Identifiability of Post-Nonlinear Mixtures

Sophie Achard and Christian Jutten, *Member, IEEE*

***Abstract*—This letter deals with the resolution of the blind source separation problem using the independent component analysis method in post-nonlinear mixtures. Using the sole hypothesis of the source independence is not obvious to reconstruct the sources in nonlinear mixtures. Here, we prove the identifiability under weak assumptions on the mixture matrix and density sources.**

***Index Terms*—Blind source separation, identifiability, independent component analysis (ICA), post nonlinear mixture.**

I. Introduction

STARTING from the observation of unknown mixtures of some independent sources, the problem of blind source separation consists of reconstructing the sources. The method of independent component analysis (ICA) allows one to find a set of combinations of the observations that are independent variables that are called the reconstructed sources. One must determine whether the reconstructed sources represent the original sources. This is referred to as the *identifiability of the mixture*. For linear mixtures, Comon [1] has shown that when there is no noise, the same number of sources and observations, and, at most, one Gaussian source, the sole hypothesis of the source independence is sufficient to guarantee that the independent reconstructed sources represent the original sources.

However, when the mixture is nonlinear, it is, in general, not identifiable. Indeed, Darmois [2][1] showed that there exist several nonlinear transformations with a nondiagonal Jacobian (which means that each output variable depends on at least two input variables), which preserve the independence of the variables.

Quite recently, Taleb and Jutten [4] suggested to define a specific nonlinear mixture called post-nonlinear (PNL). The mixture consists of a matrix and then component-wise invertible nonlinearities (see Fig. 1). They propose to solve the blind source separation problem with an independent component analysis method, which consists of looking for a separation structure made up of the composition of component-wise nonlinearities and a separation matrix, which, when applied to the observations, yields independent variables.

Manuscript received July 8, 2004; revised October 28, 2004. This work was also supported by the European BLISS project (IST-1999-14190). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Christophe Pesquet.

S. Achard is with the Laboratory of Computation and Modeling, CNRS UMR 5523, Univ. J. Fourier, 38041 Grenoble, France and also with the Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, CB2 3EB U.K. (e-mail: sa428@cam.ac.uk).

C. Jutten is with the Images and Signals Laboratory, CNRS UMR 5083, Institut National Polytechnique and Univ. J. Fourier, 38031 Grenoble, France (e-mail: christian.jutten@inpg.fr).

Digital Object Identifier 10.1109/LSP.2005.845593

[1]This reference and all the papers written by Darmois are in French and difficult to obtain, but Darmois's results are also available in the book by Kagan *et al.* [3], especially in Ch. 3.

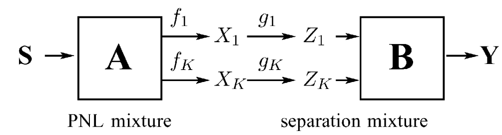

Fig. 1. PNL mixture and separation structure.

Taleb and Jutten proved that independent component analysis provides a solution to the problem of blind source separation, but their proof requires some restrictions on the density of the sources and especially that there exists a compact set on which the density is zero. They conjectured that it would be possible to remove this working assumption. In this letter, we prove that indeed, this constraint can be removed.

The letter is structured as follows. In Section II, we will give the definition of the model, the assumptions, and the main theorem. Then, the proof is detailed, Section III gives the proof of an intermediate lemma, and Section IV allows us to conclude.

II. Model, Assumptions and Theorem

A. Model

Let $S_1, \ldots, S_K$ be K **independent** random variables, let $f_1, \ldots, f_K$ be K **invertible** and differentiable functions, and $\mathbf{A}$ be an **invertible** $K \times K$ matrix. Let us define the PNL observations $X_1, \ldots, X_K$ (see left part of Fig. 1)

$$\forall i,\ 1 \le i \le K,\ X_i = f_i\left(\sum_{k=1}^{K} \mathbf{A}_{ik} S_k\right). \tag{1}$$

In this letter, we focus on the problem of blind source separation of this mixture, which consists of looking for nonlinearities $g_i, i = 1, \ldots, K$ and a matrix $\mathbf{B}$ (see right part of Fig. 1) such that $Y_1, \ldots, Y_K$ defined by (2) are independent

$$\forall i,\ 1 \le i \le K,\ Y_i = \sum_{k=1}^{K} \mathbf{B}_{ik} g_k\left(X_k\right). \tag{2}$$

