

Formulation d'indicateurs optimaux pour la détection de défauts

Jérôme ANTONI¹ Kayacan KESTEL² Cédric PEETERS² Jan HELSEN²

¹Univ Lyon, INSA-Lyon, Laboratoire Vibrations, F-69621 Villeurbanne, France

²Vrije Universiteit Brussel, Department of Mechanical Engineering, Pleinlaan 2, Elsene, Belgium

Résumé – Cet article présente une méthodologie de construction d'indicateurs optimaux pour la détection précoce de défauts dans le contexte de la surveillance vibratoire.

Abstract – The paper introduces a methodology for the design of optimal health indicators for the early detection of faults in vibration-based monitoring.

1 Introduction

La notion d'indicateur de santé est fondamentale dans le contexte de la maintenance prédictive des systèmes et des processus. Les indicateurs de santé sont des quantités scalaires qui résument l'information contenue dans des mesures, afin de permettre un suivi automatique d'un état de fonctionnement, de détecter une anomalie, de suivre une évolution, et éventuellement de déclencher une alarme [6]. Dans le contexte de la surveillance vibratoire des machines, les principaux indicateurs de santé utilisés sont par exemple la valeur efficace, la valeur crête (valeur maximale sur écart-type), le kurtosis, le coefficient d'asymétrie (skewness), le facteur de forme (valeur absolue moyenne sur écart-type), le facteur pic, etc [8]. Un des avantages des indicateurs cités est d'une part leur simplicité de calcul et, d'autre part, leur sensibilité à des changements de la densité de probabilité marginale des signaux, c'est-à-dire à une modification potentielle de l'état du système ou du processus analysé. Par ailleurs, ces indicateurs sont souvent utilisés comme paramètres pour alimenter des systèmes d'apprentissage statistique. Plusieurs travaux ont récemment remis en question ces pratiques empiriques [2]. Une première tentative consiste à définir de nouveaux indicateurs, plus "performants" selon des critères à définir. Ces approches se heurtent à la notion relative de performance, vouant ainsi à l'échec la recherche d'indicateurs universels qui remplaceraient tous les autres. Ceci est illustré par la publication de nombreux "round-robin tests" qui aboutissent à des conclusions différentes, évidemment dépendantes des données analysées. Une autre approche consiste à apprendre les indicateurs sur un jeu de données à partir de méthodes d'apprentissage machine, mais avec le risque de perdre l'intelligibilité des résultats ainsi obtenus. Enfin, une troisième approche cherche à construire des indicateurs optimaux sur la base de modélisation statistique. La contribution de cet article s'inscrit dans cette mouvance.

Nous montrons dans un premier temps comment des indicateurs optimaux pour la détection précoce de défauts peuvent être construits sur la base du couple de densités de probabilité qui caractérise les états sain et défaillant. L'optimalité est en termes de détection étant donné un taux de fausse alarme. Le fait qu'une détection précoce soit visée permet de considérablement simplifier la formulation par rapport aux travaux précurseurs publiés dans [3]. Les indicateurs ainsi formulés peuvent être comparés à un seuil de détection dont nous don-

nons l'expression asymptotique.

Nous montrons ensuite que plusieurs indicateurs classiques de la littérature sont retrouvés comme des cas particuliers de cette formulation dans le cas où l'état sain est décrit par une densité de probabilité gaussienne. Nous montrons enfin comment étendre leurs définitions dans le cas où l'état sain est décrit par une gaussienne généralisée.

L'ensemble de ces résultats est illustré sur des signaux expérimentaux.

2 Principe de construction

Soit $\mathbf{x} = (x_0, \dots, x_{L-1})^T \in \mathbb{R}^L$ une séquence de moyenne nulle de L échantillons, supposés indépendants et identiquement distribués. Soient H_0 et H_1 les états sain et défectueux, décrits par les densités de probabilité $p(\mathbf{x}|\theta_0)$ et $p(\mathbf{x}|\theta_1)$ avec θ_i , $i = 0, 1$ les vecteurs de paramètres associés. La théorie de Neyman-Pearson établit que le détecteur optimal, qui maximise la probabilité de détection étant donné un taux de fausse alarme, est donné par le rapport de vraisemblance $p(\mathbf{x}|H_1)/p(\mathbf{x}|H_0)$ [4]. Comme expliqué dans la référence [3], ceci permet de construire un indicateur de santé à partir de la fonctionnelle $I(\mathbf{x}) : \mathbb{R}^L \rightarrow \mathbb{R}$ définie par

