

CHAPITRE 7. MODELES MA.

Après les deux chapitres précédents, qui traitaient l'identification du modèle autorégressif ou de la partie autorégressive du modèle ARMA, celui-ci est consacré à l'estimation du modèle à moyenne ajustée MA (moving-average). Il pourra s'agir indifféremment d'un signal uniquement MA comme d'un signal ARMA dont la partie autorégressive a déjà été identifiée et dont seule subsiste la partie MA obtenue par filtrage inverse du signal y_t d'origine. Deux méthodes ont été développées pour déterminer les paramètres du modèle MA. La première en réalise une approximation par un modèle autorégressif d'ordre élevé, suivie par l'approximation en retour du modèle AR par un modèle MA qui sera la solution cherchée. Cette méthode s'apparente à celle établie pour le modèle MA stationnaire par DURBIN, 1959, SHANKS, 1967. La seconde méthode repose sur la décomposition de Schur de la transformée en z de la suite des matrices estimant la corrélation du vecteur Y_t projection du signal y_t . En d'autres termes, il s'agira d'utiliser une forme de modélisation MA du vecteur Y_t . Le passage de cette décomposition ou de ce modèle vectoriel au modèle MA scalaire non-stationnaire se fait par une sélection sur les lignes des matrices du modèle vectoriel, en maximisant la vraisemblance de covariances déduites des observations relativement à celles déduites du modèle. Une troisième méthode concernera un cas particulier, celui du modèle MA d'ordre zéro ! Modèle qui est tout à fait trivial dans le cas stationnaire, où le coefficient b_0 s'identifie au gain du modèle AR et n'est autre que l'écart-type de l'erreur de prédiction. Modèle cependant où la non-stationnarité du coefficient $b_0(t)$ forcera inéluctablement l'estimateur à la non-linéarité. La solution tiendra en une transformation

non-linéaire des données après laquelle un estimateur linéaire constituera un bon compromis.

1. Approximation AR long.

La méthode de modélisation MA par approximation au moyen d'un modèle autorégressif d'ordre élevé se situe dans un groupe de méthodes qui possède plusieurs concurrents. L'identification de modèles MA, de parties MA de modèle ARMA ou même de modèles complets ARMA se fait par des méthodes qu'on peut ranger en trois groupes: les méthodes par approximation, par maximum de vraisemblance et enfin par factorisation. Le premier groupe sera analysé dans ce qui suit. Le second, celui des méthodes de maximum de vraisemblance cherche à maximiser la probabilité conditionnelle des observations étant donné les paramètres à estimer. Ceci correspond à deux approches distinctes, l'une temporelle, l'autre spectrale. La première cherche donc à maximiser $p(y_0 \dots y_t / \theta)$, tandis que l'autre travaille sur le périodogramme. Ces méthodes seront décrites dans l'annexe 7. Quant aux méthodes issues de la factorisation, elles regroupent deux aspects. Le premier consiste à factoriser la covariance ou le spectre du signal pour en extraire la partie causale, à phase minimale et stable. Le second consiste à chercher une réalisation sous forme de variables d'état du modèle. On peut alors interpréter la recherche des covariances de l'état, dans ce modèle, au moyen d'une équation de Ricatti, comme une factorisation, ce qui unifie les deux aspects.

1.1. Passage AR long à MA.

Dans le cas de l'approximation par un modèle autorégressif long, une fois constatée l'équivalence entre un modèle MA d'ordre q fini et un modèle

AR d'ordre infini, la question qui se pose est de savoir comment repasser du modèle AR, approximé par un modèle d'ordre fini mais élevé, au modèle MA cherché. On trouve dans le cas du modèle stationnaire trois moyens pour cela. Le premier consiste à déterminer par division polynomiale suivant les puissances croissantes, la réponse impulsionnelle du modèle AR, et le passage au modèle MA se fait soit par troncature de la réponse impulsionnelle à ses $q+1$ premiers termes, soit par approximation de Padé (algorithme de Levinson à l'envers, c'est-à-dire en faisant décroître l'ordre de la réponse impulsionnelle jusqu'à q). Pour les modèles non-stationnaires il est clair que cette démarche ne peut s'appliquer car la réponse impulsionnelle du filtre AR est à deux indices et la division polynomiale n'a pas de sens.

Une seconde méthode utilisée dans le cas stationnaire est due à DURBIN, 1959. Elle consiste à réaliser l'inversion AR \rightarrow MA au niveau du spectre, en disant que si les spectres des deux modèles sont identiques, leurs inverses le sont aussi, et le modèle MA, devenu AR doit être une approximation autorégressive du spectre anciennement AR et devenu MA. La corrélation de ce nouveau signal est la corrélation de la suite des coefficients du modèle autorégressif long. Par l'argument d'inversion, un algorithme de Levinson appliqué à cette corrélation donne le modèle à moyenne ajustée cherché. Malgré les bonnes performances et la simplicité de cette méthode dans le cas MA stationnaire il faut malheureusement la rejeter ici pour une raison analogue à la précédente: les covariances dépendent ici de deux indices, et les spectres sont transformés en reliefs, fonctions des deux variables temps et fréquence.

La troisième méthode est la seule qui soit applicable dans le contexte non-stationnaire car elle travaille exclusivement dans le domaine temporel,

directement sur les échantillons. Une fois le modèle autorégressif identifié, celui-ci permet par un filtrage inverse du signal d'obtenir le résidu du modèle AR. Si ce modèle était exactement substituable au modèle MA, le résidu après filtrage serait l'innovation du signal, et donc l'entrée du modèle MA. On fait l'hypothèse que cette entrée approxime l'entrée réelle aussi bien que le modèle AR approxime le modèle MA. L'estimation de ce dernier est alors un problème linéaire d'identification d'un modèle dont l'entrée et la sortie sont connues, qui se fait en utilisant par exemple un critère des moindres carrés. Ceci est la démarche de SHANKS, 1967, ainsi que de STEIGLITZ, 1977, MAYNE, FIROOZAN, 1978, et KOVANLINKA, MATAUSEK, 1979. Voyons comment celle-ci se transpose dans le cas non-stationnaire.

1.1.1. Résolution par moindres carrés.

L'entrée du modèle MA, est ϵ_t , obtenue par filtrage inverse de y_t par le modèle AR identifié par une des méthodes des chapitres précédents. Le modèle s'écrit:

$$(2-102) \quad y_t = b_0(t)\epsilon_t + b_1(t-1)\epsilon_{t-1} + \dots + b_q(t-q)\epsilon_{t-q}$$

avec
$$b_i(t) = \sum_{j=0}^m b_{i,j} f_j(t)$$

On introduit alors comme pour le modèle AR un vecteur produit du signal par chaque fonction de la base, mais ici le produit est sur l'entrée. Soit E_t le vecteur obtenu:

$$(2-103) \quad E_t = [f_0(t)\epsilon_t \dots f_m(t)\epsilon_t]^T$$

Avec ce nouveau vecteur le modèle se réécrit:

$$(2-104) \quad Y_t = [E_t^T \dots E_{t-q}^T] \theta$$

où le vecteur θ contient les paramètres inconnus b_{ij} : $\theta = [b_{00} \dots b_{qm}]^T$

Cependant les erreurs sur ϵ_t font que (2-104) n'est pas satisfait. Une erreur résiduelle v_t apparaît:

$$(2-105) \quad v_t = y_t - [E_t^T \dots E_{t-q}^T] \theta$$

Si on minimise cette erreur v_t , au sens d'un critère des moindres carrés, cela signifie qu'on réalise l'estimation de θ comme une identification de système avec bruit sur la sortie. Les équations deviennent, en utilisant directement une somme temporelle pour l'espérance mathématique:

$$(2-106) \quad \sum_{t \in \Gamma} \begin{bmatrix} E_t \\ E_{t-1} \\ \cdot \\ \cdot \\ E_{t-q} \end{bmatrix} [E_t^T \dots E_{t-q}^T] \theta = \sum_{t \in \Gamma} \begin{bmatrix} E_t \\ E_{t-1} \\ \cdot \\ \cdot \\ E_{t-q} \end{bmatrix} y_t$$

La structure de la matrice multipliant θ dans (2-106) sera suivant le choix de Γ une structure de Toeplitz ou bien de produit de deux matrices de Toeplitz. Par conséquent, la résolution de (2-106) se fera par un algorithme rapide du type de ceux exposés au chapitre précédent.