Let us now denote $h_i = g_i \circ f_i, i = 1, \ldots, K$; thus, we can write $Y_1, \ldots, Y_K$ in terms of the sources as

$$\forall i,\ 1 \le i \le K,\ Y_i = \sum_{k=1}^{K} \mathbf{B}_{ik} h_k\left(\sum_{l=1}^{K} \mathbf{A}_{kl} S_l\right). \tag{3}$$

As shown in [4], the key point in the identifiability problem of PNL mixtures is to show that if $Y_1, \ldots, Y_K$ (3) are independent, then $h_i, i = 1, \ldots, K$, are **necessarily linear**.

B. Assumptions

For clarity, all the hypotheses used in the sequel are listed here.

H1) Each source appears mixed at least once in the observations.

H2) We suppose $h_1, h_2, \ldots, h_K$ differentiable and invertible (same assumptions as $f_1, f_2, \ldots, f_K$).

H3) There exists at most one Gaussian source.

H4) The joint density function of the sources is supposed differentiable, and its derivative is continuous on its support (no hypothesis is made on the support of the density functions).

Concerning hypothesis H2, assuming in addition $h_1, h_2, \ldots, h_K$ increasing functions does not involve any loss of generality since $h_1, \ldots, h_K$ are already supposed continuous and invertible (this property will be used in Section IV for the proof of the theorem). H3 is the usual hypothesis necessary for the identifiability of linear mixtures. As for H1 and H4, let us make some more detailled remarks.

Remark 1: H4 is a technical assumption and includes the case of joint density function with infinite support and with bounded support in restricting the integral on the support of the joint density. However, in the case of bounded sources, or of distributions that vanish on a compact, there already exist theoretical results [4]–[6]. For discrete-valued sources, the problem is still open, but a geometrical approach is more adequate.

Remark 2: The assumption H1 is a necessary condition. Indeed, the sources may be recovered if they appear mixed at least one time in the mixture. Since $\mathbf{A}$ is invertible, at least one nonzero entry per row and per column exists.

The simplest matrix is a diagonal matrix up to a permutation, and it implies no mixture. Consequently, $X_i = f_i\left(\sum_{k=1}^{K} \mathbf{A}_{ik} S_k\right)$ reduces to $X_i = f_i\left(\mathbf{A}_{i\sigma(i)} S_{\sigma(i)}\right)$, where σ is a permutation on $\{1, \ldots, K\}$, and the $X_1, \ldots, X_K$ are independent random variables. In that case (only using the independence assumption), it is impossible to estimate the nonlinearities f_i and to restore the sources $S_1, \ldots, S_K$.

Hence, to be able to estimate each nonlinearity f_i, the matrix $\mathbf{A}$ must actually mix the sources. If one observation only depends on one source, e.g., $X_1 = f_1(\mathbf{A}_{11} S_1)$ (without loss of generality), then this source must appear in another mixture, i.e., there exists $j \neq 1$ such that $X_j = f_j\left(\sum_{k=1}^{K} \mathbf{A}_{jk} S_k\right)$ with $\mathbf{A}_{j1} \neq 0$, and there exists $l \neq 1$ such that $\mathbf{A}_{jl} \neq 0$. If it is not satisfied, f_1 cannot be estimated, and X_1 will be considered as a source instead of S_1. Therefore, H1 can be equivalently characterized on the matrix by the following condition: For all $i, j = 1, \ldots, K$, such that $\mathbf{A}_{ij} \neq 0$, either there exists $k \neq j$ such that $\mathbf{A}_{ik} \neq 0$ or there exists $l \neq i$ such that $\mathbf{A}_{lj} \neq 0$.

C. Theorem

The identifiability is then stated by the following theorem.

Theorem 1: Let us take the model (3), with assumptions H1, H2, H3, and H4.

$Y_1, \ldots, Y_K$ are independent if and only if $h_1, \ldots, h_K$ are linear functions and $\mathbf{BA} = \mathbf{PD}$, where $\mathbf{P}$ is a permutation matrix, and $\mathbf{D}$ is a diagonal matrix.

Let us make some comments about this theorem.

