Blind Beamforming in a Cyclostationary Context using an Optimally Weighted Quadratic Cost Function

C. Vignat and P. Loubaton - Université de Marne la Vallée
Unité de Formation SPI - 2 rue de la Butte Verte
93166 Noisy le Grand Cedex e-mail: vignat@univ-mlv.fr

ABSTRACT

This paper addresses the problem of blind beamforming in a cyclostationary context. We show the equivalence between the SCORE algorithm derived by Gardner et al., and the minimization of an optimally weighted quadratic cost function. This approach allows us to justify, from a statistical point of view, the relevance of the SCORE algorithm.

Introduction

Introducing the blind beamforming problem in a cyclostationary context

The cyclostationarity property of signals appears to considerably simplify the treatment of many problems in which they are involved. Specifically, Gardner et Al. have highlighted how this property may be exploited in the context of digital communications. In this article, we address the more specific problem of blind beamforming.

In this context, a signal transmitted by some emitter is supposed to be received by an array of q sensors. In the absence of multi-path effects, the complex envelope of the continuous-time sensor array output writes:

$$\widetilde{y}(t) = h\widetilde{v}(t) + \sigma\widetilde{w}(t) \quad (1)$$

were $\widetilde{y}(t) = [\widetilde{y}_1(t)\ldots\widetilde{y}_q(t)]^T$ and $\widetilde{v}(t)$ is the complex envelope of the modulated signal. h is a constant unknown vector, representing the distortion due to propagation; its norm may be, without restriction, set to one. The $\sigma\widetilde{w}(t)$ term represents the effects of both interferences and additive noise; it is assumed to be stationary and Gaussian. If we assume moreover that $\widetilde{v}(t)$ writes

$$\widetilde{v}(t) = \sum_{m\in Z} a_m\varphi(t-mT) \quad (2)$$

where $(a_m)_{m\in Z}$ are the symbols to be transmitted, then the signal $\widetilde{v}(t)$ is cyclostationary with cyclic frequency $\frac{1}{T}$ .Denoting by $y(n)$ the array output sampled at frequency $\frac{2}{T}$, i.e. $y(n) = y(n\frac{2}{T})$, we have the following discrete time model:

$$y(n) = hv(n) + \sigma w(n) \quad (3)$$

where $v(n)$ is a discrete time, cyclostationary signal assumed to be circular, with cyclic frequency $\alpha = \frac{1}{2}$, and where $w(n)$ is a stationary Gaussian noise. The problem of blind beamforming is the following: how to reconstruct the unknown emitted sequence $v(n)$ from the observation of the received signal $y(n)$, vector h being unknown.

In the case where vector h is known, a natural solution to this problem consists in constructing an estimate signal $\widehat{v}(n)$ of $v(n)$ using a spatial filter g such as $\widehat{v}(n) = g^T v(n)$; the spatial filter is then chosen so as to minimize the quadratic cost function

$$C_{Wiener}(g) = \lim_{N\to+\infty}\frac{1}{N}\sum_{n=0}^{N-1} E\left[|\widehat{v}(n) - v(n)|^2\right] \quad (4)$$

The corresponding minimum g_{Wiener} of this function is nothing but the classical spatial Wiener filter given, up to a constant scalar factor, by $\bar{g}_{Wiener} = R_{yy}^{-1}h$, where symbol $^-$ stands for conjugation and R_{yy} denotes the "covariance matrix"

$$R_{yy} = \lim_{N\to+\infty}\left[\frac{1}{N}\sum_{n=0}^{N-1} E\left[y(n)y^*(n)\right]\right] \quad (5)$$

In practice, the observation is available from $n = 0$ to $n = N-1 < \infty$, and R_{yy} is replaced by its empirical consistent estimate $\hat{R}_{yy} = \frac{1}{N}\sum_{n=0}^{N-1} E(y(n)y^*(n))$.

