Détection et localisation séquentielle d'anomalies non-orthogonales

Lionel FILLATRE[1], Igor NIKIFOROV[1]

[1]Institut Charles Delaunay, FRE CNRS 2848, Université de technologie de Troyes,
12 rue Marie Curie, BP 10010, Troyes, France
Lionel.Fillatre@utt.fr, Igor.Nikiforov@utt.fr

Résumé – Ce papier s'intéresse à la détection et à la localisation la plus rapide possible d'anomalies non-orthogonales au moyen d'un test statistique qui traite les données par bloc de taille fixe. Il s'agit de minimiser le retard moyen de décision dans la classe des tests par bloc de données contraints par un temps moyen entre fausse alarme et une probabilité maximale de mauvaise localisation bornés. Ce retard moyen optimal n'était connu jusqu'à présent que pour des anomalies orthogonales. Les résultats majeurs de ce papier sont 1) la présentation d'un test par bloc de données qui minimise le retard moyen de détection/localisation et 2) l'équation de ce retard moyen. Des calculs numériques confirment la pertinence et la qualité des résultats théoriques.

Abstract – This work concerns the detection and isolation of non-orthogonal anomalies by using a fixed-size sample test. The mean delay of decision is minimized within the class of fixed-size sample tests satisfying a given mean time between false alarm and a given maximum false isolation probability. The asymptotical value of this delay is given. Numerical experiments confirm the relevance and quality of the theoretical results.

1 Motivation

La détection et localisation d'anomalies et de changements brusques dans des signaux aléatoires revêt une grande importance dans bon nombre d'applications relatives à la surveillance des systèmes. D'un point de vue formel, il s'agit de la généralisation du problème de détection d'un changement brusque, une hypothèse de base contre une seule hypothèse alternative (les références [1, 2, 3, 4] présentent un grand nombre de résultats dans ce domaine), au cas de plusieurs hypothèses alternatives. La contribution de ce papier consiste principalement à comparer des stratégies séquentielles et des stratégies par bloc de données de taille fixe (BTF) pour la détection et localisation d'hypothèses Gaussiennes suivant un critère de type "minimax" qui cherche à minimiser le retard moyen à la détection/localisation (voir [3]). Un algorithme séquentiel traite les données au fur et à mesure de leur collecte. Un algorithme BTF collecte un nombre fixé *a priori* d'observations puis traite ce bloc de données cumulées d'un seul coup. Les algorithmes séquentiels sont souvent théoriquement optimaux mais, en pratique, les algorithmes BTF présentent de gros avantages : ils sont souvent mieux adaptés à la collecte des données (faite par bloc de mesures), ils sont simples à implémenter et ils sont particulièrement bien adaptés aux systèmes à structure variable (si la structure du système d'observation varie, l'adaptation d'un algorithme séquentiel à ces variations est une tâche très délicate).

Le problème de détection/localisation d'anomalies volumiques dans les réseaux de télécommunications [5] illustre bien l'intérêt des approches par bloc de données. Un réseau est typiquement composé de plusieurs noeuds et chaque couple de noeuds distincts est susceptible de communiquer et de générer un flot de données dit flot Origine-Destination (OD). Les administrateurs du réseau ont la charge de router (multiplexer et orienter) ces différents flots OD sur un nombre limité de liens physiques en respectant un certain niveau de qualité (débit, retard, etc). La volumétrie de trafic sur les liens physiques du réseau est mesurée de manière routinière via le protocole Simple Network Management Protocol (SNMP). Ces mesures SNMP sont reliées aux quantités de trafic OD par une matrice de routage qui modélise la façon dont sont routés les flots OD sur les liens physiques. Cette matrice évolue dans le temps en fonction des décisions (imprévisibles) des administrateurs. Le système de mesures a ainsi une structure variable qui pénalise les méthodes purement séquentielles. Par ailleurs, la récupération des données SNMP n'est pas sans problème : perte de mesures, problème de synchronisation, information non fournie par certain routeur non-coopératif... Une approche purement séquentielle qui s'appuie sur une accumulation progressive des données peut être très perturbée par ces problèmes de collecte. Au contraire, une approche par bloc de données y sera moins sensible.