1.1.2. Résolution par modèles propres.

Notons qu'il est possible de raffiner l'estimation de θ en supposant l'existence d'un second bruit w_t additif, qui lui, affecterait l'entrée ϵ_t , les deux bruits v_t et w_t se partageant les erreurs sur le modèle et sur le signal y_t lui-même (bruits de quantification par exemple). Le modèle MA aura alors pour entrée et pour sortie deux signaux e_t et s_t dont on ne connaît qu'une version bruitée ϵ_t ou y_t :

$$(2-107) \quad s_t = [E_t^T \dots E_{t-q}^T] \theta$$

$$y_t = s_t + v_t$$

$$\epsilon_t = e_t + w_t$$

$$E(v_t) = 0, \quad E(v_t^2) = V$$

$$E(w_t) = 0, \quad E(w_t^2) = W$$

$$\text{avec } E_t = [f_0(t)e_t \dots f_m(t)e_t]^T$$

$$E_t^* = [f_0(t)\epsilon_t \dots f_m(t)\epsilon_t]^T$$

$$Y_t = [f_0(t)y_t \dots f_m(t)y_t]^T$$

$$W_t = [f_0(t)w_t \dots f_m(t)w_t]^T$$

En réécrivant (2-107):

$$(2-108) \quad [-s_t \quad E_t^T \dots E_{t-q}^T] \begin{bmatrix} 1 \\ \theta \end{bmatrix} = 0$$

on peut introduire la covariance du vecteur mesuré équivalent à (2-108):

$$(2-109) \quad R = E \left(\begin{bmatrix} -y_t \\ E_t^* \\ \cdot \\ E_{t-q}^* \end{bmatrix} [-y_t \quad E_t^* \dots E_{t-q}^*] \right)$$

Du fait que v_t et w_t sont des bruits centrés et indépendants de s_t et e_t , on obtient en tenant compte de (2-107) la valeur suivante pour R :

$$(2-110) \quad R = E \left(\begin{bmatrix} -s_t \\ E_t \\ \cdot \\ E_{t-q} \end{bmatrix} [-s_t \quad E_t^T \dots E_{t-q}^T] \right) + E \left(\begin{bmatrix} -v_t \\ w_t \\ \cdot \\ w_{t-q} \end{bmatrix} [-v_t \quad w_t^T \dots w_{t-q}^T] \right)$$

En introduisant dans (2-110) le vecteur θ à partir de (2-108), on obtient:

$$(2-111) \quad R \begin{bmatrix} 1 \\ \theta \end{bmatrix} = E \left(\begin{bmatrix} -v_t \\ W_t \\ \cdot \\ \cdot \\ W_{t-q} \end{bmatrix} \begin{bmatrix} -v_t & W_t^T & \dots & W_{t-q}^T \end{bmatrix} \right) \begin{bmatrix} 1 \\ \theta \end{bmatrix}$$

Les propriétés des bruits v_t et w_t permettent alors de remplacer dans R l'espérance mathématique par une somme ergodique ou du moins temporelle, et d'obtenir pour θ le système suivant:

$$(2-112) \quad R \begin{bmatrix} 1 \\ \theta \end{bmatrix} = \text{Diag}(V, L, W, F_0, \dots, W, F_q) \begin{bmatrix} 1 \\ \theta \end{bmatrix}$$

Dans (2-112), L désigne le nombre de points dans l'intervalle Γ de sommation, et les matrices F_i sont des blocs carrés à $(m+1)$ lignes et colonnes:

$$(2-113) \quad F_i = \sum_{t \in \Gamma} \begin{bmatrix} f_0(t-i) \\ \cdot \\ \cdot \\ f_m(t-i) \end{bmatrix} \begin{bmatrix} f_0(t-i) & \dots & f_m(t-i) \end{bmatrix}$$

On constate que si $W=0$, le système (2-112) est équivalent au système (2-106) des moindres carrés. Si W est non nul, le système à résoudre est plus complexe et s'apparente à un problème de vecteurs propres généralisés, avec une difficulté supplémentaire liée à l'existence de deux bruits différents V et W dans la matrice diagonale par blocs figurant à droite de (2-112). On ne peut donc identifier θ qu'à la condition d'introduire une contrainte liant V à W , en posant $V=k.W$ où k est un facteur connu a priori, par exemple $k=1$ ou encore k égal au rapport des variances de y_t et ϵ_t . On réécrit alors (2-112) sous la forme:

$$(2-114) \quad R \begin{bmatrix} 1 \\ \theta \end{bmatrix} = W \begin{bmatrix} k_L & & 0 \\ & F_0 & \\ 0 & & F_q \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix}$$

Cette équation se ramène à un problème ordinaire de vecteurs propres en posant:

$$(2-115) \quad u = \begin{bmatrix} k_L^{1/2} & & 0 \\ & F_0^{T/2} & \\ 0 & & F_q^{T/2} \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix}$$

$$(2-116) \quad R^* = \begin{bmatrix} k_L^{-1/2} & & 0 \\ & F_0^{-T/2} & \\ 0 & & F_q^{-T/2} \end{bmatrix} R \begin{bmatrix} k_L^{-1/2} & & 0 \\ & F_0^{-1/2} & \\ 0 & & F_q^{-1/2} \end{bmatrix}$$

où $F_0^{1/2} \cdot F_0^{T/2} = F_0$ désigne toute décomposition de F_0 comme produit d'une matrice par sa transposée. Par exemple $F_0^{1/2}$ peut être le facteur triangulaire inférieur de la décomposition de Cholesky de la matrice F_0 qui est positive. $F_0^{1/2}$ peut aussi être obtenue à partir de la diagonalisation de F_0 : $F_0 = UD^2U^T$ où U est unitaire, D diagonale, on prendra alors $F_0^{1/2} = UD$. Avec ces notations, le système (2-114) s'écrit simplement $R^* u = Wu$, ce qui montre que u est vecteur propre de R , et la nécessité pour $(R-WI)$ d'être une matrice non-négative implique pour W d'être la plus petite valeur propre (positive) de R . La méthode qui vient d'être décrite permet donc une estimation de θ , qui sera plus fine que par les moindres carrés, mais on voit que la linéarité est perdue par l'introduction de w_t^* , ce qui implique la

diagonalisation d'une matrice R^* , opération lourde et coûteuse. De plus cette méthode nécessite une hypothèse supplémentaire liant v_t et w_t . Ces limitations amènent à préférer à cette méthode la méthode plus rapide des moindres carrés. Notons encore que cette méthode peut aussi s'écrire si v_t et w_t ne sont plus ni blancs ni décorréllés entre eux.

