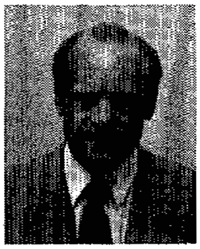
[53] Y. Nubar and R. Contini, "A minimal principle in biomechanics," *Bull. Math. Biophys.*, vol. 23, pp. 337–391, 1961.
[54] G. L. Smidt, "Biomechanical analysis of knee flexion and extension," *J. Biomechan.*, vol. 6, pp. 79–92, 1973.
[55] W. F. H. M. Mommaerts, "Muscular contraction," *Physiol. Rev.*, vol. 49, pp. 427–508, 1969.
[56] J. E. Cotes and F. Meade, "The energy expenditure and mechanical energy demand in walking," *Ergonomics*, vol. 3, pp. 97–119, 1960.
[57] H. J. Ralston, "Energy-speed relation and optimal speed during level walking," *Int. Z. Ang. Physiol.*, vol. 17, pp. 277–283, 1958.
[58] A. Cappozzo and A. Pedotti, "Modeling biological systems by optimization methods," in *Advances in Cybernetics and Systems*, vol. 3, J. Rose, Ed. New York: Gordon & Breach, 1974, pp. 1307–1318.
[59] W. N. Stainsby, "Oxygen uptake for negative work, stretching contractions by in situ dog skeletal muscle," *Amer. J. Physiol.*, vol. 230, pp. 1013–1017, 1976.
[60] A. Seireg and R. J. Arvikar, "The prediction of muscular load sharing and joint forces in the lower extremities during walking," *J. Biomechan.*, vol. 8, pp. 89–102, 1975.
[61] A. Pedotti, V. V. Krishnan, and L. Stark, "Optimization of muscle-force sequencing in human locomotion," *Math. Biosci.*, vol. 38, pp. 57–76, 1978.
[62] R. Tomović and R. Bellman, "A systems approach to muscle control," *Math. Biosci.*, vol. 8, pp. 265–277, 1970.
[63] H. Hatze, "Optimal processes of neuro-musculoskeletal control systems," in *Mathematical Modelling in Biology and Ecology*, W. M. Getz, Ed. New York: Springer, 1980, pp. 23–39.
[64] J. H. Milsum, "Control system aspects of muscular co-ordination," in *Medicine and Sport, vol.* 6: *Biomechanics II*. Basel: Karger, 1971, pp. 62–71.
[65] E. J. L. Noldus, "Optimal control structures in biological system modeling," *Quart. J. Automat. Contr.*, vol. 18, pp. 18–23, 1977.
[66] G. A. Bekey and S. M. Yamashiro, "Parameter estimation in mathematical models of biological systems," *Advanced Biomed. Eng.*, vol. 6, pp. 1–43, 1976.
[67] H. R. Weed, "The need of control concepts in bio-medical engineering," in *Proc. 7th Triennial World Congr., IFAC*, vol. 3, Helsinki, 1978, pp. 2215–2216.
[68] M. A. Arbib, *Brains, Machines and Mathematics*. New York: McGraw-Hill, 1964.
[69] H. Hatze, "An algorithm for computing higher-order derivatives of noisy experimental data sequences," CSIR Tech. Rep. TWISK 123, Pretoria, Nov. 1979.
[70] ——, *Myocybernetic Control Models of Skeletal Muscle—Characteristics and Applications*. Pretoria: Univ. South Africa Press, 1980.



Herbert Hatze was born in Vienna, Austria, on July 29, 1937. He received the electrical engineering diploma (summa cum laude) from the Technical College TGM, Vienna, Austria, in 1957, the diploma in Kinesiology (major physiology) from the Bundesanstalt fur Leibeserziehung, University of Vienna, in 1961, and the B.S. degree in mathematics and applied mathematics (cum laude), the Hons. -B.S. degree (summa cum laude), and the Ph.D. degree in applied mathematics, all from the University of South Africa, Pretoria, in 1970, 1972, and 1975, respectively.

He was a lecturer in technical mathematics and biomechanics at the Technical College Steyr, Austria, from 1962–1967, with extensive research activities in the biomechanical field. From 1968–1971, he was a Research Associate at the Human Sciences Laboratory of the Chamber of Mines of South Africa, Johannesburg. Subsequently (1971–1975) he was a lecturer in biomechanics and biostatistics at the University of Stellenbosch, South Africa, and since 1975 he has been Senior Chief Research Scientist in Biomathematics/Biomechanics at the National Research Institute for Mathematical Sciences of the CSIR, Pretoria. During the year 1976, he was Visiting Professor at the Department of Kinesiology University of Waterloo, Waterloo, Ont., Canada.

Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications

ALBERT BENVENISTE, MAURICE GOURSAT, AND GABRIEL RUGET

***Abstract*—Consider an unknown linear time-invariant system without control, driven by a white noise with known distribution. We are interested in the identification of this system, observing only the output. This problem is well known under the major assumption: the system is minimum (or maximum!) phase, in which the very popular least squares method gives an identification of the system in an autoregressive form. However, we are interested in the case where the system is nonminimum (nor maximum!) phase, i.e., we want identification of both gain and phase of the system. The literature gives only a negative result: the identification of the phase of the system is impossible in the case of a Gaussian driving noise (hence, second-order statistics are irrelevant to our problem). For a large class of other input distributions, we present an identification procedure, and give some numerical results for a concrete case origin of our study: the blind adjustment of a transversal equalizer without any startup period prior to data transmission.**

Manuscript received March 23, 1978; revised May 7, 1979 and December 10, 1979. Paper recommended by P. E. Caines, Past Chairman of the Identification Committee.
A. Benveniste is with IRISA, Université de Rennes-Beaulieu, Rennes Cedex, France.
M. Goursat is with IRIA LABORIA, Le Chesnay, France.
G. Ruget is with the Université de Paris-Sud Orsay, Orsay Cedex, France.

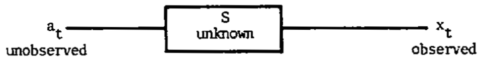
I. Introduction

The following problem arises in data communication when a receiver is obliged to achieve a blind starting phase, i.e., without the transmission by the transmitter of

0018-9286/80/0600-0385$00.75

a known (*a priori*) sequence of data (for detailed presentation, see [3], [4], [9], [10], and Section VI of this paper).

We observe the output $(x_t)_{t\in Z}$ of a time-invariant *unknown* linear system S with input $(a_t)_{t\in Z}$, (a_t) being an *unobserved* i.i.d. (independent identically distributed) random sequence with known distribution ν.

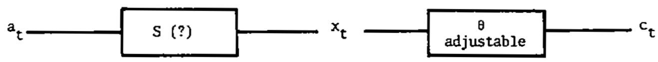


(1.1)

The problem is to restore (a_t) or, equivalently, to identify the inverse S^{-1} of S.

We neglect here a possible perturbation on the observations, but it is not a worry for two reasons: we do not need a precise identification, but only a good initial value for classical algorithms, and in data transmission we use discrete variables and the estimates will not be affected by small perturbations.

This problem is classical if S is a minimum phase system (stable with stable inverse): in this case, (a_t) is just the innovation of the sequence (x_t) (see, for example, Ljung [6]). In our case, S is a telephone channel which is generally *a nonminimum phase system*. We know that the problem of the identification of S cannot be solved when (a_t) is a Gaussian sequence; hence, in the non-Gaussian case, we cannot work with the second-order statistics (for example, by minimizing a mean-square error). Because of the instability of S^{-1}, the on-line restoration of (a_t) is impossible; we will then achieve this restoration with a delay, using the following scheme:



(1.2)

where the adjustable system θ (*equalizer*) has to be chosen such that roughly $c_t = a_{t-N}$ where N is a certain (generally unknown, but constant) delay. The method is the following.

1) We construct a functional $\mathcal{J}(\theta)$ which is minimum at the point $\theta^* = S^{-1}$; in fact, $\mathcal{J}$ is a measure of how modified is the (one-dimensional) distribution of an i.i.d. sequence when filtered by an all-pass transfer function (and we have no modification in the Gaussian case!). A particular form is close to the mutual information for the two distributions of the input (a_t) and the output (c_t) of the adjustable system.

2) We minimize $\mathcal{J}$ by using a stochastic gradient procedure. This latter point is justified by a general result on the convergence of stochastic approximation procedures with discontinuities and constraints (given in [2]); this result is needed because $\mathcal{J}$ is not smooth.

The numerical results we give show that this method has been successful for the blind adjustment of a transversal equalizer in data transmission.

The paper is organized as follows. In Section II we give a precise statement of the problem, investigate what are the special difficulties encountered in the nonminimum phase case, and describe the algorithm. Section III (here are the main theoretical results of the paper) is devoted to the analysis of a class of suited functionals $\mathcal{J}$; this analysis consists of the study of certain differential equations on the unit sphere of the infinite dimensional space $\mathbb{R}^Z$ because the inverse of a nonminimum phase system can never be exactly described by finitely many parameters (see Section II). Nevertheless, in practical cases, we shall achieve approximate identification of S^{-1} with a finite number of parameters: Section IV is devoted to the analysis of this truncation. In Section V, the reader is referred to [2] for the suited convergence result on the stochastic approximation scheme we used here. Section VI gives numerical results in the area of equalization theory.

II. Setting the Problem: Describing the Method and the Algorithm

Remark 1: Let us begin with some transfer function considerations. Let, for instance, the transfer function $S(z^{-1})$[1] be of the following form:

$$S(z^{-1}) = g\frac{P_1(z^{-1})P_2(z^{-1})}{Q(z^{-1})} \tag{2.1}$$

where g is a gain, Q and P_1 are stable monic polynomials (i.e., with zeros outside the unit circle), and P_2 is an unstable monic polynomial (with zeros inside the unit circle); because S is nonminimum phase, P_2 is not a constant. Let us expand $1/P_1$ and $1/P_2$ in the following Laurent series:

$$1/P_1(z^{-1}) = \sum_{k>0} \alpha_k z^{-k}, \quad 1/P_2(z^{-1}) = \sum_{k<0} \beta_k z^{-k},$$

both convergent outside the unit circle; then we obtain

$$1/S(z^{-1}) = 1/g\cdot Q(z^{-1})\left(\sum_{k>0}\alpha_k z^{-k}\right)\left(\sum_{k<0}\beta_k z^{-k}\right); \tag{2.2}$$

the difference with the minimum phase case lies in the fact that, because P_2 is never a constant, the last infinite series in the expression of $1/S(z^{-1})$ never disappears, so that from a theoretical viewpoint, 1) *the inverse system S^{-1} can never be described by finitely many parameters*, and 2) it is "infinitely" noncausal in the sense that restoration of the signal a_{t_0} requires knowledge of the whole sequence $(x_t)_{t\in Z}$. Consequently, *the exact inverse of S is only defined up to a time shift*. In fact, but not in the theoretical analysis, we shall truncate the Laurent series in (2.2) in order to obtain a realizable (with a delay N) approximation of $1/S(z^{-1})$ in a moving average form:

$$1/S(z^{-1}) = 1/g\cdot Q(z^{-1})\left(\sum_{0}^{N}\alpha_k z^{-k}\right)\left(\sum_{0}^{-N}\beta_k z^{-k}\right). \quad \blacksquare \tag{2.3}$$

Hence, for the theoretical analysis, we shall start directly with a description of S with infinitely many param-

[1] $z^{-1}\cdot x_t = x_{t-1}$.

eters; furthermore, we avoid the assumption of causalness, which is irrelevant in view of Remark 1. Let $(a_t)_{t\in Z}$ be a stationary sequence of i.i.d. random variables with distribution ν; the observed sequence is then given by

$$x_t=\sum_{k\in Z} s_k a_{t-k}, \quad S=(s_k)_{k\in Z}. \tag{2.4}$$

The basic assumptions on S and ν are the following:

i) the distribution ν is symmetric with finite variance,
ii) S has a finite energy, and so has S^{-1}, i.e., if we denote by

$$\Lambda=(\Lambda_{ij})_{i,j\in Z} \qquad \text{with } \Lambda_{ij}=\sum_{k\in Z} s_k s_{k+i-j}$$

the covariance operator of (x_t), then Λ is bounded and positive definite in the space l^2 where $l^2=\{(\xi_n)_{n\in Z}\colon \Sigma\xi_n^2<+\infty\}$.

