

Optimisation de critères de contraste et des performances en détection

Cédric RICHARD, Régis LENGELLÉ, Fahed ABDALLAH

Laboratoire de Modélisation et Sécurité des Systèmes, Université de Technologie de Troyes
12 rue Marie Curie, BP 2060
10010 Troyes Cedex, France
cedric.richard@utt.fr, regis.lengelle@utt.fr
fahed.abdallah@utt.fr

Résumé –

Les approches classiques de la théorie statistique de la décision conduisent au résultat fondamental que le test optimum consiste à comparer le rapport de vraisemblance à un seuil. Dans de nombreuses applications, l'implémentation d'un tel test peut s'avérer impossible, une spécification incomplète des lois de vraisemblance constituant une raison couramment rencontrée. En conséquence, on est parfois amené à reconsidérer la procédure d'élaboration d'un détecteur, en particulier en sélectionnant un critère de performance dont l'optimisation est aisée. Dans ce contexte, les critères de contraste constituent une solution souvent adoptée. L'objectif de cet article est de présenter une condition nécessaire et suffisante permettant d'identifier les critères de contraste garantissant une solution optimale au sens des théories statistiques classiques de la décision. Ce résultat est illustré par l'analyse de la pertinence de critères de contraste classiques tels que le rapport-signal-sur-bruit généralisé.

Abstract –

Statistical detection theory leads to the fundamental result that the optimum test consists of comparing any strictly monotone function of the likelihood ratio with a threshold value. In many applications, implementing such a test may be an intractable problem due to excessive time and storage requirements, or may be impossible because of incomplete specification of the conditional probability densities. Therefore, we are often led to consider a simpler procedure for designing detectors. In particular, we can use alternative design criteria such as second-order measures of quality. In this paper, a necessary and sufficient condition is given for such criteria to guarantee the best solution in the the sense of classical detection theories. This result is illustrated by discussing the relevance of well-known criteria.

1 Introduction

Lorsque l'on se place dans le cadre de la décision à hypothèses binaires, la fonction d'un détecteur est d'associer une observation à l'une des deux hypothèses en compétition. Celles-ci sont traditionnellement notées H_0 et H_1 . De façon plus formelle, soit $x \in \mathbb{R}^n$ une observation et soit $y \in \{0, 1\}$ l'indice permettant de désigner l'hypothèse dont x constitue une réalisation. La détection est accomplie au moyen d'une fonction discriminante $h(x) : \mathbb{R}^n \rightarrow \{0, 1\}$, et l'on dit qu'une erreur est commise lors de l'affectation de x si $h(x) \neq y$. On considère ici que le problème de détection peut être posé en termes probabilistes. Soit (X, Y) une paire de variables aléatoires de $\mathbb{R}^n \times \{0, 1\}$. Les observations sont supposées avoir été générées conformément aux densités de probabilités connues $p(X|Y=0)$ et $p(X|Y=1)$. Dans ce contexte, les théories de Bayes et Neyman-Pearson constituent des approches statistiques classiques pour l'élaboration de détecteurs optimaux, au sens du coût moyen minimum dans le premier cas, et de la probabilité de bonne détection maximum pour une probabilité de fausse alarme bornée supérieurement dans le

second cas [1]. On montre que ces théories conduisent au résultat fondamental que le détecteur optimal consiste à comparer le *rapport de vraisemblance* (RV), noté $L(X)$, à un seuil ν :

$$h_{\text{RV}}(X) = \begin{cases} 1 & \text{si } L(X) \triangleq \frac{p_1(X)}{p_0(X)} > \nu \\ 0 & \text{sinon.} \end{cases} \quad (1)$$

Le test (1) est appelé *test du rapport de vraisemblance* (TRV). Il est à noter que la comparaison de toute fonction strictement monotone de $L(X)$ à un seuil constitue un test équivalent au TRV dans le sens où ils ont même courbe COR. Étant donné un problème de détection, on peut donc définir une classe d'équivalence \mathcal{C}_{RV} de détecteurs optimaux, dont la particularité est de reposer sur une fonction strictement monotone du RV [2].