As h is unknown, Gardner [2] proposed the so-called SCORE algorithm remarking that the minimum g_{Wiener} of function $C_{Wiener}(g)$ coincides, up to a constant, with the minimum of the following cost function:

$$C_{score}(g) = \lim_{N\to+\infty}\frac{1}{N}\sum_{n=0}^{N-1} E\left[\left|g^T y(n) - r(n)e^{in\alpha}\right|^2\right] \quad (6)$$

where the signal $r(n)$, constructed as $r(n) = c^T y(n)$, acts as a reference signal. Vector c is supposed to be chosen non orthogonal to vector h , and may be chosen optimal in a certain sense. In the context of a finite sample observation $(y(n))_{0\le n\le N-1}$, $C_{score}(g)$ is replaced by the cost function $\frac{1}{N}\sum_{n=0}^{N-1}\left|g^T y(n) - r(n)e^{in\alpha}\right|^2$, leading to the estimate

$$\hat{g}_{Wiener} = \hat{R}_{yy}^{-1}\hat{R}_{yr}^{\alpha} \tag{7}$$

where $\hat{R}_{yr}^{\alpha}$ is the empirical estimate

$$\hat{R}_{yr}^{\alpha} = \frac{1}{N}\sum_{n=0}^{N-1} y(n)r(n)^{*}e^{-in\alpha} \tag{8}$$

of the cyclic correlation R_{yr}^{α} at frequency α.

However, the choice of a unique scalar reference signal, although motivated by computational cost considerations, is not, from a statistical point of view, fully justified. In other words, the multivariate observation $y(n)$ is reduced, in the C_{score} cost function, to a scalar signal $r(n) = c^T y(n)$, what should induce a loss of information. The purpose of this paper is to show that the reduction of $y(n)$ to a one-dimensional observation can be justified in a statistical sense.

the proposed approach

Instead of directly estimating the spatial Wiener filter, it is possible to first estimate h by some vector $\widehat{h}$ and to deduce from $\widehat{h}$ the estimate of g_{Wiener} given by $\widehat{R}_{yy}^{-1}\widehat{h}$. Under certain conditions, we shall establish that if $\widehat{h}$ is an optimal estimate in a sense to be precised, then the estimate $\widehat{R}_{yy}^{-1}\widehat{h}$ has exactly the structure (7)used by Gardner.

In order to estimate h, we remark that the correlation matrix R_{yy}^{α} verifies, from (3), the following relation:

$$R_{yy}^{\alpha} = hh^{*}R_{vv}^{\alpha} \tag{9}$$

where R_{vv}^{α} denotes the cyclic correlation coefficient of v at time 0 and cyclic frequency α. From this observation, it follows that the unknown vector h coincides, up to a constant factor, with the left singular vector of R_{yy}^{α} corresponding to its unique non zero singular value. This remark gives a number of straightforward methods to derive consistent estimates of vector h. However, in order to show the statistical relevance of the SCORE algorithm, we adapt to the present situation an approach presented by Gurelli and Nikias [5] in the context of multi-channel blind equalization.

1 Definition of a cost function and its optimal weighting

1.1 the quadratic cost function

If $x = [x_1 \ldots x_q]$ is a rxq matrix, let us denote by $F(x)$ the $\frac{q(q-1)}{2}rxq$ matrix whose block lines are the matrices $[0,\ldots,0,-x_j,0,\ldots,0,x_i,0\ldots,0]$ with $1 \le i \le j \le q$ [1]. Matrix $F(h^T)$ is in fact defined so that each of its lines are orthogonal to h^T. Moreover, it is easy to check that $Rank(F(h^T)) = q-1$, so that the rows of $F(h^T)$ form a generating family of the orthogonal of h^T. Finally, for convenience, we denote by $G(x)$ the matrix $F(x^T)^T$.

