Techniques de modélisation expérimentale à l'aide d'un micro-ordinateur

par Jean-Claude TRIGEASSOU,
I.U.T. Génie Electrique, Poitiers.

INTRODUCTION.

Cet article a pour but d'informer les enseignants de Sciences physiques des possibilités nouvelles que leur offre le micro-ordinateur pour traiter leurs résultats expérimentaux et en extraire des lois ou modèles.

Les méthodes de modélisation expérimentale qui vont être présentées sont issues de la Théorie des Systèmes et des Sciences de l'Ingénieur. Elles s'appliquent à des systèmes très divers : physiques, chimiques, mais aussi biologiques, économiques, etc. Ces méthodes, fondées sur des principes classiques, n'ont pu se développer que récemment grâce aux progrès de l'informatique et des moyens de calcul.

Après une définition de la notion de modèle, on présente dans cet exposé le principe général de la modélisation expérimentale. On s'intéresse ensuite plus particulièrement aux méthodes de minimisation d'un critère quadratique que l'on applique à un exemple.

I. MODELE - SYSTEME.

1) Définitions.

On appelle modèle d'un système physique, chimique, biologique, économique, sociologique..., un ensemble de relations mathématiques décrivant le comportement de ce système. On distingue principalement deux sortes de modèles :

— modèle de connaissance (ou modèle théorique),

— modèle de comportement (ou modèle expérimental).

Le modèle de connaissance est obtenu par une « mise en équation » du système. A partir de lois élémentaires et grâce à une suite de déductions et de raisonnements, on aboutit à une loi mathématique constituant le modèle de connaissance. L'expérimentation ne joue qu'un rôle très réduit, voire inexistant, dans l'élaboration de ce modèle.

Par contre, l'expérimentation est essentielle pour l'obtention du modèle de comportement : il traduit la (ou les) réponse (s) du système à une (ou à des) excitation (s). La loi mathématique est obtenue par des calculs portant sur les observations expérimentales.

2) Exemples élémentaires de modélisation.

Considérons un système constitué par un fil conducteur de longueur l, de section S. On se propose de déterminer la loi reliant la d.d.p. U à ses bornes à l'intensité I du courant qui le traverse.

a) MODÈLE DE CONNAISSANCE.

Grâce aux lois élémentaires de l'électrocinétique et après une mise en équations, on bâtit un modèle de connaissance

$$U = \varrho \frac{l}{S} I.$$

Néanmoins, il sera difficile, par ce procédé, d'attribuer une valeur numérique à $R = \varrho \frac{l}{S}$ car cela suppose une connaissance précise du paramètre ϱ, lui-même fonction de paramètres élémentaires tels que nombre par unité de volume et mobilité des électrons libres.

b) MODÈLE DE COMPORTEMENT.

Il est nécessaire de réaliser une expérience pour relever la caractéristique $U = f(I)$.

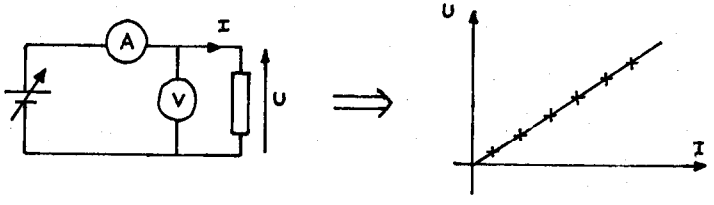

Fig. 1

L'examen de la courbe montre que dans le domaine considéré, la relation $u = f(I)$ peut être bien approchée par la loi $u = k\,I$. On fait ainsi le choix d'un type de modèle (en fait de sa structure), il reste ensuite à déterminer la valeur du paramètre k. Pour cet exemple simple, $k = \frac{\Delta U}{\Delta I} = R$.

Le modèle de comportement est donc U = RI.

Le conducteur électrique est correctement représenté par cette loi, on pourra la réutiliser pour décrire et prévoir son comportement dans un autre circuit électrique, à condition de respecter le domaine de validité initial.

c) REMARQUES.

En pratique, la frontière est moins nette entre ces deux méthodes de modélisation.

On fait souvent appel à des mesures pour estimer certains des paramètres du modèle de connaissance. De même, le modèle de connaissance permet de choisir une structure appropriée du modèle de comportement avant de déterminer ses paramètres.

Dans l'exemple précédent, modèles de connaissance et de comportement arrivent à la même conclusion : U = RI. En pratique, les lois mathématiques peuvent être très différentes : en effet, on peut correctement approcher le comportement d'un système pour une expérience donnée par une loi n'ayant aucun rapport avec celle du modèle de connaissance. De même, plusieurs lois mathématiques peuvent donner des résultats équivalents. Suivant l'objectif assigné à la modélisation, il sera donc nécessaire de procéder à une validation du modèle de comportement.

3) Notion de système.

Afin de généraliser l'exposé, on adoptera pour la suite certains termes propres à la théorie des systèmes.

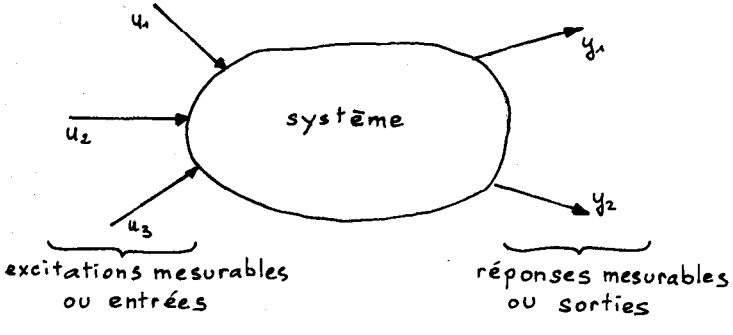

Fig. 2

Le modèle du système sera constitué par :

$$\begin{cases} y_1 = f_1(u_1, u_2, u_3, \text{paramètres}) \\ y_2 = f_2(u_1, u_2, u_3, \text{paramètres}). \end{cases}$$

La fonction $f_i(u_1...)$ est considérée dans un sens très large : fonction au sens classique, mais aussi équation différentielle, équation aux dérivées partielles, etc.

La détermination d'un modèle de comportement comprendra deux étapes essentielles :

— recherche de la structure du modèle, c'est-à-dire de la fonction $f_i(u_1...)$ permettant la meilleure description du comportement du système ;

— recherche des paramètres du modèle (on parle aussi d'identification en Théorie des Systèmes).