Dans de nombreuses applications, l'implémentation du test (1) s'avère impossible, une spécification incomplète de $p_0(X)$ et $p_1(X)$ étant un motif couramment rencontré. On est alors amené à reconsidérer la procédure d'optimisation du détecteur à seuil défini par (2), en particulier en définissant des critères différents de ceux utilisés dans le cadre

des théories statistiques classiques précédemment citées.

$$h(X) = \begin{cases} 1 & \text{si } S(X) > \nu \\ 0 & \text{sinon.} \end{cases} \quad (2)$$

Les *critères de contraste* sont souvent utilisés pour quantifier le caractère discriminant d'une statistique $S(X)$. En effet, ils ne nécessitent que la connaissance des moments conditionnels du premier et du second ordre de $S(X)$, respectivement définis par

$$\begin{aligned} m_i &\triangleq \mathbb{E}_i\{S\} \\ \sigma_i^2 &\triangleq \text{Var}_i\{S\} \end{aligned} \quad (3)$$

où $i \in \{0, 1\}$. Il convient de noter que la littérature sur les critères de contraste du type *rapport signal-sur-bruit généralisé* (RSBG) est très abondante [2, 3, 4, 5], et des résultats théoriques permettent de justifier leur mise en oeuvre [2, 6, 7, 8, 9]. Dans [10, pp. 141-3], l'objectif de l'auteur est d'unifier ces résultats en énonçant que l'utilisation d'une fonction quelconque $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$ comme critère de contraste garantit une solution optimale de la classe \mathcal{C}_{RV} . Dans cet article, nous démontrons que les critères de contraste vérifiant cette propriété doivent en fait vérifier une condition nécessaire et suffisante non-triviale, que nous exhibons.

Cet article est organisé ainsi. Dans un premier temps, nous exprimons les statistiques de détection obtenues par optimisation des critères de cette nature, en fonction du RV. La nécessité que cette fonction soit strictement monotone nous conduit alors à l'identification de l'ensemble des critères de contraste $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$ recevables. Enfin, ces éléments sont appliqués à une famille de critères de contraste particuliers englobant le RSBG afin d'en étudier la pertinence.

2 Optimisation de critères de contraste

Soit $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$ un critère de contraste pour lequel $m_i = \mathbb{E}_i\{S\}$ et $\sigma_i^2 = \text{Var}_i\{S\}$. Afin d'identifier les statistiques $S(X)$ optimum au sens de $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$, on procède à un calcul des variations de ce dernier. On obtient ainsi :

$$\delta\Psi = \frac{\partial\Psi}{\partial m_0} \delta m_0 + \frac{\partial\Psi}{\partial m_1} \delta m_1 + \frac{\partial\Psi}{\partial \sigma_0^2} \delta \sigma_0^2 + \frac{\partial\Psi}{\partial \sigma_1^2} \delta \sigma_1^2. \quad (4)$$

En utilisant le fait que $\delta m_i = \int \delta S(X) p_i(X) dX$ et que $\delta \sigma_i^2 = 2 \int (S(X) - m_i) \delta S(X) p_i(X) dX$ avec $i \in \{0, 1\}$, il en résulte :

$$\begin{aligned} \delta\Psi = \int \left[\frac{\partial\Psi}{\partial m_0} p_0(X) + \frac{\partial\Psi}{\partial m_1} p_1(X) \right. \\ \left. + 2(S(X) - m_0) \frac{\partial\Psi}{\partial \sigma_0^2} p_0(X) \right. \\ \left. + 2(S(X) - m_1) \frac{\partial\Psi}{\partial \sigma_1^2} p_1(X) \right] \delta S(X) dX. \end{aligned} \quad (5)$$