It follows from (9) that matrix $F(x^T)$verifies the other following relation:

$$F(h^T)R_{yy}^{\alpha} = 0 \tag{10}$$

This remark leads to a simple estimation scheme. First, matrix R_{yy}^{α} is estimated empirically by

$$\widehat{R}_{yy}^{\alpha} = \frac{1}{N}\sum_{n=0}^{N-1} y(n)y^{*}(n)e^{-in\alpha} \tag{11}$$

Then one estimates h (up to a constant factor) by minimizing the cost function $C(f)$ defined by:

$$C(f) = \left\| F(f^T)\widehat{R}_{yy}^{\alpha}\right\|^2 \tag{12}$$

under the non triviality constraint $\|f\| = 1$. In order to explicit the expression of $C(f)$, we introduce the row vectorization operator vec which associates to a matrix $A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}$ the row vector $vec(A) = [A_1 \cdots A_N]$. Then, it is clear that $C(f)$ can also be written as

$$C(f) = \left[vec\left(F(f)\widehat{R}_{yy}^{\alpha}\right)\right]\left[vec\left(F(f)\widehat{R}_{yy}^{\alpha}\right)\right]^{*} \tag{13}$$

Moreover, one checks easily that $vec\left(F(f)\widehat{R}_{yy}^{\alpha}\right) = f^T G(\widehat{R}_{yy}^{\alpha})$, so that $C(f)$ can be written as $C(f) = \left\|f^T G(\widehat{R}_{yy}^{\alpha})\right\|^2$. Of course, the minimum of $C(f)$ under the constraint $\|f\| = 1$ is the eigenvector associated to the smallest eigenvalue of the matrix $G(\widehat{R}_{yy}^{\alpha})G(\widehat{R}_{yy}^{\alpha})^{*}$, and is a consistent estimate of h up to a constant scalar factor.

1.2 a weighted version of the quadratic cost function

We consider in what follows a generalization of this cost function, introducing a weighting matrix W as follows

$$C_W(f) = \left[vec\left(F(f)\widehat{R}_{yy}^{\alpha}\right)\right] W \left[vec\left(F(f)\widehat{R}_{yy}^{\alpha}\right)\right]^{*} \tag{14}$$

A very important point is that W can be chosen optimally in a certain sense. In order to explain this, we first have to remark that the minimum of the above cost function is of course defined up to a constant modulus 1 scalar factor.

In order to fix which particular solution we consider, we denote by $\widehat{h}_W$ the unique solution of the above minimization problem for which $h^{*}\widehat{h}_W$ is real. Then, under mild technical assumptions, it can be shown as in [4] that the following results hold.

Theorem 1 *If $Kernel(G(R_{yy}^{\alpha})G(R_{yy}^{\alpha})^{*})$ coincides with $Kernel(G(R_{yy}^{\alpha})WG(R_{yy}^{\alpha})^{*})$ then $\widehat{h}_W$ is a consistent estimate of h*

Such a weighting matrix W will be called admissible from now on. Moreover, $\sqrt{N}\left(\widehat{h}_W - h\right)$ converges in distribution, as the number N of observations grows infinitely, toward a Gaussian random vector $\mathcal{N}(0, \sum_W)$. The matrix $\sum_W$ is called the asymptotic covariance matrix of the estimate $\widehat{h}_W$. Moreover,

Theorem 2 *Let $\Delta^{\#}$ denote the pseudo inverse matrix of matrix Δ defined as the asymptotic covariance matrix of the random vector $vec(F(h)\widehat{R}^{\alpha}_{yy})$. Matrix $\Delta^{\#}$ is admissible and leads to an asymptotically optimal estimate in the sense that*

$$\Sigma_{\Delta^{\#}} \leq \Sigma_W \quad (15)$$

Finally, the optimal asymptotic covariance matrix $\Sigma_{\Delta^{\#}}$ is given by

$$\Sigma_{\Delta^{\#}} = \left[G(R^{\alpha}_{yy})\Delta^{\#}G(R^{\alpha}_{yy})^*\right]^{\#} \quad (16)$$

This shows that it is possible to choose optimally the weighting matrix W so as to minimize the asymptotic covariance of the estimate. At this point, one should note that the estimate $\widehat{h}_W$ introduced below satisfies the requirement: $h^*\widehat{h}_W$ is real. In practice, as h is of course unknown, it is not possible to extract $\widehat{h}_W$ but $a\widehat{h}_W$ for some unknown modulus one scalar factor a. Therefore, the estimator minimizing $C_W(f)$ for $W = \Delta^{\#}$ is in some sense optimal up to a modulus one factor.

