

Filtrage des signaux par EMD et distance de Hausdorff

Ali Komaty, Abdel-Ouahab Boudraa, Kais Khaldi et Délphine Daré

Arts et Métiers ParisTech/Ecole Navale

IRENav-EA3634, BCRM Brest, CC 600, 29240 BREST Cedex 9, France

(ali.komaty,boudra,dare)@ecole-navale.fr, kais.khaldi@gmail.com

Résumé – Dans ce papier nous proposons une nouvelle méthode de filtrage de signaux combinant la décomposition modale empirique (EMD) et la distance de Hausdorff (DH). Le signal bruité est décomposé en une somme finie de modes empiriques par l’EMD suivi par l’estimation de la densité de probabilité (PDF) de chaque mode extrait. Le filtrage consiste à reconstruire partiellement le signal avec un nombre de modes sélectionnés (dits pertinents). Le critère de sélection est une mesure géométrique basée sur la DH entre la PDF du signal et celle de chaque mode empirique. Les résultats du nouveau filtrage sont comparés à ceux de l’EMD-CMSE (Consecutive MSE) [1] et ceux de l’approche ondelettes [2] sur différents signaux et avec plusieurs rapports signal à bruit.

Abstract – In this paper, a new method for signal filtering based on the empirical mode decomposition (EMD) and the Hausdorff distance is introduced. Noisy signal is decomposed into a finite number of empirical modes by EMD followed by the estimation of the probability density function (PDF) of each extracted mode. The filtered signal is obtained by a partial reconstruction using a selected set of empirical modes (termed relevant modes). The selection criterion of modes is a geometric measure based on the HD between the PDF of the noisy signal and that of each mode. Results of the new filtering are compared to that of the EMD-CMSE (Consecutive MSE) [1] and the wavelets approach [2] on different signals and with varying signal to noise ratio values.

1 Introduction

L’un des objectifs de la Décomposition Modale Empirique (ou EMD pour Empirical Mode Decomposition) est l’interprétation physique des modes empiriques ou IMF (Intrinsic mode functions) extraits du signal à analyser [3]. Il est donc important de quantifier la pertinence (statistique, fréquentielle,...) de ces modes. Comme tout processus itératif, l’EMD est sensible aux erreurs numériques (effets de bords des enveloppes,...) et cela peut se traduire par l’extraction de modes artificiels. Par ailleurs, dans le cas de signaux bruités, certains modes sont essentiellement attribués au bruit. Il est donc utile de déterminer lesquels des modes sont pertinents ou pas pour la reconstruction partielle du signal à des fins, par exemple, de filtrage [4],[1]. Wu et Huang ont récemment étudié les caractéristiques statistiques du bruit blanc en utilisant l’EMD [5]. En se basant sur les résultats de simulations numériques, une méthode pour quantifier la signification statistique des informations véhiculées par l’IMF à partir des signaux bruités a été proposée [5]. Par contre aucune stratégie n’a été développée pour identifier les modes artificiels. Peng et al. [6] et Ayenu-Prah et Attoh-Okine [7] ont proposé une méthode pour discriminer entre les modes pertinents et non pertinents en calculant la corrélation entre les IMF et le signal original. Comme les modes sont quasi-orthogonaux, tout mode pertinent doit être bien corrélé au signal original. Pour des signaux très bruités (RSB négatifs), ces deux méthodes sont limitées dû essentiellement à la forte corrélation entre le signal original et les premières modes. Une

autre méthode a été proposée par Boudraa et Cexus qui consiste à minimiser l’écart quadratique moyen entre deux modes consécutifs appelé CMSE (Consecutive MSE) [1]. Le signal est reconstruit à partir du mode pour lequel le CMSE est minimal. Cette approche est efficace en particulier quand le CMSE présente un minimum significatif. L’objet de ce travail, est d’exploiter non pas la corrélation, qui est une mesure linéaire très sensible au bruit [7], mais plutôt une mesure de similarité plus robuste exploitant la ressemblance géométrique, basée sur la distance de Hausdorff, entre les densités de probabilité de chaque IMF et le signal original pour identifier les modes pertinents.

