QUATRIEME COLLOQUE SUR LE

TRAITEMENT DU SIGNAL ET SES APPLICATIONS

Nice 7 au 12 mai 1973

ESSAI DE FORMALISATION DES PROBLEMES DE DETECTION ADAPTATIVE.

par

G. VEZZOSI

Laboratoire d'Etude des Phénomènes Aléatoires - Bât. 210
UNIVERSITE DE PARIS-SUD - Centre d'Orsay - 91405.

RESUME :

On examine les questions soulevées par l'intervention du temps et l'incertitude sur les caractéristiques du bruit dans les problèmes de détection.

Plan :

1. Définition et conditions d'emploi des méthodes adaptatives.
2. Examen des méthodes adaptatives du point de vue théorique.
3. Enoncé des problèmes de détection où le temps intervient explicitement. Choix des espaces de décision.
4. Rappels sur la théorie des tests d'hypothèses multiples.
5. Extension de cette théorie au cas où l'on doit résoudre une suite de problèmes de test indexés par le temps.
6. Application au problème de détection où existent des incertitudes sur les caractéristiques du bruit.
7. Le cas où la probabilité d'évolution du système étudié dans la fenêtre d'analyse est très petite.
8. Application au problème de détection adaptative.

SUMMARY

ESSAI DE FORMALISATION DES PROBLEMES DE DETECTION ADAPTATIVE.

G. VEZZOSI.

1. DEFINITION ET CONDITIONS D'EMPLOI DES METHODES ADAPTATIVES.

Dans les problèmes de traitement du signal un procédé de détection ou d'estimation est adaptatif lorsqu'il ne requiert pas d'information a priori sur les caractéristiques du bruit, et qu'il évalue ces caractéristiques en cours d'exploitation de l'observation incidente par une méthode directe ou algorithmique. La seule restriction à l'emploi d'une méthode adaptative est donc déterminée par la possibilité d'évaluation des caractéristiques du bruit. En l'absence de renseignements sur les mécanismes responsables de la non stationnarité du bruit, ce qui constitue la situation la plus fréquente, la seule possibilité repose sur le calcul des moyennes temporelles. Ces moyennes ne fournissent les renseignements désirés que si elles se stabilisent au moins partiellement. C'est pourquoi la propriété de stationnarité locale du bruit |1|, (c'est-à-dire, précisément, la propriété de stabilisation partielle des moyennes temporelles du bruit dans une certaine gamme de temps d'intégration), semble être une condition nécessaire pour la possibilité même de l'adaptation.

D'autre part, dans un procédé adaptatif l'observation doit permettre simultanément l'adaptation et le traitement proprement dit.

Or l'adaptation suppose l'apprentissage des caractéristiques du bruit, et donc l'observation doit contenir en principe une "référence bruit seul" |1|, c'est-à-dire une tranche de bruit pur, connue pour telle a priori. Faute de quoi l'adaptation s'effectue non sur le bruit mais sur l'observation brute, et le récepteur obtenu ne peut pas fonctionner lorsque, par exemple, le bruit devient très faible par rapport au signal.

2. EXAMEN DES METHODES ADAPTATIVES DU POINT DE VUE THEORIQUE.

Toute tentative visant à légitimer théoriquement les procédés adaptatifs et restant dans le cadre strict de l'idée d'adaptation se heurte à une difficulté de principe : l'adaptation n'est pas une donnée des problèmes réputés adaptatifs. Elle n'est qu'un moyen empirique propre à faciliter la détection ou l'estimation. D'autres moyens existent, comme par exemple les procédés non paramétriques. Tous ces moyens ont en commun un but: résoudre l'incertitude sur les caractéristiques du bruit. Dans l'examen théorique des méthodes adaptatives, ce n'est donc pas l'idée d'adaptation, mais bien celle de cette incertitude, qui doit être retenue à titre de notion première, et une théorie des problèmes adaptatifs ne peut être qu'un cas particulier de la théorie générale des problèmes où existent des incertitudes sur les caractéristiques du bruit. Il s'agit donc de construire cette théorie générale, puis de l'appliquer aux problèmes adaptatifs en supposant le bruit pourvu de la propriété de stationnarité locale, et l'observation assortie d'une référence bruit seul.

D'autre part, les méthodes adaptatives sont employées généralement dans des situations où le temps intervient explicitement dans le processus de décision, parce que ce processus doit permettre de reconstituer l'état d'un système en évolution au cours du temps. Pour suivre cette évolution il faut prendre non une décision unique comme dans la théorie habituelle, mais une suite de décisions instantanées. Le temps joue alors d'emblée un rôle majeur, et il importe d'en tenir compte dans la construction de la théorie. La présente étude examine ces différentes questions dans le cas particulier des problèmes de détection.

3. ENONCE DES PROBLEMES DE DETECTION OU LE TEMPS INTERVIENT.

Construction des espaces de décision

La plupart des problèmes de détection s'insèrent dans le cadre général suivant. Considérons un système Σ de nature quelconque, non directement observable, susceptible d'évoluer au cours du temps, et n évènements $H_1, \ldots, H_n$ relatifs à l'état instantané du système , prescrits d'avance, s'excluant mutuellement, dont il s'agit de déceler l'apparition. Ces évènements peuvent être permanents, c'est-à-dire que leur réalisation à un instant donné entraîne avec une probabilité non infiniment petite leur réalisation dans tout intervalle de temps immédiatement postérieur : le problème revient alors à détecter la succession des états d'un système à nombre fini d'états (un exemple typique est la détection de l'état du basculeur poissonnien en présence de bruit) ; ou bien ces évènements peuvent être instantanés, auquel cas l'évolution de Σ est décrite par la succession des points d'un processus ponctuel marqué : le problème revient alors à détecter des points (un exemple typique est la détection des points d'un processus poissonnien).