$$I(\mathbf{x}) = \left\langle \ln p_1(x_n; \hat{\theta}_1) - \ln p_0(x_n; \hat{\theta}_0) \right\rangle \quad (1)$$

où $\hat{\theta}_i$ sont les estimateurs du maximum de vraisemblance, $\langle \bullet \rangle = \frac{1}{L} \sum_{n=0}^{L-1} (\bullet)$ est l'opérateur de moyennage, et tel que

$$\begin{cases} \lim_{L \rightarrow \infty} I(\mathbf{x}|H_0) = 0 & (\text{état sain}) \\ \lim_{L \rightarrow \infty} I(\mathbf{x}|H_1) > 0 & (\text{état défectueux}). \end{cases} \quad (2)$$

Les propriétés de cet indicateur ont été étudiées dans [3]. L'objectif de cette communication est d'étudier le cas précoce où l'état H_0 est légèrement perturbé pour rentrer dans l'état H_1 . Posons $\theta_0 = \theta_0$ et $\theta_1 = (\theta_1, \nu)$ où la variable de perturbation ν mesure l'écart entre H_0 et H_1 . Pour cela, considérons la densité de probabilité de l'état défectueux sous la forme

$$p_1(x; \theta_1, \nu) = p_0(x; \theta_1) \phi(x; \theta_1, \nu) \quad (3)$$

où $\phi(x_n; \theta_1, 0)$ est une fonction de modulation positive de ν ,

telle que

$$(a) \quad \phi(x_n; \theta_1, 0) = 1 \quad (4)$$

$$(b) \quad \int_{\mathbb{R}} p_0(x; \theta) \phi(x_n; \theta, \nu) dx = 1 \quad (5)$$

de manière que, par construction, $p_1(x; \theta_1, \nu)$ conserve les propriétés d'une densité de probabilité et $p_1(x; \theta, 0) = p_0(x; \theta)$.

Il s'agit alors de comparer les deux modèles

$$\begin{cases} H_0 : p_0(x; \theta) \\ H_1 : p_1(x; \theta, \nu) = p_0(x; \theta) \phi(x; \theta_1, \nu). \end{cases} \quad (6)$$

Nous avons ainsi les résultats suivants.

Lemma 2.1. *Supposons $p_0(x; \theta)$ et $\phi(x; \theta, \nu)$ deux fois dérivable par rapport à θ and ν . Alors, si (i) la paramétrisation (3) est localement identifiable et (ii) $\left\langle \frac{\partial \ln \phi}{\partial \nu}(x_n; \hat{\theta}_0, 0) \right\rangle \neq 0$, l'estimateur du maximum de vraisemblance de ν sous H_0 est*

$$\hat{\nu} = \frac{\mathcal{I}_0}{|\mathcal{I}_1|} \left\langle \frac{\partial \ln \phi}{\partial \nu}(x_n; \hat{\theta}_0, 0) \right\rangle + \mathcal{O}(\hat{\nu}^2) \quad (7)$$

où $\hat{\theta}_0$ est l'estimateur du maximum de vraisemblance de θ sous H_0 , $\mathcal{I}_0 = -\mathbb{E}_{H_0} \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_0(x; \theta_0) \right\}$ est l'information de Fisher sous H_0 et

$$\mathcal{I}_1 = -\mathbb{E}_{H_0} \begin{pmatrix} \frac{\partial^2 \ln p_0}{\partial \theta^2}(x; \theta_0, 0) & \frac{\partial^2 \ln \phi}{\partial \nu \partial \theta}(x; \theta_0, 0) \\ \frac{\partial^2 \ln \phi}{\partial \nu \partial \theta}(x; \theta_0, 0) & \frac{\partial^2 \ln \phi}{\partial \nu^2}(x; \theta_0, 0) \end{pmatrix} \quad (8)$$

la matrice d'information de Fisher sous H_1 , toute deux évaluées en $\theta_0 = \mathbb{E}_{H_0} \{ \hat{\theta}_0 | H_0 \}$.

