aligned}\mathbf{B}_f &= \left[-F_1^f+\sum_{i=2}^{L+1}F_i^f b_{i-1}\right]^H_{(N-L)\times N}\\ \mathbf{B}_b &= \left[-F_1^b+\sum_{i=2}^{L+1}F_i^b b_{i-1}\right]^H_{(N-L)\times N}\end{aligned} \tag{22}$$

Then

$$\mathbf{B}_f=\begin{bmatrix} b_L & b_{L-1} & \cdots & b_1 & -1 & 0 & \dots & 0\\ 0 & b_L & \dots & b_2 & b_1 & -1 & \dots & 0\\ & \ddots & & \ddots & & \ddots & & \\ 0 & 0 & \dots & & \dots & & \dots & -1 \end{bmatrix}^H \tag{23}$$

$$\mathbf{B}_b=\mathbf{J}\mathbf{B}_f\mathbf{J} \tag{24}$$

Using these matrices and the expression (16), it is possible to write

$$\Delta\mathbf{Y}\mathbf{b}=\begin{bmatrix} \mathbf{B}_f^H & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_b^H \end{bmatrix}\begin{bmatrix} \Delta\mathbf{y}\\ \Delta\mathbf{y}^* \end{bmatrix} \tag{25}$$

Let

$$\mathbf{B}^H=\begin{bmatrix} \mathbf{B}_f^H & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_b^H \end{bmatrix}_{2(N-L)\times 2N} \tag{26}$$

and

$$\Delta\mathbf{y}^{fb}=\begin{bmatrix} \Delta\mathbf{y}\\ \Delta\mathbf{y}^* \end{bmatrix}_{2N\times 1} \tag{27}$$

Using (25), (26) and (27), the expression (15), can be written as

$$\mathbf{Y}\mathbf{b}+\mathbf{B}^H\Delta\mathbf{y}^{fb}=\mathbf{0} \tag{28}$$

Finally, using the expressions (20) and (28) in the problem (14) results

$$\min_{(\mathbf{b},\Delta\mathbf{y}^{fb})}\left\{\left(\Delta\mathbf{y}^{fb}\right)^H\ \mathbf{P}\ \Delta\mathbf{y}^{fb}\right\} \quad \text{subject to} \tag{29}$$

$$\mathbf{Y}\mathbf{b}+\mathbf{B}^H\Delta\mathbf{y}^{fb}=\mathbf{0} \tag{30}$$

B. TRANSFORMATION OF THE CTLS PROBLEM

The problem (9) is repeated here for convenience:

$$\mathbf{b}_{ot} = \arg\min_{\mathbf{b}} \left\{ \mathbf{b}^H \sum_{k=1}^{K} \left[\mathbf{Y}_k^H \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}\right)^{-1} \mathbf{Y}_k \right] \mathbf{b} \right\} \tag{31}$$

We will introduce some modifications in the equation (31). First of all it can be verify that

$$\mathbf{Y}_k \mathbf{b} = \mathbf{B}^H \mathbf{y}_k^{fb} \tag{32}$$

Applying expression (32) into the expression (31) yields

$$\begin{aligned} &\mathbf{b}_{ot} = \\ &= \arg\min_{\mathbf{b}} \left\{ \sum_{k=1}^{K} \left[\left(\mathbf{y}_k^{fb}\right)^H \mathbf{B} \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{y}_k^{fb} \right] \right\} \\ &= \arg\min_{\mathbf{b}} \; trace \left\{ \mathbf{B} \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}^H\right)^{-1} \mathbf{B}^H \widehat{\mathbf{R}}_y^{fb} \right\} \end{aligned} \tag{33}$$

where $\widehat{\mathbf{R}}_y^{fb} = \sum_{k=1}^{K} \mathbf{y}_k^{fb} \left(\mathbf{y}_k^{fb}\right)^H$ is a correlation estimator, except for a constant.

It is worth noting that the problem (33) is the same as those obtained in [2], [3], [10], [11] where the maximum likelihood criterion is used for the estimation of the frequencies $\boldsymbol{\omega}$.

Now a subspace restriction is used to reduce the effect of the noise in the $\widehat{\mathbf{R}}_y^{fb}$matrix, following a procedure inspired by that presented in [2].