Dans [2, 6], les deux types d'algorithmes sont comparés pour des hypothèses Gaussiennes orthogonales. Chaque hypothèse caractérise le vecteur des observations de façon unique suivant son vecteur moyen (la matrice de covariance est supposée connue) qui correspond typiquement à l'anomalie à détecter/localiser. Les vecteurs moyens des différentes hypothèses sont supposés orthogonaux. Ceci correspond typiquement à la détection/localisation d'anomalies dans des canaux d'observa-

tions indépendants : une anomalie apparaît dans un seul canal d'observation et le bruit n'est pas corrélé entre les canaux. Ce papier propose deux extensions majeures des travaux [2, 6, 7]. Premièrement, il traite le cas plus général où l'anomalie peut apparaître simultanément dans plusieurs canaux (anomalie non-orthogonale). Dans l'exemple précédent sur la surveillance des réseaux de télécommunications, il est évident qu'une anomalie volumique qui apparaît dans un flot OD va affecter plusieurs mesures SNMP et les anomalies finalement détectables dans les mesures SNMP ne seront pas orthogonales. Cette situation se produit presque systématiquement en présence de paramètres de nuisance (non-)linéaires lorsque les anomalies sont détectées dans l'espace de parité (utilisation du principe d'invariance) [8]. Deuxièmement, le critère d'optimalité intègre désormais la pire probabilité de mauvaise localisation d'une anomalie, ce qui le rend davantage pertinent en pratique.

2 Position du problème

Soit $(Y_t)_{t\geq 1}$ une séquence aléatoire de variables Gaussiennes indépendantes observées de manière séquentielle :

$$\mathcal{L}(Y_t) = \begin{cases} \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 I_p) & \text{si} \quad t \leq t_0 \\ \mathcal{N}(\boldsymbol{\theta}_\ell, \sigma^2 I_p) & \text{si} \quad t > t_0 \end{cases} \tag{1}$$

où la distribution de Y_t est notée $\mathcal{L}(Y_t)$, $\mathcal{N}(\boldsymbol{\theta}, \sigma^2 I_p)$ est la distribution Gaussienne de moyenne $\boldsymbol{\theta}$ et de matrice de covariance $\sigma^2 I_p$ (σ est connue), $Y_t \in \mathbb{R}^p$, $\boldsymbol{\theta}_\ell \in \mathbb{R}^p$, $1 \leq \ell \leq K$, $K \geq 2$, $\boldsymbol{\theta}_0^T = (0, \ldots, 0)$ et I_p est la matrice identité d'ordre p. Les vecteurs $\boldsymbol{\theta}_\ell$ sont connus et ont la même norme $\|\boldsymbol{\theta}_\ell\|_2^2 = c^2$ pour tout ℓ. L'algorithme de détection/localisation fournit un couple (N, ν), basé sur les observations Y_1, Y_2,$\cdots$, où N est le temps d'arrêt où un changement de type ν est déclaré et $\nu,\ \nu \in \{1, \cdots, K\}$, est la décision finale. Soit $P^\ell_{t_0+1}$ la distribution des observations Y_1, Y_2, $\ldots$ quand $t_0 = 0, 1, 2, \ldots$, $\Pr^\ell_{t_0+1}(A)$ désigne la probabilité de l'événement A sous $P^\ell_{t_0+1}$ et $E^\ell_{t_0+1}$ désigne l'espérance mathématique sous $P^\ell_{t_0+1}$. Par convention, $P^0_\infty = P_0$ et $E_0(\cdot) = E^0_\infty(\cdot)$.