Theorem 2.2. *Sous les hypothèses précédentes, l'indicateur défini par (1) s'exprime sous la forme*

$$I(\mathbf{x}) = \frac{\mathcal{I}_0}{2|\mathcal{I}_1|} \left(\left\langle \frac{\partial \ln \phi}{\partial \nu}(x_n; \hat{\theta}_0, 0) \right\rangle \right)^2 + \mathcal{O}(\hat{\nu}^3) \quad (9)$$

et est distribué selon $\chi_1^2/(2L)$.

3 Noyau gaussien généralisé

Un modèle assez général pour décrire la densité de probabilité de l'état sain sous H_0 est la gaussienne généralisée

$$p_0(x; \theta) = \frac{p}{2\theta\Gamma(1/p)} e^{-|x/\theta|^p} \quad (10)$$

caractérisé par l'exposant $p > 0$ et le paramètre d'échelle $\theta > 0$. L'estimateur du maximum de vraisemblance de ce dernier est

$$\hat{\theta}_0 = \sqrt[p]{p \langle |x_n|^p \rangle}. \quad (11)$$

Plusieurs scénarii peuvent ensuite considérés suivant la forme de la densité de probabilité de l'état défectueux sous H_1 .

3.1 Indicateur de type M_β/M_p

Considérons la forme

$$p_1(x; \theta, \nu) = p_0(x; \theta) \left(1 + \nu(|x/\theta|^\beta - p^{-\beta/p} m_{\beta,p}) \right) \quad (12)$$

avec $0 < \beta, 0 \leq \nu \leq 1/\beta$ et

$$m_{\beta,p} = \int_{\mathbb{R}} p_0(x; 1) |x|^\beta dx = \frac{\Gamma((\beta+1)/p)}{\Gamma(1/p)}. \quad (13)$$

Dans ce cas, la modulation $\phi(x; \theta, \nu)$ applique une amplification polynomiale de degré β sur les queues de distribution de $p_0(x; \theta)$. On montre alors que

$$I(\mathbf{x}) = \frac{(\hat{s}_{\beta,p}(\mathbf{x}) - p^{\beta/p} m_{\beta,p})^2}{2p^{2\beta/p} (m_{2\beta,p} - m_{\beta,p}^2(1 + \beta^2/p))} \quad (14)$$

où

$$\hat{s}_{\beta,p}(\mathbf{x}) = \frac{\langle |x_n|^\beta \rangle}{\langle |x_n|^p \rangle^{\beta/p}} = \left(\frac{M_\beta}{M_p} \right)^\beta \quad (15)$$

est le rapport des moyennes d'ordre β et p . Le test de détection s'énonce :

“Rejeter H_0 au risque α si”

$$I(\mathbf{x}) > \frac{\chi_{1,1-\alpha}^2}{2L}. \quad (16)$$

Ce scénario permet de retrouver plusieurs cas particuliers lorsque la densité de probabilité de l'état sain est gaussienne ($p = 2$).

3.1.1 Kurtosis

Avec $\beta = 4$ et $p = 2$, il vient $m_{4,2} = 3$ et $m_{8,2} = 105$ et donc

$$I(\mathbf{x}) = \frac{(\hat{s}_{4,2}(\mathbf{x}) - 3)^2}{48}, \quad 0 \leq \nu \leq \frac{1}{3} \quad (17)$$

qui fait apparaître le kurtosis

$$\hat{s}_{4,2}(\mathbf{x}) = \frac{\langle |x_n|^4 \rangle}{\langle x_n^2 \rangle^2}. \quad (18)$$

Le kurtosis est un indicateur très utilisé en surveillance vibratoire, mais souvent de manière empirique. Il apparaît ici, sans surprise, comme une mesure de distance à la normalité. L'indicateur $I(\mathbf{x})$ qui en résulte détecte de manière indifférenciée des comportements leptokurtique et platikurtique.

3.1.2 Facteur de forme

Avec $\beta = 1$ et $p = 2$, $m_{1,2} = \sqrt{2/\pi}$ et $m_{2,2} = 1$, et

$$I(\mathbf{x}) = \frac{(\hat{s}_{1,2}(\mathbf{x}) - \sqrt{2/\pi})^2}{2(1 - 3/\pi)}, \quad 0 \leq \nu \leq \sqrt{\frac{\pi}{2}} \quad (19)$$

où

$$\hat{s}_{1,2}(\mathbf{x}) = \frac{\langle |x_n| \rangle}{\sqrt{\langle x_n^2 \rangle}} \quad (20)$$

définit le facteur de forme. Tout comme le kurtosis, il s'interprète comme une mesure de distance à la normalité [7].