. La production de ces évènements dans le système Σ entraîne au niveau du récepteur la production de signaux (aléatoires en général), auxquels s'ajoute le bruit parasite pour donner naissance à l'observation brute disponible. Nous supposons de plus :

a) que dans le cas d'évènements permanents, les signaux associés sont également permanents et ont la même durée que les évènements qui les engendrent.

b) que dans le cas d'évènements instantanés, les signaux associés sont impulsifs et de durée maximale T.

Soit $I = [T_i, T_f]$ l'intervalle de l'axe des temps où est réalisée l'expérimentation : $T_f - T_i$ est un nombre très supérieur à T.

L'observation s'écrit :

- dans le cas d'évènements permanents :

$$(1) \qquad x(u) = b(u) + \sum_{1}^{n} 1_{H_i}(u)\, S_i(u) \qquad u \in I,$$

où $1_{H_i}(u)$ est la variable booléenne associée à l'évènement : "l'évènement H_i est réalisé au temps u" ; on a évidemment :

$$1_{H_i}(u)\, 1_{H_j}(u) = 0 \qquad u \in I, \quad i \neq j .$$

- dans le cas d'évènements instantanés :

$$(2) \qquad x(u) = b(u) + \sum_{1}^{n} \int_{I} S_i(u, \theta)\, dn_i(\theta) \qquad u \in I$$

Le problème est de détecter les points du processus ponctuel (représentés par les instants de discontinuité des fonctions $n_i(\theta)$); où les états du système (représentés par les paliers des fonctions $1_{H_i}(u)$), au fur et à mesure de leur apparition au cours du temps. C'est pourquoi nous devons prendre une suite de décisions instantanées, et dès lors la première tâche doit être de préciser les instants de décision, le contenu des espaces de décision instantanée, et la part de l'observation qui doit nous permettre de prendre chaque décision instantanée. Le choix de ces éléments est arbitraire. Rien n'empêche par exemple de considérer que ces éléments ne font pas partie des données du problème, mais plutôt qu'ils sont définis de façon adaptative en cours d'exploitation de l'observation incidente. Mais la théorie devient alors fort complexe, et nous n'envisageons pas cette situation dans la suite. Nous supposerons donc que sont donnés d'avance l'ensemble $J = \{t\}$ des instants de décision, la famille $\{\Delta(t),\ t \in J\}$ des espaces de décisions instantanées, et la famille $\{I(t),\ t \in J\}$ des intervalles d'observation associés à chaque décision instantanée.

Il reste alors à préciser ces différents éléments. Pour éviter d'avoir à reconstituer à chaque instant de décision toute une partie de l'évolution antérieure du système , on peut penser dissocier temporellement les décisions relatives aux évènements produits en des instants différents. Nous pouvons par exemple prendre pour instants de décision les instants de fin d'état possible dans le cas d'évènements permanents, et les instants où les signaux associés aux points du processus ponctuel peuvent s'achever dans le cas d'évènements instantanés (ce qui revient à tolérer dans ce deuxième cas un retard à la décision systématique égal à la durée du signal le plus long, ou encore à accepter de ne poser la question de la présence des signaux qu'aux instants où ces signaux prennent fin). Les espaces de décision correspondant $\{\Delta(t),\ t \in J\}$ comprennent alors n+1 éléments $\delta_o^{\ t}$, $\delta_1^{\ t}$, ..., $\delta_n^{\ t}$, où :

• $\delta_i^{\ t}$, $i \geq 1$ désigne la décision d'accepter l'hypothèse que l'évènement H_i est en cours au temps t (cas d'évènements permanents), où qu'un point du type i est apparu au temps t-T (cas d'évènements instantanés).

• $\delta_o^{\ t}$ désigne la décision d'accepter l'hypothèse qu'aucune des circonstances précédentes n'est réalisée.

Dans cette perspective, le problème de détection initial se réduit à une suite de problèmes de tests d'hypothèses multiples, et nous sommes ramenés à la théorie des tests. Nous adopterons cette solution dans la suite, parce que c'est celle qui se présente immédiatement à l'esprit de l'opérateur suivant pas à pas le déroulement temporel de l'observation, et parce qu'elle est l'analogue de la solution adoptée dans les problèmes d'estimation récursive. Nous supposerons de plus :

• que dans le cas d'évènements permanents, tout élément de I est également un instant de fin d'état possible du système Σ. L'ensemble J des instants de décision coïncide avec I :

$$J = I = [T_i,\ T_f]$$

• que dans le cas d'évènements instantanés, les points du processus ponctuel ne peuvent apparaître qu'entre les instants T_i et T_f-T. L'ensemble J est alors l'intervalle :

$$J = [T_i + T,\ T_f]$$

La décision au temps t met en jeu l'observation antérieure comprise dans un intervalle I(t) qui est une donnée du problème. I(t) peut être l'intervalle d'observation antérieur complet :

$$I(t) = [T_i,\ t].$$

$I(t)$ est plus fréquemment l'intervalle de durée τ immédiatement antérieur à l'instant de décision actuel **t**, où τ est une constante très inférieure à la durée totale d'expérimentation $T_f - T_i$.