En pratique, il est souvent difficile de dissocier ces deux étapes.

II. MODELISATION ELEMENTAIRE.

1) Méthodes classiques.

Ces méthodes supposent un comportement parfaitement déterministe du système (erreurs de mesure nulles ou négligeables) et une excitation de type particulier.

Envisageons deux exemples pour caractériser ces méthodes.

a) THERMISTANCE.

Soit à déterminer A et B de la loi $R = A\, e^{B/T}$.

On effectue la transformation $\underset{\sim}{\ln R} = \underset{\sim}{\ln A} + \underset{\sim}{B} \times \underset{\sim}{1/T}$

soit $y = a_0 + a_1 \times x$

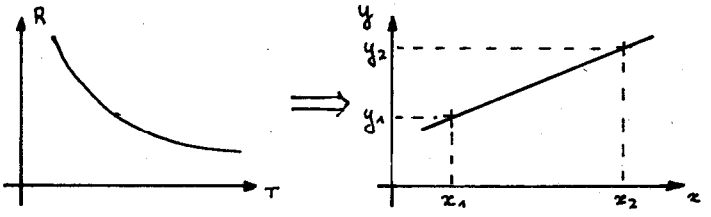

Fig. 3

On vérifie le modèle en 2 points $\{x_1, y_1\}$ et $\{x_2, y_2\}$

soit $\begin{cases} y_1 = a_1 x_1 + a_0 \\ y_2 = a_1 x_2 + a_0 \end{cases} \Rightarrow a_1, a_0 \Rightarrow A, B$

b) CHARGE D'UN CONDENSATEUR.

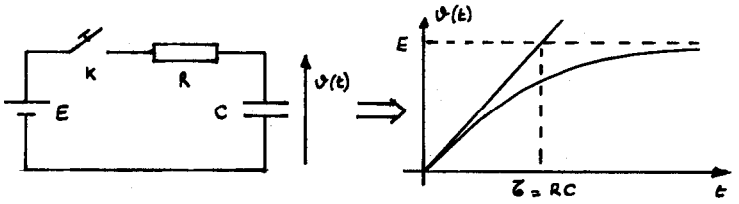

Fig. 4

Supposons $v(0) = 0$. On ferme K à $t = 0$. Alors :

$$v(t) = E(1 - e^{-t/RC}).$$

On utilise une propriété de la courbe $v(t)$: la tangente à l'origine coupe l'asymptote à l'instant $t = \tau = RC$. Donc en traçant la tangente, on a un procédé de détermination de la constante de temps du circuit.

c) REMARQUES.

Ces méthodes n'utilisent que partiellement l'information fournie par l'expérience (2 points seulement de la courbe $R = f(T)$ pour la thermistance). Elles sont inutilisables en présence d'erreurs de mesure ou lorsque le modèle est très approché (erreur de structure). Elles ne permettent la détermination des paramètres que pour des excitations particulières (essai en échelon pour le circuit RC).

Enfin elles ne donnent pas d'indication sur la qualité de l'approximation système/modèle ni sur la précision des paramètres.

2) Modélisation « manuelle » assistée par micro-ordinateur.

Les modèles traités par les méthodes élémentaires précédentes doivent rester très simples. Il n'en est plus de même lorsqu'on dispose d'un moyen de calcul informatisé.

Supposons que nous disposons d'un logiciel de simulation de modèle avec sortie graphique. Alors, si nous excitons le modèle par la même « entrée » u que le système, nous allons pouvoir comparer leurs sorties y syst et y mod et nous modifierons les paramètres du modèle jusqu'à ce que y mod « approche au mieux » y syst. Cette démarche est symbolisée par le schéma fonctionnel de la fig. 5.

Cette modélisation « manuelle » permet une approche intuitive du problème général, ainsi que la découverte de la sensibilité du modèle par rapport à la variation de ses paramètres.

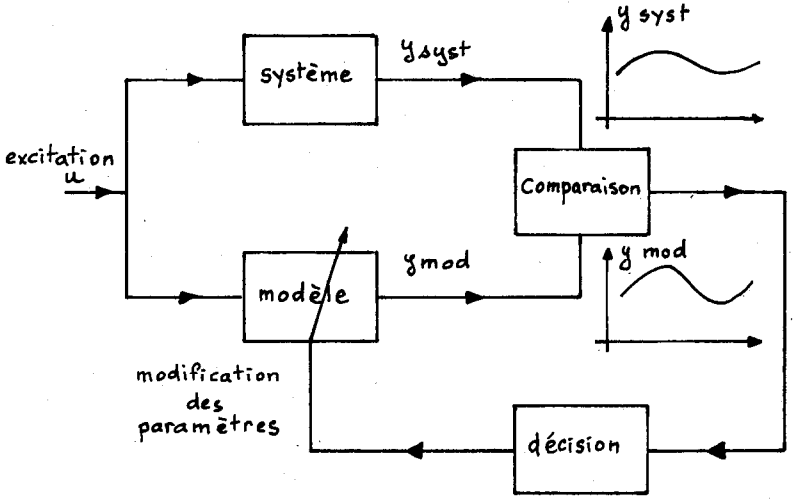

Fig. 5

Cependant, pour formatrice qu'elle soit, cette approche s'avère inadaptée en pratique :

— complexité des opérations à effectuer dès que le nombre de paramètres dépasse 1 ou 2 ;

— difficulté d'une caractérisation objective de l'approximation entre *y* syst et *y* mod.

III. DETERMINATION D'UN MODELE DE COMPORTEMENT.

1) Introduction.

Nous allons décrire les différentes étapes conduisant à l'élaboration d'un modèle de comportement.

Tout d'abord, nous définirons un critère nous permettant de qualifier de manière objective l'approximation entre système et modèle.

2) Critère.

Supposons connus les graphes *y* syst et *y* mod en fonction d'une variable indépendante *x*.

La comparaison entre *y* syst et *y* mod conduit naturellement à définir une erreur $\varepsilon = y\,\text{syst} - y\,\text{mod}$.

Le critère le plus simple pourrait être $\varepsilon = 0$, c'est-à-dire *y* syst = *y* mod pour tous les points de mesure. Malheureusement,

un tel critère est irréaliste, car il suppose à la fois des erreurs de mesure et de modèle nulles.