3.1.3 Facteur pic

La limite $p \rightarrow \infty$ avec $\beta = 2$ donne $m_{2,\infty} = 1/3$, $m_{2,\infty} = 1/5$, $\langle |x_n|^p \rangle^{1/p} \rightarrow \max(\mathbf{x})$, et $p^{2/p} \rightarrow 1$, tel que

$$I(\mathbf{x}) = \frac{45}{8} (\hat{s}_{2,\infty}(\mathbf{x}) - 1/3)^2 \quad (21)$$

où

$$\hat{s}_{2,\infty}^{-2}(\mathbf{x}) = \frac{\max(\mathbf{x})}{\sqrt{\langle |x_n|^2 \rangle}} \quad (22)$$

définit le facteur pic, un autre critère couramment utilisé en pratique. Il est à noter que, dans le formalisme proposé, l'indicateur pic correspond suppose que l'état sain est caractérisé par une densité de probabilité uniforme sur $[-\theta; +\theta]$ (cas limite de la gaussienne généralisée lorsque $p \rightarrow \infty$), hypothèse qui ne semble pas être reconnue en pratique. Plus exactement, il s'agit de comparer les deux modèles

$$\begin{cases} H_0 : p_0(x; \theta) &= \frac{1}{2\theta} \mathbb{1}[|x| \leq \theta], \\ H_1 : p_1(x; \theta, \nu) &= p_0(x; \theta) (1 + \nu(|x/\theta|^2 - 1/3)) \end{cases} \quad (23)$$

avec $0 \leq \nu \leq 3$.

3.2 Negentropie

La gaussienne généralisée permet aussi de modéliser un changement continu de comportement en modifiant la valeur de son exposant. Ceci amène à la comparaison des modèles

$$\begin{cases} H_0 : p_0(x; \theta) &= \frac{p}{2\theta\Gamma(1/p)} e^{-|x/\theta|^p}, \\ H_1 : p_1(x; \theta, \nu) &= \frac{p+\nu}{2\theta\Gamma(\frac{1}{p+\nu})} e^{-|x/\theta|^{p+\nu}}, \quad \nu > -p \end{cases} \quad (24)$$

soit $p_1(x; \theta, \nu) = p_0(x; \theta)\phi(x; \theta, \nu)$ avec

$$\phi(x; \theta, \nu) = \left(1 + \frac{\nu}{p}\right) \frac{\Gamma(1/p)}{\Gamma(\frac{1}{p+\nu})} e^{|x/\theta|^p - |x/\theta|^{p+\nu}}. \quad (25)$$

Il en résulte la définition de l'indicateur

$$I(\mathbf{x}) = \frac{(\hat{H}_p(\mathbf{x}) - \ln p - p - \Psi(1/p))^2}{2((1+p)\Psi_1(1/p) - p(1+p+p^2))} \quad (26)$$

où Ψ et Ψ_1 sont les fonctions digamma et trigamma et où

$$\hat{H}_p(\mathbf{x}) = \left\langle \frac{|x_n|^p}{\langle |x_n|^p \rangle} \ln \frac{|x_n|^p}{\langle |x_n|^p \rangle} \right\rangle \quad (27)$$

s'interprète comme la negentropie de $|x_n|^p$.

Quand $p = 2$, l'indicateur devient

$$I(\mathbf{x}) = \frac{(\hat{H}_2(\mathbf{x}) + \ln 2 + \gamma - 2)^2}{3\pi^2 - 7} \quad (28)$$

avec $\gamma \simeq 0.5772$ la constante d'Euler et

$$\hat{H}_2(\mathbf{x}) = \left\langle \frac{x_n^2}{\langle x_n^2 \rangle} \ln \frac{x_n^2}{\langle x_n^2 \rangle} \right\rangle \quad (29)$$

la negentropie de $|x_n|^2$. Bien que cet indicateur soit connu, il semble qu'il n'ait été proposé que sur des bases empiriques [1]. Le formalisme que nous lui donnons ici est nouveau et justifie son introduction de façon théorique.