1.2. Passage de l'AR long à un ARMA.

La méthode qui vient d'être décrite concernait le modèle MA, mais on peut bien sûr l'envisager pour l'identification du modèle ARMA complet. On approximera alors le modèle dont on veut estimer les coefficients a_{ij} et b_{ij} par un modèle AR d'ordre élevé, on extraira par filtrage inverse du signal y_t le résidu ϵ_t qui sera considéré comme l'entrée du modèle ARMA dont y_t est la sortie. L'introduction des deux vecteurs de projection Y_t et E_t permet d'écrire le modèle sous la forme:

$$(2-117) \quad y_t = [-Y_{t-1}^T \dots -Y_{t-p}^T \quad E_t^T \dots E_{t-q}^T] \theta$$

avec $\theta = [a_{10} \dots a_{1m} \dots a_{pm} \quad b_{00} \dots b_{qm}]$

De cette équation, on tirera le vecteur θ solution au sens des moindres carrés de l'erreur commise dans (2-117):

$$(2-118) \quad \left(\sum_{t \in \Gamma} \Phi_t^T \Phi_t \right) \theta = - \sum_{t \in \Gamma} \Phi_t^T y_t$$

avec $\Phi_t^T = [-Y_{t-1}^T \dots -Y_{t-p}^T \quad E_t^T \dots E_{t-q}^T]$

Malheureusement, il est connu dans le domaine de l'identification des systèmes entrée-sortie que cette méthode de moindres carrés est biaisée. On peut rendre l'estimation non-biaisée par une procédure dite de moindres carrés généralisés. Ceci correspond dans le cas stationnaire à plusieurs approches. Celle de STEIGLITZ, 1977, s'applique sur la réponse impulsion-

nelle du modèle et ne se transpose pas au cas non-stationnaire: il s'agit d'une procédure itérative dans laquelle on filtre la réponse impulsionnelle par l'inverse du dénominateur trouvé à l'itération précédente. Une procédure analogue est donnée par MAYNE, FIROOZAN, 1978, et consiste à travailler directement sur les signaux y_t et ϵ_t , ce qui permet d'étendre leur approche au cas non-stationnaire. Un premier modèle ARMA est identifié entre y_t et ϵ_t , avec comme partie autorégressive A_1 et comme moyenne ajustée B_1 . L'erreur d'équation obtenue sera notée $\epsilon_t^{(1)}$, c'est la différence entre $A_1 y_t$ et $B_1 \epsilon_t$. On filtre ensuite y_t par B_1 ce qui donne $y_t^{(1)}$ et c'est entre $\epsilon_t^{(1)}$ et $y_t^{(1)}$ qu'est identifié le modèle ARMA final. MAYNE, FIROOZAN, 1978, montrent qu'un tel estimateur est sans biais. La procédure de KONVALINKA, MATAUSEK, 1979, est assez voisine. A chaque itération, la partie MA précédemment obtenue permet d'extraire un signal $y_t^{(n)}$ de y_t filtré par l'inverse de la partie MA. Sur $y_t^{(n)}$ est estimé le modèle AR d'ordre élevé, et le résidu $\epsilon_t^{(n)}$ en est déduit, les parties AR et MA du nouveau modèle sont alors obtenues entre $\epsilon_t^{(n)}$ et y_t par moindres carrés. Bien qu'elle travaille sur le signal lui-même cette procédure ne peut pourtant pas se transposer au cas non-stationnaire car il n'existe pas de moyen simple de passer du modèle MA au filtre à réponse impulsionnelle infinie qui serait son inverse, c'est-à-dire tel que leur cascade transfère identiquement le signal de l'entrée à la sortie. La même restriction s'oppose à l'emploi de la méthode de MILLS, MULLIS, ROBERTS, 1980, qui utilise une estimation alternative des parties AR et MA pour des ordres croissants.

2. Factorisation MA non-stationnaire.

La famille de méthodes d'identification de modèles MA ou ARMA, qui vient d'être décrite reposait sur une approximation par un modèle

autorégressif d'ordre élevé. Celle qui va maintenant être décrite est construite autour de l'idée de factorisation spectrale, autre méthode de modélisation MA dans le cas stationnaire. On verra que le passage au cas non-stationnaire se fait par l'interprétation de cette factorisation spectrale comme un recours implicite à un modèle autorégressif dont on fait tendre l'ordre vers l'infini. Ces méthodes de factorisation spectrale sont issues des travaux de SCHUR, 1917, sur la décomposition des fonctions analytiques, à module borné. Dans le domaine du traitement du signal, la décomposition de Schur a été employée par plusieurs auteurs, mais cet emploi était implicite et n'a été reconnu que postérieurement. Il s'agit de l'algorithme donné par RISSANEN, 1973, pour la factorisation de matrices de Toëplitz par blocs, et leur inversion. C'est aussi le cas de l'algorithme donné par LE ROUX, GUEGUEN, 1976, permettant le calcul du modèle autorégressif en virgule fixe sous la forme de ses coefficients de corrélation partielle ou coefficients de réflexion. Leur algorithme, formellement équivalent à la décomposition de Schur, offre l'avantage de ne faire apparaître que des quantités possédant une interprétation simple en terme d'autocorrélations et d'intercorrélations du signal y_t et des résidus ou erreurs de prédiction. Ceci autorise une démonstration simple de la convergence du facteur obtenu vers le modèle MA, et plus particulièrement de la propriété de phase minimale du facteur (LE ROUX, GRENIER, 1980). Une démonstration algébrique de cette convergence a été élaborée par YOULA, KAZANJIAN, 1978. Il faut aussi remarquer que cet algorithme calcule les mêmes quantités intermédiaires que l'algorithme de BAUER, 1955, pour la factorisation de Cholesky d'une matrice de Toëplitz bande, mais ici les calculs ont un coût de q opérations par itération, si q est d'ordre du modèle MA, au lieu des q^2 opérations de l'algorithme de BAUER, 1955. Avant de montrer

comment la factorisation MA par décomposition de Schur se transpose dans le cadre du modèle non-stationnaire, il est nécessaire de rappeler cette décomposition dans le cas vectoriel, où se pose le problème de la normalisation du résidu. C'est ce qui va être fait maintenant.

2.1. Factorisation vectorielle.

2.1.1. Algorithme de Levinson implicite.

Le modèle autorégressif vectoriel peut s'écrire conformément à (2-119), où les A_i^+ sont les matrices du modèle direct (forward) et les A_i^- sont celles du modèle rétrograde (backward), ϵ_t^+ et ϵ_t^- sont les résidus direct et rétrograde.

$$(2-119) \quad \begin{bmatrix} I & A_1^+ & \dots & A_p^+ \\ A_p^- & \dots & A_1^- & I \end{bmatrix} \begin{bmatrix} Y_t \\ Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{bmatrix} = \begin{bmatrix} \epsilon_t^+ \\ \epsilon_t^- \end{bmatrix}$$

La minimisation de la trace de la variance de ϵ_t^+ ou de ϵ_t^- conduit au système d'équations de Yule-Walker (2-120):

$$(2-120) \quad \begin{bmatrix} I & A_1^+ & \dots & A_p^+ \\ A_p^- & \dots & A_1^- & I \end{bmatrix} \begin{bmatrix} R_0 & R_{-1} & \dots & R_{-p} \\ R_1 & & & \\ & & & R_{-1} \\ R_p & & & R_0 \end{bmatrix} = \begin{bmatrix} E_p^+ & 0 & \dots & 0 \\ 0 & \dots & 0 & E_p^- \end{bmatrix}$$

où $R_i = E(Y_t Y_{t+i}^T) = E(Y_{t-i} Y_t^T)$

et $E_p^+ = E(\epsilon_t^+ \epsilon_t^{+T})$, $E_p^- = E(\epsilon_t^- \epsilon_t^{-T})$

