

# Passerelles entre les Approches Statistiques et Géométriques de la Détection

Hervé LACRESSE, Antoine GRALL, Igor NIKIFOROV

Laboratoire de Modélisation et Sûreté des Systèmes, Université de Technologie de Troyes  
12, rue Marie Curie, BP 2060, 10010 Troyes Cedex, France

Herve.Lacresse@utt.fr, Antoine.Grall@utt.fr, Igor.Nikiforov@utt.fr

**Résumé** – Les deux principaux types d’approche des problèmes de détection et d’isolation de défaillances sont rappelés ici dans le but de mettre en évidence les liens théoriques qui existent entre celles-ci. Cette tentative d’unification permet également d’envisager des problèmes de détection d’une complexité accrue et la prise en compte de certaines informations jusque là négligées.

**Abstract** – The two main approaches of Fault Detection and Isolation (FDI) are briefly recalled here in order to put a stress on the theoretical links that might lead to a further unification of FDI. This relationship between these two approaches also allows the consideration of increasingly complex problems and might take into account some, so far, neglected information.

## 1 Formulation générale des problèmes de détection envisagés

Nous nous intéresserons à des modèles de la forme générique suivante <sup>(1)</sup>

$$\begin{cases} X_{k+1} = F(X_k, U_k, \theta) \\ Y_k = H(X_k, U_k, \theta) + \xi_k \end{cases} \quad \text{pour } k \geq 0, \quad (1)$$

où  $X_k, U_k, Y_k$  sont respectivement l’état (inconnu), une entrée du système (supposée ici connue) et la sortie (vecteur de mesures) à l’instant  $k$ ;  $\xi_k$  représente un bruit de mesure à l’instant  $k$ ,  $\theta \in \mathbf{R}^m$  est un vecteur de paramètres du modèle, considéré, en première approche, comme constant dans le temps.

Nous avons, à un instant  $n$ , une suite d’observations

$$(Y_0, \dots, Y_n, U_0, \dots, U_n)$$

où  $Y_k = Y(k\delta)$ ,  $U_k = U(k\delta)$  et  $\delta$  est le pas de discrétisation, on note aussi  $\mathcal{Y}_0^N = (Y_0, \dots, Y_n)^T$  et  $\mathcal{U}_0^N = (U_0, \dots, U_n)^T$ . Le problème général envisagé est la construction d’une règle de décision entre une hypothèse  $\mathcal{H}_0 : \theta \in \Theta_0$ , qui correspond à un fonctionnement normal du système, et une hypothèse  $\mathcal{H}_1 : \theta \in \Theta_1$ , qui correspond à un fonctionnement dégradé (assez souvent, l’hypothèse de base  $\mathcal{H}_0$  consiste «idéalement» à vérifier la nullité des paramètres formant le vecteur  $\theta$ ).

## 2 Position géométrique du problème

### 2.1 Description générale

Pour la communauté Automatique, le problème consiste à construire des relations de redondance analytique (RRA)

1. Cette écriture peut résulter de la «discrétisation» d’un modèle en temps continu

$(\mathcal{Y}_0^N, \mathcal{U}_0^N) \mapsto G(\mathcal{Y}_0^N, \mathcal{U}_0^N) = (g_1(\mathcal{Y}_0^N, \mathcal{U}_0^N), \dots, g_r(\mathcal{Y}_0^N, \mathcal{U}_0^N))$  telles que le système des  $r$  relations correspondant à

$$G(\mathcal{Y}_0^N, \mathcal{U}_0^N) = 0 \quad (2)$$

caractérise un fonctionnement normal et

$$G(\mathcal{Y}_0^N, \mathcal{U}_0^N) \neq 0 \quad (3)$$

caractérise une défaillance <sup>(2)</sup>. Le traitement effectif du problème de décision se fait alors en visant, non pas, bien sûr, la vérification exacte de ces relations, mais en examinant l’espérance  $\mathbf{E}_\theta(G(\mathcal{Y}_0^N, \mathcal{U}_0^N))$  pour comparer cette valeur à un seuil à partir duquel on décide qu’il y a défaillance, cela par des moyens statistiques qu’on peut qualifier de «classiques».