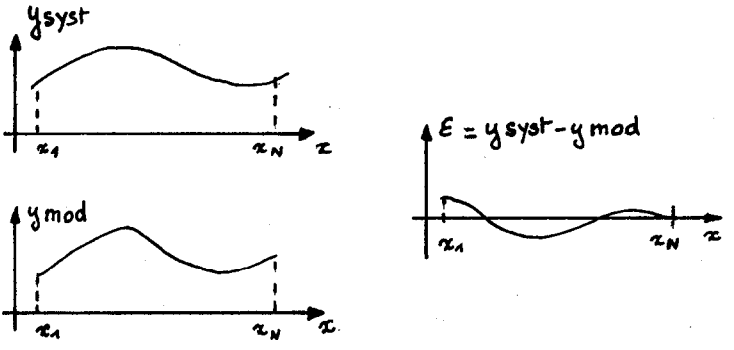

Fig. 6

A cause de ces erreurs inhérentes aux sciences expérimentales, le critère utilisé ne pourra être δ nul, mais on recherchera à rendre δ minimal. Le critère doit tenir compte de l'ensemble des erreurs, c'est pour cette raison que l'on emploie en pratique des sommes ou des intégrales.

Exemples de critères :

a) Somme de la valeur absolue de l'erreur :

$$J = \int_{X_1}^{X_N} |\varepsilon(x)|\, dx \qquad \text{ou} \qquad \sum_{i=1}^{N} |\varepsilon_i|$$

b) Critère quadratique :

$$J = \int_{X_1}^{X_N} \varepsilon^2(x)\, dx \qquad \text{ou} \qquad \sum_{i=1}^{N} \varepsilon_i^2$$

Remarque :

Le critère du type $\int_{X_1}^{X_N} \varepsilon(x)\, dx$ est à rejeter car il n'est pas significatif (fig. 7).

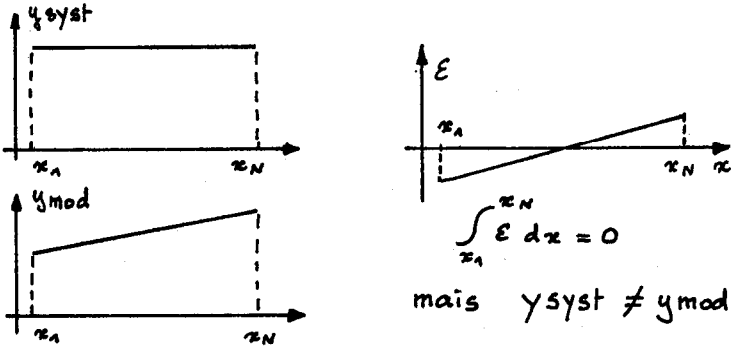

Fig. 7

3) **Etapes de la détermination d'un modèle.**

a) DÉFINITION DU SYSTÈME.

On doit d'abord définir le système, c'est-à-dire préciser les grandeurs d'entrée et de sortie, la classe des modèles capables de décrire le système ainsi que les paramètres à prendre en compte.

b) DÉFINITION D'UN CRITÈRE J.

En pratique, on utilise le critère quadratique à cause de ses propriétés mathématiques et de son interprétation statistique.

c) MINIMISATION DU CRITÈRE.

La meilleure approximaion possible entre *y* syst et *y* mod est obtenue lorsque le critère a été minimisé, tant par rapport à la structure qu'aux paramètres du modèle.

d) SENSIBILITÉ ET PRÉCISION DU MODÈLE.

Après minimisation, on testera la sensibilité de J par rapport à la variation des paramètres : cette sensibilité nous informe sur l'importance relative des différents paramètres, mais aussi sur la précision de leur détermination. Cet aspect du problème qui rejoint les statistiques ne sera pas développé dans cet exposé.

e) VALIDATION DU MODÈLE.

Sauf dans quelques cas particuliers, la minimisation du critère n'est pas suffisante pour valider le modèle. Il faut qu'il soit aussi capable de prédire le comportement du système dans un autre contexte expérimental.

4) **Minimisation d'un critère quadratique.**

On va s'intéresser au problème de la minimisation d'un critère quadratique. On peut classer les méthodes correspondantes en deux catégories :

— si le modèle est linéaire par rapport à ses paramètres, la minimisation du critère est réalisée en une seule étape, c'est la méthode des moindres carrés (ou méthode de Gauss). Elle possède une interprétation statistique qui ne sera pas développée ici ;

— si le modèle n'est pas linéaire par rapport à ses paramètres, on doit minimiser J par étapes successives. De nombreuses méthodes (dites d'optimisation) ont été développées sous l'appellation générale de Programmation Non Linéaire. On présentera d'abord une méthode très simple, puis la méthode du gradient et une de ses améliorations.

IV. METHODE DES MOINDRES CARRES OU METHODES DE GAUSS.

1) Ajustement linéaire.

Considérons un ensemble de mesures d'une grandeur y dépendant de la variable x.

Les observations y exactes sont entachées d'erreur, notons y^* ces mesures, x est supposé connu exactement.

On dispose de N couples de mesures :

$$\{y_1^*, x_1\} \text{--------} \{y_N^*, x_N\}$$

Considérons un modèle noté $\hat{y}$ tel que : $\hat{y} = a_0 + a_1 x$ où a_0 et a_1 sont des paramètres à déterminer.

Pour chaque couple $\{y_i^*, x_i\}$, on peut définir une erreur $\varepsilon_i = y_i^* - \hat{y}_i = y_i^* - a_0 - a_1 x$.

On se reportera à la fig. 8.

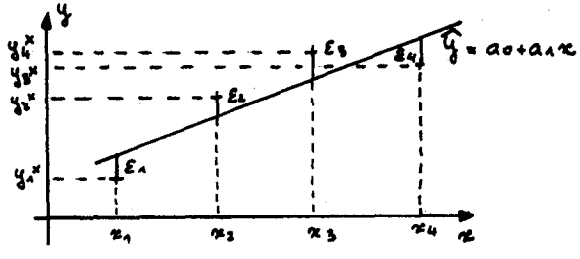

Fig. 8

On se propose de déterminer les paramètres a_0 et a_1, de telle sorte que :

$$J = \sum_{i=1}^{N} \varepsilon_i^2 = f(a_0, a_1) \text{ soit minimum.}$$

Le minimum de J est obtenu pour :

$$\left| \begin{array}{l} \dfrac{\partial J}{\partial a_0} = 0 \\ \dfrac{\partial J}{\partial a_1} = 0 \end{array} \right.$$

soit :

$$\left| \begin{array}{l} \dfrac{\partial J}{\partial a_0} = -2 \sum_1^N (y_i^* - a_0 - a_1 x_i) \times 1 = 0 \\ \dfrac{\partial J}{\partial a_1} = -2 \sum_1^N (y_i^* - a_0 - a_1 x_i) \times x_i = 0 \end{array} \right.$$

a_0 et a_1 sont solution du système linéaire :

$$\left| \begin{array}{l} a_0 \sum_1^N 1 + a_1 \sum_1^N x_i = \sum_1^N y_i^* \\ a_0 \sum_1^N x_i + a_1 \sum_1^N x_i^2 = \sum_1^N y_i^* x_i \end{array} \right.$$

2) **Cas général.**

Considérons un modèle général du type :

$$\hat{y}(x) = a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x)$$

où les fonctions $f_i(x)$ sont connues et les paramètres a_i sont à identifier. Un tel modèle est linéaire par rapport à ses paramètres.