2 Distance de Hausdorff

La distance de Hausdorff (DH) est un opérateur non linéaire qui mesure la similarité entre deux ensembles ou deux formes géométriques. Cette distance est essentiellement utilisée en traitement des images.

Soit $d(a, b) = ||a - b||$ la distance Euclidienne entre deux points a et b . La distance entre un point a et un ensemble fini de points $B = \{b_1, \dots, b_{N_b}\}$ est définie par:

$$d(a, B) = \min_{b \in B} d(a, b) = \min_{b \in B} ||a - b||, \quad (1)$$

La distance entre deux ensembles finis de points $A = \{a_1, \dots, a_n\}$ et $B = \{b_1, \dots, b_{N_b}\}$ est définie par:

$$h(A, B) = \max_{a \in A} d(a, B) = \max_{a \in A} \min_{b \in B} d(a, b) \quad (2)$$

$$= \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (3)$$

De la même manière $h(B, A)$ est donnée par:

$$h(B, A) = \max_{b \in B} d(b, A) = \max_{b \in B} \min_{a \in A} d(b, a) \quad (4)$$

$$= \max_{b \in B} \min_{a \in A} \|b - a\|, \quad (5)$$

Enfin, l'opérateur DH est défini comme suit:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (6)$$

Ainsi, deux ensembles sont proches en terme de DH si chaque point de chaque ensemble est proche de certains points de l'autre ensemble. Nous exploitons dans ce contexte la DH comme une mesure de similarité non pas entre deux formes géométriques mais entre les PDF de deux signaux 1D.

3 Approche proposée

3.1 Pourquoi l'approche géométrique et la DH?

Soit un signal $y(t)$ noyé dans un bruit blanc Gaussien $n(t)$ comme suit

$$x(t) = y(t) + n(t). \quad (7)$$

En environnement très bruité (par exemple $RSB \leq 2$), la PDF de $x(t)$ est largement dominée par celle du bruit $n(t)$. Comme en pratique la PDF de $y(t)$ est inconnue, une mesure de similarité statistique telle que la Divergence de Kullback-Leibler (DKL) ne peut pas s'appliquer. De plus étant une mesure très sensible à la PDF du signal, la DKL entre deux distributions peut être grande même si les deux distributions ont des formes géométriques proches (Fig. 1). La figure 2 montre la superposition des PDF du signal Heavysine bruité et des modes correspondants. On remarque que le mode qui représente le mieux le signal original est le 8^{ème} mode (Fig. 3) car les deux PDF ont des formes géométriques très proches. Par contre au sens de la DKL, le mode 8^{ème} n'est pas pertinent car sa distance au signal original est importante (Fig. 1). Pour y remédier, nous proposons une approche géométrique qui mesure la DH entre les PDF de deux signaux 1D (Fig. 3). Dans ce cas une PDF est vue comme une "forme géométrique". Notre choix s'est porté sur la DH car il s'agit d'une mesure très sensible aux valeurs extrêmes (Fig. 4). En effet, si deux ensembles de points A et B sont semblables (tous les points sont parfaitement superposés) sauf un seul point dans B qui est loin de tout point de A alors la distance est déterminée par ce point et est grande, par suite cette distance va nous permettre de mettre en évidence combien la courbe de la PDF de l'IMF est pointue et étroite, allure bien différente d'une forme Gaussienne. Tout ceci est obtenu sans aucune hypothèse sur la loi de distribution des probabilités.

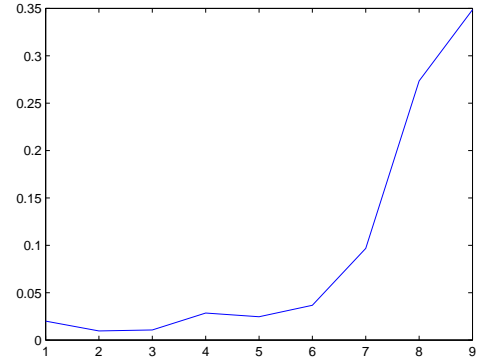


FIG. 1: DKL en fonction de l'ordre des modes pour le signal Heavysine (RSB=2dB).