### 2.2 Illustration

On peut illustrer ce type d’approche par le traitement du modèle statique suivant (nous verrons plus bas qu’il ne s’agit pas nécessairement d’une vision restrictive par rapport à la forme (1) lorsque (1) est linéaire)

$$Y = HX + \theta + \xi \quad (4)$$

où  $Y$  correspond aux mesures,  $\theta$  à une défaillance éventuelle et  $\xi$  à un bruit gaussien de moyenne nulle. Les relations de redondance analytique sont dans ce cas des relations du type « $G(Y) = 0$ » (en l’absence d’une entrée  $U$ ) et on constate que ce type de relation est obtenu grâce à des résultats classiques d’algèbre linéaire en cherchant une base du noyau à gauche de la matrice  $H$  qui définira une projection  $W$  telle que

$$WH = 0$$

On souhaite alors construire un test pour déterminer si  $\mathcal{H}_0 : \mathbf{E}(WY) = 0$  ou bien  $\mathcal{H}_1 : \mathbf{E}(WY) \neq 0$ , ce que l’on

2. On peut voir ces relations comme résultant de l’élimination de l’état du système entre  $\mathcal{Y}_0^N$  et  $\mathcal{U}_0^N$

réalise en utilisant un test de type « $\chi^2$ » sur  $\|WY\|^2$  (si  $\xi$  est un bruit gaussien,  $WY$  reste une statistique gaussienne par transformation linéaire). Il s'agit bien de déterminer, pour un niveau  $\alpha$  souhaité pour le test, le seuil  $\lambda$  tel que l'inégalité  $\|WY\|^2 \geq \lambda$  permet de penser que  $\mathcal{H}_1$  est vérifiée.

### 3 Approches statistiques

On considère le cas général où l'ensemble d'observations  $\mathcal{Y}_1^n$  est issu d'une distribution de probabilité  $P_\theta$  quelconque,  $\theta \in \Theta \subseteq \mathbf{R}^m$ . L'hypothèse de base peut se mettre sous la forme assez générale

$$\mathcal{H}_0 = \{\theta \in \omega : \eta(\theta) = 0\} \quad (5)$$

avec  $\omega \subsetneq \Theta$ , ce qui, par l'intermédiaire des composantes  $\eta_i$  ( $1 \leq i \leq r$ ) de la fonction  $\eta : \Theta \rightarrow \mathbf{R}^r$  définit l'ensemble  $\omega$  par un système de  $r$  relations. Cette formulation a été proposée par Wald<sup>(3)</sup>[1]. La contre-hypothèse est définie par

$$\mathcal{H}_1 = \{\theta \in \Theta_1 = \{\Theta \setminus \omega\}\} \quad (6)$$

Pour établir une analogie avec l'approche géométrique et les RRA,  $\eta(\theta) = 0$  caractérise un fonctionnement normal ( $\mathcal{H}_0$ ) et  $\eta(\theta) \neq 0$  caractérise une défaillance ( $\mathcal{H}_1$ ). Wald a démontré (sous certaines conditions de régularité imposées sur les fonctions  $\eta_1, \dots, \eta_r$  et les fonctions supplémentaires  $\eta_{r+1}, \dots, \eta_m$ ) qu'il est possible de construire un test de choix, entre  $\mathcal{H}_0$  et  $\mathcal{H}_1$ , asymptotiquement ( $n \rightarrow \infty$ ) *le plus rigoureux* (asymptotically most stringent test) dans la classe  $\mathcal{K}_\alpha$  des tests de niveau  $\alpha$ .

Rappelons qu'un test qualifié de **plus rigoureux**, parmi des tests de niveau  $\alpha$  construits pour un même problème statistique, est un test  $\delta^*$  qui vérifie :

$$\sup_{\theta \in \Theta_1} \{\beta(\theta) - \beta_{\delta^*}(\theta)\} \leq \sup_{\theta \in \Theta_1} \{\beta(\theta) - \beta_\delta(\theta)\} \quad (7)$$

pour tout test  $\delta$  de niveau  $\alpha$ , en notant  $\theta_\delta$  la fonction puissance d'un test et  $\beta$  l'enveloppe supérieure de toutes les fonctions puissances.