Un algorithme BTF se décrit comme suit : un bloc de m mesures est constitué (m est fixé) et, ensuite, une fonction de décision basée sur ce bloc de données choisit une hypothèse parmi $\mathcal{H}_0 : \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$, $\mathcal{H}_1 : \{\boldsymbol{\theta} = \boldsymbol{\theta}_1\}$, $\ldots$ et $\mathcal{H}_K : \{\boldsymbol{\theta} = \boldsymbol{\theta}_K\}$. La récolte des blocs de données est arrêtée dès qu'une décision $\overline{\nu}$ est prise en faveur d'une hypothèse $\mathcal{H}_{\overline{\nu}} : \{\boldsymbol{\theta} = \boldsymbol{\theta}_{\overline{\nu}}\}$ avec $\overline{\nu} > 0$. La règle de décision $(\overline{N}, \overline{\nu})$ s'écrit finalement :

$$\begin{aligned} \overline{N} &= \inf_{n\geq 1} \left\{ n\, m : \overline{S}_{(n-1)m+1} \geq h \right\} \\ \overline{\nu} &= d(Y_{\overline{N}-m+1}, \ldots, Y_{\overline{N}}) \end{aligned} \tag{2}$$

où

$$\overline{S}_k = \max_{1\leq \ell \leq K} S_k(\ell),$$

$$S_k(\ell) = \frac{1}{\sigma^2} \sum_{t=k}^{k+m-1} Y_t^T \boldsymbol{\theta}_\ell - \frac{m\, c^2}{2\sigma^2}$$

et h est un réel (seuil de décision). Soit $\mathcal{K}_{(\gamma,b)}$ la classe des algorithmes séquentiels (N, ν) de détection/localisation qui satisfont les contraintes suivantes

$$\begin{aligned} &\mathbb{E}_0(N) \geq \gamma\,, \\ &\max_{1\leq \ell \leq K} \max_{1\leq j\neq \ell \leq K} \sup_{t_0\geq 0} \Pr^\ell_{t_0+1}(\nu = j | N > t_0) \leq b. \end{aligned}$$

L'objectif est de construire un algorithme BTF au sein de la classe $\mathcal{K}_{(\gamma,b)}$ qui minimise le retard moyen maximal de détection/localisation :

$$\bar{\tau} \stackrel{\text{def}}{=} \sup_{t_0\geq 0, 1\leq \ell \leq K} \mathbb{E}^\ell_{t_0+1}(N - t_0 \mid N > t_0).$$

3 Stratégie BTF optimale

Soit $\delta_{i,j} = \frac{1}{2}\|\mathbf{e}_i - \mathbf{e}_j\|_2^2$ la distance entre deux "anomalies unitaires" où $\mathbf{e}_i = \boldsymbol{\theta}_i/c$ et $\mathbf{e}_i \neq \pm\mathbf{e}_j$ pour tout $1 \leq i \neq j \leq K$. Les nombres réels $\{\delta_{i,j}\}_{1\leq i\neq j\leq K}$ décrivent la "géométrie" mutuelle des hypothèses. Soit $\bar{\delta}_{\rm d} = \min_{1\leq j\leq K} \delta_{0,j} = 1/2$, $\bar{\delta}_{\rm i} = \min_{1\leq i\neq j\leq K} \delta_{i,j}$ (clairement, $0 < \bar{\delta}_{\rm i} < 2$) et $\omega^2 = c^2/\sigma^2$. On suppose qu'une "zone d'indifférence", interdite pour les autres anomalies, est définie autour de chaque extrémité des vecteurs $\mathbf{e}_i$. Cette hypothèse permet d'éviter des situations très défavorables où deux (ou plusieurs) hypothèses sont très proches et pénalisent trop fortement l'algorithme BTF. Formellement, cette zone correspond à une calotte sphérique définie sur la sphère unité de dimension p. Selon [9], lorsque la dimension p tend vers l'infini, le nombre de calottes sphériques, dont le rayon est inférieur à l'unité, nécessaires pour recouvrir la sphère unité tend vers l'infini, ce qui montre que cette hypothèse n'est pas tellement sévère. Le théorème suivant donne le retard moyen de détection/localisation minimum atteignable avec une stratégie BTF (2). Il précise également les valeurs optimales des paramètres associés m^* et h^*.