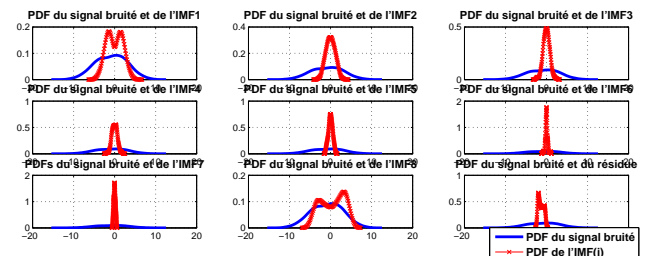


FIG. 2: Superposition des PDF du signal Heavysine (RSB=-10dB) et de ses modes.

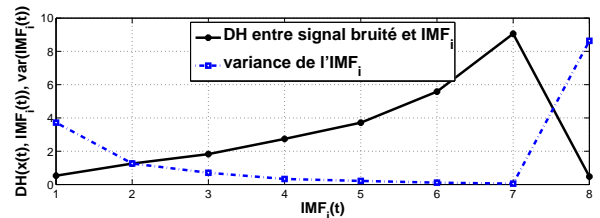


FIG. 3: DH en fonction de l'ordre des modes pour le signal Heavysine (RSB=2dB).

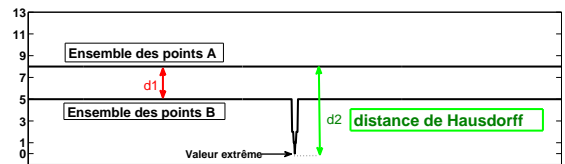


FIG. 4: Exemple illustrant la sensibilité de la DH aux valeurs extrêmes. La distance d_1 (en rouge) est la DH quand $\{B\}$ ne contient pas de valeur extrême. d_2 (en vert) est la DH quand $\{B\}$ contient une valeur extrême. La DH est déterminée par les valeurs extrêmes.

3.2 Sélection des IMF par DH

Le signal $x(t)$ est décomposé en IMF comme suit:

$$x(t) = \sum_{i=1}^N \text{IMF}_i(t) + r_N(t) \quad (8)$$

où N est le nombre de modes (profondeur de la décomposition), $\text{IMF}_i(t)$ est le i^{eme} mode et $r_N(t)$ est le résidu de la décomposition. On estime les PDF de $x(t)$ et de ses modes par une approche basée sur l'estimation par noyau [8]. Les PDF calculées sont considérées comme des sous-ensembles d'un espace métrique (Euclidien) et le degré d'isométrie est mesuré par la DH. Cette mesure est calculée entre la PDF de $x(t)$ et de chaque IMF comme suit:

$$\text{dh}(i) = H[\text{pdf}(x(t)), \text{pdf}(\text{IMF}_i(t))]. \quad (9)$$

Le premier mode à sélectionner est celui où la DH commence à décroître après le premier maximum local. On note par k_{dh} l'indice de ce mode. Finalement, le signal reconstruit (filtré) $\tilde{y}(t)$ s'écrit

$$\tilde{y}(t) = \sum_{i=k_{dh}}^N \text{IMF}_i(t) + r_N(t). \quad (10)$$

La figure 3 montre un exemple de variation de la DH en fonction de l'indice des modes. La courbe croît jusqu'à l'apparition des IMF décrivant $y(t)$, puis elle décroît jusqu'à atteindre son minimum. Dans ce cas $k_{dh} = 8$ et le signal reconstruit sera: $\tilde{y}(t) = \text{IMF}_8(t) + r_N(t)$. Dans certains cas la courbe peut croître de nouveau, mais cette croissance est due au fait que les derniers modes sont de très faibles amplitudes et par conséquent leurs variances sont petites par rapport à celle du signal original (Fig. 3). Toutes les valeurs sont concentrées dans un petit intervalle d'où une courbe de la PDF de plus en plus pointue, ce qui montre l'augmentation de la distance (décroissance de la variance). Donc on peut dire que la variance est inversement proportionnelle à la DH en terme de comparaison entre les PDF.