(2.5)

The problem is: *we know the distribution ν of the input, we observe the sequence (x_t), and we want to restore the transmitted (and unknown) sequence (a_t) of the data; in other words, we want to identify the inverse system S^{-1}* [which exists in view of assumption (2.4)]. *We have here what is often called a "blind deconvolution problem."*

Remark 2: Suppose for a moment that we have a minimum phase system S, i.e., both S and S^{-1} are causal and stable. Our identification problem then becomes a very simple subcase of a well-known problem (see Ljung [6, condition 6]) and the transmitted sequence (a_t) is the innovation of (x_t) given by

$$a_t=\lambda(x_t-\mathbb{E}(x_t|x_{t-1},x_{t-2},\cdots)) \tag{2.6}$$

where $\mathbb{E}(\cdot|\cdot)$ is the least squares estimation operator and the constant λ adjusts the variance. In this case, it is only a question of whitening the output (x_t), which is well known and done by using only the second-order statistics. In our case (where S is nonminimum phase), with the aid of the second-order statistics, we can identify the amplitude spectrum (i.e., the gain) of S, but the identification of S is impossible. Second-order techniques will give us a factorization of the spectrum of the output, but not the factorization giving the sequence (a_t). To summarize, we can note that

1) *we cannot solve our problem with the second-order statistics (for example, by minimizing a mean-square error),*

2) *the problem has no solution when ν is a Gaussian distribution.*

Therefore, we shall assume from now on that ν is not Gaussian, and use this property with more than the second-order statistics.

Remark 3: Furthermore, note that, because the distribution ν is symmetric, the whole sequence $(-a_t)$ has the same law as (a_t), so we cannot distinguish the desired system $S^{-1}=(\hat{s}_k)$ from the opposite one $-S^{-1}=(-\hat{s}_k)$ (we will see in Section III-G how to remove this drawback in the applications).

In view of the preceding remarks, the best we can achieve is to solve the following:

Problem: Construct a system θ (the equalizer) so that the global system $T=\theta\circ S$ is $\pm$ identity, except for a possible delay (i.e., $T=(t_k)$ with $|t_{k_0}|=1$ for some index k_0 and $t_k=0$ otherwise).

(2.7)

Now we start with characterizing the solutions θ for (2.7). In order to do that, we give a lemma (perhaps classical), which characterizes what are the distributions for i.i.d. sequences that are not modified when filtered through a system with energy 1.

Lemma 2.1: Consider a sequence $(a_t)_{t\in Z}$ of i.i.d. random variables with distribution ν, ν being symmetric with finite variance. Assume that there is a sequence $T=(t_k)_{k\in Z}$ with at least two nonzero terms such that $\Sigma_{k\in Z}t_k^2=1$, and the distribution of the random variable

$$c=\sum_{k\in Z} t_k a_{-k}$$

is still ν. Then ν is a Gaussian distribution.

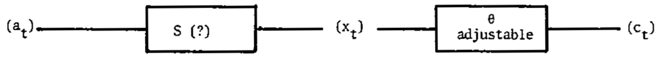
Proof: See [4].

A corollary of this result is the following characterization of the solution θ for the problem (2.7).

Theorem 2.2: Consider a system $\theta=(h_k)_{k\in Z}$ such that the distribution of the random variable $c=\Sigma_{k\in Z}h_k x_{-k}$ is still ν. Denote by $T=\theta\circ S$ the global system, and assume that the distribution ν is non-Gaussian. Then $T=\pm$ identity except for a possible delay [in the sense of (2.7)].

Proof: Consider the global system $T=(t_k)$ with $t_k=\Sigma_{p\in Z}h_p s_{k-p}$. The distribution of c being ν, we have $\mathbb{E}(c^2)=\mathbb{E}(a_t^2)$, which gives, using the independence of the random variables a_t, $\Sigma_{k\in Z}t_k^2=1$. Lemma 2.1 applied to T gives the result. ■

The previous theorem shows that, to obtain the solution, we have to adjust the tap weights of the equalizer θ in such a way that the instantaneous distribution of the output c_t of the equalizer converges to the input distribution ν.



We can now explain the method for adjusting the coefficients of the equalizer.

Step 1: Definition of the Functional to be Minimized: In a classical way, we get identification algorithms by the minimization of a mean-square error; however, this method does not work here, so we have to construct a special functional. Theorem 2.1 leads us for this point to use the output of the equalizer; so we will try to realize

$$\min_{\theta\in l^2} \mathcal{J}(\theta) \qquad \text{where } \mathcal{J}(\theta)=\mathbb{E}(\Psi(c_t)),\ c_t=\sum_{k\in Z} h_k x_{t-k} \tag{2.8}$$

where Ψ is an even function (ν is symmetric) $\mathbb{R}\to\mathbb{R}$ to be chosen such that the unique local (and hence, global) minima of $\mathcal{J}$ are $\pm S^{-1}$, except for a possible delay. This is the most difficult and interesting part of the work; this will be done in Section III where we define a class of functions Ψ corresponding to a large class of distributions ν.

Step 2: Justification of the Truncating: In the applications, the minimization is done with finite dimensional equalizers, and (2.8) becomes

$$\min_{\theta=(h_k)_{-N<k<N}} \mathcal{J}(\theta), \qquad \mathcal{J}(\theta)=\mathbb{E}(\Psi(c_t)),\ c_t=\sum_{k=-N}^{+N} h_k x_{t-k}; \tag{2.9}$$

we show in Section IV that the minima we obtain with (2.9) are close to the desired ones.

Step 3: Minimization of $\mathcal{J}$ by a Stochastic Gradient Procedure: Let ψ be the derivative of Ψ; with (2.9) we get formally

$$\operatorname{grad}\mathcal{J}(\theta)=\mathbb{E}(X_t\psi(c_t)) \tag{2.10}$$

where ${}^tX_t=(x_{t+N},\cdots,x_t,\cdots,x_{t-N})$ is the random vector giving c_t as in (2.9). For minimizing $\mathcal{J}$, we use a stochastic gradient procedure, the study of which is done in Section V, referring to a general suitable convergence result (given in [2]) for that sort of stochastic approximation algorithm; the result gives us a rough convergence of θ^t to a local minimum of $\mathcal{J}$, with the following algorithm:

$$\begin{cases} \theta^{t+1}=\theta^t-\tau X_t\psi(c_t); & \theta^t=\left(h_k^t\right)_{-N<k<N},\ \theta^0 \text{ given} \\ c_t=\displaystyle\sum_{k=-N}^{+N} h_k^t x_{t-k}, & {}^tX_t=(x_{t+N},\cdots,x_t,\cdots,x_{t-N}) \end{cases} \tag{2.11}$$

and τ is a small parameter (step of the gradient algorithm) to be chosen. The convergence to S^{-1} or $-S^{-1}$ depends on the initial value θ^0 for (2.11). The "good" domain for θ^0 may be roughly known with the hazy information we have in the applications (see Section III-G). After the convergence of (2.11), we will have

$$c_t=a_{t-n_0}, \qquad n_0=\text{identification delay}. \tag{2.12}$$

III. Analysis of the Functional $\mathcal{J}$: Measurement of the Distortion of the Distribution of an I.I.D. Sequence Filtered by a Linear System with Energy 1

A. Notations

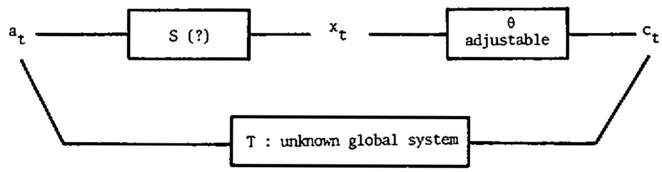
Let us denote by

$$A=(a_{-t})_{t\in Z},\quad X=(x_{-t})_{t\in Z},\quad C=(c_{-t})_{t\in Z} \tag{3.1}$$

the whole sequences with reversed time. Using the convolution product notation $*$, let us denote by

$$T=\theta*S,\quad t_k=\sum_{p\in Z} h_p s_{k-p} \tag{3.2}$$

the global system



Let l^2 denote the Hilbert space $\mathbb{R}^Z$ provided with the inner product $\langle\cdot,\cdot\rangle$ and the associated norm $\|\cdot\|$ defined by

$$\begin{aligned} l^2&=\left\{(s_k)_{k\in Z}:\ \Sigma s_k^2<+\infty\right\}; \\ \langle R,S\rangle&=\Sigma r_k s_k; \\ \|R\|^2&=\langle R,R\rangle \end{aligned} \tag{3.3}$$

and let s^2 be the unit sphere of l^2.

B. Using the Global System

Of course, T defined in (3.2) need not *a priori* to have finite energy (i.e., belong to l^2). But, with assumption (2.5, ii), we obtain that

$$T=\theta*S \text{ belongs to } l^2 \text{ if and only if } \theta \text{ belongs to } l^2. \tag{3.4}$$

We obtain (3.4) thanks to assumption ii) of (2.2) which gives us

$$\alpha_1\|\theta\|^2\leqslant\langle\theta,\Lambda\theta\rangle=\|T\|^2\leqslant\alpha_2\|\theta\|^2 \tag{3.5}$$

where the α_i's are finite positive constants and the equality is given by

$$\langle\theta,\Lambda\theta\rangle=\langle\theta,\check{S}*S*\theta\rangle=\langle\theta*S,S*\theta\rangle=\|T\|^2 \tag{3.6}$$

with

$$\check{S}=(s_{-k})_{k\in Z}. \tag{3.7}$$

Let us introduce the functional $\mathcal{V}$ on l^2 defined by

$$\mathcal{V}(T)=\mathbb{E}(\Psi(c)) \qquad \text{where } c=\sum_{k\in Z} t_k a_{-k}.^2 \tag{3.8}$$

Hence, we have

$$\mathcal{V}(T)=\mathcal{J}(\theta) \qquad \text{if } T=\theta*S \text{ is the global system given by } \theta. \tag{3.9}$$

Instead of choosing the function Ψ in order to have the local minima of $\mathcal{J}$ at the points $\pm S^{-1}$, we will choose it for having the local minima of $\mathcal{V}$ at $T=\pm I$. The second choice works thanks to (3.9). We have here one of the basic ideas of the method because we can construct $\mathcal{V}$ for two reasons: we explicitly know the desired minima and the input distribution ν. Let us begin with the following (over)simple example.

Example: Let us take the simplest case where $\mathbb{P}(a_t=\pm 1)=1/2$ and the global system $T=(t_k)$ has energy 1 and only two nonzero coefficients, which are, hence, equal to

[2] We write c in place of c_0.