Pour que l'on ait $\delta\Psi = 0$ indépendamment de δS , le terme [...] doit être nul. On obtient ainsi l'expression de la statistique $S(X)$ maximisant le critère $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$, qu'il

est possible de réécrire en fonction du RV :

$$\begin{aligned} S(X) = -\frac{1}{2} \frac{\frac{\partial\Psi}{\partial m_0} + \frac{\partial\Psi}{\partial m_1} L(X)}{\frac{\partial\Psi}{\partial \sigma_0^2} + \frac{\partial\Psi}{\partial \sigma_1^2} L(X)} \\ + \frac{m_0 \frac{\partial\Psi}{\partial \sigma_0^2} + m_1 \frac{\partial\Psi}{\partial \sigma_1^2} L(X)}{\frac{\partial\Psi}{\partial \sigma_0^2} + \frac{\partial\Psi}{\partial \sigma_1^2} L(X)}. \end{aligned} \quad (6)$$

La statistique précédente est équivalente au rapport de vraisemblance si et seulement si il s'agit d'une fonction strictement monotone de $L(X)$. L'évaluation de la dérivée première de $S(X)$ par rapport à $L(X)$ conduit au résultat suivant :

$$\begin{aligned} \frac{dS}{dL}(X) \\ = \frac{(m_1 - m_0) \frac{\partial\Psi}{\partial \sigma_0^2} \frac{\partial\Psi}{\partial \sigma_1^2} + \frac{1}{2} \left(\frac{\partial\Psi}{\partial \sigma_1^2} \frac{\partial\Psi}{\partial m_0} - \frac{\partial\Psi}{\partial \sigma_0^2} \frac{\partial\Psi}{\partial m_1} \right)}{\left(\frac{\partial\Psi}{\partial \sigma_0^2} + \frac{\partial\Psi}{\partial \sigma_1^2} L(X) \right)^2}. \end{aligned} \quad (7)$$

On constate donc que $S(X)$ défini par (6) est une fonction strictement monotone de $L(X)$ si et seulement si le numérateur de (7) est différent de 0. Ce résultat conduit directement à la proposition suivante.

Proposition 1. *L'optimisation d'un critère de contraste $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$ conduit à un test équivalent au RV si et seulement si il vérifie la condition suivante*

$$\begin{aligned} (m_1 - m_0) \frac{\partial\Psi}{\partial \sigma_0^2} \frac{\partial\Psi}{\partial \sigma_1^2} \\ + \frac{1}{2} \left(\frac{\partial\Psi}{\partial \sigma_1^2} \frac{\partial\Psi}{\partial m_0} - \frac{\partial\Psi}{\partial \sigma_0^2} \frac{\partial\Psi}{\partial m_1} \right) \neq 0. \end{aligned} \quad (8)$$

Il convient de constater que cette équation aux dérivées partielles ne peut être résolue aisément. En revanche, celle-ci peut être exploitée afin d'étudier la pertinence de certains critères de contraste couramment utilisés. Ceci fait l'objet de la section suivante.

3 Analyse de critères de contraste particuliers

3.1 Critères de contraste de la forme $\Psi((m_1 - m_0)^2, \alpha\sigma_1^2 + (1 - \alpha)\sigma_0^2)$

On considère les critères de contraste $\Psi_\alpha(u, v)$ tels que $u = (m_1 - m_0)^2$ et $v = \alpha\sigma_1^2 + (1 - \alpha)\sigma_0^2$, où α désigne un réel de l'intervalle $[0, 1]$. Afin de discuter de la pertinence des critères de cette nature, on fait appel à la Proposition 1. Celle-ci nécessite le calcul des dérivées partielles de Ψ par rapport à m_i et σ_i^2 :

$$\frac{\partial\Psi}{\partial m_0} = -2\sqrt{u} \frac{\partial\Psi}{\partial u} \quad \frac{\partial\Psi}{\partial m_1} = 2\sqrt{u} \frac{\partial\Psi}{\partial u} \quad (9)$$

$$\frac{\partial\Psi}{\partial \sigma_0^2} = (1 - \alpha) \frac{\partial\Psi}{\partial v} \quad \frac{\partial\Psi}{\partial \sigma_1^2} = \alpha \frac{\partial\Psi}{\partial v}. \quad (10)$$