On dispose de N observations :

$$\begin{array}{c} \{y_1^*, f_0(x_1), \ldots, f_n(x_1)\} \\ \hline \{y_N^*, f_0(x_N), \ldots, f_n(x_N)\} \end{array}$$

pour chaque observation, $\varepsilon_i = y_i^* - \hat{y}_i$.

Déterminons $\{a_0 \cdots a_n\}$ de telle sorte que $J = \sum_1^N \varepsilon_i^2$ soit minimum.

Cette condition est réalisée si $\dfrac{\partial J}{\partial a_j} = 0$ $\left| \begin{array}{l} j \text{ variant de 0} \\ \text{à } n. \end{array} \right.$

On est conduit à résoudre un système linéaire de $n+1$ équations à $n+1$ inconnues $\{a_0 \cdots a_n\}$, soit :

j variant de 0 à n | $a_0 \sum_{i=1}^{N} f_0(x_i) f_j(x_i) + \cdots + a_j \sum_{i=1}^{N} f_j^2(x_i) + \cdots$

$$\cdots + a_n \sum_{i=1}^{N} f_n(x_i) f_j(x_i) = \sum_{i=1}^{N} y^*(x_i) f_j(x_i).$$

Pour $n + 1 > 3$ la résolution d'un tel système nécessite des méthodes numériques (exemple : méthode du Pivot), donc un traitement informatique.

3) Application.

a) AJUSTEMENT POLYNÔMIAL.

Dans le cas où $f_j(x) = x^j$ (avec $f_0(x) = 1$) $\hat{y}(x)$ est un polynôme d'ordre n.

Ce type de modèle est utilisé pour effectuer du lissage de courbe. On peut l'utiliser par exemple pour remplacer une courbe d'étalonnage par une loi mathématique se prêtant mieux à une automatisation des calculs qu'un graphe. Le lecteur pourra trouver un programme BASIC pour le modèle polynômial en [1].

b) MODÈLES NON LINÉAIRES.

Prenons l'exemple de la thermistance $R = A\,e^{B/T}$. Le modèle de R n'est pas linéaire par rapport à A et B. Par contre, en effectuant la transformation : $\ln R = \ln A + B \times 1/T$, on obtient un modèle linéaire par rapport à ln A et B.

c) EQUATIONS DIFFÉRENTIELLES.

Soit :

$$\frac{d\,y(t)}{dt} + a_0\,y(t) = b_0\,u(t)$$

a_0 et b_0 sont les coefficients à déterminer d'après les observations de l'excitation $u(t)$ et de la réponse $y(t)$. Grâce à la méthode d'Euler, on peut transformer cette équation différentielle en système récurrent :

$$\frac{y_{n+1} - y_n}{h} \simeq \frac{d\,y(t)}{dt} = b_0 u_n - a_0 y_n \text{ (avec } h \to 0) \text{ et } t = nh$$

soit : $y_{n+1} = A\,y_n + B\,u_n$ | $b_0 = \dfrac{B}{h}$; $a_0 = \dfrac{1 - A}{h}$.

Le modèle y_{n+1} est linéaire par rapport à A et B, donc on peut lui appliquer la méthode des moindres carrés. Le lecteur trouvera un exemple d'application en [1].

V. MINIMISATION D'UN CRITERE QUADRATIQUE PAR PROGRAMMATION NON LINEAIRE.

1) Introduction.

L'intérêt de la Méthode de Gauss est de permettre la minimisation d'un critère quadratique en une seule étape. Au moyen de transformations appropriées ou grâce à des approximations, on peut transformer certains modèles non linéaires en des modèles équivalents linéaires. Malheureusement, en sciences physiques, la majorité des problèmes traités sont non linéaires et on doit envisager la minimisation du critère quadratique par d'autres procédés qui ne peuvent être mis en œuvre que par des moyens informatiques. Ces méthodes, qui permettent une recherche de l'optimum par approximations successives, sont regroupées sous l'appellation générale de Programmation Non Linéaire ou méthodes d'optimisation.

2) Exemple de modèle non linéaire.

Considérons le système suivant (fig. 9).

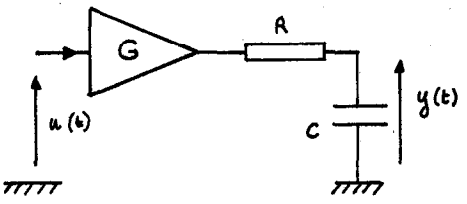

Fig. 9

Il symbolise par exemple un amplificateur de gain G avec coupure des fréquences hautes par le filtre RC.

Sa mise en équation conduit à l'équation différentielle :

$$\frac{d\,y(t)}{dt} + \frac{1}{RC}\,y(t) = \frac{G}{RC}\,u(t)$$

ou :

$$\frac{d\,y(t)}{dt} + a_0\,y(t) = b_0\,u(t)$$

dans le cas général.

On se propose de déterminer a_0 et b_0 à partir de la mesure de y et u. On peut travailler soit dans le domaine temporel, soit dans le domaine fréquentiel.

a) DOMAINE TEMPOREL.

Dans le cas général, l'excitation $u(t)$ est quelconque. On mesure $u(t)$ et $y^*(t)$. Connaissant $u(t)$, et pour des valeurs quelconques de a_0 et b_0, on peut calculer grâce à une méthode numérique (exemple : RUNGE-KUTTA 4) la réponse du modèle, soit $\hat{y}(t) = f(u(t), a_0, b_0)$. Un tel modèle est évidemment non linéaire et la minimisation du critère :

$$J = \int_{t_0}^{t_f} [y^*(t) - \hat{y}(t)]^2 \, dt$$

devra faire appel à des méthodes de P.N.L.