Our remaining work consists thus in computing matrices Δ and $\Delta^{\#}$ in order to obtain, according to the preceding theorems, the optimal weighting matrix as $W = \Delta^{\#}$, as well as the structure of the corresponding estimate of h when the noise plus interference contribution converges toward zero.

2 Derivation of the optimal weighting matrix

We first express matrix $\Delta = cov\left[vec(F(h)\widehat{R}^{\alpha}_{yy})\right]$, where the symbol cov stands, in order to simplify the exposition, for the asymptotic covariance matrix. For this, we use the relation $vec(ABC) = vec(B)(A^T \otimes C)$ for any three matrices A, B and C. Thus

$$\Delta = [G(h) \otimes I_q]\left[cov\left(vec(\widehat{R}^{\alpha}_{yy})\right)\right][G(h) \otimes I_q]^* \quad (17)$$

(recall that $G(h) = F(h)^T$) . We have first to compute $cov\left(vec(\widehat{R}^{\alpha}_{yy})\right)$. In this aim, we recall the following formulas which generalize the Bartlett formulas [3] in the case of stationary process:

$$cov\left[\widehat{R}^{\alpha}_{k,l}\widehat{R}^{\alpha}_{k',l'}\right] = \sum_{u\in Z} R^{(0)}_{k,k'}(u)R^{(0)}_{l,l'}(u)^* e^{-j\alpha u}$$
$$+\sum_{u\in Z} R^{(\alpha)}_{k,k'}(u)R^{(\alpha)}_{l,l'}(u)^* e^{-j\alpha u}$$
$$+\sum_{u\in Z} C_{k,k',l,l'}(u,u,0)e^{-j\alpha u} \quad (18)$$

with $C_{k,k',l,l'}(u,u,0) = cum[y_k(u), y^*_l(u'), y^*_{k'}(0), y_{l'}(0)]$. Using Parseval identity, this can be rewritten in the frequency domain as:

$$cov\left[R^{\alpha}_{k,l}R^{\alpha}_{k',l'}\right] = \frac{1}{2\pi}\int_0^{2\pi} S^{(0)}_{k,k'}(\omega+\alpha)S^{(0)}_{l,l'}(\omega)e^{j(\alpha+\omega)}d\omega$$

$$+\frac{1}{2\pi}\int_0^{2\pi} S^{(\alpha)}_{k,k'}(\omega+\alpha)S^{(\alpha)}_{l,l'}(\omega)^* e^{j(\alpha+\omega)}d\omega$$
$$+\sum_{u\in Z} C_{k,k',l,l'}(u,u,0)e^{-j\alpha u} \quad (19)$$

Let us set $T_1 = \frac{1}{2\pi}\int_0^{2\pi} S^{(0)}(\omega+\alpha)^T S^{(0)}(\omega)d\omega$,
$T_2 = \frac{1}{2\pi}\int_0^{2\pi} S^{(\alpha)}(\omega+\alpha)^T S^{(\alpha)}(\omega)^* d\omega$ and
$T_3 = \left(\overline{h}h^T \otimes hh^*\right)\sum_{u\in Z} C_v(u,u,0)e^{-j\alpha u}$
with $C_v(u,u,0) = cum[v(n), v(n)^*, v(0)^*, v(0)]$.

Then, putting all pieces together and using the Gausianity property of $w(n)$, it follows immediately that

$$cov\left(vec\left(\widehat{R}^{\alpha}_{yy}\right)\right) = T_1 + T_2 + T_3 \quad (20)$$

Therefore,

$$\Delta = [G(h) \otimes I_q][T_1 + T_2 + T_3][G(h) \otimes I_q]^* \quad (21)$$

Let us first evaluate the contribution of term T_2. For this, we remark that

$$[G(h)^* \otimes I_q]\, T_2\, [G(h) \otimes I_q] = \quad (22)$$
$$\frac{1}{2\pi}\int_0^{2\pi} G(h)^* S^{(\alpha)}(\omega+\alpha)^T G(h) \otimes S^{(\alpha)}(\omega)^* d\omega$$