4 Résultats

Nous testons le filtrage par EMD-DH sur quatre signaux (Doppler, Blocks, Heavysine, Bumps) pour plusieurs valeurs de RSB et comparons les résultats à ceux des ondelettes [7] et du CMSE [3]. Nous présentons les résultats pour des RSB de 2dB et de -5dB. Les ondelettes utilisées sont les symmlets de Daubechies d'ordre 5 suivi d'un seuillage doux [7]. La figure 2 présente le résultat d'estimation des PDF du signal Heavysine et de ses modes (avec un RSB=2dB). Le résultat du calcul de la DH pour ce signal est représenté par la figure 3. Le signal est reconstruit à partir du huitième mode. La figure 5 montre les résultats de filtrage des signaux Heavysine et Doppler par EMD-DH et par ondelettes. Pour le signal Doppler les résultats des deux méthodes sont comparables (Fig. 5b) mais pour le Doppler, la méthode DH surpasse celle des ondelettes (Fig. 5a). Les

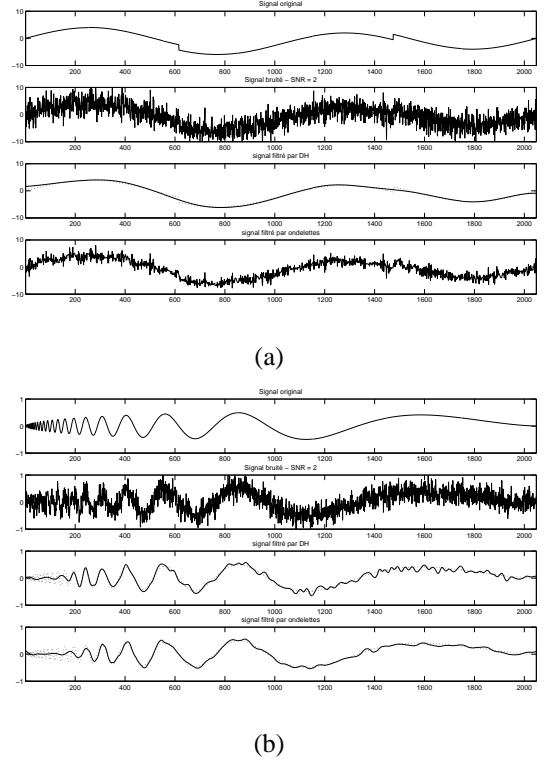


FIG. 5: Signaux Heavysine (a) et Doppler (b): non bruités, bruités et filtrés par ondelettes et EMD-DH (RSB=2dB).

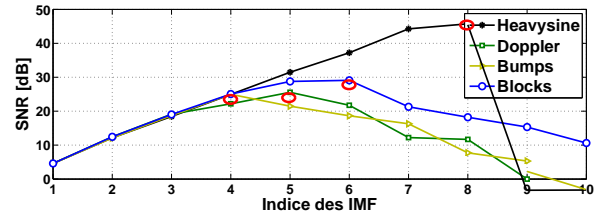


FIG. 6: Variation des différentes valeurs de RSB en fonction du nombre de modes sélectionnés pour la reconstruction des 4 signaux (en rouge, le k_{dh} pour les 4 signaux)

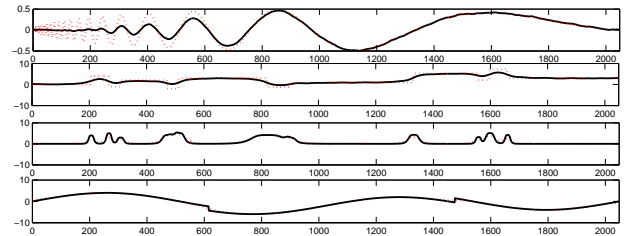


FIG. 7: Signaux débruités par EMD-DH pour un RSB=-5dB (pointillé rouge: les signaux originaux).