Consider the singular value decomposition of the matrix $\widehat{\mathbf{R}}_y^{fb}$ [7], that is $\widehat{\mathbf{R}}_y^{fb} = \mathbf{U}\Sigma\mathbf{U}^H$ where $\mathbf{U} = [\mathbf{u}_1 ... \mathbf{u}_{2N}]$; $\mathbf{U}^H\mathbf{U} = \mathbf{I}$; $\Sigma = diag\left(\lambda_1 ... \lambda_{2N}\right)$ and $\mathbf{u}_i$ are the singular vectors and λ_i are the singular values ordered from the largest to the smallest.

In the case $\mathbf{n}_k = \mathbf{0}$ and for $K \to \infty$, it is shown in [7] that $\lambda_i = 0$ for $2M + 1 \le i \le 2N$. Imposing this condition on $\widehat{\mathbf{R}}_y^{fb}$ produces $\widehat{\mathbf{R}}_y^{fbM} = \mathbf{U}_M \Sigma_M \mathbf{U}_M^H$ where [2]

$$\mathbf{U}_M = [\mathbf{u}_1 \; ... \; \mathbf{u}_{2M}] \; ; \; \Sigma_M = diag\left(\alpha_1 \; ... \; \alpha_{2M}\right)$$

$$\alpha_i = \frac{1}{\lambda_i} \left(\lambda_i - \frac{\sum_{n=2M+1}^{2N} \lambda_n}{2N - 2M} \right)^2 \; ; \; i = 1, \; ... \; , 2M \tag{34}$$

Let $\mathbf{U}_M \left(\Sigma_M\right)^{0.5} = \mathbf{V}$. It can be verified that $\mathbf{B}^H\mathbf{V} = [\mathbf{S}_1\mathbf{b} \; ... \; \mathbf{S}_{2M}\mathbf{b}]$ where

$$\mathbf{S}_i = \begin{bmatrix} v_{L+1,i} & \cdots & v_{1,i} \\ \vdots & \ddots & \vdots \\ v_{N,i} & \cdots & v_{N-L,i} \\ v_{N+1,i} & \cdots & v_{N+L+1,i} \\ \vdots & \ddots & \vdots \\ v_{2N-L,i} & \cdots & v_{2N,i} \end{bmatrix} \tag{35}$$

Substituting these results into the equation (33) yields

$$\begin{aligned} &\mathbf{b}_{ot} = \\ &= \arg\min_{\mathbf{b}} \; trace \left\{ \mathbf{B} \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^H \widehat{\mathbf{R}}_y^{fbM} \right\} \\ &= \arg\min_{\mathbf{b}} \; trace \left\{ \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{V} \mathbf{V}^H \mathbf{B} \right\} \\ &= \arg\min_{\mathbf{b}} \; trace \left\{ [\mathbf{S}_1\mathbf{b} \; ... \; \mathbf{S}_{2M}\mathbf{b}]^H \left(\mathbf{B}^H \mathbf{P}^{-1} \mathbf{B}\right)^{-1} \right. \\ &\qquad \left. [\mathbf{S}_1\mathbf{b} \; ... \; \mathbf{S}_{2M}\mathbf{b}] \right\} \end{aligned} \tag{36}$$

Let $\mathbf{B}^H\mathbf{P}^{-1}\mathbf{B} = \mathbf{G}^H\mathbf{G}$ where $\mathbf{G}$ is the Cholesky decomposition of $\mathbf{B}^H\mathbf{P}^{-1}\mathbf{B}$. Then the problem (36) can be written as

$$\begin{aligned} \mathbf{b}_{ot} = \arg\min_{\mathbf{b}} \; trace &\left\{ [\mathbf{S}_1\mathbf{b} \; ... \; \mathbf{S}_{2M}\mathbf{b}]^H \right. \\ &\left. \mathbf{G}^{-1} \left(\mathbf{G}^H\right)^{-1} [\mathbf{S}_1\mathbf{b} \; ... \; \mathbf{S}_{2M}\mathbf{b}] \right\} \end{aligned} \tag{37}$$

Finally, the last problem can expressed as

$$\mathbf{b}_{ot} = \arg\min_{\mathbf{b}} \left\{ \mathbf{b}^H \mathbf{D}^H \mathbf{D} \mathbf{b} \right\} \quad \text{subject to} \quad b_0 = -1 \tag{38}$$

where

$$\mathbf{D} = \begin{bmatrix} \left(\mathbf{G}^H\right)^{-1} \mathbf{S}_1 \\ \vdots \\ \left(\mathbf{G}^H\right)^{-1} \mathbf{S}_{2M} \end{bmatrix}_{4M(N-L)\times(L+1)} \tag{39}$$

and the minimization in the problem (38) is subjected to $b_0 = -1$ in agreement with the definition of the vector $\mathbf{b}$.