Le test proposé est fondé sur la statistique suivante :

$$\tilde{\Lambda}(\mathcal{Y}_1^n) = n \eta(\hat{\theta}(n))^T \tilde{\mathcal{F}}_r^{-1}(\hat{\theta}(n)) \eta(\hat{\theta}(n)), \quad (8)$$

où  $\hat{\theta}(n)$  est l'estimation des composantes de  $\theta$  qui maximise la fonction de vraisemblance  $f_\theta(\mathcal{Y}_1^n)$ , où la matrice carrée  $\tilde{\mathcal{F}}_r$  est obtenue à partir de la matrice  $\Sigma = J\mathcal{F}^{-1}J^T$  en supprimant ses  $m - r$  dernières lignes et ses  $m - r$  dernières colonnes.  $J$  est la matrice jacobienne de l'application  $\tilde{\eta} : \theta \mapsto (\eta(\theta)^T, \eta_{r+1}(\theta), \dots, \eta_m(\theta))^T$ ,  $\mathcal{F}$  est la matrice d'information de Fisher obtenue à partir de  $\mathcal{Y}_0^N$ . Le test  $\delta^*$  fondé sur (8) est, également, *le plus puissant* dans la classe des tests de niveau  $\alpha$  ayant une fonction puissance  $\beta(\theta)$  *constante* sur chaque surface  $S_c \in \mathcal{S}$ , avec la famille de surfaces  $\mathcal{S} = \{S_c\}$  définie par

$$S_c = \left\{ \theta / \eta(\theta)^T \tilde{\mathcal{F}}_r^{-1}(\theta) \eta(\theta) = c \right\}$$

3. La théorie de Wald exige, également, l'existence de  $m - r$  fonctions supplémentaires  $\eta_i = \eta_i(\theta_1, \dots, \theta_m)$ ,  $i = r + 1, \dots, m$  pour établir un difféomorphisme de l'ensemble des paramètres sur lui-même à l'aide des fonctions  $\eta_1, \dots, \eta_m$ .

ce qui signifie qu'il vérifie :

$$\beta_{\delta^*}(\theta) = \beta_{\delta^*}(\theta'), \forall \theta \in \Theta_1, \forall \theta' \in \Theta_1 \quad (9)$$

et

$$\beta_{\delta^*}(\theta) \geq \beta_\delta(\theta), \forall \theta \in \Theta_1 \quad (10)$$

pour tout test  $\delta$  de niveau  $\alpha$  donné, ayant une fonction puissance constante sur les surfaces de la famille  $\mathcal{S}$ .

## 4 Passerelles entre les deux types d'approche

### 4.1 Principes généraux

Par analogie avec la notion de RRA rappelée au paragraphe 2, on peut parler de *relations de redondance paramétrique* (RRP) pour désigner les équations du type  $\eta(\theta) = 0$  mises en place au paragraphe 3.

Le passage des RRA aux RRP est relativement simple dans le cas où le bruit  $\xi_0^N$  intervient dans la forme interne<sup>(4)</sup> des RRA à travers des fonctions linéaires. Il existe donc un lien entre les deux types d'approches et nous examinerons la manière dont on peut compléter les relations de redondance analytique obtenues pour un système du type (1) afin d'obtenir la situation décrite par Wald<sup>(3)</sup>. Nous le ferons d'abord dans le cas où les observations  $\mathcal{Y}_0^N$  dépendent *linéairement* de l'état  $X_0$  du système (on sait alors procéder à l'élimination de l'état par des opérations d'algèbre linéaire), puis dans le cas plus général où le système (1) est issu d'un système dynamique où les observations dépendent de manière *polynomiale* de l'état (on dispose alors de techniques d'élimination de l'état reposant sur des outils de calcul différentiel, comme les bases de Gröbner). La théorie de Wald peut alors déboucher sur la construction d'un test optimal pour les problèmes décrits par ces modèles.

Les RRP ou les RRA évoquées jusqu'ici sont des contraintes d'*égalité* (statistique) et une étape supplémentaire consistera à prendre en compte des contraintes de type *inégalité* sur les paramètres  $\theta$  des modèles considérés, afin de pouvoir prendre en compte l'information contenue dans les bornes éventuelles que l'on peut déterminer pour certains paramètres de nuisance : il est en effet «classique» de procéder à l'élimination d'un paramètre parce que sa valeur est inconnue à un instant donné, cependant la réalité physique du problème envisagé permet de fixer des bornes «raisonnables» à ce paramètre (altitude d'un aéronef, puissance d'un moteur...). Des investigations sur cet aspect du problème ([3]) ont déjà montré la pertinence de la prise en compte d'informations «de bornes» pour l'amélioration des règles de décision usuelles.