$\cos\alpha$ and $\sin\alpha$, respectively, for some $\alpha\in[0,\pi/4]$ (this suffices by a symmetry argument); let us choose $\Psi(x)=x^2-|x|$; we obtain $\mathcal{V}(T)=\mathcal{V}(\alpha)=\mathbb{E}((a_i\cos\alpha+a_j\sin\alpha)^2-|a_i\cos\alpha+a_j\sin\alpha|)=1/2((\cos\alpha+\sin\alpha)^2-(\cos\alpha+\sin\alpha))+1/2((\cos\alpha-\sin\alpha)^2-(\cos\alpha-\sin\alpha))=1-\cos\alpha$, which strictly increases when α goes from 0 to $\pi/4$. We have obtained a very simple case in which we achieved a measurement of the distortion of the distribution of the output c_t with respect to the input distribution (here Bernoulli) (see Remark 4 below). ■

Now the general approach is the following:

1) choice of the function Ψ and characterization of the admissible input distributions ν in order that the only local minima of $\mathcal{V}$ become $\pm I$ except for some delay,

2) analysis of the *steepest descent lines* (*s.d.l.*) *of* $\mathcal{V}$,

3) analysis of the steepest descent lines of $\mathcal{J}$.

We shall make use of the framework of dynamical systems theory (vector fields and one-parameter flows on smooth Hilbertian manifolds), for which we refer the reader to the Appendix I.

Let us start with the first point by introducing the two vector fields:

$$\begin{cases} V_T=-\mathbb{E}(A\psi(c)), & T\in l^2 \\ \bar{V}_T=-(\mathbb{E}(A\psi(c))-T\cdot\mathbb{E}(c\psi(c))), & T\in s^2 \end{cases} \tag{3.10}$$

with ψ the derivative of Ψ, $A=(a_{-t})$ [see (3.1)], $c=c_0=\Sigma t_k a_{-k}$, and s^2 is the unit sphere of l^2. Formally, the vector fields $(V_T)_{T\in l^2}$ and $(V_T)_{T\in s^2}$ are, respectively, the opposite of the gradient of $\mathcal{V}$ and of the gradient of the restriction of $\mathcal{V}$ to s^2. Because we deal with gradients in an infinite dimensional vector space, we give precise conditions (for the proof, see [4]) which ensure that this formal statement is indeed correct.

Assumptions

(3.11): The function ψ is odd, $\mathcal{C}^1$, and such that[3]

$$\sup_{\|T\|<K}\mathbb{E}(\psi'(\Sigma t_k a_{-k}))^2<\infty \qquad \text{for every } K<\infty$$

and ν has a fourth moment.

(3.12): $\psi=\tilde{\psi}+\psi_1$ where $\tilde{\psi}$ satisfies (3.11) and ψ_1 is piecewise constant with finitely many discontinuities; ν has a fourth moment and bounded density with respect to the Lebesgue measure.

(3.13): (3.11) is in force, and $\psi\in\mathcal{C}^3$ with

$$\sup_{\|T\|<K}\mathbb{E}\big(\psi^{(3)}(\Sigma t_k a_{-k})\big)^2<\infty \qquad \text{for every } K<\infty.$$

(3.14): (3.12) and (3.13) are in force; furthermore, $d\nu/dx$ is $\mathcal{C}^2$ with bounded second derivative. ■

Lemma 3.1: i) Under (3.11) [*respectively, (3.12)*], *the vector field* (V_T) *is locally bounded in* l^2 (*respectively,* $l^2-\{0\}$) *and is the opposite of the gradient field of* $\mathcal{V}$; $(\bar{V}_T)$ *is locally bounded and is the opposite of the gradient field of the restriction to* s^2 *of* $\mathcal{V}$.

ii) Furthermore, under (3.13) [*respectively (3.14)*], (V_T) *is locally Lipschitz in* l^2 (*respectively* $l^2-\{0\}$) *and* $(\bar{V}_T)$ *is locally Lipschitz.*

Example: We give an example where (3.11) is verified. Suppose that the second-order moments of ν exist and let $\hat{\nu}$ denote the Fourier transform of ν. We get

$$\hat{\nu}(t_k u)=1-\frac{E(a^2)}{2!}t_k^2u^2+\frac{E(a^4)}{4!}t_k^4u^4 +\cdots+(-1)^n\frac{E(a^{2n})}{(2n)!}(t_ku)^{2n}+0(t_ku)^{2n}.$$

The Fourier transform of the law of $c=\Sigma t_k a_{-k}$ with $T=(t_k)$ is

$$\hat{\mu}(u)=\prod_k\hat{\nu}(t_ku)=1-\frac{E(a^2)}{2!}\Big(\sum_k t_k^2\Big)u^2 +\left(\frac{E(a^2)^2}{2!2!}\sum_{i,j}t_i^2t_j^2+\frac{E(a^4)}{4!}\sum_k t_k^4\right)u^4 +\cdots+(-1)^nP_n(t_k)u^{2n}+0(u^{2n})$$

where $P_m(t_k)$ is a convergent series with a homogeneous $2m$-degree term with respect to the coefficients (t_k) and satisfies $|P_m(t_k)|\leqslant Q_m(\|T\|^2)$ where Q_m is an m-degree polynomial. Therefore, the moments of c with an order $\leqslant 2n$ exist and are uniformly bounded with the condition $\|T\|\leqslant K<+\infty$. It is clear that we have (3.11) if $\psi(x)$ grows at most as $|x|^n$ for $x\to\infty$. ■

Hence, under (3.13) or (2.14), the vector fields (V_T) and $(\bar{V}_T)$ define uniquely the one-parameter flow of their integral curves. Let us go back to the analysis of $\mathcal{V}$; the reason for introducing the field $(\bar{V}_T)_{T\in s^2}$ is the fact that Lemma 2.1 suggests that the functional $\mathcal{V}$ should be analyzed in spherical coordinates. From now on, assumptions (3.11)–(3.14) are in force.

C. Analysis of the Restriction of $\mathcal{V}$ *to* s^2*: Measurement of the Distortion of a Product Distribution by a Rotation*

In order to calculate the spherical partial derivatives of $\mathcal{V}$, let us fix a pair (i,j), $i\neq j$ and a fixed sequence $(t_k)_{k\neq i,j}$ such that

$$R^2=1-\sum_{k\neq i,j}t_k^2>0.$$

For $\alpha\in[0,2\pi]$, let $T_\alpha\in s^2$ denote the system with coefficients $(t_k)_{k\neq i,j}$ and $t_i^\alpha=R\cos x$, $t_j^\alpha=R\sin x$.

(3.15): Let $(\partial/\partial\alpha_{ij})\mathcal{V}(T)$ denote the derivative of $\alpha\to\mathcal{V}(T_\alpha)$ at the point $T_\alpha=T$.

We have the following formula (see Appendix II).

Lemma 3.2: Let μ *denote the distribution of the random variable* $\Sigma_{k\neq i,j}t_ka_{-k}$ *and* $\psi^\mu(x)=\int\psi(x-y)\mu(dy)$. *We have*

$$\frac{\partial}{\partial\alpha_{ij}}\mathcal{V}(T_\alpha)=2R\int_0^\infty\int_0^\infty(y\psi^\mu(Rx)-x\psi^\mu(Ry))P_\alpha(dx,dy)$$

where $P_\alpha(dx,dy)$ *is the transformation of* $\nu\times\nu$ *by a rotation with angle* α.

[3] $\psi',\psi'',\psi^{(3)}$ are successive derivatives of ψ.

In view of Lemma (3.2), we see that $(\partial/\partial\alpha_{ij})\mathcal{V}(T_\alpha)=0$ for $\alpha=k\pi/4$ $(k\in\mathbb{Z})$, using the invariance of P_α by $(x,y)\to(y,x)$ in that case. Hence, we have the following.

Corollary 3.3: If all the nonzero coefficients of $T\in s^2$ have the same module, then T is a stationary point of the restriction of $\mathcal{V}$ to s^2.

With this result, we have that the set of the systems $\pm I$ is given by stationary points of $\mathcal{V}$. It is not enough; we now have to choose ψ such that all the local (hence, global) minima of $\mathcal{V}$ will only be given by $\pm I$. In this situation, the other stationary points of $\mathcal{V}$ will be saddle points and $\mathcal{V}$ will look like an egg box.

Using the symmetries of the problem, we only consider the case $0<\alpha<\pi/4$ where we want $(\partial/\partial\alpha_{ij})\mathcal{V}(T_\alpha)>0$. This will be done for the two families of sub- and super-Gaussian distributions we define below.

Definition: i) We shall say that ν is a *sub-Gaussian* distribution if either ν is uniform on $[-d,+d]$ or if $\nu(dx)=Ke^{-g(x)}dx$ with K a constant and g an even function such that $g(x)$ and $g'(x)/x$ are strictly increasing on $\mathbb{R}_+$.

ii) We shall say that ν is *super-Gaussian* if $\nu(dx)=Ke^{-g(x)}dx$ with g an even function such that $g(x)$ is strictly increasing and $g'(x)/x$ strictly decreasing on $\mathbb{R}_+$.

Examples: $\nu(dx)=Ke^{-|x|^\gamma}dx$ is super-Gaussian for $\gamma<2$, Gaussian for $\gamma=2$, and sub-Gaussian for $\gamma>2$, whereas the limiting case $\gamma=+\infty$ gives the uniform distribution.

Lemma 3.4: Assume that ν is sub-Gaussian (respectively super-Gaussian) and ψ is odd, twice differentiable except at the origin, such that

$$\psi(0_+)\le 0,\quad \psi''\ge 0\quad \text{on } (0,\infty)$$
$$(\text{respectively } \psi(0_+)\ge 0,\quad \psi''\le 0\quad \text{on}(0,\infty)),$$

one at least of the two inequalities being strict. Then

$$\frac{\partial}{\partial\alpha_{ij}}\mathcal{V}(T_\alpha)>0 \qquad \text{for } 0<\alpha<\pi/4.$$

Corollary 3.5 (Main Theorem): Under the assumptions of Lemma 3.4, the only local (hence, global) minima of the restriction of $\mathcal{V}$ to s^2 are $\pm I$ (except for a possible delay), and they are the unique stable attractors of the flow of his steepest descent lines (s.d.l.).

The proofs of Lemma 3.4 and Theorem 3.5 are in Appendix II. Note that, because we work on infinite dimensional manifolds, knowing the minima of $\mathcal{V}$ does not ensure that the s.d.l. converge to those minima; classically, this further convergence result requires properties like "positive definiteness" of $\mathcal{V}$; here, we use a special argument.

Remark 1: Crest Lines: In view of Lemma 3.2, we see that the hyperplanes $t_i=t_j$ are the crest lines of $\mathcal{V}$. Fig. 1 shows the steepest descent lines of the restriction of $\mathcal{V}$ to s^2 in the quadrant $t_k\ge 0$ for every k.

Remark 2: Example of Function ψ: In the applications, we choose ψ as simple as possible. For example,

$$\psi(x)=x-\gamma \qquad \text{for } x>0 \text{ in the sub-Gaussian case}$$
$$\psi(x)=-x+\gamma \qquad \text{for } x>0 \text{ in the super-Gaussian case} \tag{3.16}$$

where $\gamma>0$ is arbitrary.

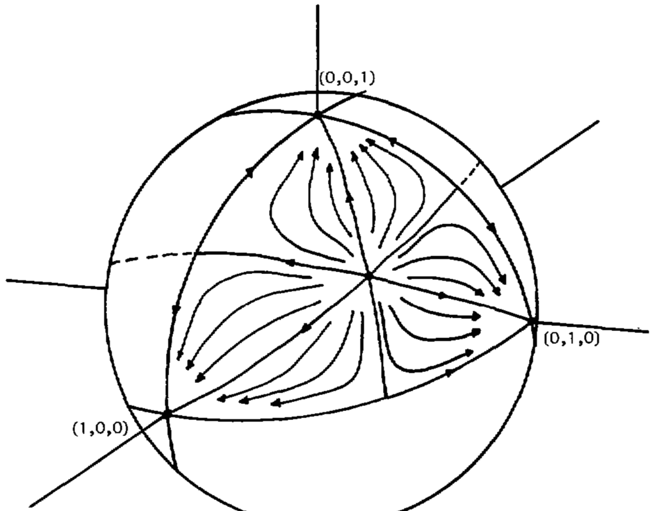


Fig. 1.