4. RAPPELS SUR LA THEORIE DES TESTS D'HYPOTHESES MULTIPLES.

La théorie des tests d'hypothèses multiples est un cas particulier de la théorie de la décision, et il est commode d'adopter le formalisme général de cette théorie, fondé sur la considération des trois ensembles Δ (espaces des décisions), Θ (espaces des paramètres) et $\mathfrak{X}$ (espaces des observations) pour présenter les problèmes de tests. Dans la théorie des tests, Δ contient un nombre fini d'éléments δ_o, δ_1, ..., δ_n, et à tout élément de l'espace paramétrique on peut faire correspondre un élément de Δ qui s'interprète comme "la décision correcte relative à ce paramètre". Pour rendre compte de cette particularité, il est commode de considérer le paramètre θ comme le couple d'un paramètre décisionnel η, $\eta = 0, 1, \ldots, n$, dont la connaissance permet de prendre la décision sans erreur, et d'un paramètre indésirable $\theta\eta$ qui ne possède pas cette propriété. On a alors :

$$\theta = (\eta\,,\ \theta_\eta\,),$$

et l'espace paramétrique Θ s'écrit :

$$\Theta = \bigcup_{\eta=0}^{n} \eta \times \Theta_\eta$$

Les problèmes soulevés par la définition de l'optimalité ne sont pas essentiellement différents de ceux que l'on rencontre en théorie de la décision. Partant de l'idée qu'une stratégie ne peut être jugée que par ses conséquences, nous sommes conduits à spécifier une application $L(\delta,\ \theta)$ de $\Delta \times \Theta$ dans R qui exprime la perte associée à la décision δ lorsque θ est la valeur du paramètre, et à caractériser toute stratégie

$$L_{ij}(\theta_j) = 1 - \delta_{ij} \qquad i, j = 0, 1, \ldots, n \quad ,$$

δ_{ij} : symbole de Kronecker ,

l'optimalité au sens de Neyman-Pearson moyen revient à déterminer la stratégie qui maximise la probabilité de détection moyenne :

$$(6) \qquad \beta(S) = \sum_{1}^{n} \pi'_n \int_{\Theta_n} \mu_n(d\theta_n) \int_{\mathfrak{X}} P_n(dx|_{\theta_n}) \; S(\delta_n|_x)$$

sous la contrainte :

$$(7) \qquad \alpha(S) = \int_{\Theta_o} \mu_o(d\theta_o) \int_{\mathfrak{X}} P_o(dx|_{\theta_o}) \; S(\delta_o|_x) \leq \alpha$$

qui concerne la probabilité de fausse alarme moyenne.

Quant à l'optimalité au sens de Neyman-Pearson conditionnel elle revient à maximiser la probabilité de détection moyenne $\beta(S)$ sous la contrainte :

$$(8) \qquad \alpha(S|_{\theta_o}) = \int_{\mathfrak{X}} P_o(dx|_{\theta_o}) \quad S(\delta_o|_x) \leq \alpha$$

qui concerne la probabilité de fausse alarme conditionnelle.

Tant que la perte est du type zéro-un (et plus généralement tant qu'elle est indépendante des paramètres indésirables), le point de vue de Neyman-Pearson moyen conduit aux calculs classiques de rapports de vraisemblance. Par contre, le point de vue conditionnel ne conduit pas au calcul de ces quantités, et la détermination des stratégies optimales est en fait un problème où manquent les méthodes générales. On peut montrer cependant que s'il existe une loi a priori dans l'espace des la perte dans ces situations. Ce moment d'ordre un s'écrit :

$$(3) \quad \rho(S) = E\left[L \,\middle|\, \bigcup_{\eta=1}^{n} H_\eta\right] = \sum_{1}^{n} \pi'_\eta \int_{\Theta_\eta} \mu_\eta(d\theta_\eta) \int_{\mathfrak{X}} P_\eta(dx|_{\theta_\eta}) \times \cdots$$

$$\cdots \times \left[\sum_{0}^{n} L(\delta_i, \eta, \theta_\eta)\, S(\delta_i|_x)\right]$$

expression dans laquelle :

• π'_η, $\eta = 1, 2, \ldots, n$ est la probabilité a priori de l'hypothèse H_η conditionnelle à l'évènement :

$$\bigcup_{\eta=1}^{n} H_\eta .$$

• $\mu_\eta(\cdot)$ est la loi a priori du paramètre indésirable θ_η.

• $P_\eta(\cdot|\theta_\eta)$ est la loi qui régit l'observation dans l'hypothèse H_η lorsque θ_η est la valeur du paramètre indésirable correspondant.

• $S(\delta_i|_x)$ désigne la probabilité de prendre la décision δ_i lorsqu'on a observé x. Les n+1 fonctions $S(\delta_o|_\cdot)$, ..., $S(\delta_n|_\cdot)$ constituent la stratégie S.

Il reste maintenant à préciser la contrainte, et sur cette question deux conceptions sont possibles, également plausibles et dignes d'intérêt pratiquement, et qui aboutissent à deux conceptions de l'optimalité dans les problèmes de test où H_o est composite (c'est-à-dire ou Θ_o contient plus d'un élément). En effet :

a) ou bien l'on considère que les pertes associées aux différentes valeurs du paramètre indésirable dans l'hypothèse H_o sont comparables, c'est-à-dire que l'on tolère de grandes valeurs du risque conditionnel E $\left[L_{|H_o,\ \theta_o}\right]$ lorsque ces valeurs sont compensées par d'autres valeurs très petites. Nous sommes alors conduits à imposer la contrainte sur le risque moyen :

$$(4) \qquad E\left[L_{|H_o}\right] = \int_{\Theta_o} E\left[L_{|H_o,\ \theta_o}\right] \mu_o(d\theta_o) \leq R \ ,$$

où $\mu_o(\cdot)$ est la loi à priori du paramètre indésirable θ_o.