Si on prolonge alors la matrice des R_i dans (2-120) en ajoutant q colonnes

avant et après cette matrice, on obtient un système tel que (2-121):

$$(2-121) \begin{bmatrix} I & A_1^+ & \dots & A_p^+ \\ A_p^- & \dots & A_1^- & I \end{bmatrix} \begin{bmatrix} R_q & & R_{-q} & R_{-q-p} \\ R_{q+p} & R_q & & R_{-q} \end{bmatrix} = \begin{bmatrix} E_q^+ & \dots & E_1^+ & E_0^+ & 0 & \dots & 0 & \Delta_1^+ & \dots & \Delta_q^+ \\ \Delta_q^- & \dots & \Delta_1^- & 0 & \dots & 0 & E_0^- & E_1^- & \dots & E_q^- \end{bmatrix}$$

On observe immédiatement en se reportant à (2-119) que les E^+ , E^- , Δ^+ , Δ^- , sont des intercorrélations entre ϵ_t^+ , ϵ_t^- et Y_{t+1} :

$$(2-122) \begin{cases} E_i^+ = E(\epsilon_t^+ \cdot Y_{t+i}^T) \\ E_i^- = E(\epsilon_t^- \cdot Y_{t-i}^T) \end{cases}$$

$$(2-123) \begin{cases} \Delta_i^+ = E(\epsilon_t^+ \cdot Y_{t-p-i}^T) \\ \Delta_i^- = E(\epsilon_t^- \cdot Y_{t+p+i}^T) \end{cases}$$

On suppose maintenant que Y_t est un signal vectoriel à moyenne ajustée (MA) tel que (2-124), les corrélations R_i s'annulent dès que $|i| > q$.

$$(2-124) Y_t = B_0 \epsilon_t + \dots + B_q \epsilon_{t-q}$$

Lorsque dans l'algorithme de Levinson on cherche à réaliser le passage à l'ordre autorégressif $p+1$, l'équation (2-121) devient, avec la restriction sur les R_i introduite par le fait que Y_t soit MA, et en indiquant par p toutes les quantités liées au modèle d'ordre p :

$$(2-125) \quad \begin{bmatrix} I & A_1^+ & \dots & A_p^+ & 0 \\ 0 & A_p^- & \dots & A_1^- & I \end{bmatrix} \begin{bmatrix} R_q & R_0 & R_q & 0 \\ 0 & R_q & R_0 & R_{-q} \end{bmatrix} = \begin{bmatrix} E_q^+(p) & \dots & E_1^+(p) & E_0^+(p) & 0 & \dots & 0 & \Delta_1^+(p) & \dots & \Delta_q^+(p) & 0 \\ 0 & \Delta_q^-(p) & \dots & \Delta_1^-(p) & 0 & \dots & 0 & E_0^-(p) & E_1^-(p) & \dots & E_q^-(p) \end{bmatrix}$$

Il est alors clair que l'algorithme de Levinson qui s'écrit:

$$(2-126) \quad \begin{cases} K_{p+1}^+ = -\Delta_1^+(p)(E_0^-(p))^{-1} \\ K_{p+1}^- = -\Delta_1^-(p)(E_0^+(p))^{-1} \end{cases}$$

$$(2-127) \quad \begin{bmatrix} I & A_1^+(p+1) & \dots & A_{p+1}^+(p+1) \\ A_{p+1}^-(p+1) & \dots & A_1^-(p+1) & I \end{bmatrix} = \begin{bmatrix} I & K_{p+1}^+ \\ K_{p+1}^- & I \end{bmatrix} \begin{bmatrix} I & A_1^+ & \dots & A_p^+ & 0 \\ 0 & A_p^- & \dots & A_1^- & I \end{bmatrix}$$

peut aussi s'écrire en faisant uniquement appel au second membre de (2-125), c'est à dire en combinant (2-126) et (2-127):

$$(2-128) \quad \begin{bmatrix} E_q^+(p+1) & \dots & E_0^+(p+1) & 0 & \Delta_1^+(p+1) & \dots & \Delta_q^+(p+1) \\ \Delta_q^-(p+1) & \dots & \Delta_1^-(p+1) & 0 & E_0^-(p+1) & \dots & E_q^-(p+1) \end{bmatrix} = \begin{bmatrix} I & K_{p+1}^+ \\ K_{p+1}^- & I \end{bmatrix} \begin{bmatrix} E_q^+(p) & \dots & E_0^+(p) & \Delta_1^+(p) & \dots & \Delta_q^+(p) & 0 \\ 0 & \Delta_q^-(p) & \dots & \Delta_1^-(p) & E_0^-(p) & \dots & E_q^-(p) \end{bmatrix}$$

2.1.2. Obtention du modèle MA.

L'algorithme ainsi décrit est la décomposition de Schur de l'autocorrélation de Y_t . Les conditions initiales ($p=0$) sont prises comme: $E_1^+(0) = \Delta_1^-(0) = R_1$ et $E_1^-(0) = \Delta_1^+(0) = R_1$. Le modèle MA s'obtient alors en faisant tendre p vers l'infini. Le modèle AR d'ordre infini étant alors équivalent au modèle MA(q), on peut identifier ϵ_t et $\epsilon_t^+(\omega)$. En reprenant

l'interprétation des E_i^+ , et en tenant compte de (2-124), on obtient:

$$(2-129) \quad E_i^+(\omega) = E(\epsilon_t Y_{t+1}^T) = E(\epsilon_t \epsilon_t^T) B_i^T$$

On en déduit: $B_i = (E_i^+(\omega))^T (E_0^+(\omega))^{-1}$ mais ici $B_0 = I$, car le résidu ϵ_t n'est pas normalisé avec une variance égale à I . Pour normaliser correctement le résidu, il faut poser $\epsilon_t = L \epsilon_t^*$ et identifier ϵ_t^* à $\epsilon_t^*(\omega)$, alors:

$$(2-130) \quad E(\epsilon_t \epsilon_t^T) = L \cdot E(\epsilon_t^* \epsilon_t^{*T}) \cdot L^T = L \cdot E_0^+(\omega) \cdot L^T$$

Le choix de L se fait donc en posant:

$$(2-131) \quad E_0^+(\omega) = (E_0^+)^{1/2} (E_0^+)^{T/2}$$

ce qui est obtenu par toute décomposition de $E_0^+(\omega)$ en un produit d'une matrice par sa transposée, par exemple la décomposition de Cholesky de $E_0^+(\omega)$, mais aussi tout autre décomposition qui s'en déduit par produit par une matrice unitaire. Prenant alors $L = (E_0^+)^{-1/2}$, on obtient:

$$(2-132) \quad B_i = (E_i^+(\omega))^T (E_0^+(\omega))^{-T/2}$$

Ceci appelle deux remarques. La première est que la présentation qui vient d'être faite ne constitue pas une démonstration de la convergence vers la valeur correcte des B_i (voir YOULA, KAZANJIAN, 1978, et LE ROUX, GRENIER, 1980), mais elle présente l'avantage de mettre en évidence l'interprétation en termes d'intercorrélation de cet algorithme, ce qui est indispensable pour en réaliser l'extension non-stationnaire. La seconde remarque est que le modèle MA vectoriel n'est pas unique. En multipliant à droite les B_i par une matrice unitaire, on obtient un nouveau modèle qui est encore une factorisation de la corrélation de Y_t . Cette absence d'unicité sera très gênante dans le cas non-stationnaire où l'on utilisera un modèle vectoriel intermédiaire non-unique pour trouver un modèle scalaire unique. Il faudra

pour cela une phase de sélection de la solution correcte dans un sous-espace de solutions.

