



HAL
open science

Cooperation of multi-layer perceptrons for angles of arrival estimation

Nadine Rondel, Gilles Burel

► **To cite this version:**

Nadine Rondel, Gilles Burel. Cooperation of multi-layer perceptrons for angles of arrival estimation. IEE Fourth International Conferences on Artificial Neural Networks, Jun 1995, Cambridge, United Kingdom. pp.323-328, <10.1049/cp:19950576>. <hal-03223338>

HAL Id: hal-03223338

<https://hal.science/hal-03223338v1>

Submitted on 17 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

COOPERATION OF MULTI-LAYER PERCEPTRONS FOR ANGLES OF ARRIVAL ESTIMATION

N D M Rondel ^{†‡} & G Y Burel [†]

[†] Thomson Broadband Systems, France

[‡] SEFT, France

IEE 4th Int. Conf. on Artificial Neural Networks, June 26-28th, 1995, Cambridge, UK

ABSTRACT

A novel method for estimating the Angles Of Arrival (AOA) of multiple plane waves with a narrow-band array of sensors is presented here. This method is based on the cooperation of two Multi-Layer Perceptrons (MLP). The elements of the covariance matrix of the signals are input to the first MLP (a classical three-layer perceptron). Then, taking profit of the gross estimations provided by this first neural net, we refine them by using a new kind of perceptron: the Constrained Structured Weights (CSW) perceptron, which we demonstrate finds the Maximum Likelihood Estimation (MLE) of the AOA. Unlike the classical MLE approach, this method presents the great advantage of being computationally inexpensive. Finally some results and comparisons are presented.

INTRODUCTION

The Angles of Arrival problem has raised and is still raising considerable interest in the array-processing community Schmidt (1), Stoica and Nehorai (2) and (3).

The signals impinging upon the antenna may be either the waves reflected by some objects around a radar or a sonar, or the waves emitted by active sources (if we are dealing with radio-astronomy, for instance).

Consider a linear array of m equi-spaced sensors (see figure (1)). Due to the planar properties of the waves we consider, the phase delay between two successive sensors induced by a wave arriving at angle θ is:

$$\omega = \frac{2\pi d}{\lambda} \sin \theta \quad (1)$$

where λ is the wavelength. By choosing the inter-sensor distance d such that $d = \frac{\lambda}{2}$, we finally get: $\omega = \pi \sin \theta$. Let n be the number of sources, and N the number of snapshots (or observations). Using the narrow-band assumption, the AOA problem reduces to the parameters estimation of the following model:

$$\mathbf{y}_t = A(\Omega)\mathbf{x}_t + \mathbf{b}_t \quad (2)$$

where

- \mathbf{y}_t is the $m \times 1$ vector received on the m -sensor antenna,

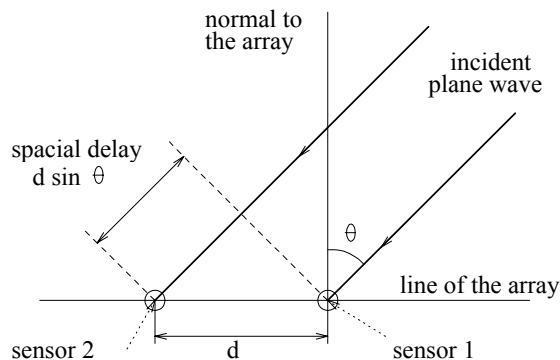


Figure 1: Phase delay induced by a plane wave impinging upon a linear antenna

- $t \in [1, N]$ is the index of the snapshot,
- \mathbf{b}_t is the $m \times 1$ noise vector, and \mathbf{x}_t is the $n \times 1$ source vector,
- $A(\Omega)$ is the $m \times n$ matrix of the n steering-vectors: $A(\Omega) = [\mathbf{a}(\omega_1), \dots, \mathbf{a}(\omega_n)]$, and $\mathbf{a}(\omega_k) = [1, e^{j\omega_k}, \dots, e^{j(m-1)\omega_k}]^T$

We wish to estimate parameter Ω given by $\Omega = [\omega_1, \dots, \omega_n]^T$, where ω_k determines the k -th angle of arrival. We will consider \mathbf{b}_t and \mathbf{x}_t as unknowns, and we make the following usual hypotheses: $N > m > n$; \mathbf{b}_t is a white gaussian noise vector; the sources are complex variables, independent from the noise, but might be correlated to each other; the number of sources, n , is assumed to be known. Various methods for estimating n are provided in the literature (see for instance Wax and Kailath (4)).