Remark 3: A Connection with the Mutual Information: Let us consider the case where $\nu(dx)=Ke^{-g(x)}dx$, and g has a third derivative which is >0 (respectively <0) on $\mathbb{R}_+$ in the sub-Gaussian (respectively super-Gaussian) case. In both cases, we can choose $\psi=g'$ and, hence,

$$\mathcal{V}(T)=\mathbb{E}(-\log f(c)),$$
$$c=\sum t_k a_{-k},$$
$$\nu(dx)=f(x)\,dx \tag{3.17}$$

is a functional which measures the distortion of the distribution of c with respect to the original ν; ν_c being the distribution of c, (3.17) becomes

$$\mathcal{V}(T)=-\int\log\frac{d\nu}{d\nu_c}\,d\nu_c-\int\log\frac{d\nu_c}{dt}\,d\nu_c, \tag{3.18}$$

i.e., $\mathcal{V}$ is the sum of the entropy of ν_c and the mutual information of ν with respect to ν_c. It can be seen that this choice is in a certain sense optimal near the minima of the restriction of $\mathcal{V}$ to s^2. However, this functional cannot measure the desired distorsion for general non-Gaussian distributions.

Remark 4: Back to the Simple Example: Note that the Bernoulli distribution used in the simple example is neither sub- nor super-Gaussian in our sense; in fact, analyzing the case of only two coefficients was oversimple. Indeed, there are other local minima than the desired ones if we look at $\mathcal{V}(T)=\mathbb{E}(c^2-|c|)$ when a_t is a Bernoulli sequence. But a central limit argument shows that when T is far from $\pm I$, then c is approximately Gaussian, which gives in that case

$$\mathcal{V}(T)=1-(2\pi)^{-1/2}\int xe^{-x^2/2}dx=1-(2/\pi)^{1/2}>0$$
$$=\mathcal{V}(\pm I)=\min_{\|T\|=1}\mathcal{V}(T)$$

so that $\mathcal{V}$ appears as well-conditioned "far" from the desired identity systems. On the other hand, the theory of large deviations indicates that insufficiently stable local minima can be ignored by stochastic approximation procedures [1]. Finally, although there is no proof for ensuring the convergence of the procedure we use in that case, the theoretical results we obtain in the sub- or super-Gaussian cases justify in a very satisfactory way the validity of the procedure when used for discrete approximations of sub- or super-Gaussian distributions. The experiments of Section VI enforce this claim, where such approximations are used.

Conclusion about Section III-C: In Section II, we have indicated that Gaussian distributions ν are characterized as being the only distributions such that $\nu \times \nu$ is invariant under a rotation (Lemma 2.1 generalizes this point). For the sub- or super-Gaussian distributions, we have here given a functional which can measure how $\nu \times \nu$ was modified through the action of a rotation; note that this functional is not the mutual information (which is a classical measure of the distance between two distributions), although it is, in some cases, connected with it, as was indicated in Remark 3.

D. Study of the Function $\mathcal{V}$ in l^2

In order to use the previous results, this study is done with spherical coordinates:

$$T \in l^2 - \{0\} \leftrightarrow (\rho, \overline{T}), \ \rho > 0, \ \overline{T} \in s^2, \text{ and } T = \rho \overline{T}.$$

For the spherical partial derivatives, Lemma 3.2 still gives the formula for $(\partial/\partial\alpha_{ij})\mathcal{V}(\rho\overline{T}_\alpha)$ with $R^2 = \rho^2 - \Sigma_{k \neq i,j} t_k^2$, and we can apply Lemma 3.4. Thus, we have only to study the radial derivatives of $\mathcal{V}$. We always take the assumptions (3.13) or (3.14) for the integral curves $(T_s)_{s \in \mathbf{R}}$ to exist in $l^2 - \{0\}$ (see Appendix II). We then have the following result.

Theorem 3.6: Under the assumptions of Lemma 3.4 for the sub-Gaussian case with $\psi(x) = -\gamma\, sign(x) + \tilde{\psi}(x)$, if the pair $(\gamma, \tilde{\psi})$ is such that

$$\int x\tilde{\psi}(x)\nu(dx) = \gamma \int |x| \nu(dx), \tag{3.17}$$

then the local minima of $\mathcal{V}$ in l^2 are given and only given by the systems $\pm I$, and these systems are the only stable attractive points of the flow $(T_s)_{s \in \mathbf{R}}$. On the other hand, there is no possible choice for ψ in the super-Gaussian case.

Proof: See Appendix II.

Fig. 2 shows the s.d.l. of $\mathcal{V}$ on l^2 for two coordinates. The radial component of (V_T) is entering for $\|T\| > \rho_c$ and exiting for $\|T\| < \rho_c$.

Remark 5: Example of Function ψ: For illustrating Theorem 3.6, we give here the simplest choice of ψ we use in data communications:

$$\psi(x) = x - \gamma \qquad \text{for } x > 0 \quad \text{with } \gamma = \frac{\mathbb{E}(a_t^2)}{\mathbb{E}|a_t|}. \tag{3.18}$$

Taking again algorithm (2.11), we obtain the following identification procedure:

$$\theta^{t+1} = \theta^t - \tau X_t(c_t - \gamma \operatorname{sign}(c_t)), \tag{3.19}$$

which is exactly the algorithm proposed and experimented by Sato [11], for which we give here a complete proof.

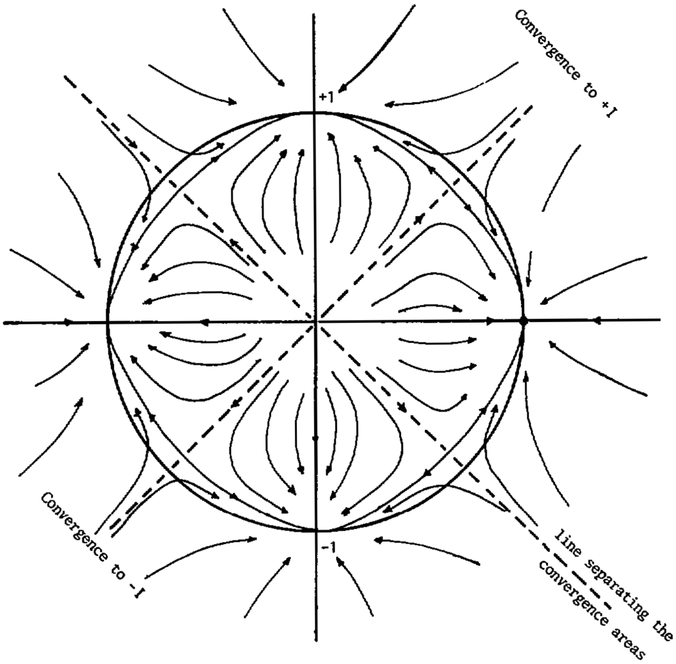


Fig. 2.

E. Study of the Steepest Descent Lines of $\mathcal{J}$

We achieve now the study of $\mathcal{J}$ which is the base of our algorithm. At this point, it is clear from (3.4) and (3.9) that all the local (hence, global) minima of $\mathcal{J}$ are (only) given by $\pm S^{-1}$ (except for a possible delay) and $\mathcal{J}$ keeps, with a distortion, the shape of an egg box. It remains to give precisely the behavior of the s.d.l. of $\mathcal{J}$.

Theorem 3.7: Let ψ and ν be as in Theorem 3.6. Then the flow $(\theta_s)_{s \in \mathbf{R}}$ of the s.d.l. of $\mathcal{J}$ exists, and its stable limit points are the systems $\pm S^{-1}$ (up to a time shift); moreover, the set of the crest lines of $\mathcal{J}$ (trajectories θ_s that converge to a saddle point of $\mathcal{J}$) is a countable union of manifolds with codimension 1, and hence is an exceptional set.

Proof: See Appendix II.

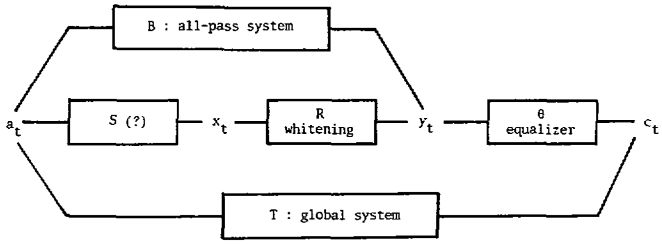
Let us denote by $(T_s)_{s \in \mathbf{R}} = (\theta_s * S)_{s \in \mathbf{R}}$ the trajectories followed by the global system when the equalizer follows (θ_s). It is seen in the proof of Theorem 3.7 that (T_s) is the flow of the s.d.l. of $\mathcal{V}$ when *l^2 is provided with the metric associated with* Λ^{-1} [see (2.5)]:

$$\|T\|^2_{\Lambda^{-1}} = \langle T, \Lambda^{-1} T \rangle.$$

When Λ is ill-conditioned (least eigenvalue $\ll$ greatest eigenvalue), this causes a loss in the efficiency of the algorithm that can be removed by using a conjugate gradient procedure. We describe now an example of such a procedure, which has the further advantage of permitting an identification in both the sub- *and* super-Gaussian case.

F. Using a Whitening Filter: An Algorithm with Constraints

Let us insert between S and the equalizer θ an arbitrary (but fixed) whitening filter R, obtaining the following identification scheme:



where R is such that (y_t) is a white noise in the second-order sense $(\mathbb{E}(y_t y_s) = \delta_{t-s} \cdot \mathbb{E}(a_t^2))$; note that (y_t) is *not*, in general, an i.i.d. sequence. Such a system exists, but it is not unique; it is realized, for example, with (2.6). It is straightforward that, since the system B is an all-pass transfer function, the convolution by B is an isometry of l^2: $\|\theta\| = \|T\|$ if $T = \theta * B$. Now, when (θ_s) follows an s.d.l. of $\mathcal{J}$, the global system (T_s) follows an s.d.l. of $\mathcal{V}$, l^2 being provided with the metric given by the inverse of the covariance of (y_t), i.e., *the usual metric* (the minima of $\mathcal{J}$ being $\pm B^{-1}$).

Moreover, if (θ_s) follows an s.d.l. of $\mathcal{J}$ in s^2, then (T_s) follows an s.d.l. of the restriction of $\mathcal{V}$ to s^2 (provided with the usual metric). Descending along the s.d.l. of the restriction of $\mathcal{J}$ to s^2 is realized by a stochastic gradient algorithm with the constraint $\|\theta\| = 1$; after truncating, the algorithm is similar to (2.11):

$$\begin{cases} \theta^{t+1} = \lambda_t\big(\theta^t - Y_t \psi(c_t)\big), & \theta^t = \big(h_k^t\big)_{-N<k<N} \\ c_t = \sum\limits_{-N}^{N} h_k^t y_{t-k}, & Y_t = (y_{t-N}, \cdots, y_t, \cdots, y_{t-N}) \\ \lambda_t \in \mathbb{R} \text{ is chosen such that } \|\theta^{t+1}\| = 1. \end{cases} \tag{3.20}$$

Since $1 - \lambda_t = 0(\tau^2)$, (3.20) is the stochastic approximation algorithm with a constraint associated with the gradient field of the restriction of $\mathcal{J}$ to s^2, which is equal to

$$\bar{V}_\theta = -(\mathbb{E}(Y\psi(c)) - \theta \cdot \mathbb{E}(c\psi(c))), \qquad \theta \in s^2$$

(see [2]). The experiments confirm that (3.20) may be more efficient.

Remark 6: With this identification scheme, our algorithm appears as a pure phase-recovering procedure.