Dans cette perspective, la stratégie optimale est celle qui minimise l'index (1) sous la contrainte (4), et peut être dénommée <u>optimale au sens de Neyman-Pearson moyen</u>.

b) ou bien l'on considère que ces pertes ne sont pas comparables ; alors nous sommes conduits à imposer une contrainte sur le risque conditionnel $E\left[L_{|H_o,\ \theta_o}\right]$:

$$(5) \qquad E\left|L_{|H_o,\ \theta_o}\right| \leq R \qquad \theta_o \in \Theta_o \ ,$$

où la constante R que l'on pourrait d'ailleurs fort bien remplacer dans une théorie plus générale, par une fonction de θ_o, désigne le niveau de risque conditionnel admissible.

Dans cette perspective, la stratégie est celle qui minimise l'index (1) sous la famille de contraintes (5), et peut être dénommée <u>optimale au sens de Neyman-Pearson conditionnel</u>.

Aucune des deux conceptions précédentes ne met en jeu la probabilité a priori de l'hypothèse H_o, et la seconde ne met pas en jeu la loi a priori $\mu_o(\cdot)$ du paramètre indésirable θ_o. Les stratégies optimales au sens de Neyman-Pearson moyen et de Neyman-Pearson conditionnel ne sont donc pas identiques en général, puisque celle-ci requiert moins d'éléments que celle-là.

Dans le cas fréquent en pratique où la perte est du type 0-1, c'est-à-dire où :

par la perte moyenne qu'elle entraîne. Cette moyenne est prise sur l'ensemble des situations qui peuvent se présenter. Dans le cas où le paramètre peut être assimilé à un élément aléatoire, le procédé permet d'associer à toute stratégie un nombre réel unique qui s'interprète comme l'index de performances de cette stratégie, d'ordonner l'ensemble des stratégies en fonction de la valeur de cet index, et de définir ainsi la notion d'optimalité (optimalité au sens de Bayes).

Dans les problèmes de traitement du signal, s'il semble légitime de probabiliser le paramètre à cause des fluctuations qu'il subit à chaque épreuve, par contre il paraît moins justifié de caractériser les stratégies par la perte moyenne qu'elles entraînent, parce que les différentes composantes de la perte ne sont pas toutes comparables, ce qui enlève tout sens physique à l'opération mathématique permettant de définir le risque moyen.

Dans les problèmes de détection par exemple, les fausses alarmes ont généralement des conséquences totalement différentes des non détections et des erreurs de classification, de sorte que les pertes afférentes à ces deux types de situations ne sont pas comparables ou interchangeables.

Nous pouvons alors appliquer le point de vue de Neyman Pearson, c'est-à-dire effectuer la moyenne des pertes en traitant séparément les situations non comparables, obtenir ainsi une perte moyenne à plusieurs composantes, choisir l'une des composantes pour index de performances, et éliminer les composantes restantes en leur imposant seulement de ne pas dépasser un certain niveau admissible.

Par exemple dans les problèmes de test évoqués plus haut les pertes correspondant aux situations H_1, ..., H_n sont généralement exprimables dans la même unité, ce qui permet d'adopter pour index de performances le moment d'ordre un de paramètres indésirables Θ_o telle que la stratégie du rapport de vraisemblance correspondante satisfait la contrainte (8) avec une égalité, alors cette stratégie est optimale au sens de Neyman-Pearson conditionnel. Lorsque les conditions d'application de ce résultat sont satisfaites, le point de vue conditionnel conduit donc au calcul de rapports de vraisemblance. Mais il s'agit de rapports fictifs, où la loi a priori dans l'hypothèse H_o n'est pas la loi réelle, mais une loi artificielle permettant de satisfaire la contrainte (8) avec une égalité.

5. EXTENSION DE LA THEORIE PRECEDENTE AU CAS OU L'ON DOIT RESOUDRE UNE SUITE DE PROBLEMES DE TEST INDEXES PAR LE TEMPS.

Soit J un intervalle de l'axe des temps de durée finie.

Nous devons résoudre une suite de problèmes de test définis par les éléments :

$$\{\Delta(t),\ \Theta(t),\ \mathcal{X}(t),\ L^t,\ t \in J\},$$

que nous supposons connus dès le début de l'expérimentation.

Soient δ^t, θ^t, x^t respectivement la décision, le paramètre et l'observation au temps t. Comme précédemment, il est commode de dissocier le paramètre θ^t en un paramètre décisionnel η^t et un paramètre indésirable $\theta^t_{\eta^t}$, de sorte que nous pouvons écrire l'espace paramétrique $\Theta(t)$ sous la forme :

$$\Theta(t) = \bigcup_{\eta^t=0}^{n} \eta^t \times \Theta_{\eta^t}(t)$$

Le paramètre décisionnel η^t peut aussi se représenter par la suite des variables booléennes associées aux évènements $\eta^t = 1$, $\eta^t = 2$, ..., $\eta^t = n$. Ces évènements peuvent être permanents ou instantanés. Nous noterons ces variables booléennes $\{1_{H_i}(t),\ i = 1, 2, ..., n\}$ ou $\{d\nu_i(t),\ i = 1, 2, ..., n\}$ suivant le cas.