3.3 Indicateur de répartition

Le formalisme développé permet également la comparaison des états sain et défectueux sur la base des fonctions de répartition associées. Pour cela, posons

$$p_1(x; \theta, \nu) = p_0(x; \theta) (1 + \nu (\mathbb{1}\{|x/\theta| \geq u\} - 2F_0(u) + 1)) \quad (30)$$

$\nu \leq 1/F_0(u)$, où $F_0(u)$ est la fonction de répartition de $p_0(x; 1)$. La modulation ainsi appliquée à $p_0(x; \theta)$ réalise une amplification des queues de distribution pour $|x/\theta| > u$. Lorsque p_0 est la gaussienne généralisée d'exposant p , l'indicateur qui compare les deux modèles s'exprime

$$I(\mathbf{x}) = \frac{(\langle \mathbb{1}\{|x_n/\hat{\theta}_0| > u\} \rangle - 2(1 - F_0(u)))^2}{4((2F_0(u) - 1)(1 - F_0(u)) - 2u^2 p_0(u; 1)^{2/p})}. \quad (31)$$

L'indicateur s'interprète comme la distance quadratique entre $\mathbb{P}(|x/\theta| > u|H_0) = 2(1 - F_0(u))$ et l'estimation $\hat{\mathbb{P}}(|x/\theta| > u|H_1) = \langle \mathbb{1}\{|x_n/\hat{\theta}_0| > u\} \rangle$.

4 Application

Trois indicateurs parmi ceux introduits précédemment sont illustrés ici sur des signaux vibratoires acquis sur un banc de test. Les données proviennent de la base référencée dans [5] et correspondent à des mesures successives sur un palier à roulements jusqu'à l'apparition d'un défaut et dégradation avancée du système. L'objectif est de réaliser une détection du défaut aussi précoce que possible. La première mesure disponible est prise comme référence pour caractériser l'état sain H_0 et les autres mesures sont systématiquement testées contre H_1 . L'état sain est ici modélisé par une gaussienne généralisée d'exposant $p \simeq 1.65$; bien que la densité de probabilité soit très proche d'une gaussienne, ce réglage s'avère indispensable pour pouvoir placer un seuil de détection correct qui n'engendre pas d'excès de fausses alarmes par rapport au risque fixé *a priori*. Dans le cas présent, d'excellents résultats ont été obtenus avec l'indicateur (14) fondé sur le rapport

$$\hat{s}_{3,p}(\mathbf{x}) = \left(\frac{M_3}{M_{1.65}} \right)^3, \quad (32)$$

l'indicateur (26) fondé sur la negentropie d'ordre 1.65 et l'indicateur (31) fondé sur le quantile $F_0(u) = 0.99$. L'évolution chronologique des trois indicateurs au cours du test est montrée dans la figure 1; une détection franche est clairement visible par dépassement de seuil au risque 0.1% vers les deux tiers du test.

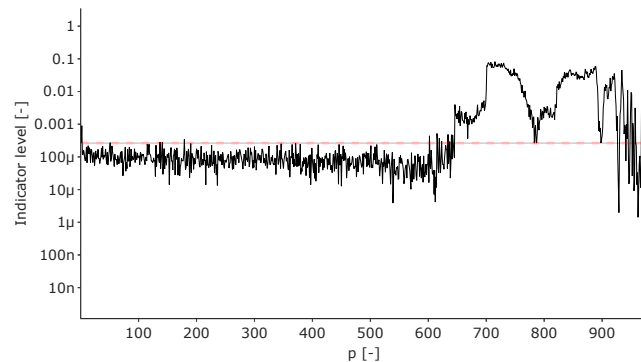
5 Conclusion

L'objectif de cette communication est de proposer une méthodologie de construction d'indicateurs optimaux pour la détection précoce de défauts. Bien que plusieurs indicateurs connus soient ainsi retrouvés comme particuliers, les enseignements de cette approche n'en sont pas moins riches. D'abord, l'optimalité s'entend seulement pour la détection à taux de fausse alarme constant. En particulier, ceci n'implique pas que les mêmes indicateurs de santé soient généralement et également valables pour le suivi d'évolution ou pour le pronostic; il