En substituant (9) et (10) dans (8), on aboutit directement à la conclusion que l'optimisation des critères considérés conduit à un test de la classe \mathcal{C}_{RV} si et seulement si :

$$\sqrt{u} \frac{\partial\Psi}{\partial v} \left[\alpha(1 - \alpha) \frac{\partial\Psi}{\partial v} - \frac{\partial\Psi}{\partial u} \right] \neq 0. \quad (11)$$

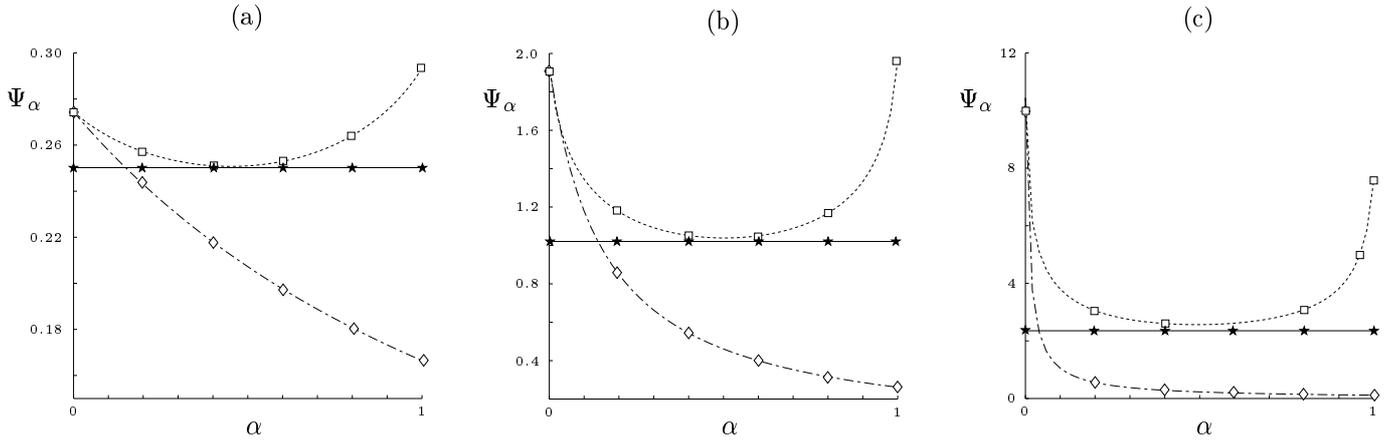


Figure 1: Détection d'un signal déterministe connu noyé dans un bruit blanc gaussien. Les figures ci-dessous présentent le RSBG de la statistique maximisant ce critère de contraste (□), du RV (◇) et du filtre adapté (★), en fonction de α . Le rapport signal-sur-bruit est égal à -0.6 dB en (a), 0 dB en (b), et 0.35 dB en (c).

L'existence de la statistique (6) indique que $\partial\Psi/\partial\sigma_0^2$ et $\partial\Psi/\partial\sigma_1^2$ ne peuvent être simultanément égaux à 0. En utilisant (10), ceci signifie que $\partial\Psi/\partial v \neq 0$. Dans ces conditions, (11) devient

$$\alpha(1-\alpha)\frac{\partial\Psi}{\partial v} \neq \frac{\partial\Psi}{\partial u}. \quad (12)$$

Pour tout $\alpha \in]0, 1[$, l'équation (12) peut être résolue en posant $u = u' + v'$ et $v = \alpha(1-\alpha)(u' - v')$. Ceci conduit à $\Psi(u, v) \neq \psi(u + v/\alpha(1-\alpha))$, où ψ est une fonction à valeurs dans \mathbb{R} . La résolution de l'équation (12) dans le cas où $\alpha \in \{0, 1\}$ conduit à $\partial\Psi/\partial u \neq 0$, ce qui signifie que Ψ doit dépendre explicitement de $(m_1 - m_0)^2$.