Remark 7: Because of Theorem (3.5), this procedure allows an identification even in the super-Gaussian case.

G. How to Converge to S^{-1} Rather than to $-S^{-1}$

We only give qualitative indications. The problem is to be able to choose a good initial value for the algorithms. For (3.20), the partition lines for the evolution of (T_t) are the crest lines of $\mathcal{V}$ for the usual metric (given in Remark 1); hence, convergence to S^{-1} is obtained for a θ^0 such that $T^0 = \theta^0 * B$ is in the half plane $(t_{\max} > 0)$ where $t_{\max}$ is the coefficient of T^0 with maximum modulus (see Fig. 1).

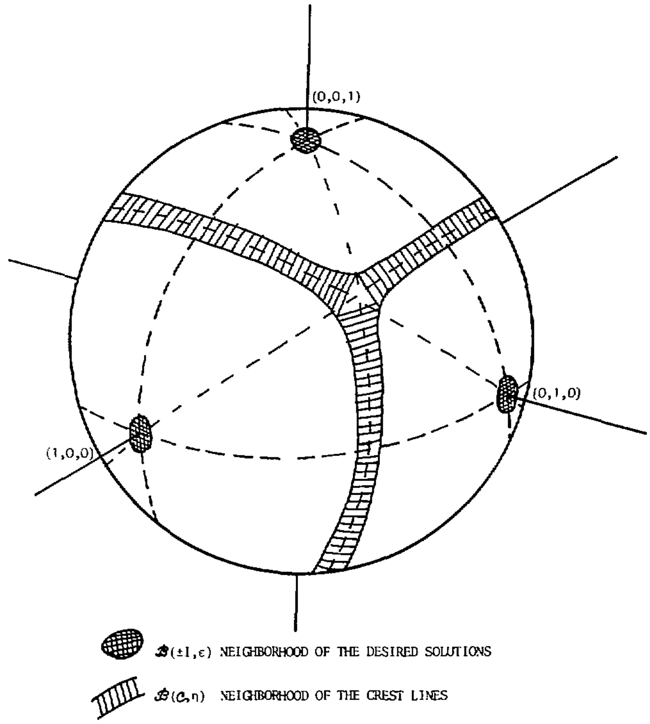


Fig. 3.

Of course, B being unknown, we cannot ensure such a condition. This may be done with a rough indication on B: in our application, we can assume that the greatest coefficient of B is positive; so we take $\theta^0 = +I$, and $T^0 = B$ satisfies the condition.

The same information on S is sufficient for our application when we use (2.11).

IV. Analysis of the Truncating

We investigate now the effect of using (2.9) in place of (2.8), with the same question for the procedure with constraint. Precise statements with corresponding proofs are available only in [4] because they are rather complicated. Here, we give only: 1) heuristic comments on what those results signify, and 2) the appropriate (and unusual) Lyapunov function we use in order to measure the swiftness of the convergence of T_t to $\pm I$.

A. Heuristic Comments

We refer the reader to [4] for the further conditions we require on ψ, ν, etc.

The Algorithm with Constraint: Fig. 3 shows the effect of the truncating: *all the local minima of the restriction of* $\mathcal{V}$ *to* $s^2 \cap \mathcal{T}^N$ ($\mathcal{T}^N$ is the hyperplane defined by the truncating: $\mathcal{T}^N = \{T = \theta * B$: θ has only nonzero coefficients h_k for those $|k| \leqslant N\}$) *lie either in a neighborhood* $\mathcal{B}(\pm I, \epsilon)$ *of the desired solutions or in a neighborhood* $\mathcal{B}(\mathcal{C}, \eta)$ *of the crest lines*. Of course, the minima in $\mathcal{B}(\mathcal{C}, \eta)$, if they exist, are not very stable, so that they are ignored in practice by the stochastic approximation procedure.

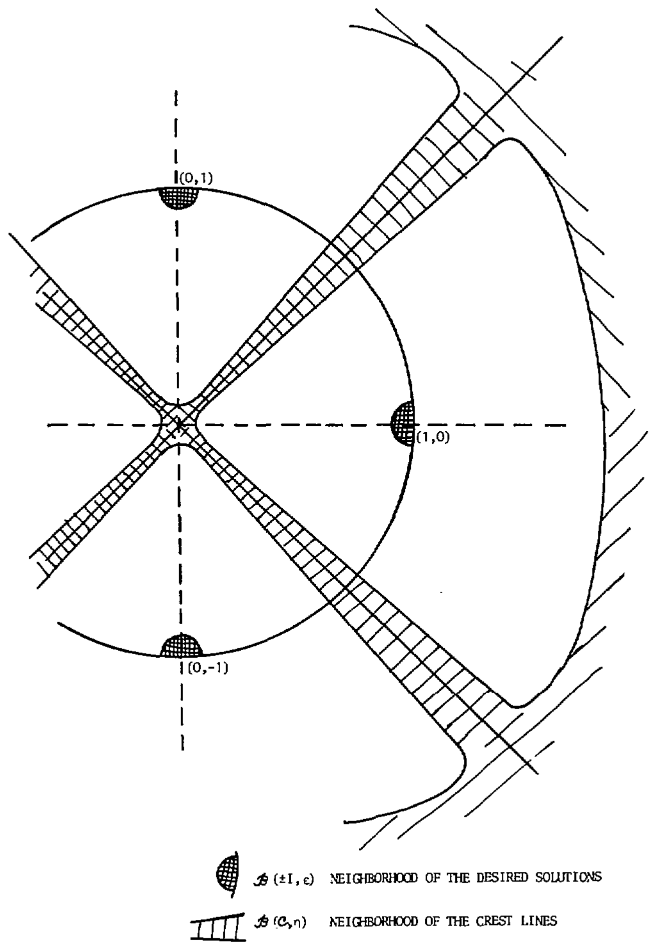


Fig. 4.

It is seen that both ϵ and η depend on the measure of the B-tail in each attraction domain of a $\pm I$ system, i.e.,

$$B(N,i)=\left(\sum_{|k-i|>N} b_k^2\right)^{1/2} \tag{4.1}$$

where the corresponding identity system is, up to a sign change, a pure delay of magnitude i. Note that $B(N,i)$ depends on N (obviously) *and* on i; hence, for fixed N, we have to choose the delay i we will have so that $B(N,i)$ is minimized and initialize the procedure with a 1 at this place in the equalizer. In practice, this is equivalent to having a good idea of the location of the greatest coefficient in the truncated estimate of B^{-1}. For precise statements, see [4].

The Algorithm Without Constraints: We give in Fig. 4 the corresponding results for the location of the minima of the restriction of $\mathcal{V}$ to $\mathcal{T}^N=\{T=\theta * S,\ \theta$ has only nonzero coefficients h_k for $|k|\leqslant N\}$. Here, ϵ and η depend on the S^{-1}-tail in the ith attraction domain of a $\pm I$ system, i.e.,

$$S(N,i)=\left(\sum_n\left(\sum_{|k-i|<N}\hat{s}_k s_{n-k}\right)^2\right)^{1/2}, \qquad S^{-1}=(\hat{s}_k) \tag{4.2}$$

where S^{-1} is here the inverse of S without delay: $(S^{-1}*S)(i)=1$ for $i=0$, $=0$ otherwise; if we denote by $S_{N,i}^{-1}$ the inverse truncated by a window of length $2N+1$ centered at i, then (4.2) reduces to

$$S(N,i)=\|S*(S^{-1}-S_{N,i}^{-1})\|. \tag{4.3}$$

[Note that when S is a unit gain, i.e., an all-pass system, then (4.2) reduces to the form in (4.1).] Once more, a good idea of the location of the greatest coefficient in a truncated estimated of S^{-1} (i.e., an idea of the delay) will increase the accuracy of the estimate when N is fixed.

B. An Appropriate Lyapunov Function

Standard Lyapunov functions like $F(s)=\|T_s-I\|_{P^{-1}}^2$ where P^{-1} is some positive definite operator are irrelevant here. We have to use Lyapunov functions relevant to the analysis in spherical coordinates we have pursued in Section III.

For the algorithm with constraint, we work with

$$\bar{\delta}(s)=\|T_s-\operatorname{sign}(t_i)\cdot I_i\|^2, \qquad s\geqslant 0,\ T_s\in s^2 \tag{4.4}$$

where $T_0=(t_k)_{k\in Z}$, $|t_i|=\max_{k\in Z}|t_k|$, and I_i is an i-delay system (with 1 at the ith coefficient and 0 elsewhere).

For the algorithm without constraints, we analyze separately the motion on the radius and the motion of the radius, the latter for which we use

$$\delta(s)=\|\overline{T}_s-\operatorname{sign}(t_i)\cdot I_i\|^2, \quad \overline{T}_s=T_s/\|T_s\|\in s^2 \tag{4.5}$$

with the same meaning for t_i and I_i as above. For the motion on the radius, we analyze with further details the curve $\rho=\rho_c$ (see Fig. 2) where the vector has no radial component.

V. Analysis of the Stochastic Approximation Schemes (2.11) and (3.20)

This analysis is based on a general convergence result for stochastic approximation procedures for which we refer the reader to [2] where the particular example we have here is analyzed in Section III.

VI. Numerical Results

A. Origin of the Problem, Preliminary Remarks

The situation is (roughly speaking) the following: an emitter transmits a sequence (a_t) of data through a channel $S=(s_k)$; the receiver observes the output (x_t) given by

$$x_t=\sum_{k=-N}^{+N} s_k a_{t-k}. \tag{6.1}$$

In fact, S includes the channel, together with some additional filters at the emitter and the receiver, so that $S=(s_k)$ represents the global impulse response sampled at rate Δt. The unknown data are viewed as random variables, independent and uniformly distributed over some

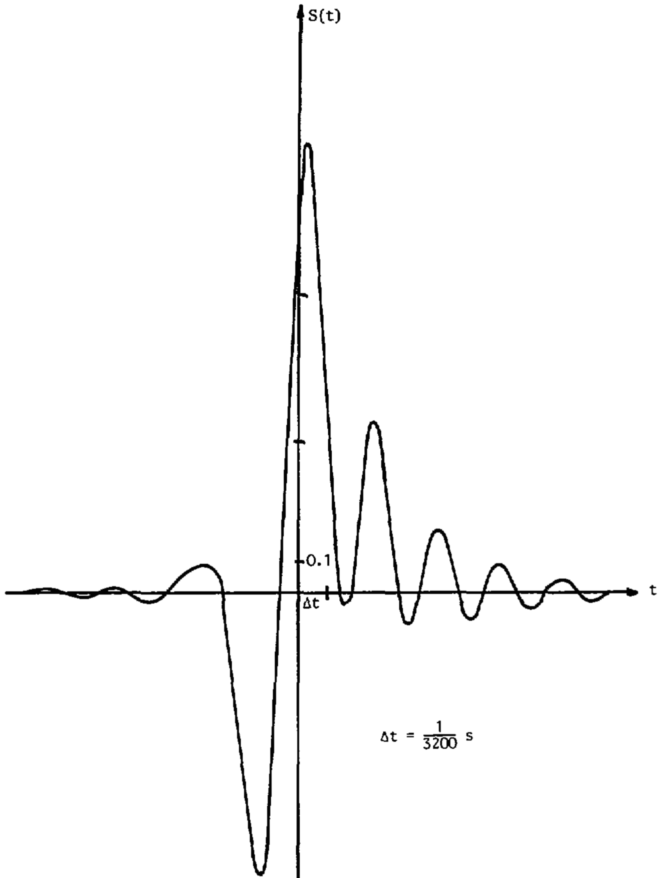


Fig. 5. Impulse response of the channel.

finite set E (for example, $E=\{\pm 1, \pm 3\}$ or $\{\pm 1, \pm 3, \pm 5, \pm 7\}$). In order to increase the binary output of the channel, we can take: 1) Δt_1 smaller than Δt or 2) a large number of values for E.