The paper is organized as follows: we first present the classical methods used to solve this type of problem, their advantages and limitations. Thereafter, a novel method using the cooperation of a classical multi-layer perceptron and a CSW perceptron is presented. It is shown how the results provided by the first perceptron can be used as a starting point to our novel MLE-like method. Finally, the results given by both methods are compared to those of classical methods, then analysed and discussed before a conclusion is eventually drawn.

A VIEW OVER PREVIOUS APPROACHES

The most quoted AOA estimation method is proba-

bly the MULTiple SIGNAL Classification (MUSIC, see Schmidt (1), Stoika and Nehorai (2) and (3), Xu and Buckley (5)). MUSIC is indeed quick and efficient; it is nonetheless limited in precision and resolution when a small amount of data only is available (N too small), and, most of all, it is inefficient in the case of correlated signals (see for instance Stoika and Nehorai (2)). Basically, the MUSIC method consists in computing the $m \times m$ correlation matrix of the signals, in order to extract the m eigenvectors and their associated eigenvalues. Assuming n , the number of sources is known, the n eigenvectors associated to the n largest eigenvalues define the so-called “signal subspace”, while the $m - n$ remaining eigenvectors span the so-called “noise subspace”. The orthogonality properties of these eigenvectors lead to consider a real function of the variable ω , which is defined as the norm of the projection of the steering vector $\mathbf{a}(\omega)$ onto the “noise subspace”. The inverse of this function is spotted for all ω . The n maxima are the estimation of the $\omega_i, i \in [1, n]$.

Another sub-optimal method, namely ESPRIT (for Estimation of Signal Parameters via Rotational Invariance Techniques), was developed shortly after MUSIC, and is also based on eigenvalue decompositions of the correlation matrix of the signals (see Roy *et al* (6) and Roy (7) for more details). ESPRIT, though superior to MUSIC in most cases, cannot easily handle the AOA problem when dealing with correlated signals.

On the other hand, the MLE method is known (see Stoika and Nehorai (2) and (3)) as an excellent estimator, and it is not handicapped in the case of partially correlated signals. Unfortunately, MLE presents the serious drawback of being computationally intensive. Indeed, the MLE method roughly consists in building a statistical model of the behavior of the signals; a criterion is defined, that is the log-likelihood of this model. All possible values for the unknown parameters are tested, and the estimate is the parameter vector that maximizes the log-likelihood. It is this compulsory test of *all* possible values that makes MLE a time consuming method. Therefore, though sub-optimal, the MUSIC method is generally preferred to MLE.

A NEW METHOD

A rough estimation

The first neural net (see figure (2)) receives exactly the same data as a MUSIC or an ESPRIT estimator, that is to say the $m \times m$ covariance matrix of the signals $R = E\{\mathbf{y}_t \mathbf{y}_t^H\}$. In practice, this matrix is unknown and we use an estimate of the covariance matrix $\hat{R} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}_t \mathbf{y}_t^H$. Matrix R is hermitian, thus it can be represented by m^2 reals. This leads us to create an MLP with m^2 inputs and n outputs. Finally, the hidden layer possesses L neurons, with L being a “nice” compromise between n and m^2 . The

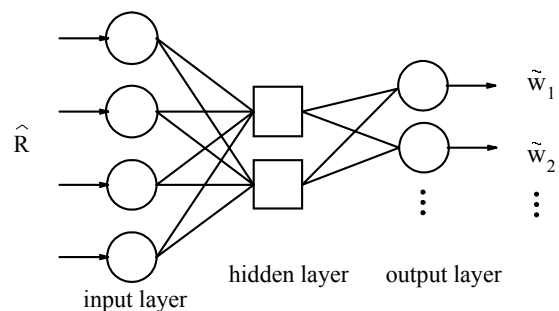


Figure 2: The structure of the first MLP

network is trained (see Burel (8)) to provide an estimation $\tilde{\omega}_k$ of ω_k at the k -th output neuron. It has to be noted that the first part of the system is composed of, say, q three-layer perceptrons, each of which is specifically adapted to the n -source estimation problem (where $n \in [1, q]$). These q networks are trained separately, in such a way that the outputs ω_k are sorted into strict increasing order within the training set. It is the estimation \hat{n} of the number of sources (we remind that this estimation is not exposed here) that decides which neural net (namely the \hat{n} -output network) is to be utilized. The training base is computed for angles θ_k varying from -60 to 60 degrees. For example here, in the case of the 2-source network, angle $\omega_1 = \pi \sin \theta_1$ is computed for θ_1 varying from -60° to 59° , and ω_2 for θ_2 varying from $\theta_1 + 1$ to 60° . Therefore, $60 \times (2 \times 60 + 1) = 7260$ couples are needed to describe the whole training set. In order to facilitate learning, the training data are computed for “high” SNR (25 dB) and number of snapshots ($N = 40$) parameter values.