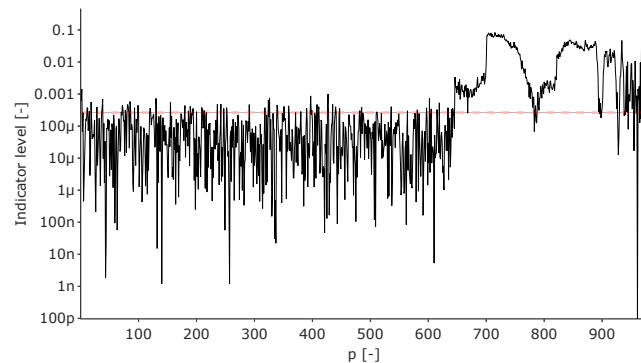
est important du point de vue pratique de faire cette distinction, qui n'est pas toujours reconnue dans la littérature spécialisée. Ensuite, l'optimalité est relative à un modèle probabiliste de l'état sain *et* de l'état défectueux. Là encore, ce constat est rarement énoncé de manière explicite et il convient de rappeler que la plupart des indicateurs de santé de la littérature ne sont optimaux que dans le cas où l'état sain est caractérisé par une gaussienne. Cette communication a montré comment d'autres scénarii peuvent assez facilement être considérés via une gaussienne généralisée. Il en résulte évidemment de nouveaux indicateurs comme le rapport des moyennes d'ordre M_β/M_p et la negentropie H_p d'ordre p .

Références

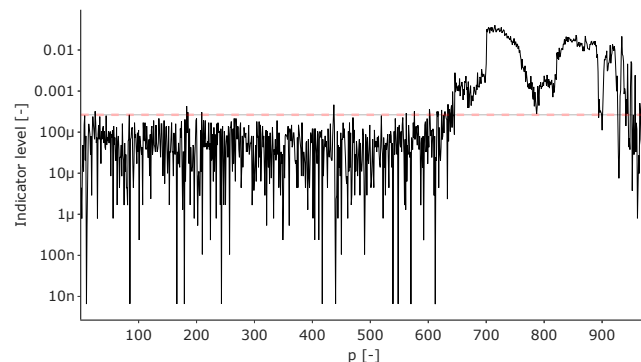
- [1] J. ANTONI : The infogram : Entropic evidence of the signature of repetitive transients. *Mechanical Systems and Signal Processing*, 74:73–94, 2016. Special Issue in Honor of Professor Simon Braun.
- [2] J. ANTONI : A Critical Overview of the “Filterbank-Feature-Decision” Methodology in Machine Condition Monitoring. *Acoustics Australia*, 49(2):177–184, juin 2021.
- [3] J. ANTONI et P. BORGHESANI : A statistical methodology for the design of condition indicators. *Mechanical Systems and Signal Processing*, 114:290–327, 2019.
- [4] M. BASSEVILLE et I. V. NIKIFOROV : *Detection of abrupt changes : theory and application*. Prentice Hall information and system sciences. Prentice Hall, 1993.
- [5] J. LEE, H. QIU, G. YU et J. LIN : Rexnord technical services, bearing data set, nasa ames prognostics data repository (<http://ti.arc.nasa.gov/project/prognostic-data-repository>), nasa ames research center, moffett field, ca. 2007.
- [6] R.B. RANDALL : *Vibration-based Condition Monitoring : Industrial, Aerospace and Automotive Applications*. Wiley, 2011.
- [7] D. WANG : Some further thoughts about spectral kurtosis, spectral l2/l1 norm, spectral smoothness index and spectral gini index for characterizing repetitive transients. *Mechanical Systems and Signal Processing*, 108:360–368, 2018.
- [8] J. ZHU, T. NOSTRAND, C. SPIEGEL et B. MORTON : Survey of condition indicators for condition monitoring systems. *In Annual Conference of the PHM Society*, volume 6, 2014.



(a) Indicateur fondé sur $M_3/M_{1.65}$



(b) Negentropie d'ordre 1.65



(c) Indicateur fondé sur le quantile 0.99

FIGURE 1 : Évolution des indicateurs au cours du test. L'hypothèse H_0 est caractérisée par une gaussienne généralisée d'exposant $p = 1.65$. Le seuil de détection $\chi_1^2/(2L)$ est fixé au risque $\alpha = 0.1\%$