Le lecteur trouvera cet exemple traité avec programme BASIC en [1].

b) DOMAINE FRÉQUENTIEL.

On excite le système par des tensions sinusoïdales pour différentes fréquences, c'est-à-dire que l'on trace la courbe de réponse en fréquence de l'amplificateur, soit $\varrho^* = f(\omega)$.

Avec la notation imaginaire :

$$j\omega \underline{Y} + a_0 \underline{Y} = b_0 \underline{U}$$

soit :

$$\underline{k} = \frac{\underline{Y}}{\underline{U}} = \frac{b_0}{a_0 + j\omega} = \{\varrho, \theta\}.$$

Supposons, pour simplifier, que l'on ne s'intéresse qu'au module ϱ. On définit alors un modèle non linéaire :

$$\hat{\varrho}(\omega, a_0, b_0) = \frac{b_0}{\sqrt{a_0^2 + \omega^2}}.$$

Si on considère N couples de mesures :

$$\{\varrho_i^*, \omega_i\}$$

on peut définir un critère quadratique :

$$J = \sum_{i=1}^{N} \left(\varrho^*_{(\omega i)} - \frac{b_0}{\sqrt{a_0 + \omega_i^2}} \right)^2.$$

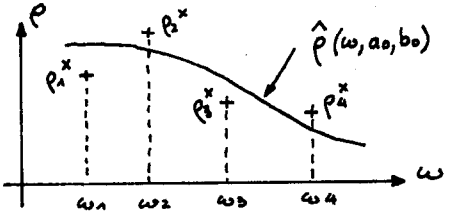

Fig. 10

On se propose de déterminer a_0 et b_0 pour que le critère J soit minimal.

3) **Méthode séquentielle.**

Pour l'exemple précédent, donnons une représentation au voisinage du minimum de $J = f(a_0, b_0)$ (fig. 11).

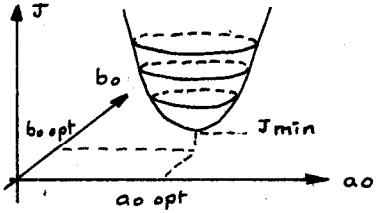

Fig. 11

Comme la représentation à 3 dimensions n'est pas commode, on utilise aussi la représentation par lignes de niveau de J (fig. 12).

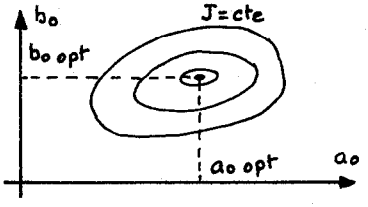

Fig. 12

Une méthode élémentaire de recherche de J_{min} est dite méthode séquentielle. A partir d'un couple initial $\{a_{0\,init}, b_{0\,init}\}$, on cherche J_{min} (ou J optimal) par étapes (fig. 13).

1 b_0 = cte recherche du min local de $J = f(a_0)$

2 a_0 = cte recherche du min local de $J = f(b_0)$

3 etc.

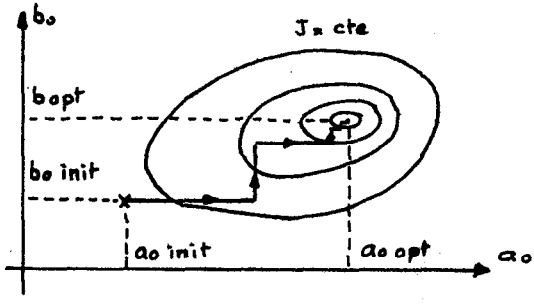

Fig. 13

On peut remarquer qu'il s'agit de la méthode employée pour rechercher l'équilibre d'un pont en alternatif (minimum de son) en réglant une résistance et une capacité, par exemple.

Recherche du minimum local :

Soit : b_0 = cte, a_0 variable.

Supposons J décroissant pour a_0 croissant.

On s'impose un incrément Δa de test. La recherche du minimum local conduit à l'algorithme de la fig. 14.

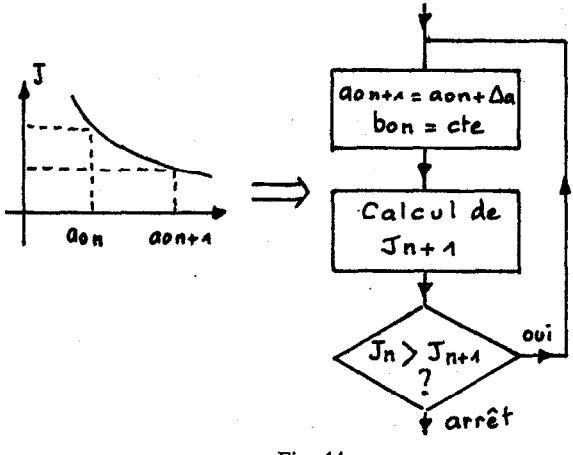

Fig. 14

Cette méthode est très simple à mettre en œuvre sur micro-ordinateur. Elle se généralise à n paramètres, cependant elle est relativement lente car un seul paramètre est testé par étape. La lenteur est due aussi au choix de l'incrément : on peut le choisir

grand, loin du minimum (rapidité), puis le diminuer lorsqu'on s'en approche (précision).

4) Méthode du gradient.

a) PRINCIPE DE LA MÉTHODE.

Considérons d'abord le problème de la recherche du minimum d'une fonction $f(x)$ (voir fig. 15).

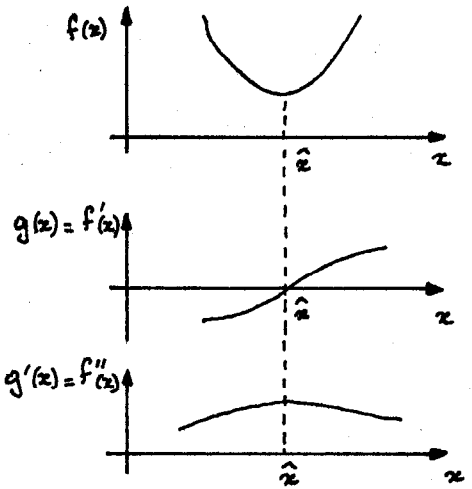

Fig. 15

Définissons $g(x) = f'(x)$.

Chercher le minimum de $f(x)$ est équivalent à chercher le zéro de $g(x)$.

Considérons l'algorithme suivant :

soit : x_n la valeur de x à l'étape n.