Une stratégie S est maintenant une suite de stratégies élémentaires :

$$S = \{ S^t, \ t \in J \}$$

où chaque stratégie élémentaire S^t est une probabilité de transition |2| opérant entre les ensembles $\mathcal{X}(t)$ et $\Delta(t)$.

Pour appliquer le point de vue de Neyman-Pearson dans cette situation, nous pouvons considérer que les pertes subies lorsque l'un quelconque des évènements $H_1, \ldots, H_n$ se trouve réalisé aux différents instants de décision s'ajoutent, tandis que les pertes subies lorsque l'évènement H_o est réalisé ne s'ajoutent pas et doivent être traitées individuellement.

L'index de performances $\rho(S)$ de la stratégie S s'écrit alors :

• dans le cas d'évènements permanents

$$(9) \qquad \rho(S) = E \int_J \left[\sum_1^n {}_i \, 1_{H_i}(t) \, L^t\right] dt \quad ,$$

• dans le cas d'évènements instantanés

$$(10) \qquad \rho(S) = E \int_J \left[\sum_1^n {}_i \, L^t \, d\nu_i(t)\right] \quad ,$$

où nous supposons satisfaites les conditions assurant l'existence presque sûre des intégrales temporelles (9) et (10).

A ces deux expressions de l'index il faut adjoindre le système de contraintes :

$$(11) \qquad E\left[L^t \,|_{\eta^t = 0}\right] \leq R, \qquad t \in J,$$

si l'on adopte le point de vue de Neyman-Pearson moyen, et le système de contraintes :

$$(12) \qquad E\left[L^t \,|_{\eta^t=0,\ \theta^t_o}\right] \leq R, \quad \theta^t_o \in \Theta_o(t), \qquad t \in J,$$

si l'on adopte le point de vue conditionnel.

Revenons aux expressions (9) et (10) de l'index, et admettons qu'il est licite d'intervertir les deux opérations d'espérance mathématique et d'intégration temporelle. Les expressions (9) et (10) deviennent alors :

$$(9)' \qquad \rho(S) = \int_J \left[\Sigma_{1\,i}^{n}\ E(1_{H_i}(t)\ L^t)\right] dt$$

$$(10)' \qquad \rho(S) = \int_J \left[\Sigma_{1\,i}^{n}\ E(d\nu_i(t)\ L^t)\right]$$

Considérons les intégrants de (9)' et (10)'. Les variables $1_{H_i}(t)$ et $d\nu_i(t)$ étant des variables de pile au face, on a identiquement :

$$(13) \qquad E\left[1_{H_i}(t)\ L^t\right] = Pr\left[1_{H_i}(t) = 1\right] \times E\left[L^t \,|_{1_{H_i}(t)=1}\right]$$

$$(14) \qquad E\left[d\nu_i(t)\ L^t\right] = Pr\left[d\nu_i(t) = 1\right] \times E\left[L^t \,|_{d\nu_i(t)=1}\right]$$

où $Pr\left[1_{H_i}(t) = 1\right]$ et $Pr\left[d\nu_i(t) = 1\right]$ désignent la probabilité a priori que l'hypothèse H_i soit réalisée au temps t.

$$Pr\left[1_{H_i}(t)\right] = \pi_i(t)$$

$$Pr\left[d\nu_i(t) = 1\right] = \rho_i(t)\ dt$$

D'autre part, rappelons que étant donnés un espace de probabilité Ω, n évènements $A_1, \ldots, A_n$ deux à deux disjoints de cet espace, et une variable aléatoire X définie sur Ω, on a identiquement :

$$(15)\qquad \sum_{1}^{n} \Pr(A_i)\, E[X|_{A_i}] = [\sum_{1}^{n} \Pr(A_i)]\, E[L|_{\bigcup_{i=1}^{n} A_i}]$$

Compte tenu de (13), (14) et (15) les expressions (9)' et (10)' deviennent :

$$(9)''\qquad \rho(S) = \int_J [\sum_{1}^{n} \pi_i(t)]\, E[L^t|_{\bigcup_{i=1}^{n} \{1_{H_i}(t)=1\}}]\, dt$$

$$(10)''\qquad \rho(S) = \int_J [\sum_{1}^{n} \rho_i(t)]\, E[L^t|_{\bigcup_{i=1}^{n} \{d\nu_i(t)=1\}}]\, dt$$

Or les moments conditionnels $E[L^t|_{\bigcup_{i=1}^{n} \{1_{H_i}(t)=1\}}]$ et $E[L^t|_{\bigcup_{i=1}^{n} \{d\nu_i(t)=1\}}]$ ne sont autres que les index de performances de la stratégie élémentaire S^t. L'index $\rho(S)$ s'écrit donc comme l'intégrale des index élémentaires $\rho(S^t)$:

$$(9)'''\qquad \rho(S) = \int_J [\sum_{1}^{n} \pi_i(t)]\, \rho(S^t)\, dt$$

$$(10)'''\qquad \rho(S) = \int_J [\sum_{1}^{n} \rho_i(t)]\, \rho(S^t)\, dt$$

et donc le problème de minimiser $\rho(S)$ sous les contraintes (11) et (12) revient à minimiser en chaque point t l'index $\rho(S^t)$ sous la contrainte :

$$E[L^t|_{\eta^t=o}] \leq R$$

dans le point de vue de Neyman-Pearson moyen, et sous la contrainte : $E[L^t|_{\eta^t=o,\ \theta_o^t}] \leq R$

dans le point de vue de Neyman-Pearson conditionnel.