In the sequel, the estimations of the true AOA coming out of this first MLP will be denoted $\hat{\Omega} = [\tilde{\omega}_1, \dots, \tilde{\omega}_n]^T$.

A refined estimation

The original idea is to refine the gross estimation $\hat{\Omega}$ through the optimization of a likelihood criterion on a gradient descent neural network: a maximum likelihood criterion is optimal in the sense of performance, but it is slow. The velocity is enhanced here, through a gradient descent algorithm. Furthermore, the initialization at a point close to the solution (using the rough estimate $\hat{\Omega}$ from the first MLP) brings two advantages: (1) the way to the solution is short, thus the system is even quicker; (2) it avoids being trapped in local minima. A general view of this second network is shown on figure (3). The network is a three-layer perceptron, where weights and neurons can take complex values. The N snapshots taken from the array outputs are presented randomly one after the other to the input layer. The input and output layers are both composed of m units. The output layer is forced to reproduce the $m \times 1$ input vector as closely as possible. We would like to point out that the hidden layer contains n neurons and that

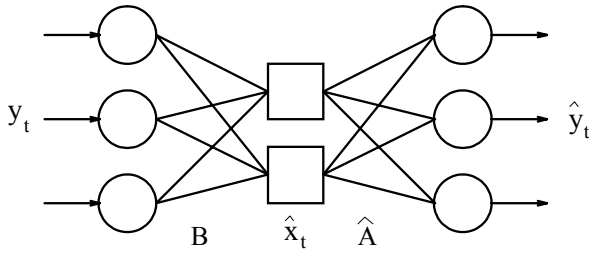


Figure 3: *The Constrained Structured Weights (CSW) MLP*

we called $\hat{\mathbf{x}}_t$ the hidden layer vector, so as to show the similarities between $\hat{\mathbf{y}}_t = \hat{A}\hat{\mathbf{x}}_t = A(\hat{\Omega})\hat{\mathbf{x}}_t$ and equation (2), where \hat{A} is *constrained* to represent the steering matrix $A(\hat{\Omega})$ corresponding to the parameter vector $\hat{\Omega}$. $\hat{\mathbf{x}}_t$ is generated out of observation vector \mathbf{y}_t , using a linear transformation: $\hat{\mathbf{x}}_t = B\mathbf{y}_t$. The error criterion to minimize is defined as:

$$e_{MS} = \frac{1}{2}E\{|\hat{\mathbf{y}}_t - \mathbf{y}_t|^2\}, \quad (3)$$

where \mathbf{y}_t and $\hat{\mathbf{y}}_t$ are the $m \times 1$ input and output vectors respectively, and $t \in [1, N]$. We prove (see appendix A) that minimizing e_{MS} is equivalent to optimizing the Maximum Likelihood criterion. Therefore, the second part of our system is theoretically equivalent to an MLE search.

In this CSW perceptron, the output layer does *not* provide any estimation of Ω , but a means to compute an error criterion. What is new in this network, is that it is the *weights* that contain the wanted information; the structure of the network can be specified as follows: the matrix of the weights joining the hidden layer to the output layer is constrained to take the form of a matrix of steering vectors $\mathbf{a}(\hat{\omega}_k)$, $k \in [1, N]$, where only the $\hat{\omega}_k$ may vary. This *output matrix* $\hat{A} = A(\hat{\Omega})$ is initialized with the $\hat{\Omega}$ previously estimated, and the matrix of the weights joining the first layer to the hidden layer (the *input matrix* B) is simply initialized as the pseudo-inverse of the output matrix $A(\hat{\Omega})$. During learning, the output matrix $A(\hat{\Omega})$ will be updated with respect to the constrained structure of the weights.