For the restoration of the transmitted sequence, the solution consists of inserting before the detection an equalizer, following the scheme of Section II (see [10]). Classically, the tap weights of this equalizer are adjusted for minimizing the mean-square error between the input and the output of the equalizer, using a stochastic approximation procedure. This technique requires the knowledge of the transmitted sequence which is, of course, not available in practical situations; this drawback is avoided in practice using: 1) a settling phase during which the emitter transmits an *a priori* known sequence allowing mean-square equalization, and 2) the detected signals in place of the true ones in the mean-square equalization algorithm after this settling phase, thus allowing a certain degree of adaptation when the channel is slowly varying; this procedure is satisfactory if we have only one emitter and one receiver. However, if we have several receivers hearing the same emitter (multipoint communication) and if one receiver starts during the transmission (after a local break, for example), it is not desirable for the emitter to interrupt the data transmission in order to transmit the known sequence, so that the receiver has to inverse the system S, observing only the output: this is exactly the problem stated in Sections I and II. Now we shall illustrate our method with a simple example in this area.

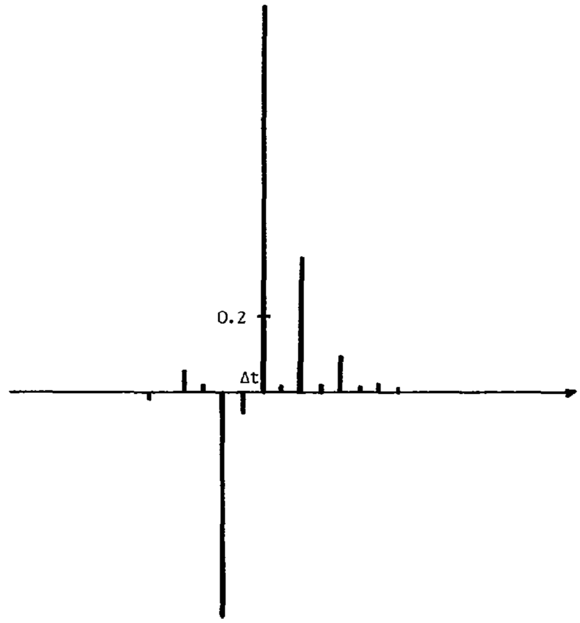


Fig. 6. Sampled impulse response.

B. Examples, Results

Typical Telephone Channel: Fig. 5 shows the impulse response of a typical telephone channel; the first negative peak implies that this channel is clearly nonminimum phase. Fig. 6. gives the sampled response for $\Delta t = 1/3200$ s. The data to be transmitted are equally distributed with 8 possible values: $E=\{\pm 1, \pm 3, \pm 5, \pm 7\}$, achieving a binary output of 9600 bits/s. The additive noise on the output is a Gaussian white noise with variance $2\cdot 10^{-2}$. The equalizer has 21 tap weights, and the procedure we have used is given by (3.19):

$$\theta^{t+1}=\theta^t-\tau X_t(c_t-\gamma\,\mathrm{sign}(c_t)),$$
$$\gamma=\frac{\mathbb{E}a^2}{\mathbb{E}|a|}=5.25. \tag{6.2}$$

Meaning of the Figures:

Fig. 7: Exact inverse S^{-1}, estimated with known data and without noise.

Fig. 8: Tap weights of the equalizer after convergence of (6.2).

Fig. 9: Evolution of the number of errored data; the output of the equalizer is quantized to the nearest possible value to produce the estimated data. Without an equalizer, we obtain about 80 percent of errored data ①, whereas the evolution of the error rate at the output of the equalizer is shown in ②.

Fig. 10: Evolution of the mean-square error; note that the amount of necessary data for having a restoration without error (about 3500 in Fig. 9) depends on the choice of the small parameter τ; our purpose here was not to

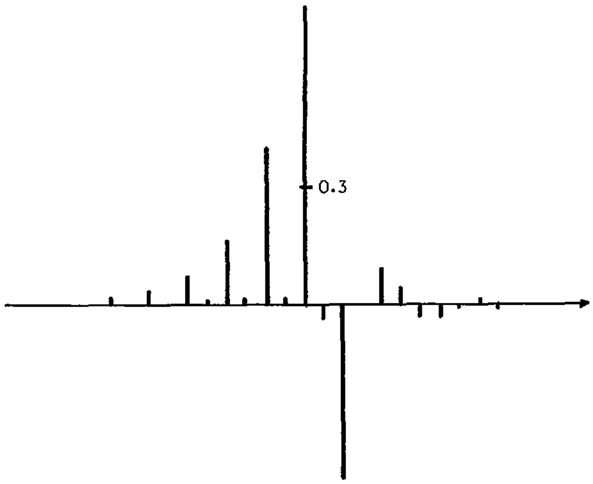


Fig. 7. "Exact" inverse (with the knowledge of the data).

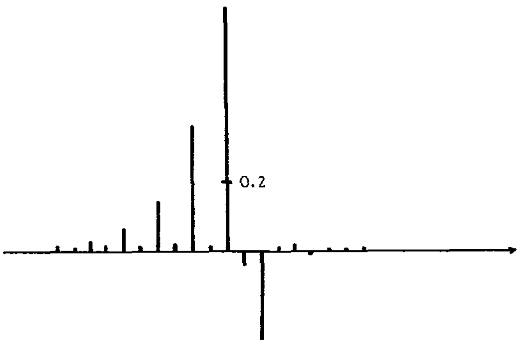


Fig. 8. Identified inverse.

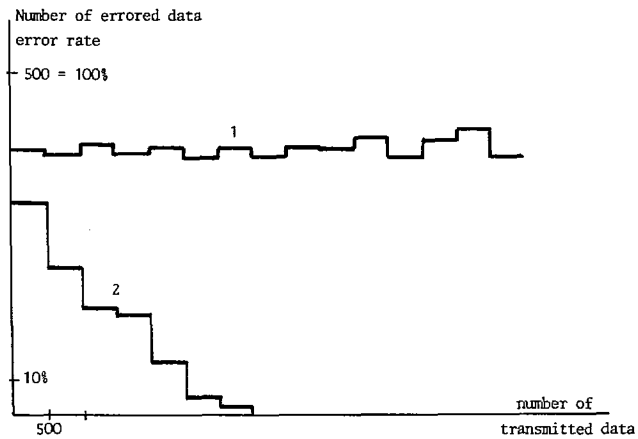


Fig. 9. Evolution of the error rate. ① output of the channel. ② output of the equalizer.

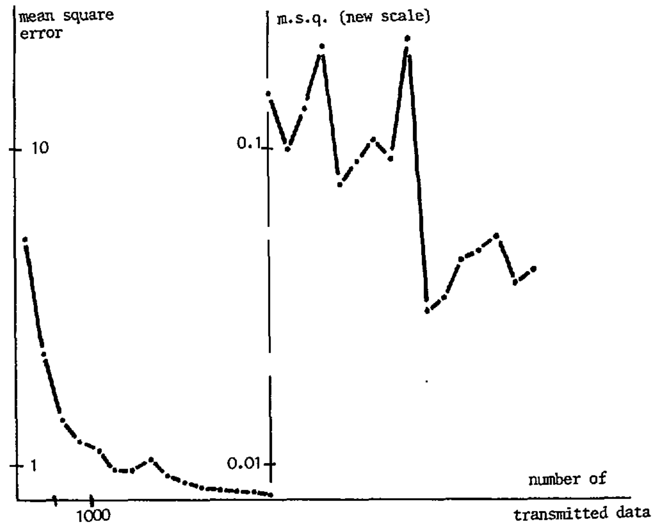


Fig. 10. Evolution of the mean-square error (output of the equalizer).

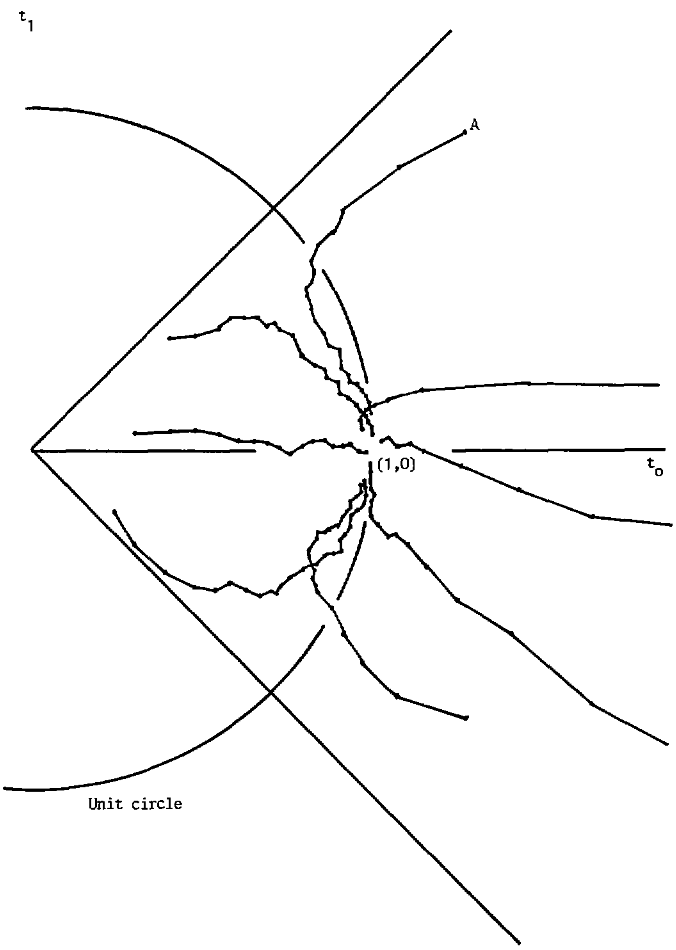


Fig. 11. Some trajectories of the stochastic gradient algorithm (two coefficients of the global line showing the steepest descent lines).

determine an optimal value or τ (we have taken τ about 10^{-3} or 10^{-4}).

Fig. 11: Some trajectories of the global channel (together with the equalizer) for two significant coordinates and different initial values; this figure is to be compared with Fig. 2.

Fig. 12: Three difference simulations with the same initial value: the figure gives a good idea of the behavior (of two coordinates) of the trajectories of the global system around a steepest descent line.

Remark: We refer the reader to [4] for a more realistic experimentation. We give there numerical results obtained with a procedure like (6.2) for the blind adjustment of a baseband equalizer operating in a 16 level-quadrature amplitude modulation system (QAM 16) with double sampling. The presentation is a little more complicated so we

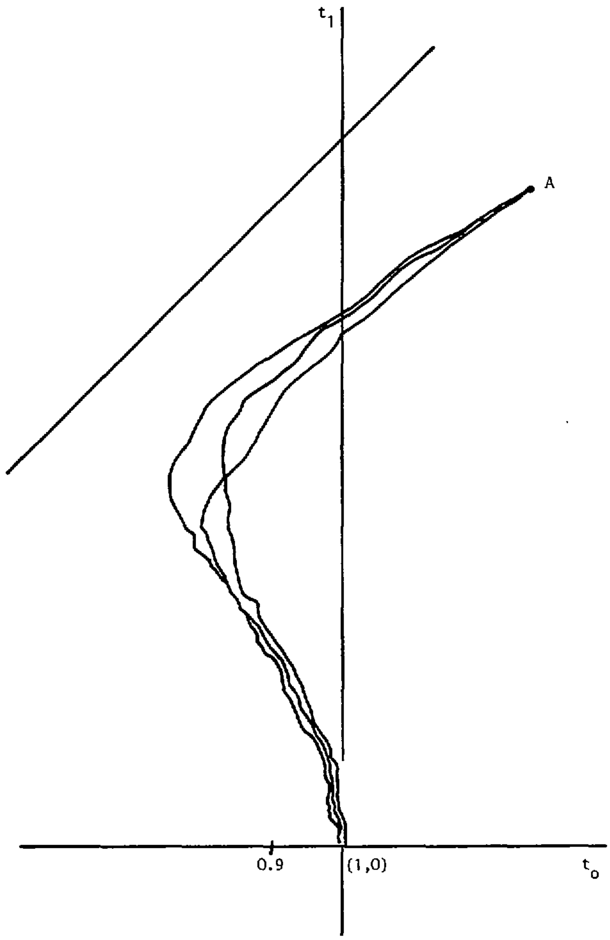


Fig. 12. Three trajectories for the same initial value.

have avoided it here; the numerical results are better (we obtain a faster convergence).