Définissons : $x_{n+1} = x_n - \mathrm{K}g(x_n)$ avec $\mathrm{K} > 0$:

— si : $x_n > \hat{x}$, $g(x_n) > 0$

alors :

$x_{n+1} < x_n$ et $f(x) \searrow$

— si : $x_n < \hat{x}$, $g(x_n) < 0$

alors :

$x_{n+1} > x_n$ et $f(x) \searrow$

Donc, si la valeur de K est correctement choisie, x_{n+1} se rapproche de $\hat{x}$.

Cet algorithme porte le nom de méthode du gradient :

$$x_{n+1} = x_n - K f'(x_n) \qquad \text{avec} \qquad K > 0$$

($K < 0$ correspond à la recherche d'un maximum).

Il nécessite le calcul de $f'(x_n)$. Si le calcul direct n'est pas possible, on approche $f'(x_n)$ par l'expression :

$$f'(x_n) \simeq \frac{f(x_{n+h}) - f(x_n)}{h} \qquad \text{avec} \qquad h \to 0$$

b) APPLICATION.

Dans l'exemple précédent, $\hat{\varrho}(\omega) = \dfrac{b_0}{\sqrt{a_0^2 + \omega^2}}$ et :

$$J = \sum_1^N \left(\varrho^*(\omega_i) - \frac{b_0}{\sqrt{a_0^2 + \omega_i^2}} \right)^2 \qquad \Rightarrow J = f(a_0, b_0),$$

donc on doit calculer $\dfrac{\partial J}{\partial a_0}$ et $\dfrac{\partial J}{\partial b_0}$.

On obtient :

$$\left| \begin{array}{l} \dfrac{\partial J}{\partial a_0} = -2 \sum_{i=1}^N (\varrho_i^* - \varrho_i) \dfrac{\partial \hat{\varrho}_i}{\partial a_0} \\ \dfrac{\partial J}{\partial b_0} = -2 \sum_{i=1}^N (\varrho_i^* - \varrho_i) \dfrac{\partial \hat{\varrho}_i}{\partial b_0} \end{array} \right.$$

avec :

$$\left| \begin{array}{l} \dfrac{\partial \hat{\varrho}_i}{\partial a_0} = -\dfrac{b_0 a_0}{(a_0^2 + \omega_i^2)^{3/2}} \\ \dfrac{\partial \hat{\varrho}_i}{\partial b_0} = \dfrac{1}{\sqrt{a_0^2 + \omega_i^2}}. \end{array} \right.$$

Finalement, l'algorithme à 2 variables s'écrit :

$$\left| \quad a_{0\,n+1} = a_{0\,n} - K \left(\frac{\partial J}{\partial a_0} \right)_n \right.$$

$$b_{0\,n+1} = b_{0\,n} - K \left(\frac{\partial J}{\partial b_0} \right)_n$$

On a représenté sur la fig. 16 l'évolution du point $\{a_{0\,n}, b_{0\,n}\}$: il se déplace sur la ligne de plus grande pente, donc plus rapidement que par la méthode séquentielle.

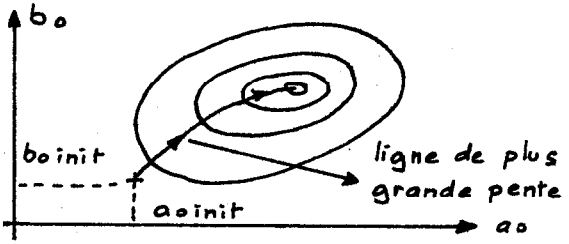

Fig. 16

c) CHOIX DE K.

Un des problèmes posé par l'emploi de la méthode du gradient consiste dans le choix optimal de K. Si K est trop grand, on a un risque d'instabilité (oscillations autour de l'optimum).

Si K est trop faible, la convergence est assurée, mais elle sera très lente (elle se ralentit au voisinage du minimum car les dérivées tendent alors vers 0).

5) Méthode du gradient de type Newton.

a) PRINCIPE DE LA MÉTHODE.

On peut améliorer la méthode du gradient en utilisant un algorithme plus perfectionné pour la recherche du zéro de $g(x) = f'(x)$ (fig. 16).

Employons la formule de Taylor :

$$g(x_{n+h}) = g(x_n) + h\, g'(x_n) + \cdots$$

Cherchons l'accroissement h tel que $x_{n+h} = \hat{x}$; alors :

$$g(x_{n+h}) = 0 \quad \text{et} \quad h = -\frac{g(x_n)}{g'(x_n)}$$

soit :

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}.$$

Il s'agit tout simplement de la méthode de Newton pour la recherche du zéro d'une fonction $g(x)$. En revenant à $f(x)$, l'algorithme s'écrit :

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}.$$

Il nécessite le calcul de $f'(x_n)$ et $f''(x_n)$. En pratique, on peut se dispenser du calcul exact de $f''(x_n)$ comme on le verra par la suite.

Si on compare cet algorithme à celui du gradient, on s'aperçoit que $K = 1/f''(x_n)$, il s'agit d'une méthode de gradient, mais à K variable. Si $f''(x_n)$ est calculé de manière très approchée, l'algorithme convergera malgré tout car il reste du type gradient.

Remarque n° 1.

Supposons que $f(x)$ soit de type parabolique au voisinage de $\hat{x}$. Alors :

$$\begin{cases} f(x) = a(x-\hat{x})^2 + b \\ f'(x) = 2a(x-\hat{x}) \\ f''(x) = 2a \end{cases}$$

l'algorithme devient : $x_{n+1} = x_n - \dfrac{2a(x_n-\hat{x})}{2a} = \hat{x}$, c'est-à-dire que, quelque soit x_n, $x_{n+1} = \hat{x}$.

En pratique, même si l'hypothèse parabolique n'est pas vérifiée, on constate une grande rapidité de convergence.

Remarque n° 2.

Considérons la fig. 17.

— si $x_{init} < x_3$ alors $x_{n+1} = x_n - \dfrac{f'(x_n)}{f''(x_n)}$,

avec $K = \dfrac{1}{f''(x_n)} > 0$, donc convergence ;

— si $x_{init} > x_3$ alors $K = \dfrac{1}{f''(x_n)} < 0$,

et x_{n+1} se rapproche de $\hat{x}_2$ (maximum), donc diverge.

Par contre, l'algorithme du gradient convergerait vers $\hat{x}_1$ pour $x_{init} < \hat{x}_2$.