But, $S^{(\alpha)}(\omega+\alpha)^T = \overline{h}h^T S^{(\alpha)}_{vv}(\alpha+\omega)$. As $F(h^T)h = 0$, it follows that $G(h)^*\overline{h} = 0$. Therefore, the term T_2 has no contribution in Δ. Similarly, contribution of term T_3 vanishes. On the other hand, $S^{(0)}(\omega)^T = \overline{h}h^T S^{(0)}_{vv}(\omega) + \sigma^2 S^{(0)}_{ww}(\omega)$. Hence, remarking again that $G(h)^*\overline{h} = 0$, it follows that

$$\begin{aligned}\Delta &= \frac{1}{2\pi}\int_0^{2\pi} G(h)^*\sigma^2 S^{(0)}_{ww}(\omega)G(h) \quad (23)\\ &\qquad \otimes\left(hh^* S^{(0)}_{vv}(\omega) + \sigma^2 S^{(0)}_{ww}(\omega)\right)d\omega\end{aligned}$$

i.e.

$$\Delta = \sigma^2\Delta_0 + \sigma^4\Delta_1 \quad (24)$$

where $\Delta_0 = G(h)KG(h)^* \otimes hh^*$
with $K = \frac{1}{2\pi}\int_0^{2\pi} S^{(0)}_{ww}(\omega)S^{(0)}_{vv}(\omega)d\omega$.

From now on, we consider only the case where the signal to noise ratio is favorable. In this case, one may approximate Δ by $\sigma^2\Delta_0$. On the other hand, it can be shown that Δ_0 is admissible, so that

$$\Sigma_{\Delta^{\#}} = \sigma^2\left[G\left(R^{\alpha}_{yy}\right)\Delta_0^{\#}G\left(R^{\alpha}_{yy}\right)\right]^{\#} + o(\sigma^2) \quad (25)$$

and that, up to terms in $o(\sigma^2)$, this coincides with the asymptotic covariance of the estimate $\widehat{h}_{\Delta_0^\#}$. Therefore, as $\sigma^2 \to 0$, the estimate $\widehat{h}_{\Delta\#}$ may be approximated in a relevant way by $\widehat{h}_{\Delta_0^\#}$.

We now study the structure of $\widehat{h}_{\Delta_0^\#}$. In this aim, we remark that $\widehat{h}_{\Delta_0^\#}$ minimizes the following quadratic form $f^T G\left(\widehat{R}^\alpha_{yy}\right)\Delta_0^\# G\left(\widehat{R}^\alpha_{yy}\right)^* \overline{f}$ under the constraint $\|f\| = 1$. Let us denote by $\widehat{Q}$ the associated matrix $\widehat{Q} = G\left(\widehat{R}^\alpha_{yy}\right)\Delta_0^\# G\left(\widehat{R}^\alpha_{yy}\right)^*$. Clearly,

$$\Delta_0^\# = (G(h)KG(h)^*)^\# \otimes \frac{hh^*}{\|h\|^4} \qquad (26)$$

On the other hand, matrix K is positive definite. Therefore, the column spaces of both $(G(h)KG(h)^*)^\#$ and $G(h)$ coincide, and are $(q-1)$ dimensional. If we denote by $V(h)$ a $\frac{q(q-1)}{2}$x$(q-1)$ isometric matrix for which $Range(V(h)) = Range(G(h))$, matrix $(G(h)KG(h)^*)^\#$ can be written as $(G(h)KG(h)^*)^\# = V(h)LL^*V(h)^*$ for some invertible $(q-1)$x$(q-1)$ matrix L. Therefore, matrix $\Delta_0^\#$ writes $\Delta_0^\# = V(h)LL^*V(h)^* \otimes \frac{hh^*}{\|h\|^4}$

Moreover, it is straightforward that

$$G\left(\widehat{R}^\alpha_{yy}\right)V(h)L \otimes \frac{h}{\|h\|^2} = G\left(\widehat{R}^\alpha_{yy}\frac{h}{\|h\|^2}\right)V(h)L \quad (27)$$