VII. Conclusion

An identification procedure has been presented for nonminimum phase systems without control, the input being a non-Gaussian white noise with sub- or super-Gaussian distribution; an original functional to be minimized has been presented for achieving this, together with the corresponding stochastic gradient and conjugate gradient procedures. This procedure has been successfully applied in adjusting a transversal equalizer in data communication without the transmission of an (*a priori*) known sequence.

Appendix I
Some Facts About Dynamical Systems on s^2

We refer the reader to [5] for more general results. Let us begin with the simple case of ordinary differential equation in $\mathbb{R}^n$. Let

$$x' = V(x) \quad x(0) = x_0 \tag{I.1}$$

be a differential equation in $\mathbb{R}^n$ with V: $\mathbb{R}^n \to \mathbb{R}^n$ locally Lipschitz and time-independent; (I.1) has a unique solution $(x(s))_{s\geqslant 0}$ which can also be extended to $\mathbb{R}_-$, obtaining an *integral curve* $(x(s))_{s\in\mathbb{R}}$. Usually, $(V(x))_{x\in\mathbb{R}^n}$ is called a *vector field* (a map that assigns a vector to each point), and the whole set of integral curves $x(s)$ with all possible x_0's is called the (*one-parameter*) *flow* of the integral curves of this vector field.

Now let us go back to the unit sphere of l^2: s^2 is a $\mathcal{C}^\infty$ = manifold on the Hilbert space l^2. Using the canonical immersion of s^2 in l^2, we can identify the tangent space of s^2 at $x \in s^2$ with the hyperplane in l^2 which is orthogonal to x, say, $x^\perp$. Then a vector field on s^2 is a map $x \to V(x)$ where $V(x) \in x^\perp$; we say that V is locally Lipschitz if this map is locally Lipschitz with respect to the underlying metric. A curve $(x(s))_{s\in\mathbb{R}}$ in s^2 is a map from $\mathbb{R}$ into s^2; if this map is differentiable, then the derivative $x'(s)$ is an element of the tangent space of s^2 at $x(s)$ (i.e., $x(s)^\perp$). Now by a solution to the differential equation (II.1) on $s^2(V(x)$ now being a locally Lipschitz vector field on s^2), we mean a curve $(x(s))_{s\in\mathbb{R}}$ such that $x'(s) = V(x(s))$ and $x(0) = x_0$. Allowing x_0 to take all possible values on s^2, we get the one-parameter flow of the integral curves of the vector field $V(x)$. We say that a point x^* is a (*stable*) *attractor* of this flow if $x(s) \to x^*$ for every x_0 belonging to some neighborhood of x^*.

For example, if $V(x) = -\operatorname{grad}\mathcal{J}(x)$ $(= -(d/dx)\mathcal{J}(x))$ is the gradient field of a functional $\mathcal{J}$: $s^2 \to \mathbb{R}$, then $(x(s))_{s\in\mathbb{R}}$ is nothing but the flow of the steepest descent lines of $\mathcal{J}$. Note that, although this is the case in finite-dimensional spaces, all the local minima of $\mathcal{J}$ need not be attractors of the flow of these steepest descent lines: this property requires further conditions on $\mathcal{J}$ like "positive definiteness" near those local minima.

One could go through the same considerations with l^2 instead of s^2.

Appendix II
Proofs of the Statements of Section III

Proof of Lemma 3.2: Let us take $c_{ij} = c - t_i a_{-i} - t_j a_{-j} = \sum_{k\neq i,j} t_k a_{-k}$; c_{ij} is independent of a_i and a_j; with μ being the distribution of c_{ij}, we get

$$\mathcal{V}(T_\alpha) = \mathbb{E}(\Psi(a_i \cdot R\cos\alpha + a_j R\sin\alpha + c_{ij}))$$

and

$$\frac{\partial}{\partial\alpha_{ij}}\mathcal{V}(T_\alpha) = R\,\mathbb{E}((a_j\cos\alpha - a_i\sin\alpha) \cdot\psi(a_i R\cos\alpha + a_j R\sin\alpha + c_{ij}))$$

$$= R\int\int(Y\cos\alpha - X\sin\alpha) \cdot\psi^\mu(RX\cos\alpha + RY\sin\alpha)\nu(dX)\nu(dY). \tag{II.1}$$

We set $x = X\cos\alpha + Y\sin\alpha$ and $y = Y\cos\alpha - X\sin\alpha$, and let $\mathbb{P}_\alpha(dx,dy)$ be the law of (x,y), X and Y being independent with distribution ν. Then (II.1) becomes

$$\frac{\partial}{\partial \alpha_{ij}} \mathcal{V}(T_\alpha) = R \int \int y \psi^\mu(Rx) \mathbb{P}_\alpha(dx, dy). \quad \text{(II.2)}$$

But ψ being odd and ν symmetric, ψ^μ is odd; on the other hand, $\mathbb{P}_\alpha(dx,dy)$ is invariant under the maps $(x,y) \to (-x,-y)$ and $(x,y) \to (-y,x)$. These remarks, together with (II.2), give the result.

Proof of Lemma 3.4: We consider only the sub-Gaussian case, the same proof works for the super-Gaussian one by changing some inequalities. In order to use Lemma 3.2, we want to show that

$$\chi^\mu(x) = \psi^\mu(x)/x \quad \text{is strictly increasing on } \mathbb{R}_+ = (0, +\infty). \quad \text{(II.3)}$$

If $t_k = 0$ for $k \neq i,j$, we get $\psi^\mu = \psi$ and (II.3) immediately. Otherwise, we know that μ is *unimodal*, i.e., $\mu(dx) = h(x)dx$ with h an even function which strictly decreases on $\mathbb{R}_+$. With the choice of ψ as in Lemma 3.4, we get ψ of the following form:

$$\psi(x) = -\gamma \operatorname{sign}(x) + \tilde{\psi}(x) \quad \text{(II.4)}$$

where $\tilde{\psi}$ is the smooth part of ψ and γ is the jump at the origin. We have

$$\psi^\mu(x) = \psi * h(x) = \psi_1^\mu + \psi_2^\mu,$$

$$\psi_1^\mu(x) = -2\gamma \int_0^x h(t)\, dt,$$

$$\psi_2^\mu = \tilde{\psi} * h(x); \quad \text{(II.5)}$$

therefore,

$$(\psi_1^\mu)'(x) = -2\gamma h(x) \quad \text{is strictly increasing on } \mathbb{R}_+ \text{ for } \gamma > 0. \quad \text{(II.6)}$$

h being even and strictly decreasing on R_+, $(\psi_2^\mu)'' = \tilde{\psi}'' * h$ gives

$$(\psi_2^\mu)'' \geqslant 0 \text{ on } \mathbb{R}_+ \text{ (and } > 0 \text{ if } \psi'' > 0). \quad \text{(II.7)}$$

With (II.6) and (II.7), we obtain $(\psi^\mu)'' > 0$ on R_+, which, together with $\psi^\mu(0) = 0$, gives (II.3).

Now, using Lemma 3.2, the result we want is

$$\int \int_{\{x>0, y>0\}} xy(\chi(x) - \chi(y)) \mathbb{P}_\alpha(dx, dy) > 0, \quad \text{for } 0 < \alpha < \frac{\pi}{4}, \quad \text{(II.8)}$$

where χ is strictly increasing on R_+, and we shall apply (II.8) to the function χ^μ defined in (II.3). In polar coordinates, (II.8) becomes

$$\int_0^\infty \rho^3 F(\alpha,\rho)\, d\rho > 0 \qquad \text{for } 0 < \rho, \ 0 < \alpha < \frac{\pi}{4}$$

where

$$F(\alpha,\rho) = \int_0^{\pi/2} \sin\theta \cos\theta (\chi(\rho\cos\theta) - \chi(\rho\sin\theta)) \cdot f(\rho\cos(\alpha+\theta)) f(\rho\sin(\alpha+\theta))\, d\theta$$

where $f = K \cdot e^{-g}$ is the density of ν. Using the fact that $g'(x)/x$ is increasing, we get $\theta \to g(\rho\cos\theta) + g(\rho\sin\theta)$ is decreasing on $(0, \pi/4)$ and increasing on $(\pi/4, \pi/2)$. Using this fact, and comparing the points $\theta_1 = (\pi/4) + \theta'$ and $\theta_2 = (\pi/4) - \theta'$ for $\theta' \in (0, \pi/4)$, we easily get $F(\alpha,\rho) > 0$ and (II.8).

Proof of Theorem 3.5: Let $(\overline{T}_s)$ be a trajectory of the flow with $\overline{T}_s = (t_k(s))_{k \in Z}$. The equations $\{t_k = \text{constant for } k \neq i,j\}$ and $\{t_i^2 + t_j^2 = R^2\}$ give curves that are orthogonal to the hyperplanes $\{t_i/t_j = \text{constant}\}$; hence, with Lemma 3.4, we get

$$\begin{cases} \text{for every } (i,j),\ s \to \dfrac{|t_i(s)|}{|t_j(s)|} \\ \quad \text{is strictly decreasing, increasing,} \\ \text{or constant, respectively, with} \\ \quad \dfrac{|t_i(0)|}{|t_j(0)|} < 1,\ > 1, \text{ or } = 1. \end{cases} \quad \text{(II.9)}$$

Now consider $\overline{T}_0 \in s^2$, $\overline{T}_0 = (t_i)$ with t_i maximum for $i \in I_n = \{i_1, \cdots, i_n\}$. We want to show that

$$\lim_{s \to \infty} \overline{T}_s = \overline{T}_\infty \quad \text{in } l^2\text{-strong} \quad \text{(II.10)}$$

where $\overline{T}_\infty = (t_i(\infty))$ is such that

$$t_i(\infty) = \frac{1}{\sqrt{n}} \operatorname{sign}(t_i) \quad \text{for } i \in I_n, \qquad = 0 \quad \text{for } i \notin I_n. \quad \text{(II.11)}$$

For $n = 1$, we get a stable attractor for $\overline{T}_\infty$, while for $n > 1$, we get a saddle point, which is not a stable attractor. This will finish the proof. Thus, let us go back to (II.10) and (II.11). Thanks to (II.9), the function $\rho(s) = \max_{k \in Z} |t_k(s)|$ is increasing and bounded by 1 ($\|\overline{T}_s\|^2 = 1$), and therefore has a limit $\rho(\infty)$; (II.9) also gives the convergence of $(1/\rho(s)) \cdot \overline{T}_s$ to the system $\sqrt{n} \cdot \overline{T}_\infty$ in l^2-weak. Thus, $\rho(\infty) = \sqrt{n}$ and $\overline{T}_s \to \overline{T}_\infty$ in l^2-weak; finally, $\|\overline{T}_s\| = 1$ gives the strong convergence.