2.1.3. Forme normalisée.

Il est à noter que la normalisation du résidu ϵ_t qui vient d'être exécutée après convergence de l'algorithme calculant les E_i^+ , peut aussi se réaliser à chaque itération. Cette normalisation a été introduite pour le modèle AR vectoriel par MORF, VIEIRA, LEE, KAILATH, 1978, et dans la décomposition de Schur par DELSARTE, GENIN, KAMP, 1979. Il a été montré par GAMBOTTO, 1979, que si on lui adjoint une normalisation de Y_t telle que $R_0 = E(Y_t Y_t^T) = I$, alors tous les coefficients des E_i^+ sont compris entre -1 et +1, ce qui généralise au cas vectoriel l'algorithme en virgule fixe de LEROUX, GUEGUEN, 1976. Un autre avantage de cette version normalisée est d'utiliser un seul coefficient de réflexion qui sera noté X_p , au lieu de K_p^+ et K_p^- . L'algorithme repose sur la décomposition de $E_0^+(p)$ en $(E_0^+(p))^{1/2} (E_0^+(p))^{T/2}$, et s'écrit comme (2-133) et (2-135) où les étoiles dénotent les quantités normalisées:

$$(2-133) \quad \begin{bmatrix} E_q^{*+}(p+1) & \dots & E_0^{*+}(p+1) & 0 & \Delta_1^{*+}(p+1) \dots \Delta_q^{*+}(p+1) \\ \Delta_q^{*-}(p+1) \dots \Delta_1^{*-}(p+1) & 0 & E_0^{*-}(p+1) & \dots & E_q^{*-}(p+1) \end{bmatrix} =$$

$$\Theta(p+1) \begin{bmatrix} E_q^{*+}(p) & \dots & E_0^{*+}(p) & \Delta_1^{*+}(p) \dots \Delta_q^{*+}(p) & 0 \\ 0 & \Delta_q^{*-}(p) \dots \Delta_1^{*-}(p) & E_0^{*-}(p) & \dots & E_q^{*-}(p) \end{bmatrix}$$

$$(2-134) \quad \Theta(p+1) = \begin{bmatrix} (I - X_{p+1}^T X_{p+1})^{-1/2} & 0 \\ 0 & (I - X_{p+1}^T X_{p+1})^{-1/2} \end{bmatrix} \begin{bmatrix} I & X_{p+1} \\ X_{p+1}^T & I \end{bmatrix}$$

$$(2-135) \quad X_{p+1} = -\Delta_1^{*+}(p) (E_0^{*-}(p))^{-1}$$

La détermination d'un modèle MA se résume donc à quatre principes

simples: mettre en équivalence le modèle MA avec un modèle autorégressif d'ordre infini, normaliser ce modèle AR, en déduire un algorithme où le modèle AR reste implicite, mettant l'accent sur les intercorrélations signal-résidu, interpréter ces intercorrélations comme les coefficients d'un modèle MA. C'est à partir de ces principes qu'a pu être étendue au cas non-stationnaire cette méthode de factorisation MA.

2.2. Factorisation scalaire non-stationnaire.

2.2.1. Equivalence "AR long" et MA non-stationnaire.

Mettre cette démarche en oeuvre dans la modélisation MA non-stationnaire suppose donc de trouver un équivalent autorégressif au modèle dont les $b_i(t)$ sont les coefficients, conformément à (2-102). Si le modèle autorégressif cherché possède un gain variable $g(t)$, il s'écrira:

$$(2-136) \quad y_t = - \sum_{i=1}^{\infty} a_i(t-i) y_{t-i} + g(t) \epsilon_t$$

La comparaison de (2-102) et (2-136) met en évidence l'égalité de $g(t)$ et $b_0(t)$ du fait que les ϵ_t sont les innovations du signal y_t . L'identification des deux modèles s'obtient aisément en insérant (2-102) dans (2-136):

$$(2-137) \quad b_0(t) \epsilon_t = \sum_{i=1}^{\infty} a_i(t-i) \sum_{j=0}^q b_j(t-i-j) \epsilon_{t-i-j}$$

L'indépendance des ϵ_t entre eux implique alors la relation (2-138) entre les coefficients AR et MA:

$$(2-138) \quad a_i(t) = - \sum_{j=1}^{\min(i,q)} a_{i-j}(t+j) \frac{b_j(t)}{b_0(t)}$$

La condition initiale de cette récurrence sur les a_i est $a_0(t)=1$. Si on suppose que le modèle autorégressif a un gain constant, on trouve par le même calcul que les $a_i(t)$ vérifient également la récurrence (2-138), mais avec comme condition initiale $a_0(t)=(b_0(t))^{-1}$. L'équivalence entre modèle MA d'ordre q et modèle AR d'ordre infini suppose deux conditions dans le cas non-stationnaire qui sont déduites du calcul précédent et n'apparaissent pas dans le cas stationnaire. La première est que la normalisation $a_0(t)=1$ doit être abandonnée. La seconde est que la base de fonctions sur laquelle s'expriment les $b_i(t)$ n'est pas la même que celle associée aux $a_i(t)$.

L'abandon de la normalisation de a_0 est rendu nécessaire par l'emploi qui est fait du modèle AR. Celui-ci n'est pas utilisé explicitement, mais seul intervient le résidu ϵ_t de ce modèle, et par son intercorrélacion avec le signal d'origine y_t . Il est alors indispensable que ce résidu soit normé pour qu'on puisse en déduire la valeur des $b_i(t)$, et nous venons de le voir, ceci implique un coefficient a_0 variant avec le temps t . Ceci complique beaucoup l'algorithme MA non-stationnaire.

Quant à la base de fonctions sur laquelle s'expriment les $a_i(t)$, il est clair à partir de (2-138), qu'elle diffère de celle des $b_i(t)$. Si par exemple ces derniers s'expriment sur la base $(1, t, t^2, \dots, t^m)$, c'est-à-dire si ce sont des polynômes, les $a_i(t)$ seront des fractions rationnelles de t et non des polynômes, la nature de la base change donc. Si $b_0(t)$ est constant, les $a_i(t)$ redeviennent des polynômes, la nature de la base est donc inchangée, mais (2-138) montre que le degré du polynôme $a_i(t)$ n'est en général pas égal à m mais croît avec i , et vaut mi . On peut aussi ramener le cas des $a_i(t)$ fractions rationnelles à ce cas polynomial mais avec un degré très grand. Le même raisonnement peut se faire pour toutes les bases "ordinaires" de fonc-

tions: le passage des $b_i(t)$ aux $a_i(t)$ se fait non seulement par l'accroissement de l'ordre q à l'infini, mais aussi par l'accroissement du degré m de la base à l'infini. Fort heureusement, ces deux effets n'ont pas la même amplitude, et l'accroissement du degré m se fait approximativement linéairement avec i , or les $a_i(t)$ tendent vers 0 quand i croît, et donc l'erreur que l'on ferait en approximant le degré par une valeur M finie devient négligeable en valeur absolue dès que i dépasse plusieurs fois m . Le principe de la méthode sera donc de recourir à un modèle AR non-stationnaire implicite de degré $M > m$. Les contraintes du temps de calcul limiteront M à deux ou trois fois m .