### 4.2 Illustration de ces liens sur un exemple

Considérons le cas où le système (1) est linéaire avec la forme suivante

$$\begin{cases} X_{k+1} &= AX_k + BU_k + F_x \theta_k \\ Y_k &= CX_k + DU_k + F_y \theta_k + \xi_k \end{cases} \quad (11)$$

4. i.e. obtenue en remplaçant  $\mathcal{Y}_0^N$  par son expression en fonction de  $\theta_0^N$ ,  $\mathcal{U}_0^N$  et  $\xi_0^N$  après élimination de l'état.

où nous supposons toujours l'entrée  $U_k$  connue à chaque instant. La réécriture de l'équation de mesure

$$Y_k = CX_k + DU_k + F_y\theta_k + \xi_k$$

entre les instants  $k = 0$  et  $k = N$  permet de mettre (11) sous la forme suivante ([2]) :

$$\mathcal{Z}_0^N \triangleq \mathcal{Y}_0^N - \mathcal{M}_N(B, D)\mathcal{U}_0^N = \mathcal{O}_N X_0 + \mathcal{M}_N(F_x, F_y)\theta_0^N + \xi_0^N$$

avec

$$\mathcal{O}_N(A, C) = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^N \end{pmatrix}$$

la matrice d'observabilité liée à  $A$  et  $C$  et  $\mathcal{M}_N(B, D)$ ,  $\mathcal{M}_N(F_x, F_y)$  des matrices de Toeplitz adéquates. Nous sommes alors ramené à une équation statique de la forme

$$Z = HX + \theta + \xi \quad (12)$$

où  $X$  correspond à l'état,  $\xi$  au bruit et  $\theta$  à une défaillance éventuelle. Le problème de décision envisagé est

$$\mathcal{H}_0 = \{Z \sim \mathcal{N}(HX, I_N)\}$$

contre

$$\mathcal{H}_1 = \{Z \sim \mathcal{N}(HX + \theta, I_N), \theta \neq 0\} \quad (13)$$

Ce problème statistique est laissé invariant par le groupe des transformations liées aux différents états  $X$  possibles

$$G = \{T_X : Z \mapsto Z + HX\}$$

et l'on obtient une statistique invariante (i.e. qui permet de construire une règle de décision sans utiliser l'état inconnu  $X$ ) lorsqu'on construit une projection  $W$  telle que

$$WH = 0 \text{ et } WW^T = I \quad (14)$$

satisfaisant une conditions «naturelle» de rang liée à la dimension du noyau à gauche de  $H$ .

L'approche proposée dans [1] peut alors être mise en place en définissant l'ensemble

$$\omega = \{\theta / W\theta = 0\} \quad (15)$$

L'application  $\eta$  mentionnée plus haut est alors définie par  $\theta \mapsto \eta(\theta) = W\theta$  et, si les lignes de  $W$  sont indépendantes, on peut compléter les  $r$  relations correspondant à  $\eta(\theta) = 0$  pour suivre la démarche de Wald par  $N - r$  autres relations. L'estimation de  $\theta$  au sens du maximum de vraisemblance est dans ce type de cas donnée trivialement par  $\hat{\theta} = Z$  et l'approche proposée débouche sur l'utilisation de la statistique

$$T(Z) = \|WZ\|^2 \quad (16)$$

si l'on veut obtenir un test ayant les propriétés d'optimalité énoncées dans [1].

On peut montrer que, sous les conditions mentionnées plus haut, le choix de la projection  $W$  n'influence pas la qualité d'un test fondé sur  $T(Z)$ , qui est, en fait, le test utilisé classiquement dans les approches par redondance analytique.

La deuxième condition imposée dans (14) peut en effet être vue ici comme un élément de «confort» calculatoire, car si l'on choisit une autre projection  $W' = AW$  avec  $W$  vérifiant (14) et  $A$  une matrice carrée quelconque, la démarche suivie précédemment conduit également à l'utilisation de la statistique  $T(Z) = \|WZ\|^2$ .

On obtient ainsi sur cet exemple une justification statistique théorique de l'approche par RRA déjà évoquée pour ce type d'exemple dans 2.2.

## 5 Conclusion

Les perspectives entrouvertes ici laissent penser qu'il est possible d'obtenir une cohérence théorique accrue entre les approches géométriques et statistiques des problèmes de détection, tout en se montrant plus systématique dans la sélection de la partie considérée comme pertinente des informations disponibles pour construire des règles de décision.

## Références

- [1] A.Wald. *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*. Trans. Amer. Math. Soc., 54:426-482,1943.
- [2] M.Basseville. *Information Criteria for Residual Generation and Fault Detection and Isolation*. Automatica, 33(5):783-803,1997
- [3] H.Lacresse, A.Grall. *Fault Detection in the Presence of a Bounded Nuisance Parameter*. ACC2001, Arlington, Virginie, USA, 23-25 juin 2001, pp 2700-2701