Proof of Theorem 3.6: For fixed $T \in s^2$, we study the function

$$v(\rho) = \mathcal{V}(\rho T)\rho \geqslant 0. \quad \text{(II.12)}$$

Let ν_c denote the distribution of $c = \sum t_k a_{-k}$, so that $v(\rho) = \mathbb{E}(\Psi(\rho c))$ and

$$v'(\rho) = \mathbb{E}(c\psi(c)) = \int x \tilde{\psi}(\rho x) \nu_c(dx) - \gamma \int |x| \nu_c(dx)$$

$$v^{(3)}(\rho) = \int x^3 \tilde{\psi}''(\rho x) \nu_c(dx) \geqslant 0$$

because $\tilde{\psi}'' \geqslant 0$ on $\mathbb{R}_+$. Therefore, the function v' is convex with $v'(0) \leqslant 0$ and strictly convex when $v'(0) = 0$. But with (3.17), we get $v''(0)=\psi'(0)\cdot\mathbb{E}(a^2)$ which is strictly negative when $\gamma=0$. Therefore, anyway,

there exists $\rho_c>0$ unique such that

$$v'(\rho_c)=0, \quad v''(\rho_c)>0. \tag{II.13}$$

Consequently, the radial component of the vector field (V_T) is outgoing for $\rho<\rho_c$ and entering for $\rho>\rho_c$. Finally, the fact that the only local minima of $\mathcal{V}$ are the systems $\pm I$ results from Lemma 3.4 (motion of the radius), which remains valid, and from the equality $\rho_c=1$ for $T=\pm I$.

We achieve the convergence of the flow (T_s) in the following way: 1) as in Theorem 3.5, $(1/\|T_s\|)\cdot T_s\to T_\infty$ in l^2 where T_∞ is given by (II.11) and 2) $\|T_s\|\to 1$ thanks to (II.13) and the fact that $\rho_c=1$ when $T=\pm I$.

Note that in the super-Gaussian case, there is a contradiction between the conditions required on ψ in this case and the needed inequality $v^{(3)}(\rho_c)\geqslant 0$, so that the systems $\pm I$ may only be saddle points for $\mathcal{V}$ in l^2.

Proof of Theorem 3.7: We work in the sub-Gaussian case. Let $(W_T)_{T\in l^2}$ be the vector field defined by

$$W_T=(S*\check{S})*V_T=-(S*\check{S})*\mathbb{E}(\Lambda\psi(c)). \tag{II.14}$$

(W_T) is $\mathcal{C}^1$ thanks to (3.4), (3.9), and ii) of Lemma 3.1. Let us define the vector field $(V_\theta)_{\theta\in l^2}$ by

$$W_T=S*V_\theta \qquad \text{for } T=S*\theta \tag{II.15}$$

or, equivalently,

$$V_\theta=-\mathbb{E}(X\psi(c)) \qquad \text{for } c=\Sigma t_k a_{-k}. \tag{II.16}$$

(V_θ) is $\mathcal{C}^1$ and opposite to the gradient field of $\mathcal{J}$ in l^2. Hence, the flow (θ_s) of the integral curves of (V_θ) is nothing but the flow of the steepest descent lines of $\mathcal{J}$, and (II.15) shows that when the equalizer follows one of these steepest descent lines, then the global system follows an integral curve of the vector field (W_T); on the other hand, (3.6) implies that $W_T=\Lambda V_T$ so that the field W_T simply appears as the opposite of the gradient field of $\mathcal{V}$ in l^2 when provided with the metric (equivalent to the ordinary one) associated with the operator Λ^{-1}. Thus, *the flow (T_s) appears as the flow of the steepest descent lines of $\mathcal{V}$, l^2 being endowed with this new metric.*

It is a general situation that the change of metric (with another equivalent one) does not modify the behavior of the s.d.l. (although the minima and saddle points do not change, the crest lines are, of course, modified). Let us give a quick explanation: let T_0 be the initial point for $(T_s)_{s\geqslant 0}$ and $(\tilde{T}_s)_{s\geqslant 0}$ [respectively the integral curves of (V_T) and (W_T)]. Let us consider the path $(\tilde{T}_s)_{0\leqslant s\leqslant t}$: there exists $\sigma(t)<\infty$ such that both paths $(\tilde{T}_s)_{0\leqslant s\leqslant t}$ and $(T_s)_{0\leqslant s\leqslant\sigma(t)}$ have the same length with respect to the metric associated with Λ^{-1} so that, $(\tilde{T}_s)$ being an s.d.l. for this metric, we have

$$\mathcal{V}(T_{\sigma(t)})\geqslant\mathcal{V}(\tilde{T}_t). \tag{II.17}$$

The path $(T_s)_{0\leqslant s<\infty}$ is of finite length for both metrics, and we have the same property for $(T_s)_{0\leqslant s<\sigma(\infty)}$ with $\sigma(\infty)=\lim\sigma(s)$, and thus for $(\tilde{T}_s)_{0\leqslant s<\infty}$. Therefore, $\tilde{T}_\infty=\lim\tilde{T}_s$ exists and is necessarily a stationary point of $\mathcal{V}$, which gives the desired result. Note that T_s may converge to $\pm I$ and $\tilde{T}_s$ to a saddle point: this is in accord with (II.17) because we cannot ensure that $\sigma(\infty)=\infty$.

References

[1] R. Azencott and G. Ruget, "Mélanges d'équations différentielles et grands écarts à la loi des grands nombres," *Z. Wahrscheinlichkeitsth.*, vol. 38, pp. 1–54, 1977.

[2] A. Benveniste, M. Goursat, and G. Ruget, "Analysis of stochastic approximation schemes with applications to data communications algorithms," to be published.

[3] A. Benveniste and M. Goursat, "Une méthode adaptative d'identification non-gaussienne...," *Colloque GRETSI*, pp. 83/1–83/8, Apr. 1977.

[4] A. Benveniste, M. Bonnet, M. Goursat, C. Macchi, and G. Ruget, IRIA LABORIA Rep. 325, Sept. 1978.

[5] S. Lang, *Introduction aux Variétés Différentiables.* Paris: Dunod, 1967.

[6] L. Ljung, "Consistency of the least squares identification method," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 779–781, Oct. 1976.

[7] ——, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 539–551, Oct. 1977.

[8] ——, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551–575, Oct. 1977.

[9] C. Macchi, J. P. Jouannaud, and O. Macchi, "Récepteurs adaptatifs pour transmission de données à grande vitesse," *Ann. Télécom.*, vol. 30, no. 9–10, 1975.

[10] J. G. Proakis, "Advances in equalization for intersymbol interference," in *Advances in Communication Systems*, vol. 4. New York: Academic, 1975, pp. 124–194.

[11] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun.*, pp. 679–682, June 1975.

[12] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II. New York: Wiley.

Albert Benveniste was born in Paris, France, in 1949. He was graduated from the Ecole des Mines de Paris in 1971.

From 1971 to 1973 he was with the Centre d'Automatique de l'Ecole des Mines (Fontainebleau); from 1974 to 1976 he was with the IRIA (Rocquencourt); since 1976 he has been with the IRISA (Rennes), and is actually a consultant at the CNET (National Research Center for Telecommunications). After some work on probability theory (Markov processes and ergodic theory) for his "Thèse d'Etat" (Paris, 1975), his interests moved towards the area of applied mathematics (automatic control, identification and recursive algorithms, signal processing, speech and image coding, and data communication systems).

Maurice Goursat was born in 1945. He graduated in mathematics from the University of Poitiers and received a 3rd Cycle Thesis from the University of Paris VI in 1971.

Since 1969 he has been a Researcher at the Institut (National at 01/01/1980) de Recherche d'Informatique et d'Automatique; his main interest was first in numerical analysis and is now in automatic control, signal processing, and stochastic control (with the Systems Theory Group of INRIA).

Gabriel Ruget was born in Albi, France, in 1944. He studied physics and pure mathematics at Ecole Normale Supérieure, Paris, rue d'Ulm, and then worked in analysis and geometry at Centre National de la Recherche Scientifique (Paris).

After his "thèse d'Etat" (Orsay, 1969), he taught in Tunis and Paris. In 1974 his interests moved towards the field of applied mathematics (probability and statistics with applications to neurophysiology, automatic control, signal processing, and communications). In 1975 he occupied a Chair at the University of Rennes; in 1976 he was the Director of the Institut de Recherches en Informatique et Systèmes Aléatoires (a laboratory in Rennes, associated with C.N.R.S.). He now occupies a Chair of Mathematics at Université Paris-Sud. His present interests include computations about rare events (large deviations theory), fast simulations, and problems of propagation in epidemiology.

Feedback System Design: The Fractional Representation Approach to Analysis and Synthesis

C. A. DESOER, FELLOW, IEEE, RUEY-WEN LIU, JOHN MURRAY, AND RICHARD SAEKS, FELLOW, IEEE

***Abstract*—The problem of designing a feedback system with prescribed properties is attacked via a fractional representation approach to feedback system analysis and synthesis. To this end we let *H* denote a ring of operators with the prescribed properties and model a given plant as the ratio of two operators in *H*. This, in turn, leads to a simplified test to determine whether or not a feedback system in which that plant is embedded has the prescribed properties and a complete characterization of those compensators which will "place" the feedback system in *H*. The theory is formulated axiomatically to permit its application in a wide variety of system design problems and is extremely elementary in nature requiring no more than addition, multiplication, subtraction, and inversion for its derivation even in the most general settings.**

I. Introduction

INTUITIVELY, the linear feedback system design process may be broken down into three steps: modeling, analysis, and synthesis; each of which may be carried out via a multiplicity of time and frequency domain techniques. In engineering practice, however, the three steps are loosely matched to one another. The purpose of the present paper is to use fractional representation models to the analysis and synthesis of feedback systems. Here, if one desires to design a system with prescribed properties the given plant is initially modeled as a quotient of two operators, each of which has the desired properties. Once such a model has been specified a similar model may be formulated for the feedback system constructed from that plant which, in turn, may be used to determine whether or not the feedback system has the desired properties. Moreover, the set of compensators which will cause the feedback system to have the prescribed properties may be completely characterized in terms of such a model. As such, by choosing a model for the plant which is matched to the design criteria the analysis and synthesis processes for a feedback system may be greatly simplified.

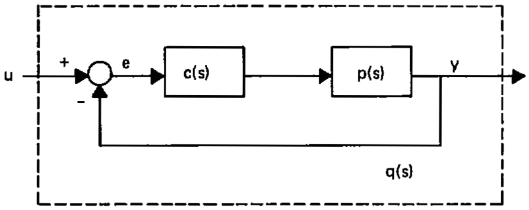


Fig. 1. Single-variate control system.

These ideas are illustrated by the following derivation of the set of stabilizing compensators for the single variate control system of Fig. 1.

We say that a transfer function $p(s)$ is *exponentially stable* (exp. stable) if $p(s)$ is a *proper rational* function with poles having *negative* real parts. Although the plant may naturally be modeled as a quotient of coprime polynomials [16],[19] $p(s)=a(s)/b(s)$ since our ultimate goal is a

Manuscript received April 2, 1979; revised October 15, 1979. Paper recommended by B. Francis, Chairman of the Linear Systems Committee. The work of C. A. Desoer was supported in part by the Joint Services Electronics Program at the University of California, Berkeley, under AFOSR Contract 76-C-0100. The work of R.-W. Liu was supported in part by ONR Contract 78-C-0444. The work of J. Murray and R. Saeks was supported in part by the Joint Services Electronics Program, Texas Tech University, under ONR Contract 76-C-1136.

C. A. Desoer is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

R.-W. Liu is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

J. Murray and R. Saeks are with the Department of Electrical Engineering, Texas Tech University, Lubbock, TX 79409.

0018-9286/80/0600-0399$00.75 ©1980 IEEE