Théorème 1 *Considérons le modèle (1). Soit $(\overline{N}, \overline{\nu})$ l'algorithme BTF de détection/localisation (2). Alors, le retard minimal de détection/localisation au sein de la classe $\mathcal{K}_{(\gamma,b)}$ et les paramètres optimaux h^*, m^* de l'algorithme BTF atteignant ce retard optimal sont donnés par :*

$$\begin{aligned} \bar{\tau}^* &\lesssim \frac{4\ln\gamma}{\omega^2}, \\ h^* &\sim 2\ln\gamma, \\ m^* &\sim \frac{2\ln\gamma}{\omega^2} \end{aligned} \tag{3}$$

$$\textit{quand} \quad \min\left\{\bar{\delta}_{\rm i}^2; \frac{\bar{\delta}_{\rm i}}{2}\right\} \ln\gamma \gtrsim^+ \ln b^{-1} \ \textit{ et } \ b^{-1} \to +\infty \tag{4}$$

où $x \gtrsim^+ y$ est équivalent à $x \geq y(1+|o(1)|)$ quand $y \to +\infty$.

La comparaison du retard moyen de détection/localisation du Théorème 1 avec les bornes théoriques optimales données dans [3] montre que la stratégie BTF est sous-optimale dans le cas d'hypothèses Gaussiennes non-orthogonales mais que la perte d'optimalité est clairement bornée.

4 Résultats numériques

Cette section compare les formules asymptotiques pour le retard moyen de détection/localisation et les résultats numériques d'une optimisation numérique non-asymptotique de l'algorithme BTF. En effet, le choix optimal des paramètres $m = m(\gamma, b)$ et $h = h(\gamma, b)$ de l'algorithme BTF peut se réduire au problème d'optimisation suivant :

$$(\hat{m}, \hat{h}) \quad = \quad \arg\min_{m,h} \ \bar{\tau}(m,h) \qquad (5)$$
$$\text{sous contrainte} \qquad \overline{T}(m,h) = \gamma.$$

Une minimisation simultanée du délai moyen de détection sous les deux contraintes γ et b est impossible. Pour cette raison, dans le problème d'optimisation décrit par l'équation (5), la contrainte active est γ. La probabilité de fausse localisation $\overline{P} = \overline{P}(\hat{m}, \hat{h})$ est calculée comme une fonction de $(\hat{m}, \hat{h})$ obtenus à partir de (5).

Comparons l'équation asymptotique du délai moyen de détection/localisation avec les résultats de l'optimisation numérique non-asymptotique (5) de l'algorithme BTF et les résultats numériques obtenus dans [6]. La comparaison est présentée sur la figure 1 (gauche), resp. (droite), pour les paramètres suivants : $\omega^2 = c^2 = 1$, $\sigma = 1$, $p = 25$ et $K = 1$, resp. $K = 10$. Dans le premier cas, $K = 1$, l'hypothèse alternative est donnée par $\boldsymbol{\theta}_1^T = (c, 0, \ldots, 0)$ et, dans le second cas, $K = 10$, les hypothèses alternatives $\boldsymbol{\theta}_\ell^T = (0, \ldots, 0, c, 0, \ldots, 0)$ sont orthogonales puisque le seul élément non nul est le ℓ-ème, $1 \leq \ell \leq 10$. Le délai moyen maximum $\bar{\tau}(\hat{m}(\gamma), \hat{h}(\gamma))$ obtenu par l'optimisation numérique (5), comme une fonction de γ, est appelé le délai BTF moyen maximum "exact". Le pire délai BTF moyen "exact" (avec $\operatorname{ess\,sup}$) est obtenu dans [6] au moyen d'une optimisation numérique. Cette courbe est seulement valable quand les hypothèses alternatives sont orthogonales (voir Fig. 1). La borne asymptotique inférieure présentée dans [3] et le délai BTF moyen maximum asymptotique $\bar{\tau}^*$ de l'équation (3), comme fonctions de γ, sont également présentés sur la figure 1. Dans le cas de $K = 10$ hypothèses alternatives, une borne conservative supérieure $\widetilde{\beta}$ pour la probabilité maximale de fausse localisation est également tracée sur la figure 1 (droite). Cette borne n'est pas détaillée dans ce papier pour éviter une surcharge technique. Elle permet cependant de mettre en évidence l'évolution des performances en fausse localisation des méthodes employées vis-à-vis du temps moyen entre fausse alarme.