If we put

$$r(n) = \frac{h^*}{\|h\|^2}y(n) \qquad (28)$$

it is clear that $\widehat{R}^\alpha_{yy}\frac{h}{\|h\|^2} = \widehat{R}^\alpha_{yr}$ where it is understood that

$$\widehat{R}^\alpha_{yr} = \frac{1}{N}\sum_{n=0}^{N-1} y(n)r(n)^* e^{-in\alpha} \qquad (29)$$

Therefore, $\widehat{Q} = G\left(\widehat{R}^\alpha_{yy}\right)V(h)LL^*V(h)^*G\left(\widehat{R}^\alpha_{yy}\right)^*$. By the particular structure of the operator G, we have $Rank(G\left(\widehat{R}^\alpha_{yr}\right)) = q-1$. Hence, it is clear that $Rank(\widehat{Q}) \leq q-1$, and that, unless with probability 0, $Rank(\widehat{Q}) = q-1$. Thus, the above minimization problem has, as unique solution, the vector of the Kernel of $\widehat{Q}$, i.e. vector $\widehat{R}^\alpha_{yr}$.

3 Conclusion

As a conclusion, we may remark that the optimal weighting approach consists in the following steps:

1- the reference signal should be constructed using (28). As however vector h is unknown, one has to construct a consistent initial estimate $\widehat{h_0}$ and to form the estimated reference signal $\widehat{r}(n)$, defined by analogy with (28) by $\widehat{r}(n) = \frac{\widehat{h}_0^*}{\|\widehat{h_0}\|^2}y(n)$

2- an estimate $\widehat{R}^\alpha_{yr}$ can then be computed following (29), replacing $r(n)$ by its estimate $\widehat{r}(n)$

3- one estimates $\widehat{h}$ by $\widehat{h} = \frac{\widehat{R}^\alpha_{y\widehat{r}}}{\left\|\widehat{R}^\alpha_{y\widehat{r}}\right\|}$

One can show that replacing h by its consistent estimate does not modify the asymptotic covariance matrix.

4- finally the corresponding estimate $\widehat{g}_W$ of g_W is given by:

$$\widehat{g}_W = \frac{\widehat{R}^{-1}_{yy}\widehat{R}^\alpha_{y\widehat{r}}}{\left\|\widehat{R}^\alpha_{y\widehat{r}}\right\|} \qquad (30)$$

and exactly coincides with the solution given by SCORE algorithm in the case where the "scalar reference" signal coincides with $\widehat{r}(n)$ (see [2] for more details). Therefore, our results demonstrate that, in a certain sense, the use of a well chosen scalar reference signal in the context of the SCORE algorithm is quite motivated from a statistical point of view.

One may finally remark that, if the initial estimate $\widehat{h_0}$ is built as the greatest left singular vector of $\widehat{R}^\alpha_{yy}$, relation (30) rewrites simply as $\widehat{h} = \frac{\widehat{R}^\alpha_{yy}\widehat{h_0}}{\left\|\widehat{R}^\alpha_{yy}\widehat{h_0}\right\|}$, which may be viewed as one step of the power algorithm for the extraction of an eigenvector of matrix $\widehat{R}^\alpha_{yy}$.

References

[1] H. Liu, G. Xu and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems", in Proc. ICASSP 1994, vol 4, pp. 581-584

[2] W. Gardner, "Cyclostationarity in Communications and Signal Processing", IEEE Press, 1994

[3] B. Porat, "Digital Processing of Random Signal", Prentice Hall Inc., 1994

[4] K. Abed-Meraïm, J.F. Cardoso, A. Gorokhov, P. Loubaton, E. Moulines, "On Subspace Methods for Blind Identification of Single-Input Multiple-Output FIR Systems", submitted to IEEE Signal Processing

[5] M. Gurelli and C. Nikias, "A New Eigenvector Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals", in Proc. ICASSP 1993, vol 4, pp. 448-451