Nous constatons ainsi, que avec nos hypothèses, le problème d'optimaliser la décision globale revient à optimaliser chaque décision instantanée, ce qui nous ramène au paragraphe précédent.

6. APPLICATION AU PROBLEME DE DETECTION OU EXISTENT DES INCERTITUDES SUR LES CARACTERISTIQUES DU BRUIT.

Pour appliquer la théorie précédente au problème de détection initial, nous devons revenir aux équations (1) et (2) donnant l'observation, et définir l'espace des observations $\mathfrak{X}(t)$ et l'espace paramétrique $\Theta(t)$.

Comme on l'a indiqué plus haut, la décision au temps t met en jeu l'observation antérieure comprise dans un intervalle I(t). L'ensemble $\mathfrak{X}(t)$ est donc l'ensemble des formes possibles de l'observation dans cet intervalle.

Le paramètre décisionnel η^t intervient explicitement sous sa forme booléenne dans les expressions (1) et (2) de l'observation. En effet, transcrit sous forme booléenne, ce paramètre n'est autre que la suite $\{1_{H_i}(t),\ i = 1, 2, \ldots, n\}$ dans le cas d'évènements permanents, et la suite $\{dn_i(t-T),\ i = 1, 2, \ldots, n\}$ dans le cas d'évènements instantanés.

Il reste alors à préciser le paramètre indésirable associé à la réalisation de l'hypothèse H_{η^t} au temps t. La définition de ce paramètre est relativement arbitraire, puisqu'il s'agit de choisir dans l'expression de l'observation la part de cette observation que l'on choisit d'appeler "paramètre" au lieu de "réalisation d'un élément aléatoire". Le choix est libre. Une solution qui parait satisfaisante consiste à considérer comme des paramètres les différents éléments qui définissent les propriétés statistiques du bruit dans la fenêtre d'analyse I(t), la valeur du signal associé à l'hypothèse H_{η^t}, et la valeur et la position de tous les autres signaux ou portions de signaux qui apparaissent dans la fenêtre I(t) et dont on ne teste pas la présence au temps t.

Moyennant cette définition du paramètre indésirable la probabilité de transition $P_{\eta^t}(\cdot|\theta^t_{\eta^t})$ qui régit l'observation x^t dans l'hypothèse H_{η^t} n'est autre que la loi conditionnelle du bruit seul, décentrée par les différents signaux qui peuvent se présenter. Dans la suite nous admettons que cette loi est une loi de Gauss. Elle est donc définie par la donnée d'une fonction de covariance, élément d'un champ de covariances $\Gamma(t)$.

Pour faciliter l'écriture de l'espace paramétrique, il est commode d'introduire les notations suivantes : nous désignerons par J(t) l'ensemble des instants de décision autres que l'instant t dont les signaux associés sont contenus au moins partiellement dans la fenêtre I(t). Soit u un point quelconque de J(t). Aux valeurs 1, 2, ..., n de η^u est associé un signal qui est un élément d'un espace $\Sigma_{\eta^u}(u)$. Ce signal est un nombre dans le cas d'évènements permanents, et une fonction dans le cas d'évènements instantanés. Dans ce deuxième cas nous noterons $\Sigma^{I(t)}_{\eta^u}(u)$ l'ensemble de la restriction des signaux à la fenêtre d'analyse I(t). Enfin nous conviendrons qu'à la valeur zéro du paramètre η^u est associé un signal identiquement nul.

L'espace $\Theta_{\eta^t}(t)$ des paramètres indésirables associés à l'hypothèse H_{η^t} s'écrit alors immédiatement. On a :

a) dans le cas d'évènements permanents :

$$\Theta_{\eta^t}(t) = \Gamma(t) \times \Sigma_{\eta^t}(t) \times \prod_{u \in J(t)} \left\{ \bigcup_{\eta^u=o}^{n} \eta^u \times \Sigma_{\eta^u}(u) \right\}$$

b) dans le cas d'évènements instantanés :

$$\Theta_{\eta^t}(t) = \Gamma(t) \times \Sigma^{I(t)}_{\eta^t}(t) \times \prod_{u \in J(t)} \left\{ \bigcup_{\eta^u=o}^{n} {}^{u} \times \Sigma^{I(t)}_{\eta^u}(u) \right\}$$

L'espace des paramètres indésirables dans l'hypothèse H_{η^t} comprend donc un facteur qui résulte de notre incertitude sur les caractéristiques du bruit, un facteur qui résulte de notre incertitude sur la valeur du signal, et un facteur qui résulte de la possibilité d'évolution du système Σ dans la fenêtre d'analyse $I(t)$ et exprime la dimension temporelle du problème considéré.

Munis de l'espace des observations $\mathfrak{X}(t)$ et de l'espace paramétrique $\Theta(t)$:

$$\Theta(t) = \bigcup_{\eta^t=o}^{n} \eta^t \times \Theta_{\eta^t}(t),$$

nous pouvons appliquer la théorie du paragraphe 4, et concevoir les stratégies optimales pour les points de vue de Neyman-Pearson moyen et conditionnel. Cependant, la complexité des espaces de paramètres indésirables $\Theta_o(t)$, $\Theta_1(t)$, ..., $\Theta_n(t)$ rend très difficile la détermination effective de ces stratégies.