2.2.2. Signal vectoriel intermédiaire.

Comment estimer le modèle AR quand la contrainte $a_0(t)=1$ n'existe plus. Il faut tout d'abord restreindre les solutions possibles en normalisant le résidu ϵ_t de telle sorte que sa variance soit 1. On peut alors écrire les conditions d'orthogonalité qui avaient conduit à la solution (2-32) sous forme d'équations du type de Yule-Walker, mais l'insertion dans le vecteur θ des paramètres $a_{00} \dots a_{0m}$ ne permet pas d'obtenir une solution unique: on obtiendra (3-139) en écrivant l'orthogonalité de ϵ_t avec $Y_{t-1} \dots Y_{t-p}$:

$$(2-139) \quad \theta^T \cdot E \begin{pmatrix} Y_t \\ Y_{t-1} \\ \cdot \\ \cdot \\ Y_{t-p} \end{pmatrix} = [Y_t^T Y_{t-1}^T \dots Y_{t-p}^T] \theta = [S_0 \dots S_M \ 0 \dots 0]$$

Le système ne possède pas de solution unique, car il y a $(p+1)(M+1)$ inconnues et $p(M+1)$ équations. Il y aura donc $M+1$ solutions dont chacune pourra être obtenue en posant dans le second membre de (2-139) $S_i \neq 0$ et

$S_j=0$, pour $j \neq i$. Ceci impose à toute solution de l'équation (2-139) d'être combinaison linéaire des lignes de la matrice $[A_0^* \dots A_p^*]$ des coefficients du modèle autorégressif de Y_t , dans sa forme normalisé par $E(\epsilon_t^* \epsilon_t^{*\top})=I$. La normalisation du modèle non-stationnaire se ramène donc à la recherche d'un vecteur L à $(M+1)$ composantes tel que:

$$(2-140) \quad [a_{00} \dots a_{pm}] = L^{\top} [A_0^* \dots A_p^*]$$

2.2.3. Algorithme implicite.

Il s'agit maintenant de rendre implicite le rôle du modèle autorégressif. Ceci se fait sans difficulté au moyen de la factorisation de Schur du signal vectoriel Y_t , par l'algorithme normalisé décrit en (2-133) à (2-135). Le seul point délicat est l'utilisation de cette méthode stationnaire sur Y_t un signal non-stationnaire. On utilisera en effet pour les autocorrélations R_i un estimateur ergodique (somme temporelle). Est-il légitime de le faire ? La réponse, positive, nous est donnée par le modèle autorégressif implicite. C'est l'invariance dans le temps de ce modèle qui permet de faire dans (2-121) ou dans (2-125) une somme temporelle à gauche et à droite. On pourrait objecter qu'on ne recherche qu'une seule solution invariante, et non $M+1$. Rien n'assurerait que les M solutions auxiliaires introduites sont associées à des vecteurs invariants, puisque chacune constitue une régression d'une composante de Y_t sur le passé $Y_{t-1} \dots Y_{t-p}$. Cependant un simple examen de (2-139) montre que l'invariance du vecteur des a_{ij} permet sa mise en facteurs dans toute somme temporelle, et que la structure du second membre est conservée. La solution reste donc combinaison linéaire des $M+1$ solutions obtenues, et l'objection tombe. En d'autres termes, chacune des $M+1$ solutions n'a aucune signification pour Y_t mais leur ensemble définit un sous-espace dans lequel figure la solution cherchée. Le

seul problème qui subsistera est la recherche de L donnant la direction optimale dans ce sous-espace. Mais avant d'éclaircir ce point, il est souhaitable de préciser la façon dont la solution $b_{00} \dots b_{qm}$ se déduira de L.

2.2.4. Passage au modèle MA non-stationnaire.

Si on suppose que la décomposition de Schur associée à Y_t a été effectuée, et en tenant compte de la relation $\epsilon_t = L^T \epsilon_t^*$, on interprétera les E_i comme des intercorrélations:

$$(2-141) \quad L^T E_i = L^T \Sigma E(\epsilon_t^* Y_{t+i}^T) = \Sigma E(\epsilon_t Y_{t+i}^T)$$

Utilisant alors l'expression de Y_t en fonction de y_t , et la nature MA de y_t , on en déduit:

$$(2-142) \quad E_i^T L = \Sigma_t \begin{bmatrix} f_0(t+i) \\ \cdot \\ \cdot \\ f_m(t+i) \end{bmatrix} E(\epsilon_t Y_{t+i}^T)$$

$$= \Sigma_t \begin{bmatrix} f_0(t+i) \\ \cdot \\ \cdot \\ f_m(t+i) \end{bmatrix} b_i(t)$$

D'où:

$$(2-143) \quad E_i^T L = \Sigma_t \begin{bmatrix} f_0(t+i) \\ \cdot \\ \cdot \\ f_m(t+i) \end{bmatrix} [f_0(t) \dots f_m(t)] \begin{bmatrix} b_{i0} \\ \cdot \\ \cdot \\ b_{im} \end{bmatrix}$$

Soient, alors les matrices F_i :

$$(2-144) \quad F_{i,t} = \begin{bmatrix} f_0(t+i) \\ \cdot \\ \cdot \\ f_m(t+i) \end{bmatrix} [f_0(t) \dots f_m(t)]$$

L'équation (2-143) admet comme solution (2-145) où le signe # désigne la pseudo-inverse, $(A^\# = (A^T A)^{-1} A^T)$:

$$(2-145) \quad \begin{bmatrix} b_{i0} \\ \cdot \\ \cdot \\ b_{im} \end{bmatrix} = F_{i,t}^\# E_i^T L = B_i L$$

Ainsi est calculé le modèle MA non-stationnaire dès que L est connu, à partir de la factorisation de Y_t , vecteur des projections de y_t sur la base $[f_0(t) \dots f_m(t)]$. Une autre solution était possible à partir de (2-143) utilisant la matrice S qui sélectionne les m+1 premières lignes des matrices E_i :

$$(2-146) \quad \begin{bmatrix} b_{i0} \\ \cdot \\ \cdot \\ b_{im} \end{bmatrix} = \begin{bmatrix} f_0(t+i) \\ \cdot \\ \cdot \\ f_m(t+i) \end{bmatrix} [f_0(t) \dots f_m(t)]^{-1} S E_i^T L$$

Bien que la première solution soit plus satisfaisante pour l'esprit, il ne semble pas que l'abandon des (M-m) dernières composantes des intercorrélations sur la base nuise à la qualité de l'estimateur obtenu. On peut aussi noter que la base des M+1 fonctions sur lesquelles s'effectue la modélisation AR implicite c'est-à-dire la décomposition de Schur, peut ne pas être un sur-ensemble de la famille $[f_0(t) \dots f_m(t)]$. Si l'étude de cette base $[f_0(t) \dots f_m(t)]$ montre qu'il existe une base $[f_0^*(t) \dots f_m^*(t)]$ sur laquelle l'approximation AR(∞) du modèle MA(q) est meilleure, il est possi-

ble d'utiliser cette seconde base pour modéliser Y_t . Il suffit alors de remplacer dans la définition (2-144) des matrices F_i la matrice colonne des $f_i(t)$, $i \in [0, M]$ par la colonne des $f_i^*(t)$.