The simplified version of a complex-valued neuron is presented in figure (4). Moreover, the following

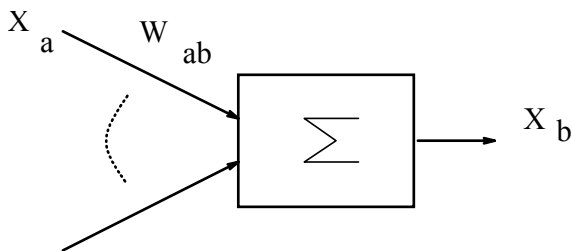


Figure 4: *The complex-valued neuron*

equation yields the propagation rule from neurons a to neuron b :

$$X_b = \sum_a W_{ab}X_a \quad (4)$$

Classically, the minimization of the error criterion e_{MS} requires the computation of a gradient in order to renew the weights after each iteration, as:

$$\Delta W_{ab} = -\alpha \frac{\partial e_{MS}}{\partial W_{ab}}, \quad (5)$$

where W_{ab} is the weight from neuron a to neuron b . For the input matrix weights, equation (5) is easy to use and the result of the computation (see Appendix B) of $\frac{\partial e_{MS}}{\partial W_{ab}}$ is:

$$\frac{\partial e_{MS}}{\partial W_{ab}} = \delta_b X_a^*, \quad (6)$$

where

$$\delta_b = \frac{\partial e_{MS}}{\partial X_b} = \sum_{k \in \text{succ}(b)} \delta_k W_{bk}^*, \quad (7)$$

where $\text{succ}(b)$ is the set of neurons on the layer succeeding to b and z^* stands for the complex conjugate of z .

Computing the gradient of the output matrix is a bit more delicate, because of the special constrained structure. It should be noted that, as far as the output matrix is concerned, the only parameters are the $\hat{\omega}_a$; thus the output gradient can be expressed as (see Appendix C):

$$\frac{\partial e_{MS}}{\partial \hat{\omega}_a} = \text{Re} \left(\sum_b \left(\frac{\partial e_{MS}}{\partial W_{ab}} \right) \left(\frac{\partial W_{ab}}{\partial \hat{\omega}_a} \right)^* \right), \quad (8)$$

where $\frac{\partial e_{MS}}{\partial W_{ab}}$ follows equation (6) and $\delta_b \triangleq \frac{\partial e_{MS}}{\partial X_b}$ is calculated using equation (3) with $X_b = \mathbf{y}_{tb}$. We get $\delta_b = \hat{\mathbf{y}}_{tb} - \mathbf{y}_{tb}$ here, where \mathbf{y}_{tb} (resp. $\hat{\mathbf{y}}_{tb}$) stands for the b -th component of \mathbf{y}_t (resp. $\hat{\mathbf{y}}_t$). Thus, for the output neuron b :

$$\frac{\partial e_{MS}}{\partial W_{ab}} = (\hat{\mathbf{y}}_{tb} - \mathbf{y}_{tb}) X_a^* \quad (9)$$

Moreover, the specific shape of the antenna gives $W_{ab} = e^{j(b-1)\hat{\omega}_a}$, thus we can write:

$$\frac{\partial W_{ab}}{\partial \hat{\omega}_a} = j(b-1)W_{ab} \quad (10)$$

Eventually, equations (8), (9) and (10) are combined to give the final expression for the output gradient:

$$\frac{\partial e_{MS}}{\partial \hat{\omega}_a} = \text{Re} \left(\sum_b (\hat{\mathbf{y}}_{tb} - \mathbf{y}_{tb}) X_a^* j(b-1)W_{ab}^* \right) \quad (11)$$

After each iteration, the $\hat{\omega}_a$ are renewed as:

$$\Delta \hat{\omega}_a = -\alpha \frac{\partial e_{MS}}{\partial \hat{\omega}_a}, \quad (12)$$

Then, all the steering-vectors $\mathbf{a}(\hat{\omega}_a)$ forming the constrained structured output matrix are renewed in their turn. For our experiments, we used the same α as in equation (5).

Before reviewing the results, we would like to stress the fact that, as for the first MLP, this CSW perceptron uses only the information present in the covariance matrix. Thus, all the methods exposed thereafter are provided with the same information and the difference in the performances simply lies in a better or lesser use of it.