7. LE CAS OU LA PROBABILITE D'EVOLUTION DU SYSTEME Σ DANS LA FENETRE D'ANALYSE $I(t)$ EST TRES PETITE.

Cette complexité résulte essentiellement du troisième facteur des espaces paramétriques (16) et (17), c'est-à-dire de la possibilité d'évolution du système Σ dans l'intervalle de traitement $I(t)$. Par possibilité d'évolution, nous entendons le fait que conditionnellement à l'hypothèse $\eta^t = i$, les paramètres décisionnels autres que η^t et engendrant des signaux dans la fenêtre d'analyse $I(t)$ peuvent prendre des valeurs différentes de i dans le cas d'évènements permanents, et différentes de zéro dans le cas d'évènements instantanés.

Dans la plupart des situations concrètes, le système Σ peut évoluer dans l'intervalle $I(t)$, mais la probabilité de cette évolution est très petite. Nous allons montrer qu'il est légitime de négliger dans ce cas tous les problèmes soulevés par la possibilité d'évolution du système Σ dans l'intervalle $I(t)$.

a) Plaçons-nous conditionnellement à une certaine valeur i du paramètre décisionnel η^t ($i = 0, 1, \ldots, n$) et introduisons l'évènement E_i de l'espace paramètrique $\Theta_i(t)$ qui indique l'absence d'évolution du système Σ dans la fenêtre $I(t)$. E_i est défini de la façon suivante :

- s'il s'agit de détecter des états, E_i est l'évènement que tous les éléments de la famille des paramètres décisionnels intervenant au temps t :

$$\{\eta^u,\ u \in J(t)\}$$

sont égaux à i.

- s'il s'agit de détecter des points, E_i est l'évènement que tous les éléments de la famille des paramètres décisionnels intervenant au temps t :

$$\{\eta^u,\ u \in J(t)\}$$

sont égaux à zéro.

b) Revenons maintenant aux expressions de l'index et des contraintes pour la stratégie instantanée S^t. On a pour expression de l'index :

$$\rho(S^t) = E\left[L^t \Big| \bigcup_{i=1}^{n} \{\eta^t = i\}\right] = \sum_{1}^{n} {}_i\ \pi_i{}'(t)\ E\left[L^t \Big|_{\eta^t = i}\right] \tag{18}$$

où $\pi_i{}'(t)$ désigne la probabilité a priori que $\eta^t = i$ lorsque l'évènement $\bigcup\limits_{i=1}^{n} \{\eta^t = i\}$ est réalisé, et pour expression des contraintes :

$$E\left[L^t \Big|_{\eta^t = o}\right] \leq R \tag{19}$$

dans le point de vue moyen, et

$$(20) \qquad E\left[L^t \big|_{\eta^t=o,\ \theta_o^t}\right] < R \qquad \theta_o^t \in \Theta_o(t)$$

dans le point de vue conditionnel.

c) Décomposons maintenant les moments du type $E\left[L^t \big|_{\eta^t=i}\right]$ qui interviennent dans les expressions (18) et (19) en les conditionnant par les évènements E_i et leurs contraires $\overline{E_i}$:

$$(21) \quad E\left[L^t \big|_{\eta^t=i}\right] = Pr(E_i)\, E\left[L^t \big|_{\eta^t=i,\ E_i}\right] + Pr(\overline{E_i})\, E\left[L^t \big|_{\eta^t=i,\ \overline{E_i}}\right].$$

En portant (21) dans (18) et (19), ces expressions deviennent :

$$(22) \quad \rho(S^t) = \sum_{1}^{n}{}_i \ \pi'_i(t)\, Pr(E_i)\, E\left[L^t \big|_{\eta^t=i,\ E_i}\right] + \sum_{1}^{n}{}_i \ \pi'_i(t)\, Pr(\overline{E_i})\, E\left[L^t \big|_{\eta^t=i,\ \overline{E_i}}\right]$$

$$(23) \quad E\left[L^t \big|_{\eta^t=o}\right] = Pr(E_o)\, E\left[L^t \big|_{\eta^t=o,\ E_o}\right] + Pr(\overline{E_o})\, E\left[L^t \big|_{\eta^t=o,\ \overline{E_o}}\right].$$

Semblablement, remplaçons la contrainte (20) par le couple de contraintes :

$$(24) \quad \begin{cases} E\left[L^t \big|_{\eta^t=o,\ E_o,\ \theta_o^t}\right] \leq R, & \theta_o^t \in E_o \\ E\left[L^t \big|_{\eta^t=o,\ \overline{E_o},\ \theta_o^t}\right] \leq R, & \theta_o^t \in \overline{E_o} \end{cases} .$$

L'intérêt des transformations précédentes est de séparer dans les expressions des pertes et des contraintes une partie où l'hypothèse est figée dans toute la tranche temporelle I(t), et où donc tout se passe comme si le système Σ n'évoluait pas au cours du temps, d'une partie qui concentre tous les effets de l'évolution temporelle.

d) Supposons maintenant la perte L^t majorée par un nombre positif M, la probabilité de E_o, E_1, ..., E_n extrêmement faible, et considérons les expressions (22) et (23). Alors, quelle que soit la stratégie considérée, la contribution aux pertes moyennes de la partie qui résulte de l'évolution de Σ dans l'intervalle I(t) est négligeable et en tous cas majorée par le nombre :

$$\max\left[\Pr(\overline{E_o}), \Pr(\overline{E_1}), \ldots, \Pr(\overline{E_n})\right] \times M,$$

que l'on peut rendre arbitrairement petit.