2.2.5. Réduction du signal vectoriel au signal scalaire.

Le seul point restant à préciser est le calcul du vecteur L permettant de passer du modèle vectoriel de Y_t au modèle scalaire non-stationnaire. Il faut pour cela réintroduire l'information: le vecteur Y_t est obtenu à partir du signal y_t à moyenne ajustée, d'ordre q , ou encore, le modèle AR implicite n'est pas quelconque, mais approxime l'inverse d'un modèle MA. Une façon de le faire est de contraindre les matrices de covariances du signal y_t ou Y_t à être égales à la même matrice calculée théoriquement à partir du modèle MA (qui dépend de L). On ne pourra pas en réalité obtenir l'égalité entre la matrice mesurée et la matrice théorique, c'est-à-dire reconstruite à partir du modèle, du fait des approximations (il faut bien arrêter la récurrence sur p à un certain point, et de plus M étant limité supérieurement, l'évolution temporelle du modèle AR n'est pas exactement celle de l'inverse du modèle MA). Il suffira alors de maximiser la vraisemblance de l'hypothèse d'égalité des deux matrices. ANDERSON, 1958, donne l'expression de la vraisemblance V de l'hypothèse $R=R^*$ où R et R^* sont deux matrices de covariance de dimension n , calculées sur N points:

$$(2-147) \quad V = e^{nN/2} (\text{Det}(R.(R^*)^{-1}))^{N/2} e^{-\text{Tr}(R.(R^*)^{-1})/N}$$

ou en prenant le logarithme de V :

$$(2-148) \quad \text{Log } V = N/2.(n + \text{LogDet}(R.(R^*)^{-1}) - \text{Tr}(R.(R^*)^{-1}))$$

La relation (2-148) peut être rendue plus symétrique sous la forme:

$$(2-149) \quad \text{Log } V = N/2.(n + \text{LogDet}((R^*)^{-1/2}.R.(R^*)^{-T/2}) - \text{Tr}((R^*)^{-1/2}.R.(R^*)^{-T/2}))$$

Dans le problème posé ici, R sera la matrice de covariance théorique, estimée à partir des $b_i(t)$, et R^* la même matrice mais mesurée sur y_t . Trois choix (et même plus) sont possibles pour cette matrice de covariance:

Choix 1: $R = E(Y_t Y_t^T)$

Choix 2: $R = E \left(\begin{bmatrix} Y_t \\ \cdot \\ \cdot \\ Y_{t-q} \end{bmatrix} [Y_t \dots Y_{t-q}] \right)$

Choix 3: $R = E \left(\begin{bmatrix} Y_t \\ \cdot \\ \cdot \\ Y_{t-q} \end{bmatrix} [Y_t^T \dots Y_{t-q}^T] \right)$

Leur efficacité va en croissant de 1 à 3, mais aussi leur complexité liée à la taille de la matrice R. Bien entendu, l'espérance est dans R^* remplacée par une somme temporelle, ce qui contraint, vu la non-stationnarité de y_t , à effectuer la même somme temporelle dans R. Chacun des ces trois choix conduit à une matrice R qui est une fonction quadratique des b_{ij} et donc des λ_i , les composantes du vecteur L. Il suffit pour s'en convaincre de calculer la matrice R_k où:

$$(2-150) \quad R_k = \sum_t E(Y_t Y_{t+k}^T)$$

En définissant $F_m(t)$ et $F_M(t)$ (ce dernier à partir des $f_i(t)$, mais aussi si on le souhaite à partir des $f_i^*(t)$ comme pour la formule (2-146)):

$$(2-151) \quad F_m(t) = \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix} \quad \text{et} \quad F_M(t) = \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_M(t) \end{bmatrix}$$

on écrira:

$$(2-152) \quad R_k = \sum_t F_M(t) E(y_t y_{t+k}) F_M^T(t+k) \\ = \sum_t F_M(t) \left[\sum_{s=0}^{q-k} F_m(t-s) B_s L L^T B_{s+k} F_m(t-s) \right] F_M^T(t+k)$$

où on a utilisé la relation (2-145), définissant B_i . Ce calcul permet donc d'écrire, en développant suivant les composantes λ_i de L , et en notant B_{ij} la j -ème colonne de B_i :

$$(2-153) \quad R_k = \sum_{i=0}^M \sum_{j=0}^M \lambda_i \lambda_j R_k(i, j)$$

avec

$$(2-154) \quad R_k(i, j) = \sum_t F_M(t) \left[\sum_{s=0}^{q-k} F_m(t-s) B_{si} B_{s+k, j} F_m(t-s) \right] F_M^T(t+k)$$

On pourra d'ailleurs symétriser l'expression de $R_k(i, j)$ relativement à i et j en posant:

$$(2-155) \quad R_k(i, j) = \sum_t F_M(t) \left[\sum_{s=0}^{q-k} F_m(t-s) (B_{si} B_{s+k, j} + B_{sj} B_{s+k, i}) F_m(t-s) \right] F_M^T(t+k)$$

On peut donc écrire la matrice $(R^*)^{-1/2} R (R^*)^{-T/2}$ figurant dans le critère de vraisemblance comme:

$$(2-156) \quad C = (R^*)^{-1/2} R (R^*)^{-T/2} = \sum_{i, j=0}^M \lambda_i \lambda_j C_{ij}$$

Les matrices C_{ij} dépendent uniquement de $R_k(i, j)$, $k \in [0, q]$ et de R^* mesuré. Le problème est maintenant ramené à la maximisation de la quantité $\text{LogDet}(C) - \text{Tr}(C)$ où C est une matrice carrée, fonction quadratique des λ_i :

$$(2-157) \quad V(\lambda_0 \dots \lambda_M) = \text{LogDet}(C) - \text{Tr}(C)$$

$$\text{avec} \quad C(\lambda_0 \dots \lambda_M) = \sum_{i, j=0}^M \lambda_i \lambda_j C_{ij} \quad (C_{ij} = C_{ij}^T = C_{ji})$$

La présence du logarithme et du déterminant rendent cependant impossible cette maximisation de façon directe. Il faudra recourir à un algorithme du type Newton-Raphson pour la recherche du maximum de la fonction. Il est pour cela nécessaire de calculer son gradient et son Hessien. Le calcul du gradient ne pose pas de difficulté pour $\text{tr}(C)$:

$$(2-158) \quad \frac{\partial}{\partial \lambda_i} \text{Tr}(C) = \text{Tr} \left(\sum_{j=0}^M 2\lambda_j C_{ij} \right)$$

Pour le terme en $\text{LogDet}(C)$, on introduira une variation du premier ordre $\Delta\lambda_i$ et on aura:

$$(2-159) \quad \frac{\partial}{\partial \lambda_i} \text{LogDet}(C) = \lim_{\Delta\lambda_i \rightarrow 0} \frac{1}{\Delta\lambda_i} (\text{LogDet}(C_{\Delta\lambda_i}) - \text{LogDet}(C))$$

On désigne par $C_{\Delta\lambda_i}$ la valeur prise par le critère C quand λ_i est modifié de $\Delta\lambda_i$. Le calcul montre en posant:

$$C_{ij}^* = C^{-1/2} C_{ij} C^{-1/2}$$

$$C_i = 2 \sum_{j=0}^M C_{ij}$$

$$C_i^* = 2 \sum_{j=0}^M C_{ij}^*$$

que la dérivée partielle en λ_i s'écrit:

$$(2-160) \quad \frac{\partial}{\partial \lambda_i} \text{LogDet}(C) = \lim_{\Delta\lambda_i \rightarrow 0} \frac{1}{\Delta\lambda_i} \text{LogDet}(I + \Delta\lambda_i C_i^*)$$

Un calcul analogue avec des variations $\Delta\lambda_i$ et $\Delta\lambda_j$ montre, après de lourdes manipulations, et en se servant du théorème de Schwarz rendant le calcul des dérivées partielles indépendant de l'ordre dans lesquelles elles se font, que le Hessien se calcule comme:

$$(2-161) \quad \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \text{LogDet}(C) = \lim_{\Delta\lambda_i \rightarrow 0, \Delta\lambda_j \rightarrow 0} \frac{1}{\Delta\lambda_i \Delta\lambda_j} \text{LogDet}(I + \Delta\lambda_i \Delta\lambda_j (2C_{ij}^* - C_i^* C_j^*))$$