EXPERIMENTAL RESULTS

In order to underline the performances of this novel AOA system, experiments are carried out by the cascaded real-neuron-MLP + CSW-perceptron pair, as well as by MUSIC, ESPRIT and MLE. In the experiments above, the parameters are set as follows: $n = 2$ sources, $m = 5$ sensors, $N = 15$ snapshots, and $SNR = 10$ dB for each signal. θ_2 varies from -50° to 50° , while θ_1 is fixed at 0° , but unknown.

The first perceptron is a 25-input, 2-output MLP, with 10 hidden neurons. The CSW-perceptron has 5 inputs, 5 outputs, and 2 hidden neurons. The networks are very small, therefore their use does not require much computing power: for one run, the networks need no more time than a MUSIC run: 0.4 seconds on a Sun workstation against 30 minutes for MLE (the MLE method requires the sampling of the whole working space). The number of iterations for the gradient descent of the CSW perceptron is 20.

To illustrate the two-source case, the discrepancy $\frac{1}{2} \left(\sqrt{E\{(\hat{\omega}_1 - \omega_1)^2\}} + \sqrt{E\{(\hat{\omega}_2 - \omega_2)^2\}} \right)$ is depicted as a function of θ_2 (θ_1 is fixed at 0°). Each point of the average discrepancy is statistically computed over 60 runs. Figure (5) shows the discrepancies obtained for uncorrelated signals.

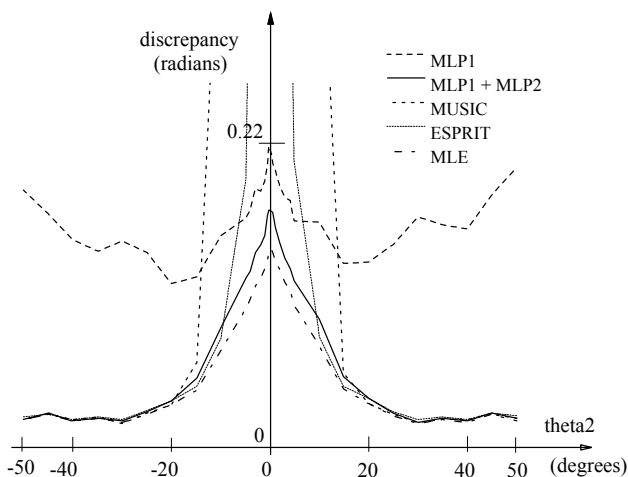


Figure 5: *Discrepancies for MUSIC, ESPRIT, MLP1, MLP1+MLP2, and MLE (uncorrelated signals)*

When the angles are quite apart, it can be noticed that all methods are alike: MUSIC, ESPRIT and the CSW-perceptron almost reach the performances of MLE. However, when θ_2 comes closer and closer to θ_1 , MUSIC soon fails and finds only one angle (in our experiments, MUSIC maxima merge for angles

closer than 12 degrees). Then, ESPRIT fails in its turn and provide spurious angles. Finally, for angles closer than 5 degrees, only MLE and our fine estimation MLP can still locate the angles. Moreover, it is important to note that (1) the CSW-perceptron improves dramatically the estimations established by the preceding MLP, except in the case where the interval between θ_1 and θ_2 is smaller than two degrees: in this region, the CSW perceptron only slightly improves the performances of the first MLP; (2) the performances of the networks pair are always very comparable to those of MLE. Figure (6) repre-

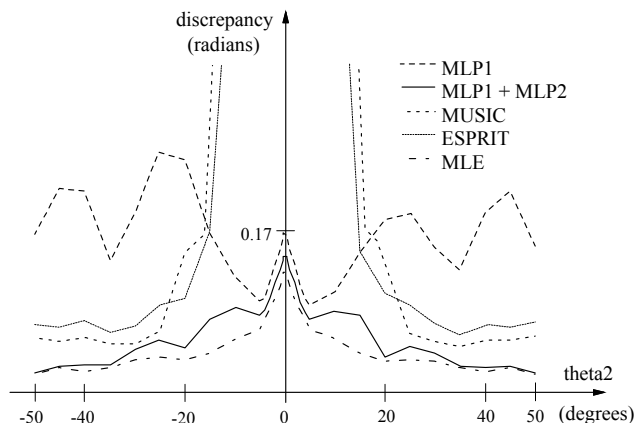


Figure 6: *Discrepancies for MUSIC, ESPRIT, MLP1, MLP1+MLP2, and MLE (strongly correlated signals)*

sents the discrepancies obtained when the signals are strongly correlated: the correlation of the sources is:

$$E\{\mathbf{x}\mathbf{x}^H\} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \quad (13)$$

All other parameters are kept alike. It is interesting to note that the failure of MUSIC and ESPRIT in the case of close angles is more obvious here. On the other hand, the cooperation of neural networks still shows low discrepancies, even when θ_1 and θ_2 are extremely close together (at a 0.3° distance interval, for instance).