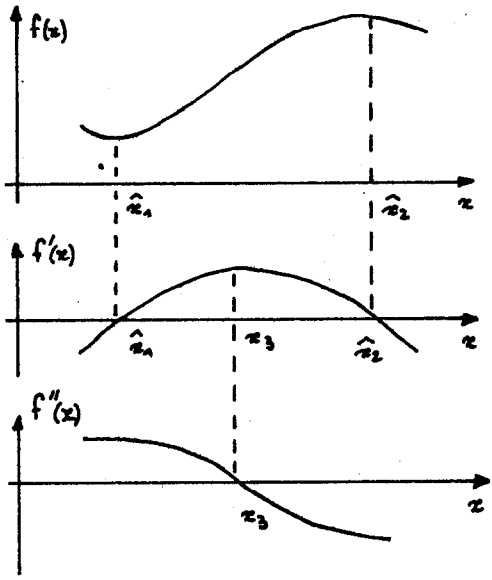

Fig. 17

On constate donc un risque de divergence de l'algorithme si $x_{initial}$ est mal choisi.

En pratique, on utilisera cet algorithme en raison de sa grande rapidité de convergence, malgré un volume de calcul plus important et un risque de divergence si l'estimation initiale est mal conditionnée.

b) APPLICATION.

J est une fonction des deux variables a_0 et b_0.

Donc on doit obtenir simultanément $J'_{a_0} = 0$ et $J'_{b_0} = 0$ en employant la méthode de Newton. C'est-à-dire que l'on doit à chaque étape calculer les accroissements Δa et Δb vérifiant le système linéaire :

$$J'a_0(\hat{a_0}, \hat{b_0}) = 0 = J'_{a_0}(a_{0\,n}, b_{0\,n}) + \Delta a\, J''_{a_0, a_0} + \Delta b\, J''_{a_0, b_0} + \cdots$$

$$J'a_0(\hat{a_0}, \hat{b_0}) = 0 = J'_{b_0}(a_{0\,n}, b_{0\,n}) + \Delta a\, J''_{b_0, a_0} + \Delta b\, J''_{b_0, b_0} + \cdots$$

on obtient ensuite :

$$\left| \begin{array}{l} a_{0\,n+1} = a_{0\,n} + \Delta a \\ b_{0\,n+1} = b_{0\,n} + \Delta b. \end{array} \right.$$

On doit calculer les dérivées secondes J''_{a_0, a_0}, J''_{a_0, b_0}, J''_{b_0, b_0},

$$J''_{a_0, a_0} = \frac{\partial^2 J}{\partial {a_0}^2} = 2 \sum_{i=1}^{N} \frac{\partial \hat{Q}_i}{\partial a_0} \cdot \frac{\partial \hat{Q}_i}{\partial a_0} - 2 \sum_{i=1}^{N} (Q_i^* - \hat{Q}_i) \frac{\partial^2 \hat{Q}_i}{\partial {a_0}^2}.$$

Au voisinage du minimum $Q_i^* \simeq \hat{Q}_i$, alors on néglige le 2e terme, ce qui évite le calcul de $\frac{\partial^2 \hat{Q}}{\partial {a_0}^2}$.

Soit :

$$\left| \begin{array}{l} J''_{a_0, a_0} \simeq 2 \sum_{1}^{N} \left(\frac{\partial \hat{Q}_i}{\partial a_0} \right)^2 \\ J''_{a_0, b_0} = J''_{b_0, a_0} \simeq 2 \sum_{1}^{N} \frac{\partial \hat{Q}_i}{\partial a_0} \cdot \frac{\partial \hat{Q}_i}{\partial b_0} \\ J''_{b_0, b_0} \simeq 2 \sum_{1}^{N} \left(\frac{\partial \hat{Q}_i}{\partial b_0} \right)^2 \end{array} \right.$$

Pour cet exemple, on aurait pu calculer exactement les dérivées secondes. Dans le cas général, cette approximation évite le calcul de dérivées secondes numériques du modèle qui sont très imprécises.

L'algorithme n'est plus exactement du type Newton, mais il s'avère néanmoins nettement plus rapide que l'algorithme du type gradient.

c) INITIALISATION.

Nous avons montré dans le cas monodimensionnel les risques de divergence de cet algorithme pour une mauvaise initialisation. Cela signifie en pratique qu'il faudra réaliser une prémodélisation par des méthodes élémentaires. Dans le cas du filtre RC, on utilisera par exemple les propriétés classiques telles que fréquence de coupure, pente des asymptotes, etc.

Si cette information initiale n'est pas suffisante, on pourra l'affiner par une recherche de type gradient.

Toutes les méthodes de recherche peuvent converger vers des minimums locaux au lieu du minimum absolu. Pour l'éviter, on

pourra effectuer plusieurs recherches correspondant à des initialisations différentes.

6) **Exemple d'application.**

On a étudié les algorithmes gradient et Newton, sur un exemple pour lequel on connaissait les valeurs exactes de a_0 et b_0 ($a_0 = 1$; $b_0 = 1$). On a d'abord comparé la rapidité de convergence des 2 algorithmes à partir de la même estimation initiale $\{a_{0\,init} = 1{,}1,\ b_{0\,init} = 0{,}95\}$ et pour N = 12 points de mesure connus sans erreur. Les résultats sont consignés sur la fig. 18.