First, it is easy to prove that if $h_1, \ldots, h_K$ are linear and $\mathbf{BA} = \mathbf{PD}$, then $Y_1, \ldots, Y_K$ are independent. Thus, it remains to prove the converse, that is, *if* $Y_1, \ldots, Y_K$ *are independent, then* $h_1, \ldots, h_K$ *are linear and* $\mathbf{BA} = \mathbf{PD}$. However, it is sufficient to prove that the independence of $Y_1, \ldots, Y_K$ implies the linearity of $h_1, \ldots, h_K$. Indeed, when the functions $h_1, \ldots, h_K$ are linear, the model is simply a linear model, and by applying the results of Comon [1], we deduce that $\mathbf{BA} = \mathbf{PD}$.

The proof of theorem 1 is done according to the following scheme.

In Section III, we first prove the next lemma 1, in which H1 is replaced by H1'. This assumption H1' is just to process but leads to define a smaller set of mixing matrices $\mathbf{A}$. The difference of mixing matrices defined by H1' and H1 correspond to a particular case where one row with only one nonzero entry exists. This special case will be studied in Section IV for completing the proof of theorem 1.

III. Lemma 1

Lemma 1: Let us suppose the following.

H1') There exist at least two nonzero elements in each row of $\mathbf{A}$.

Let us take the model (3), with assumptions H1', H2, H3, and H4. If $Y_1, \ldots, Y_K$ are independent, then $h_1, \ldots, h_K$ are linear functions.

A. Outlines

Equivalently, we can prove *by reductio ad absurdum* that if there exists at least one nonlinear function among the h_m functions, then for any invertible matrix $\mathbf{B}$, $Y_1, \ldots, Y_K$ cannot be independent.

A natural characterization of independent variables is given by the equality of their joint density and the product of their marginal densities, but in practice, this criterion needs to be expressed in a different way.

Thus, in Section III-B, we present an explicit characterization of nonindependence based on mutual information. Indeed, mutual information is minimal and equal to zero when calculated for independent random variables.

Finally, in Section III-C, we will show that if there exists at least one nonlinear function among the h_m functions, then mutual information calculated for $Y_1, \ldots, Y_K$ defined in (3) is nonzero.

B. Characterization of Independence

The characterization of independence used in the sequel comes from theorem 3.4 in [7] (which is recalled below). It gives a characterization of the minimum of mutual information over the set Π, where $\Pi = \{f : \forall i, \int f d\mu_{-i} = p_{Y_i}\}$, and $\int f d\mu_{-i}$ denotes the integral of f with respect to all the coordinates except the ith coordinate. It is clear that $p_{Y_1,\ldots,Y_K} \in \Pi$ as well as $\prod p_{Y_i}$.

When the integrals exist, we define δ

$$\delta(w) = \int w \log\left(\frac{w}{\prod_{i=1}^{K} p_{Y_i}}\right) d\mu, \qquad w \in \Pi.$$

However, δ can be rewritten as

$$\delta(w) = \int w \log(w) d\mu - \int w \log \left(\prod_{i=1}^{K} p_{Y_i} \right) d\mu$$

and the last term in the difference can be expressed as follows:

$$\begin{aligned}
&\int w \log \left(\prod_{i=1}^{K} p_{Y_i} \right) d\mu \\
&= \sum_{j=1}^{K} \int w \log(p_{Y_j}) d\mu = \sum_{j=1}^{K} \int \left(\int w d\mu_{-j} \right) \log(p_{Y_j}) d\mu_j \\
&= \sum_{j=1}^{K} \int p_{Y_j} \log(p_{Y_j}) d\mu_j = \sum_{j=1}^{K} \int \prod_{i=1}^{K} p_{Y_i} \log(p_{Y_j}) d\mu.
\end{aligned}$$

The last expression uses the fact that $\int p_{Y_j} d\mu_j = 1$.

Then, it follows that for all $w \in \Pi$

$$\delta(w) = \int w \log(w) d\mu - \int \left(\prod_{i=1}^{K} p_{Y_i} \right) \log \left(\prod_{i=1}^{K} p_{Y_i} \right) d\mu.$$

δ represents, in fact, the Kullback–Leibler divergence between w and $\prod_{i=1}^{K} p_{Y_i}$. It is known [8] that δ satisfies $I(Y_1, Y_2, \ldots, Y_K) = \delta(p_{Y_1, Y_2, \ldots, Y_K})$, where I denotes the mutual information.