CONCLUSION

A neural system for Angles Of Arrival Estimation has been presented in this paper. The system is composed of two MLPs: a three-layer perceptron is first fed with the normalized correlation matrix of the received signals, assuming that the number of sources is known. It outputs a "rough" estimation of the parameters to estimate, which are the AOA. Thereafter, a complex-weight complex-neuron perceptron is initialized with the preceding rough estimation, and is trained during a limited number of iterations so as to minimize an output error criterion. This MLP is dedicated in the sense that its weights are constrained to take the form of a steering matrix, thus following closely the mathematical model of the

problem. It implies that the error to minimize is theoretically the same as the minimization involved by the MLE algorithm. Called the Constrained Structured Weights (CSW) perceptron, this fine estimation MLP proves experimentally to perform as expected, that is to say in a very comparable way to MLE, while its speed is comparable to that of MUSIC.

Experiments where angles are very close together or when signals are strongly correlated to each other (where both MUSIC and ESPRIT approaches fail) can however be handled by the CSW approach, and show the interest of the approach.

REFERENCES

1. Schmidt R O, Nov 1981, "A signal subspace approach to multiple emitter location and spectral estimation", Ph.D. dissertation, Stanford University, Stanford, CA
2. Stoica P and Nehorai A, May 1989, "MUSIC, maximum likelihood, and Cramér-Rao bound", *IEEE Trans. ASSP*, **37**, (5), 720-741
3. Stoica P and Nehorai A, Dec 1990, "MUSIC, maximum likelihood, and Cramér-Rao bound: further results and comparisons", *IEEE Trans. ASSP*, **38**, (12), 2140-2150
4. Wax M and Kailath T, Apr 1985, "Detection of signals by information theoretic criteria", *IEEE Trans. ASSP*, **33**, 387-392
5. Xu X L and Buckley K M, Oct 1992, "Bias analysis of the MUSIC location estimator", *IEEE Trans. ASSP*, **40** (10), 2559-2569
6. Roy R, Paulraj A and Kailath T, 1986, "Estimation of signal parameters via rotational invariance techniques-ESPRIT", *Proc IEEE ICASSP*, **4**, 2495-2498
7. Roy R H, 1987, "ESPRIT, Estimation of signal parameters via rotational invariance", PhD dissertation, Stanford Univ, Stanford, CA
8. Burel G, 1991, "Réseaux de neurones en traitement d'images: des modèles théoriques aux applications industrielles", PhD dissertation, Université de Bretagne Occidentale, Brest, France

APPENDIX A

MLE estimator

The equation ruling our antenna is $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, where \mathbf{A} represents $A(\Omega)$, the matrix of the steering-vectors. Let us consider the noise $\mathbf{b} = \mathbf{y} - \mathbf{A}\mathbf{x}$. \mathbf{b} is a gaussian variable with zero mean and its covariance

matrix is $\Gamma = \sigma^2 I$ by definition. Thus the conditional probability of the noise vector is:

$$p(\mathbf{b}/\hat{A}, \hat{\mathbf{x}}, \hat{\sigma}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \hat{A}\hat{\mathbf{x}})^H \Gamma^{-1}(\mathbf{y} - \hat{A}\hat{\mathbf{x}})\right)}{(2\pi)^{m/2} \sqrt{|\Gamma|}} \quad (14)$$

The definition of Γ yields $\Gamma^{-1} = \frac{1}{\hat{\sigma}^2} I$ and $|\Gamma| = \hat{\sigma}^{2m}$.
Thence:

$$p(\mathbf{b}/\hat{A}, \hat{\mathbf{x}}, \hat{\sigma}) = \frac{1}{(2\pi)^{m/2} \hat{\sigma}^m} \exp\left(-\frac{1}{2\hat{\sigma}^2} |\mathbf{y} - \hat{A}\hat{\mathbf{x}}|^2\right) \quad (15)$$