Cette figure confirme plusieurs choses : les courbes BTF "exactes" sont proches de la courbe asymptotique et les résultats obtenus sont pertinents par rapport aux résultats précédemment publiés dans [6]. Naturellement, le pire délai BTF moyen "exact" obtenu dans [6] correspond à une critère plus pessimiste (avec $\operatorname{ess\,sup}$) et, pour cette raison, la courbe du délai BTF moyen maximum (5) se situe à gauche de la courbe du pire délai BTF moyen "exact" de [6].

Considérons le cas des hypothèses alternatives non-orthogonales. Le rapport signal-sur-bruit vaut $\omega^2 = 1$, $\sigma = 1$, $p = 25$ et $K = 10$. Les vecteurs $\boldsymbol{\theta}_\ell$ sont définis comme précédemment sauf que $\boldsymbol{\theta}_2^T = (0.4472, 0.8944, 0, \ldots, 0)$ et $\boldsymbol{\theta}_4^T = (0, 0, 0.2873, 0.9568, 0, \ldots, 0)$ ce qui conduit à $\bar{\delta}_{\mathrm{i}} = 0.5528$. Les résultats sont présentés dans la figure 2. Cette figure montre que 1) la courbe BTF "exacte" est proche de la courbe asymptotique et 2) la borne conservative pour la pire probabilité de fausse localisation $\widetilde{\beta}$ reste relativement importante (même pour de larges valeurs de γ) à cause de la non-orthogonalité des hypothèses alternatives.

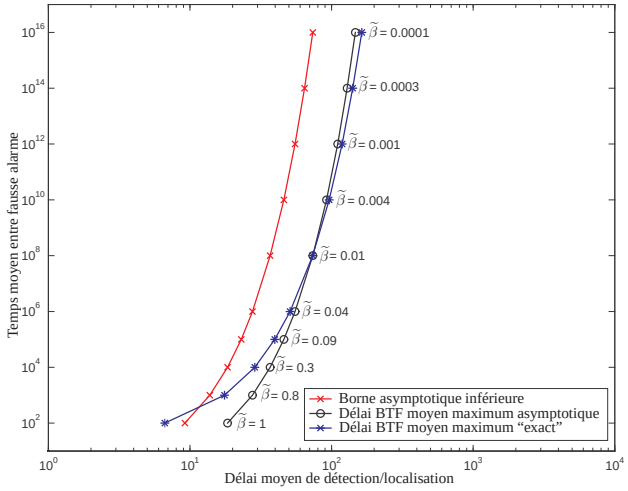

FIG. 2 – Comparaison entre les résultats d'une optimisation numérique de l'algorithme BTF et la borne asymptotique dans le cas de $K = 10$ hypotheses non-orthogonales : borne asymptotique inférieure (croix) ; délai BTF moyen maximum asymptotique (cercles) ; délai BTF moyen maximum "exact" (étoiles).

5 Conclusion

Ce papier propose un algorithme BTF pour la détection/localisation d'hypothèses Gaussiennes non-orthogonales lorsque l'instant de changement est multiple de la taille du bloc de données. Les performances de cet algorithme dépendent de la géométrie (en terme de distance de Kullback-Leibler) entre les hypothèses. Cet algorithme est sous-optimal mais la perte d'optimalité est asymptotiquement bornée.