3.2 Rapport signal-sur-bruit généralisé

En conservant les notations introduites précédemment, le RSBG est défini par $\Psi_\alpha(u, v) = u/v$. Ce critère de contraste est fréquemment utilisé et la littérature faisant mention de ses propriétés est abondante [2, 3, 4, 5]. En particulier, on peut noter que $\Psi_\alpha(\lambda S + \mu) = \Psi_\alpha(S)$ pour tout $\lambda \in \mathbb{R}^*$, $\mu \in \mathbb{R}$. Ainsi, étant donné un problème de détection, on peut définir une classe d'équivalence de détecteurs dont la statistique maximise le RSBG. On note celle-ci $\mathcal{C}_{\text{RSBG}}$.

En considérant la condition d'admissibilité (12), il est aisé de constater que la maximisation du RSBG conduit à un test équivalent au TRV, c'est à dire $\mathcal{C}_{\text{RSBG}} \subset \mathcal{C}_{\text{RV}}$. Ce résultat peut être complété en donnant l'expression de toute statistique conjointement optimale au sens du RSBG et des théories statistiques classiques de la décision, résultat que l'on obtient à partir de la relation (6) après avoir évalué $\partial\Psi_\alpha/\partial m_i$ et $\partial\Psi_\alpha/\partial\sigma_i^2$:

$$S(X) = \lambda \frac{L(X) - 1}{\alpha(L(X) - 1) + 1} + \mu, \quad \lambda \in \mathbb{R}^*, \quad \mu \in \mathbb{R}. \quad (13)$$

Afin d'illustrer ce dernier résultat, on considère le scénario suivant

$$\begin{cases} H_0 : x = w \\ H_1 : x = s + w, \end{cases} \quad (14)$$

où s est un signal déterministe connu défini sur l'intervalle $\{1, \dots, N\}$, et w désigne un bruit blanc, gaussien, de

moyenne nulle et de covariance R . La décision optimale entre H_0 et H_1 est prise en comparant le RV donné ci-dessous à un seuil ν .

$$L(X) = \exp(-d^2/2) \exp(T(X)). \quad (15)$$

Dans (15), $T(X) = X^T R^{-1} s$ est appelé *filtre adapté* (FA) et $d^2 = s^T R^{-1} s$ représente le rapport-signal-sur-bruit de sortie du FA [6]. Afin de comparer le RSBG de $T(X)$ et $L(X)$ avec le RSBG de la statistique $S_\alpha(X)$ maximisant Ψ_α , nous allons à présent déterminer $\Psi_\alpha(T)$ et $\Psi_\alpha(L)$, pour tout $\alpha \in [0, 1]$. On démontre aisément que $E_0\{T\} = 0$, $E_1\{T\} = d^2$ et $\text{Var}_i\{T\} = d^2$. Ainsi, on a :

$$\Psi_\alpha(T) = d^2 \quad (16)$$

pour tout $\alpha \in [0, 1]$. Le RV est distribué selon une loi log-normale, et l'on a $E_0\{L\} = e^{d^2/2}$, $E_1\{L\} = e^{3d^2/2}$, $\text{Var}_0\{L\} = e^{d^2}(e^{d^2} - 1)$ et $\text{Var}_1\{L\} = e^{3d^2}(e^{d^2} - 1)$. En conséquence

$$\Psi_\alpha(L) = \frac{e^{d^2} - 1}{\alpha e^{2d^2} + (1-\alpha)} \quad (17)$$

pour tout $\alpha \in [0, 1]$. Ces résultats peuvent être comparés au RSBG de la statistique suivante

$$S_\alpha^*(X) = \frac{L(X) - 1}{\alpha(L(X) - 1) + 1} \quad (18)$$

qui maximise Ψ_α conformément à la relation (13). Dans la mesure où $E_i\{S_\alpha^*\}$ et $\text{Var}_i\{S_\alpha^*\}$ ne peuvent pas être déterminés de façon analytique dans le cadre du scénario (14), $\Psi_\alpha(S_\alpha^*)$ a été estimé grâce à des simulations. Le paramètre N a été fixé à 1 et 10.000 réalisations des hypothèses H_0 et H_1 ont été utilisées. Le RSBG de S_α^* , L et T sont représentés en fonction de α sur la Fig. 1, pour diverses valeurs de d^2 . Il apparaît que $\Psi_\alpha(S_\alpha^*)$ est supérieur à $\Psi_\alpha(L)$ et $\Psi_\alpha(T)$, ce qui est conforme à la théorie. On note également que $\Psi_0(S_0^*) = \Psi_0(L) = \exp(d^2) - 1$ car $S_0^*(X) = L(X) - 1$. Bien évidemment, les trois statistiques étudiées conduisent à des structures de décision de la classe d'équivalence \mathcal{C}_{RV} .