Thus the log-likelihood is defined as:

$$\begin{aligned} L_1 &= \log p(\mathbf{b}/\hat{A}, \hat{\mathbf{x}}, \hat{\sigma}) \\ &= -\frac{m}{2} \log(2\pi) - m \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} |\mathbf{y} - \hat{A}\hat{\mathbf{x}}|^2 \end{aligned} \quad (16)$$

We wish to maximize the log-likelihood of the noise; that is, to maximize:

$$L_2 = -m \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} |\mathbf{y} - \hat{A}\hat{\mathbf{x}}|^2 \quad (17)$$

Since the maximization should be reached for all values of $\hat{\sigma}$, the optimization reduces to the minimization of:

$$L_3 = |\mathbf{y} - \hat{A}\hat{\mathbf{x}}|^2 \quad (18)$$

which should be applied to the whole set of N snapshots \mathbf{y}_t , $t \in [1, N]$. Finally it appears that MLE consists in the minimization of:

$$L = \sum_{t=1}^N |\mathbf{y}_t - \hat{A}\hat{\mathbf{x}}_t|^2 \quad (19)$$

Vector $\hat{A}\hat{\mathbf{x}}_t$ belongs to the subspace generated by \hat{A} , called $span\{\hat{A}\}$. Hence the best choice for $\hat{\mathbf{x}}_t$ is such that $\hat{A}\hat{\mathbf{x}}_t$ is the projection of \mathbf{y}_t on $span\{\hat{A}\}$. To conclude, MLE is equivalent to the minimization of:

$$\sum_{t=1}^N |\mathbf{y}_t - P_{\hat{A}}(\mathbf{y}_t)|^2 \quad (20)$$

where $P_{\hat{A}} = \hat{A}(\hat{A}^H \hat{A})^{-1} \hat{A}^H$ is the matrix of orthogonal projection on $span\{\hat{A}\}$, and \hat{A} is the estimation of A .

Constrained-structured MLP

The MLP depicted on figure (3) is such that $\hat{\mathbf{y}}_t = \hat{A}\hat{\mathbf{x}}_t$ where \hat{A} presents the same shape as the steering-vectors; thus $\hat{\mathbf{y}}_t \in span\{\hat{A}\}$. As previously said, the network tries to minimize

$$e_{MS} = \frac{1}{2} E\{|\hat{\mathbf{y}}_t - \mathbf{y}_t|^2\}, \quad (21)$$

which proves that the best solution is to chose $\hat{\mathbf{y}}_t = P_{\hat{A}}(\mathbf{y}_t)$. This demonstrates that MLE and the constrained-structured MLP search for the minimum of the same

criterion. Hence they lead to the same solution in theory.

APPENDIX B

By convention the derivation of a real r by a complex z is chosen as:

$$\frac{\partial r}{\partial z} = \frac{\partial r}{\partial z^R} + j \frac{\partial r}{\partial z^I}, \quad (22)$$

where z^R and z^I are the real and imaginary parts of z , respectively.

Property 1: if a real e_{MS} depends upon a complex z according to the structure presented on figure (7), where x_1, \dots, x_n are complex, function h is derivable with respect to the real and the imaginary parts of its argument and g_1, \dots, g_n are holomorphic functions, then we have:

$$\frac{\partial e_{MS}}{\partial z} = \sum_k \frac{\partial e_{MS}}{\partial x_k} \left(\frac{\partial x_k}{\partial z} \right)^* \quad (23)$$

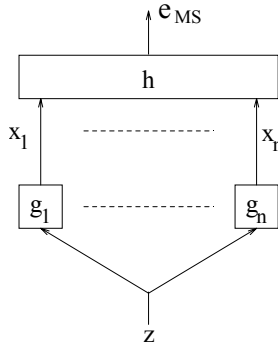


Figure 7: Dependence relation between e_{MS} and z

Demonstration: e_{MS} is such that:

$$\begin{aligned} de_{MS} &= \sum_k \left(\frac{\partial e_{MS}}{\partial x_k^R} dx_k^R + \frac{\partial e_{MS}}{\partial x_k^I} dx_k^I \right) \\ &= \sum_k \operatorname{Re} \left\{ \frac{\partial e_{MS}}{\partial x_k} dx_k^* \right\} \\ &= \operatorname{Re} \left\{ \sum_k \frac{\partial e_{MS}}{\partial x_k} dx_k^* \right\} \end{aligned} \quad (24)$$

Therefore we can also write:

$$de_{MS} = \operatorname{Re} \left\{ \frac{\partial e_{MS}}{\partial z} dz^* \right\} \quad (25)$$