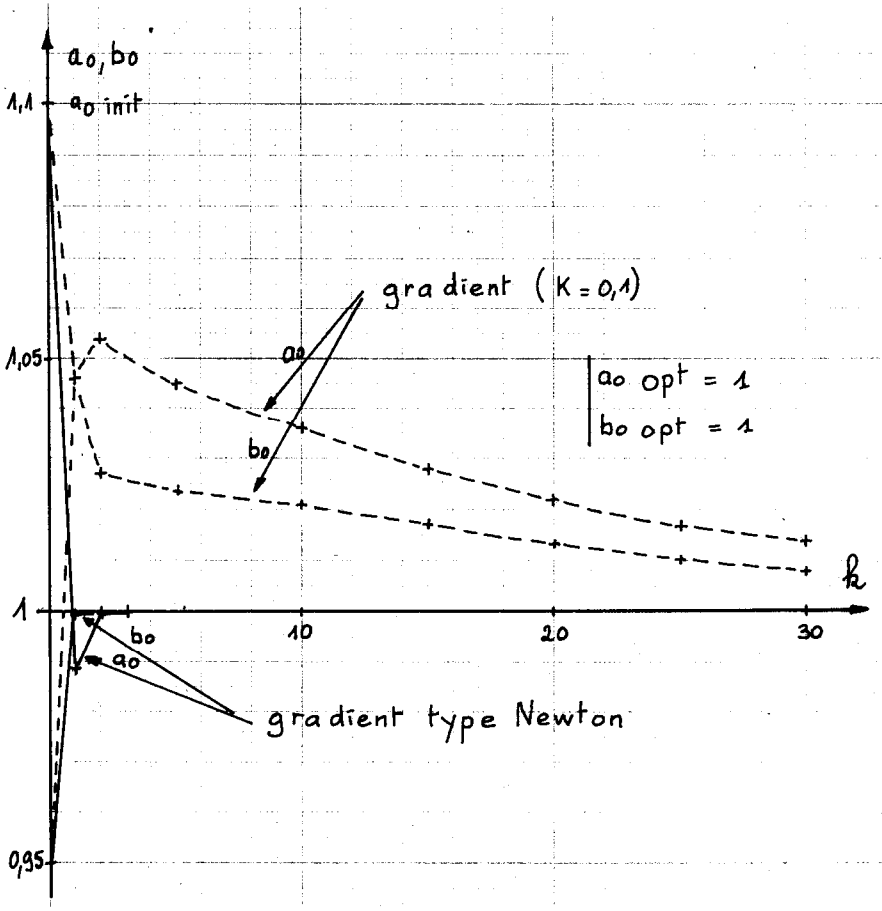

Fig. 18

Pour K = 0,1, l'algorithme du gradient converge, mais très lentement.

Pour K = 0,2, il oscille autour du minimum. Par contre, l'algorithme de type Newton (avec approximation de la dérivée seconde) converge en 3 étapes.

On a ensuite appliqué ce deuxième algorithme à des mesures avec erreur (fig. 19).

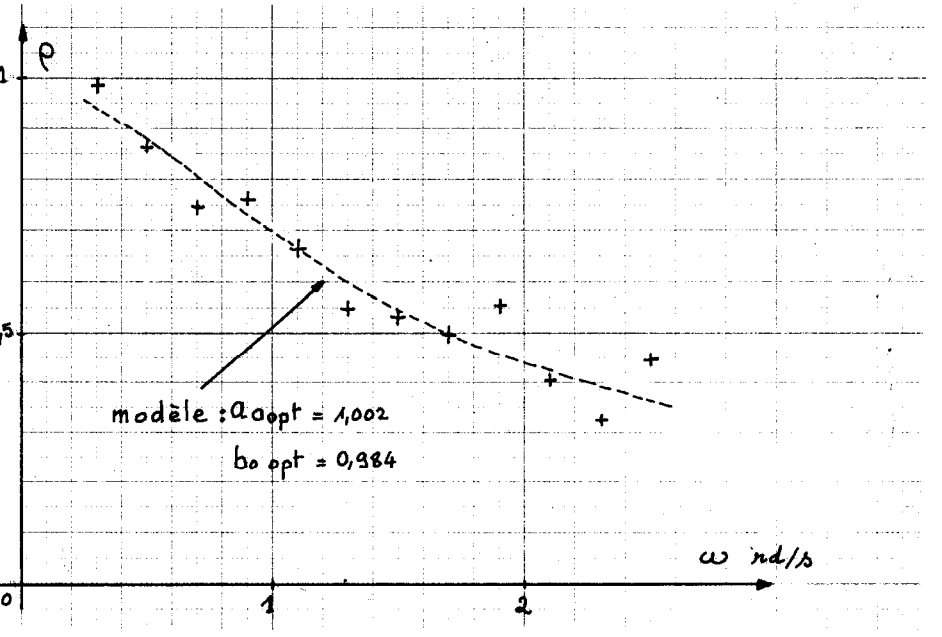

Fig. 19

La courbe en pointillé de la fig. 19 correspond au modèle obtenu au bout de 3 itérations. Les résultats sont consignés dans le tableau suivant :

	Estimation initiale	1re étape	2e étape	3e étape
a_0	0,9	0,975	0,999	1,002
b_0	1,1	0,977	0,983	0,984
J	0,204	0,0312	0,0306	0,0306

Le lecteur intéressé trouvera les programmes BASIC correspondants avec de plus amples détails de calcul en [1].

CONCLUSION.

Le sujet traité étant très vaste, cet exposé n'a pas la prétention d'être exhaustif. En particulier des points importants on été seulement signalés ou passés sous silence : sensibilité du critère par rapport aux paramètres, précision, propriétés statistiques. Le lecteur intéressé par ces sujets pourra trouver les compléments nécessaires en bibliographie.

Les exemples proposés sont du niveau post-baccalauréat (classes prép., D.E.U.G., I.U.T., B.T.S.) pour la partie théorique, par contre les programmes peuvent être exploités dès la terminale. Ces exemples sont principalement du domaine de l'électricité, mais on trouvera facilement d'autres applications en Sciences physiques, particulièrement en mécanique, pour le traitement des résultats expérimentaux.

BIBLIOGRAPHIE

PROGRAMMES EN BASIC.

[1] « *Traitements numériques et gestion des données expérimentales* ». Atelier Informatique. Journées de l'U.d.P. Strasbourg, octobre 1984.

MÉTHODES NUMÉRIQUES.

[2] « *Méthodes de calcul numérique* ». J.-P. NOUGIER. Masson 1983.

MODÉLISATION EXPÉRIMENTALE (IDENTIFICATION).

[3] « *Cours d'introduction aux méthodes d'identification* ». M. CUENOD et J.-L. FATTO. Albin Michel 1973.

[4] « *Identification des processus* ». L. POVY. Dunod 1975.

[5] « *Automatique des systèmes linéaires.* Tome 2 : *Identification* ». P. DE LARMINAT et Y. THOMAS. Flammarion 1977.

[6] « *Identification des processus par la méthode du modèle* ». J. RICHALET, A. RAULT et R. POULIQUEN. Gordon and Breach 1971.

[7] « *System identification, parameter and state estimation* ». P. EYKHOFF. Wiley 1974.

PROGRAMMATION NON LINÉAIRE.

[8] « *Introduction à la pratique de l'optimisation* ». L. PUN. Dunod 1972.

[9] « *Commande optimale des processus.* Tome 2 : *Programmation non linéaire et ses applications* ». R. BOUDAREL, J. DELMAS et P. GUICHET. Dunod 1968.