résultats pour RSB=2dB sont résumés dans le tableau 1. A l'exception du signal Doppler, l'approche EMD-DH donne de meilleurs résultats que le CMSE ou les ondelettes. La figure 7 montre les résultats du filtrage des signaux: Blocks, Bumps, Heavysine et Doppler pour un RSB=-5dB. Les RSB à la sortie (signaux filtrés) sont présentés dans le tableau 2. Ces résultats montrent la robustesse du critère même en environnement très bruité. Ces résultats montrent également que le filtrage par reconstruction partielle est possible pourvu qu'il y ait un critère de sélection robuste des modes pertinents. Cette robustesse est vérifiée par les résultats de la figure 6 qui montre que pour chaque signal le maximum du RSB a lieu à l'ordre k_{dh} (l'indice du premier mode dans la reconstruction partielle) confirmant ainsi les résultats obtenus par notre critère. Nous remarquons que pour chaque signal la courbe du RSB ne présente qu'un seul maximum. Ce type de filtrage est attractif dans la mesure où la décomposition est auto-adaptative comparée aux méthodes à base de noyau. De plus, le critère utilisant la DH proposé ne nécessite pas d'a priori sur la PDF du signal à analyser. Un autre résultat de cet travail est l'exploitation de la DH utilisée exclusivement en images pour le traitement des signaux 1D.

5 Conclusion

Dans ce papier une nouvelle méthode de débruitage des signaux (stationnaires ou non) basée sur la reconstruction partielle des modes empiriques extraits par EMD est présentée. Un nouveau critère de sélection des modes empiriques pertinents basée sur la distance de Hausdorff est introduit. Ce critère géométrique a démontré une bonne robustesse en environnement très bruité en terme de sélection de modes. Une comparaison a été faite avec des méthodes classiques telles que l'approche ondelettes et a montré que ce critère donne les meilleurs performances en terme de RSB. Les perspectives de ce travail concernent le test du filtrage EMD-DH sur une large classe de signaux synthétiques et réels, et en présence de d'autres types de bruits (coloré, . . .).

	<i>Signal</i>			
	<i>Doppler</i>	<i>Blocks</i>	<i>Heavysine</i>	<i>Bumps</i>
	RSB [dB]	RSB [dB]	RSB [dB]	RSB [dB]
Bruit	2	2	2	2
Ondelettes	28,78	11,94	14,47	18,76
EMD-CMSE	11,13	11,94	11,18	19,86
EMD-DH	26.89	25.01	41.86	22.61

TAB. 1: RSB des résultats de filtrage pour les signaux tests contaminés par un bruit blanc Gaussien (RSB=2dB).

	<i>Signal</i>			
	<i>Doppler</i>	<i>Blocks</i>	<i>Heavysine</i>	<i>Bumps</i>
	RSB [dB]	RSB [dB]	RSB [dB]	RSB [dB]
Bruit	-5	-5	-5	-5
EMD-DH	14.67	17.01	31.24	14.37

TAB. 2: RSB des résultats de filtrage pour les signaux tests contaminés par un bruit blanc Gaussien (RSB=-5dB).

References

- [1] A.O. Boudraa and J.C. Cexus, "EMD-Based Signal Filtering," *IEEE Trans. Instrum. Mea.*, vol. 56, pp. 2196-2202, 2007.
- [2] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, No. 3, pp. 613-627, 1995.
- [3] N.E. Huang et. al., "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Royal Society of London*, vol. 454, pp. 903-995, 1998.
- [4] P. Flandrin, P. Goncalves and G. Rilling, "Detrending and denoising with EMD," *EUSIPCO*, Wien, pp. 1581-1584, 2004.
- [5] Z. Wu and N.E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Royal Society of London, Series A*, vol. 460, Issue 2046, pp. 1597-1611, 2004.
- [6] Z.K. Peng, W. Tse Peter and F.L. Chu, "Comparison study of improved Hilbert-Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Sig. Proc.*, vol. 19, pp. 974-988, 2005.
- [7] X. AyenuPrah and N.O. Attoh-Okine, "A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 2, no. 1, pp. 1-24, 2010.
- [8] A.W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, 1997.