Then, theorem 3.4 in [7] claims the following.

Theorem 3.4.) *$\widetilde{w}$ minimizes $\int w \log w d\mu$, $w \in \Pi$, if and only if $\int (\log \widetilde{w} + 1) \alpha d\mu = 0$ for all integrable function α satisfying $\int \alpha d\mu_{-i} = 0$, $i = 1, \ldots, K$.*

As $Y_1, \ldots, Y_K$ are defined by (3), i.e., with at least one nonlinear invertible function h_m and $\mathbf{B}$ an invertible fixed matrix, the second term in δ, $\int \prod_{i=1}^{K} p_{Y_i} \log \prod_{i=1}^{K} p_{Y_i} d\mu$, is a constant. Thus, by applying theorem 3.4 to δ, $I(Y_1, Y_2, \ldots, Y_K)$ will be nonzero if and only if there exists a function α such that

$$\int \alpha d\mu_{-i} = 0, \quad \text{for all } i = 1, \ldots, K \tag{4}$$

and

$$\int \alpha \log(p_{Y_1, \ldots, Y_K}) d\mu \neq 0. \tag{5}$$

Here, we have applied theorem 3.4 by taking $\widetilde{w} = p_{Y_1, \ldots, Y_K}$. Let us notice that $\int \alpha(\log \widetilde{w} + 1) d\mu = \int \alpha \log \widetilde{w} d\mu$ because of the definition of α.

For simplicity, the construction of α will be detailed in the two-dimensional case, i.e., $K = 2$. This assumption graphically represents how to construct α. Following the scheme of the proof, the extension to a more general case when $K > 2$ is obvious.

C. Construction of α

In this section, the construction of α is explained, and we will give an example to complete the proof.

1) Let us recall the properties that have to be verified by α in the case $K = 2$:
   - The support of α is included in the support of the joint density of $Y_1, \ldots, Y_K$ [so that the integral (5) is well defined].
   - $\int \alpha d\mu_{-i} = 0$ for all $i = 1, \ldots, K$. In our specific case, with $K = 2$, these relations can be expressed as $\int \alpha d\mu_{-1} = \int \alpha dy_2 = 0$ and $\int \alpha d\mu_{-2} = \int \alpha dy_1 = 0$.
   - Equation (5) verified by α and taking into account the model (3) writes

$$\begin{aligned}
&\int \log p_{Y_1, Y_2}(y_1, y_2) \alpha(y_1, y_2) dy_1 dy_2 \\
&\qquad = \int \Phi(u_1, u_2) \alpha(\mathbf{B}(u_1, u_2)) |\det \mathbf{B}| du_1 du_2 \neq 0
\end{aligned}$$

   where Φ is defined by

$$\Phi(u_1, u_2) = \log \frac{\prod_{i=1,2} p_{S_i} \left(\sum_{j=1,2} \mathbf{A}_{ij}^{-1} h_j^{-1}(u_j) \right)}{|h_1'(h_1^{-1}(u_1)) h_2'(h_2^{-1}(u_2))|}. \tag{6}$$

2) Integrating twice by part on the last integral, we obtain the equation shown at the bottom of the page, where $\mathbf{y} = (y_1, y_2)$, $\partial_1 \partial_2 \Phi(v_1, v_2) = \partial^2 \Phi / \partial v_1 \partial v_2$, and $\mathcal{A}(v_1, v_2) = \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \alpha(\mathbf{B}(u_1, u_2)) |\det \mathbf{B}| du_1 du_2$.

   Let us denote $\Phi(v_1, v_2) = \log \Theta(v_1, v_2)$.

3) The key to the proof is now to study the function Θ

$$\Theta(u_1, u_2) = \frac{\prod_{i=1,2} p_{S_i} \left(\sum_{j=1,2} \mathbf{A}_{ij}^{-1} h_j^{-1}(u_j) \right)}{|h_1'(h_1^{-1}(u_1)) h_2'(h_2^{-1}(u_2))|}.$$

   Indeed, following Darmois [9], we show that the function Θ cannot be expressed as a product of univariate functions.