Functions g_k are holomorphic, hence:

$$dx_k = \frac{\partial x_k}{\partial z} dz \quad (26)$$

which, combined to (24), leads to:

$$de_{MS} = \operatorname{Re} \left\{ \sum_k \frac{\partial e_{MS}}{\partial x_k} \left(\frac{\partial x_k}{\partial z} \right)^* dz^* \right\} \quad (27)$$

Identification with (25) gives:

$$\operatorname{Re} \left\{ \left[\frac{\partial e_{MS}}{\partial z} - \sum_k \frac{\partial e_{MS}}{\partial x_k} \left(\frac{\partial x_k}{\partial z} \right)^* \right] dz^* \right\} = 0 \quad (28)$$

Thus property 1 is proved, since:

$$\begin{aligned} \operatorname{Re}\{ab\} = 0 \quad \forall b &\implies \operatorname{Re}\{1.a\} = 0 \text{ and } \operatorname{Re}\{j.a\} = 0 \\ &\implies a = 0 \end{aligned} \quad (29)$$

Using the notation $\delta_b = \frac{\partial e_{MS}}{\partial X_b}$, the gradient of the complex neuron presented in figure (4) can be expressed as:

$$\frac{\partial e_{MS}}{\partial W_{ab}} = \frac{\partial e_{MS}}{\partial X_b} \left(\frac{\partial X_b}{\partial W_{ab}} \right)^* = \delta_b X_a^* \quad (30)$$

since $X_b = \sum_a W_{ab} X_a$, and using property 1 ($z = W_{ab}, n = 1, x_1 = X_b$ here).

Finally, $z = X_b$, $x_k = X_k$ and property 1 yield to:

$$\frac{\partial e_{MS}}{\partial X_b} = \sum_{k \in \operatorname{succ}(b)} \frac{\partial e_{MS}}{\partial X_k} \left(\frac{\partial X_k}{\partial X_b} \right)^* \quad (31)$$

which gives, with $X_k = \sum_b W_b X_b$:

$$\delta_b = \frac{\partial e_{MS}}{\partial X_b} = \sum_{k \in \operatorname{succ}(b)} \delta_k W_{bk}^* \quad (32)$$

APPENDIX C

The constraints imposed on the CVW perceptron lead us to introduce some modifications in the refreshment rule of the weights of the output matrix. These weights having the form $W_{ab} = e^{j(b-1)\hat{\omega}_a}$, or more generally $W_{ab} = e^{f_b(\hat{\omega}_a)}$, they depend only upon the $\hat{\omega}_a$. Thus, we can write:

$$\frac{\partial e_{MS}}{\partial \hat{\omega}_a} = \sum_b \left(\frac{\partial e_{MS}}{\partial W_{ab}^R} \frac{\partial W_{ab}^R}{\partial \hat{\omega}_a} + \frac{\partial e_{MS}}{\partial W_{ab}^I} \frac{\partial W_{ab}^I}{\partial \hat{\omega}_a} \right) \quad (33)$$

Convention (22) yields to:

$$\frac{\partial e_{MS}}{\partial W_{ab}} = \frac{\partial e_{MS}}{\partial W_{ab}^R} + j \frac{\partial e_{MS}}{\partial W_{ab}^I} \quad (34)$$

Thence we have:

$$\begin{aligned} \sum_b \left(\frac{\partial e_{MS}}{\partial W_{ab}} \right) \left(\frac{\partial W_{ab}}{\partial \hat{\omega}_a} \right)^* &= \\ \sum_b \left(\frac{\partial e_{MS}}{\partial W_{ab}^R} + j \frac{\partial e_{MS}}{\partial W_{ab}^I} \right) \left(\frac{\partial W_{ab}^R}{\partial \hat{\omega}_a} - j \frac{\partial W_{ab}^I}{\partial \hat{\omega}_a} \right) \end{aligned} \quad (35)$$

Developing the second term of equation (35) reveals that its real part is identical to the right term of equation (33). The gradient of the output matrix thus reduces to:

$$\frac{\partial e_{MS}}{\partial \hat{\omega}_a} = \operatorname{Re} \left(\sum_b \left(\frac{\partial e_{MS}}{\partial W_{ab}} \right) \left(\frac{\partial W_{ab}}{\partial \hat{\omega}_a} \right)^* \right) \quad (36)$$