Références

[1] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes : theory and application*. Prentice Hall, 1993.

[2] T. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 595–608, 2000.

[3] I. Nikiforov, "A lower bound for the detection/isolation delay in a class of sequential tests," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 3037–3046, 2003.

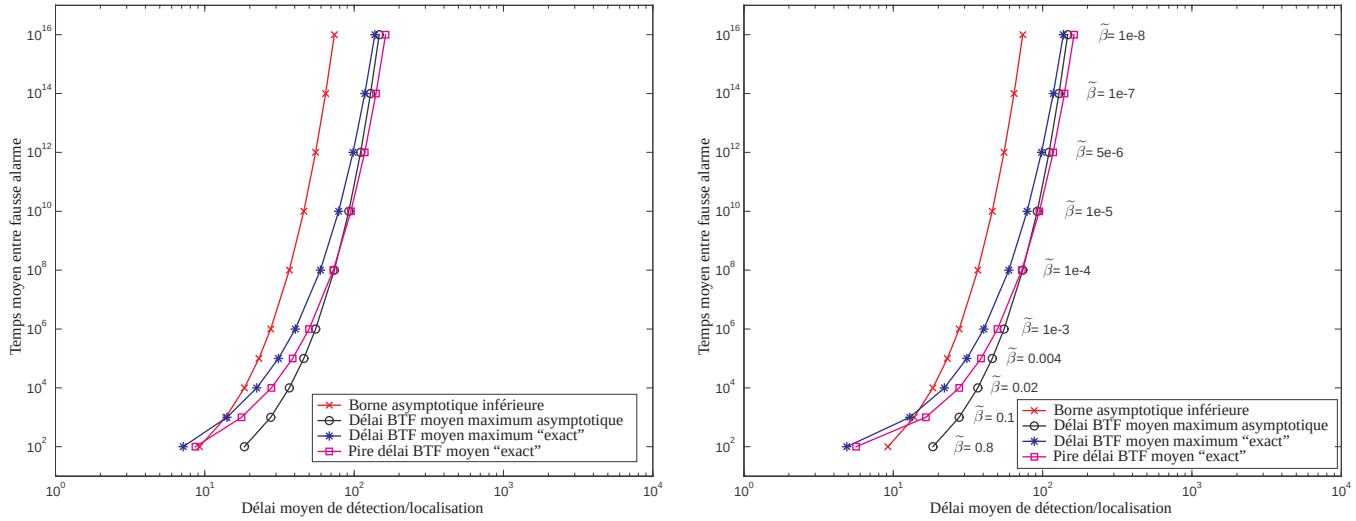

FIG. 1 – Comparaison entre les résultats d’une optimisation numérique de l’algorithme BTF et la borne asymptotique dans le cas $K = 1$ (gauche) et $K = 10$ (droite) : borne asymptotique inférieure (croix) ; délai BTF moyen maximum asymptotique (cercles) ; délai BTF moyen maximum “exact” (étoiles) ; pire délai BTF moyen “exact” (carrés).

[4] A. G. Tartakovsky, “Multidecision quickest change-point detection : Previous achievements and open problems,” *Sequential Analysis*, vol. 27, no. 2, pp. 201–231, 2008.

[5] L. Fillatre, I. Nikiforov, and S. Vaton, “Détection-localisation séquentielle d’anomalies volumiques dans un réseau de télécommunications,” in *GRETSI’07, Troyes, France, CD ROM*, septembre 2007.

[6] I. Nikiforov, “Two strategies in the problem of change detection and isolation,” *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 770–776, 1997.

[7] A. G. Tartakovsky, “Efficiency of the generalized Neyman-Pearson test for detecting changes in a multichannel system,” *Probl. Inf. Transm*, vol. 28, no. 4, pp. 341–350, 1992.

[8] L. Fillatre and I. Nikiforov, “Non-bayesian detection and detectability of anomalies from a few noisy tomographic projections,” *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 401–413, 2007.

[9] E. Weisstein, *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall, 1999.