4 Conclusion et perspectives

Ces travaux caractérisent les critères de contraste dont l'optimisation conduit à une statistique optimale au sens

des théories statistiques classiques. L'efficacité de cette approche est cependant conditionnée par le fait qu'il existe, dans la classe de statistiques où s'effectue la recherche d'un optimum, au moins une fonction strictement monotone du RV. En conséquence, le choix de cette classe est en pratique crucial, et constitue d'ailleurs une question encore ouverte à l'heure actuelle si l'on évite les considérations asymptotiques [11]. Les résultats présentés dans cet article suggèrent toutefois que si les performances des détecteurs obtenus par optimisation de divers critères de contraste pertinents sont identiques, alors il est probable que la classe de statistiques considérée contienne au moins une fonction strictement monotone du RV pour le problème traité. Cette question mérite des développements futurs, en particulier en considérant les détecteurs de type *linéaire généralisé*, dont l'optimisation au sens de tout critère de contraste $\Psi(m_0, m_1, \sigma_0^2, \sigma_1^2)$ est particulièrement aisée [5, 12, 13]. Enfin, connaissant le RV pour un problème de détection donné, nous avons montré qu'il est possible de construire un test optimal, à la fois au sens d'un critère de contraste donné et des théories statistiques classiques, comme l'illustre la Fig. (1). Il serait à présent intéressant d'étudier l'influence de cette double propriété sur la robustesse du détecteur ainsi constitué.

References

- [1] H. L. Van Trees. *Detection, estimation, and modulation theory*, vol. I, John Wiley & Sons, 1968.
- [2] B. Picinbono et P. Duvaut. Detection and contrast. *Stochastic processes in underwater acoustics* (C. Baker, ed.), Springer-Verlag, 1986.
- [3] P. Duvaut. *Contraste et détection. Application à la quantification et aux filtres de Volterra optimaux pour la détection*. Thèse de doctorat de l'Université Paris 11, Centre d'Orsay, 1987.
- [4] W. A. Gardner. A unifying view of second-order measures of quality for signal classification. *IEEE Transactions on Communications*, vol. 28, no. 6, pp. 807–816, 1980.
- [5] C. Richard. *Une méthodologie pour la détection à structure imposée*. Thèse de doctorat de l'Université de Technologie de Compiègne, 1998.
- [6] B. Picinbono et P. Duvaut. Nouvelle approche pour la détection par seuil. *Proc. neuvième colloque GRETSI*, pp. 87–91, 1983.
- [7] H. V. Poor. Robust decision design using a distance criterion. *IEEE Transactions on Information Theory*, vol. 26, pp. 575–587, 1980.
- [8] P. Rudnick. Likelihood detection of small signals in stationary noise. *Journal of Applied Physics*, vol. 32, pp. 140–143, 1961.
- [9] P. Rudnick. A signal to noise property of binary decision. *Nature*, vol. 193, pp. 604–605, 1962.
- [10] K. Fukunaga. *Statistical Pattern Recognition*. Academic Press, 1990.
- [11] L. Devroye, L. Györfi et G. Lugosi. *A probabilistic theory of pattern recognition*, Springer-Verlag, 1996.
- [12] C. Richard et R. Lengellé. Data-driven design and complexity control of time-frequency detectors. *Signal Processing*, vol. 77, pp. 37–48, 1999.
- [13] C. Richard et R. Lengellé. Nouvelle approche pour la détection dans le plan temps-fréquence. *Proc. seizième colloque GRETSI*, pp. 659–662, 1997.