On a dans les deux cas à calculer une limite qui est $\lim \text{LogDet}(I+\Delta.A)$. Un développement du déterminant suivant la première colonne de $I+\Delta A$ montre que:

$$(2-162) \quad \lim_{\Delta \rightarrow 0} \text{LogDet}(I+\Delta A) = \text{Tr}(A)$$

Les résultats (2-158) à (2-162) se condensent alors pour donner:

$$(2-163) \quad \frac{\partial}{\partial \lambda_i} V(\lambda_0 \dots \lambda_M) = \text{Tr}(C_i^*) - \text{Tr}(C_i)$$

$$\text{et} \quad \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} V(\lambda_0 \dots \lambda_M) = \text{Tr}(2C_{ij}^* - C_i^* C_j^*) - 2\text{Tr}(C_{ij})$$

Il suffit alors d'utiliser ces quantités pour former le vecteur gradient Γ et la matrice H du Hessien et de corriger la valeur L_n de L , obtenue après n itérations de Newton-Raphson: $L_{n+1} = L_n - H^{-1} \Gamma$. L_n converge alors vers la solution L d'où on déduit les b_{ij} par (2-145). On aura intérêt à démarrer cette partie de l'algorithme en choisissant des valeurs initiales L_0 voisines de l'optimum, ce qui pourra se faire en utilisant les valeurs de b_{ij} données par la méthode utilisant explicitement le modèle AR long et en calculant L au moyen de (2-145).

2.2.6. Résumé de l'algorithme de factorisation.

L'algorithme complet est donc constitué des phases suivantes, permettant d'estimer sur y_t , $t \in [0, T]$ le modèle $MA(q)$ associé à la base de fonctions $[f_0(t) \dots f_m(t)]$:

- A) Choisir M et la base $[f_0(t) \dots f_M(t)]$
- B) Calculer les vecteurs Y_t des projections de y_t
- C) Effectuer la factorisation de Schur sur les Y_t pour obtenir les E_i .

- D) Calculer les matrices F_i par (2-144)
- E) Normaliser les E_i , pour obtenir les B_i de (2-145)
- F) Calculer les R_k associés aux B_i (2-152)
- G) Les normaliser par les covariances mesurées pour obtenir les C_{ij} .
- H) Déterminer le vecteur L par la procédure de Newton-Raphson (2-163)
- I) En déduire les b_{ij} (2-145)

Il est évident que cet algorithme consomme beaucoup plus de temps de calcul et de place mémoire que l'algorithme explicite, mais sa précision est nettement supérieure comme le montrent les résultats présentés dans le chapitre 9 (Simulations). Si on ne craint pas d'augmenter encore le coût du calcul, on pourra améliorer la procédure en substituant dans l'étape B, à l'estimateur ergodique calculant les corrélations des Y_t , un estimateur développé par NEWTON, 1980, à partir des travaux sur les modèles MA scalaires de WALKER, 1961 et 1962. L'intérêt de cet estimateur est de tenir compte des corrélations au-delà de l'ordre q , alors que la troncature pure et simple ($R_k=0$ pour $k>q$) dans l'estimateur ergodique, ne garantissant pas la positivité de la corrélation, peut parfois interdire la convergence de la factorisation de Schur.

3. Gain non-stationnaire.

Il existe un cas très particulier du modèle MA, c'est le modèle $MA(0)$. Dans le cas stationnaire, un tel modèle est réduit à un gain qui s'estime en général sans difficulté. Dans le cas non-stationnaire, ce gain est une fonction du temps et il est possible de tenter de l'estimer comme le modèle MA

habituel, avec l'hypothèse qu'il s'exprime comme combinaison linéaire des fonctions $f_i(t)$ de la base. Ce gain intervient dans le modèle AR non-stationnaire, où il est responsable des fluctuations de la variance $\sigma^2(t)$ du résidu ϵ_t . Aucune des méthodes MA décrites dans ce chapitre ne permet cependant de calculer le gain $b_0(t)$ tel que $\sigma^2(t) = (b_0(t))^2$. En effet les deux méthodes décrites font appel à un modèle AR long (explicite ou implicite) alors que le gain non-stationnaire n'est pas équivalent à un modèle AR, ce qui contraint à développer une méthode spécifique pour ce cas.

Le problème étant non-linéaire, il faut se ramener par une transformation des données à un problème linéaire. Un moyen consiste à remplacer ϵ_t par $|\epsilon_t|$. La variable $|\epsilon_t|$ reste blanche, mais sa moyenne devient non nulle et vaut:

$$(2-164) \quad E(|\epsilon_t|) = \sigma_t (2/\pi)^{1/2} = b_0(t) (2/\pi)^{1/2}$$

Il est alors possible de calculer $b_0(t)$ approximant au sens des moindres carrés cette moyenne, le critère sera:

$$(2-165) \quad J = \sum_t (|\epsilon_t| - (2/\pi)^{1/2} b_0(t))^2$$

Ce qui conduit en introduisant les $f_i(t)$ au système:

$$(2-166) \quad \left(\sum_t \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix} \right) \begin{bmatrix} f_0(t) \dots f_m(t) \end{bmatrix} \begin{bmatrix} b_{i0} \\ \cdot \\ \cdot \\ b_{im} \end{bmatrix} = (\pi/2)^{1/2} \sum_t \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix} \cdot |\epsilon_t|$$

Bien que le gain $b_0(t)$ ainsi calculé soit la plupart du temps satisfaisant, on peut regretter qu'il lui soit possible de s'annuler, et même d'être négatif, par exemple si $f_i(t) = t^i$, $b_0(t)$ devient un polynôme dont la valeur peut sur certains intervalles de temps être en dessous de 0. Un

moyen d'y remédier consiste à poser que le gain $b_0(t)$ est en réalité une exponentielle:

$$(2-167) \quad b_0(t) = e^{g(t)} \quad \text{avec} \quad g(t) = \sum_{i=0}^m g_i f_i(t)$$

Il s'agit ici aussi de rendre le problème linéaire, ce qui peut se faire en retenant comme signal $\text{Log}|\epsilon_t|$. L'espérance mathématique devient:

$$(2-168) \quad E(\text{Log}|\epsilon_t|) = \text{Log}(b_0(t)) - \frac{C + \text{Log}2}{2}$$

où C est la constante d'Euler, $C = 0.577 \ 215 \ 665$. Le critère d'estimation sera:

$$(2-169) \quad J = \sum_t (\text{Log}|\epsilon_t| + \frac{C + \text{Log}2}{2} - g(t))^2$$

Ce qui donne en utilisant l'expression de $g(t)$ sur la base de fonctions $f_i(t)$:

$$(2-170) \quad \left(\sum_t \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix} \begin{bmatrix} f_0(t) \dots f_m(t) \end{bmatrix} \right) \begin{bmatrix} g_0 \\ \cdot \\ \cdot \\ g_m \end{bmatrix} = (\pi/2)^{1/2} \sum_t \begin{bmatrix} f_0(t) \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix} (\text{Log}|\epsilon_t| + \frac{C + \text{Log}2}{2})$$

Il est clair que ces deux estimateurs (2-166) et (2-170) bien que pris au sens des moindres carrés, ne sont pas optimaux car ni $|\epsilon_t|$, ni $\text{Log}|\epsilon_t|$ ne sont gaussiens, contrairement à ϵ_t . Cependant, dans la pratique, la dynamique de $b_0(t)$ ou $g(t)$, qui est le rapport entre les amplitudes extrêmes prises par $b_0(t)$ ou $g(t)$, est réduite, restant de l'ordre de 1 et ces estimateurs sont amplement suffisants.