La recherche de la stratégie optimale au sens de Neyman-Pearson moyen, c'est-à-dire de la stratégie qui minimise (22) avec une contrainte sur (23), peut donc s'effectuer en négligeant cette partie, c'est-à-dire en faisant comme si le système Σ n'avait pas d'évolution temporelle. L'erreur commise dans cette approximation est d'autant plus faible que la probabilité a priori de l'évolution est plus petite.

Examinons maintenant le cas du point de vue de Neyman-Pearson conditionnel. L'index étant le même que précédemment, il est justiciable de la même approximation. La contrainte (20) pose par contre un problème nouveau, parce qu'elle ne met pas en jeu $\Pr(\overline{E_o})$, de sorte qu'il est inutile d'espérer obtenir le résultat escompté en invoquant le processus limite $\Pr(\overline{E_o}) \to 0$. Observons toutefois la forme (24) de cette contrainte. Seule la première contrainte de (24) nous intéresse, parce que la seconde concerne une éventualité dont la probabilité est très petite, et qu'il parait absurde d'imposer une contrainte du type (24) sur une éventualité qui ne se produit presque jamais. Nous pouvons donc éliminer cette deuxième contrainte de la définition du point de vue de Neyman-Pearson conditionnel. Moyennant cette convention, nous sommes conduits à la même conclusion que dans le point de vue moyen : tant que la probabilité d'évolution du système Σ est très petite, il est légitime de négliger tous les problèmes soulevés par cette évolution.

Cette circonstance permet de simplifier considérablement les espaces de paramètres indésirables (16) et (17). Il devient alors inutile de distinguer le cas des évènements permanents du cas des évènements instantanés. Soit $\Sigma_i^{I(t)}$ l'ensemble des signaux apparaissant dans l'intervalle I(t) dans l'hypothèse H_i. On a :

- pour i = 1, 2, ..., n

$$\Theta_i(t) = \Gamma(t) \times \Sigma_i^{I(t)}$$

- pour i = o

$$\Theta_o(t) = \Gamma(t).$$

Nous retrouvons ainsi l'énoncé habituel de la théorie de la détection, qui apparait donc comme un cas limite de la théorie plus générale où le temps intervient explicitement.

8. APPLICATION AU PROBLEME DE DETECTION ADAPTATIVE

En nous plaçant dans le cadre de l'approximation précédente, nous pouvons alors revenir aux problèmes adaptatifs. La durée de l'intervalle d'analyse associé à chaque décision instantanée est généralement de l'ordre de grandeur du temps de stationnarité locale du bruit. L'ensemble des covariances possibles du bruit de la théorie précédente devient ainsi un ensemble de fonctions de corrélation.

A cette différence près, la théorie se déroule comme dans le cas précédent. En particulier, nous constatons qu'il existe deux conceptions de l'optimalité pour les systèmes adaptatifs, correspondant aux points de vue de Neyman-Pearson moyen et conditionnel. Ces deux conceptions aboutissent à des structures essentiellement différentes.

Dans ces structures, les fonctions de détection et d'adaptation se trouvent en général mélangées intimement, de sorte que leur séparation devient un artifice technique qui peut être poussé aussi loin qu'on le désire, sans qu'il y ait un but absolu qui puisse être définitivement atteint. D'autre part, ces structures dépendent des lois a priori. C'est pourquoi les méthodes adaptatives traditionnelles, qui dissocient nettement les fonctions d'adaptation et de détection dès le départ, et qui ne font pas appel aux lois a priori, ne sauraient être optimales en général. Cependant, plusieurs symptômes laissent prévoir que ces méthodes conduisent à des performances assez voisines des performances optimales.

En effet, il est fréquent que le bruit soit très lentement non stationnaire, de sorte que l'estimation de la fonction de corrélation locale est excellente. Considérons alors la situation asymptotique où l'estimation est effectuée avec une erreur nulle. On montre aisément que les structures optimales doivent dans ce cas calculer le rapport de vraisemblance conditionnel et le comparer à un seuil fixe dans le point de vue de Neyman-Pearson moyen, et un seuil mobile dépendant de la fonction de corrélation dans le point de vue conditionnel. L'adaptation et la détection se trouvent ainsi complètement dissociées. Sous réserve que le seuil soit ajusté correctement, on peut donc conjecturer que les méthodes adaptatives traditionnelles conduisent à des performances très voisines des performances optimales chaque fois que la fonction de corrélation du bruit peut s'estimer avec une bonne précision.

D'autre part, à cause de la diversité des types de bruit que l'on peut rencontrer, et de l'égalité approximative des fréquences d'apparition de chacun de ces types, le champ des fonctions de corrélation possibles du bruit est très vaste, et la loi a priori de la fonction de corrélation est répartie de façon très diffuse dans ce champ. La donnée de cette loi n'amène pas une information bien précise sur la forme de la fonction de corrélation à chaque épreuve ; en tous cas cette information est beaucoup moins précise que celle déduite de l'étude empirique du bruit lui-même, et cela même lorsque le bruit est relativement instable. D'où l'on déduit qu'il doit exister des récepteurs indépendants des lois a priori et dont les performances moyennes sont proches des performances optimales. Les récepteurs adaptatifs traditionnels semblent bien être un exemple de ces récepteurs privilégiés, du moins tant que l'on régule leur seuil correctement.

REFERENCES

|1| MERMOZ H. "Antennes de détection optimales et adaptatives" Collection technique et scientifique du CNET (1971).

|2| NEVEU J. "Bases mathématiques du calcul des probabilités." Masson 1967.