   In fact, let us suppose that there exist two functions β_1 and β_2 such that

$$\Theta(u_1, u_2) = \beta_1(u_1) \beta_2(u_2).$$

$$\begin{aligned}
\int \log p_{Y_1, Y_2}(\mathbf{y}) \alpha(\mathbf{y}) d\mathbf{y} &= \int \Phi(u_1, u_2) \alpha(\mathbf{B}(u_1, u_2)) |\det \mathbf{B}| du_1 du_2 \\
&= \int \partial_1 \partial_2 \Phi(v_1, v_2) \underbrace{\iint_{]-\infty, v_1] \times]-\infty, v_2]} \alpha(\mathbf{B}(u_1, u_2)) |\det \mathbf{B}| du_1 du_2}_{\mathcal{A}(v_1, v_2)} dv_1 dv_2
\end{aligned}$$

Then, denoting $\widetilde{u}_i = h_i^{-1}(u_i)$, $i = 1, 2$, this is equivalent to suppose

$$\prod_{i=1,2} p_{S_i}(\mathbf{A}_{i1}^{-1}\widetilde{u}_1 + \mathbf{A}_{i2}^{-1}\widetilde{u}_2) = \prod_{i=1,2} (\beta_i(u_i)|h_i'(\widetilde{u}_i)|). \quad (7)$$

By assumption H1' on the matrix $\mathbf{A}$, there exist also at least two nonzero terms in each row of $\mathbf{A}^{-1}$, and the assumption on the functions h_i means that there exists one function h_m such that h_m' is nonconstant. Then, using the same argument as Darmois in [9], we show that all the functions in (7) are nonzero on the whole space $\mathbb{R}$. Thus, (7) is equivalent to

$$\sum_{i=1,2} \log p_{S_i}(\mathbf{A}_{i1}^{-1}\widetilde{u}_1 + \mathbf{A}_{i2}^{-1}\widetilde{u}_2) = \sum_{i=1,2} \log(\beta_i(u_i)|h_i'(\widetilde{u}_i)|). \quad (8)$$

Then, we apply the results of Darmois concerning the solutions of equations of the form

$$F_1(U_1) + F_2(U_2) = f_1(u_1) + f_2(u_2)$$

where U_1 and U_2 are linear combinations of u_1 and u_2. The result claims that when f_1 or f_2 is nonconstant, F_1 and F_2 are quadratic. Applying the result to (8) implies that the density function of S_1 and S_2 are Gaussian, which contradicts hypothesis H3.

Finally, as $\Phi = \log\Theta$, Φ is not equal to the sum of univariate functions, we conclude that there exists a domain V on which the function $\partial_1\partial_2\Phi$ is nonzero, and its sign does not change.

4) To conclude the proof, we have to write a function α such that $\mathcal{A} > 0$ on V. For instance, in this purpose, we can choose α such that $\mathcal{A}$ is strictly positive on V and null outside V and defined by

$$\alpha(x,y) = \sin\left(\pi\frac{(x-\lambda_1)}{p_1}\right)\sin\left(\pi\frac{(y-\lambda_2)}{p_2}\right),$$
$$\text{for } (x,y) \in \left[\lambda_1 - \frac{p_1}{2}, \lambda_1 + \frac{p_1}{2}\right] \times \left[\lambda_2 - \frac{p_2}{2}, \lambda_2 + \frac{p_2}{2}\right]$$

where λ_1, λ_2, p_1, and p_2 are defined in Fig. 2.

Let us note that when $K > 2$, it is known that the results of Darmois are also true, and the function α is build with two sine functions and $K-2$ cosine functions. The lemma is then also true for $K > 2$. Fig. 2 represents the function $\alpha \circ \mathbf{B}$.

IV. Proof of Theorem 1

With lemma 1, the identifiability of post-nonlinear mixtures is proven, subject to the restriction on the mixing matrix $\mathbf{A}$, which must satisfy H1'. In order to complete the proof of theorem 1, it remains now to consider the matrices that verify H1 and not H1'. Consequently, this consists of dealing with the case where X_1 and X_2 are defined by (9)

$$X_1 = f_1(\mathbf{A}_{11}S_1), \qquad X_2 = f_2(\mathbf{A}_{21}S_1 + \mathbf{A}_{22}S_2). \quad (9)$$

Under the assumptions on the mixture and the nonlinearities and following the proof of the previous lemma (lemma 1), showing that X_1 and X_2 are dependent, i.e., there exists a stochastic link between X_1 and X_2, implies in this case that the function Φ [defined in (6)] cannot be expressed as a sum of univariate functions and that the variables Y_1 and Y_2 cannot be independent. This concludes the proof of theorem 1.

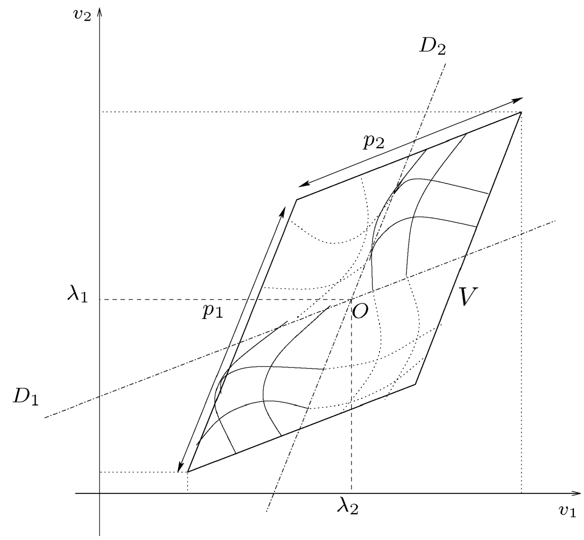

Fig. 2. Graphical representation of $\alpha \circ \mathbf{B}$ and definition of the support of $\alpha \circ \mathbf{B}$, -.- axes defined by the variable change $\mathbf{B}$: D_1 : $\mathbf{B}_{11}(v_1-\lambda_1) + \mathbf{B}_{12}(v_2-\lambda_2) = 0$ D_2 : $\mathbf{B}_{21}(v_1-\lambda_1) + \mathbf{B}_{22}(v_2-\lambda_2) = 0$.

The problem of studying the link between X_1 and X_2 is called a factor analysis problem. Darmois [10] solved this problem by studying the difference between the joint repartition function and the product of the marginals. Actually, he proved that if f_1 and f_2 are monotonous invertible functions, then the sign of the difference between the joint repartition function and the product of the marginals is strictly positive on a certain neighborhood of the space. Then, we conclude that X_1 and X_2 are dependent.

Acknowledgment

The authors would like to thank S. Sénecal and G. Fort for their enthusiasm and fruitful discussions. The authors also would like to thank D.-T. Pham for pointing out some theoretical problems.

References

[1] P. Comon, "Independent component analysis, a new concept ?," *Signal Process.*, vol. 3, no. 36, pp. 287–314, Apr. 1994.
[2] G. Darmois, "Analyse des liaisons de probabilités," in *Proc. Int. Stat. Conf.*, vol. III A, Washington, DC, 1951, pp. 231–232.
[3] A. M. Kagan, Y. V. Linnik, and C. R. Rao, *Characterization Problems in Mathematical Statistics*. New York: Wiley, 1973.
[4] A. Taleb and C. Jutten, "Sources separation in post-nonlinear mixtures," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2807–2820, Oct. 1999.
[5] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "A geometric approach for separating post nonlinear mixtures," in *Proc. Int. Workshop EUSIPCO*, vol. II, Toulouse, France, Sep. 2002, pp. 11–14.
[6] F. Theis and P. Gruber, "Separability of analytic postnonlinear blind source separation with bounded sources," in *Proc. Int. Workshop of ESANN*, Evere, Belgium, 2004, pp. 217–222.
[7] H. Joe, "Majorization, randomness and dependence for multivariate distributions," *Ann. Probabil.*, vol. 15, no. 3, pp. 1217–1225, 1987.
[8] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.
[9] G. Darmois, "Analyse générale des liaisons stochastiques," *Rev. Inst. Int. Stat.*, vol. 21, pp. 2–8, 1953.
[10] ——, "Sur certaines formes de liaisons de probabilité," in *Proc. Colloques Int. Centre Nat. Rech. Sci.*, vol. 13, Lyon, France, Jun. 1948, pp. 19–21.