# Régularisation de noyaux temporellement élastiques et analyse en composantes principales non-linéaire pour la fouille de séries temporelles

Pierre-François Marteau*

*UMR 6074 IRISA, Université de Bretagne Sud, 56000 Vannes, France
pierre-francois DOT marteau AT irisa DOT fr,
http://people.irisa.fr/Pierre-Francois.Marteau/

**Résumé.** Dans le domaine de la fouille de séries temporelles, plusieurs travaux récents exploitent des noyaux construits à partir de distances élastiques de type Dynamic Time Warping (DTW) au sein d'approches à base de noyaux. Pourtant les matrices, apparentées aux matrices de Gram, construites à partir de ces noyaux n'ont pas toujours les propriétés requises ce qui peut les rendre *in fine* impropres à une telle exploitation. Des approches émergeantes de régularisation de noyaux élastiques peuvent être mises à profit pour répondre à cette insuffisance. Nous présentons l'une de ces méthodes, $K_{DTW}$, pour le noyau DTW, puis, autour d'une analyse en composantes principales non-linéaire (K-PCA), nous évaluons la capacité de quelques noyaux concurrents (élastiques v.s non élastiques, définis v.s. non définis) à séparer les catégories des données analysées tout en proposant une réduction dimensionnelle importante. Cette étude montre expérimentalement l'intérêt d'une régularisation de type $K_{DTW}$.

## 1 introduction

Les méthodes à noyau utilisées en analyse exploratoire des données (K-PCA, K-LDA, K-CCA, etc.) ou pour traiter des tâches de classification ou de régression (machines à support vectoriel, SVM) nécessitent, dans leurs fondements, l'usage de noyaux définis (positifs ou négatifs). Pourtant, bon nombre d'études relativement récentes en fouille de données temporelles présentent des résultats produits par de telles méthodes exploitant des noyaux temporellement *élastiques* (NTE) non définis Haasdonk (2005) Zhang et al. (2010) ou régularisés par des méthodes spectrales Narita et al. (2007). L'émergence de nouvelles méthodes de régularisation pour NTE offre aujourd'hui des alternatives à l'exploitation des noyaux élastiques non définis que nous nous proposons d'évaluer de manière comparative sur des jeux de données simples mais potentiellement explicites. D'une manière générale, les procédures de régulation ont été développées pour approximer des noyaux non définis par des noyaux définis (ou semi-définis). Les premières approches appliquent directement des transformations spectrales aux matrices de Gram issues des noyaux non définis. Ces méthodes Wu et al. (2005) Chen et al. (2009) consistent à i) changer le signe des valeurs propres négatives ou décaler ces valeurs propres en utilisant la valeur de décalage minimal nécessaire pour rendre le spectre des valeurs propres

| Noyau | forme | Propriétés |
|---|---|---|
| Gaussien-Euclidien | $exp(-\|\|.,.\|\|^2/\sigma)$ | défini positif, non élastique |
| Gaussien-DTW | $exp(-DTW(.,.)/\sigma)$ | non défini, élastique |
| Matrice de corrélation la plus proche de la matrice issue du Gaussien-DTW | $NC(exp(-DTW(.,.)/\sigma))$ | défini positif, élastique |
| DTW régularisé | $K_{DTW}(.,.)$ | défini positif, élastique |
| DTW régularisé à la puissance $t$ | $K_{DTW}(.,.)^t$ | défini positif, élastique |

TAB. 1 – *Liste des noyaux analysés*

positif, et ii) reconstruire la matrice de Gram issue du noyau avec les vecteurs propres d'origine afin de produire une matrice semi-définie positive. D'autres approches sont basées sur la recherche de la matrice de corrélation (matrice symétrique positive semi-définie ayant une diagonale unitaire) la plus proche de la matrice de Gram issue du noyau non défini, la proximité étant prise au sens d'une norme (norme de Frobenius pondérée) Higham (2002).

Cependant ces procédures de *convexification* sont difficiles à interpréter géométriquement Graepel et al. (1998) et l'effet attendu du noyau d'origine non défini peut être, selon les études, soit perdu ou pour le moins atténué par ces méthodes agissant directement sur le spectre matriciel, soit encore minime voir négatif comparativement à l'exploitation directe de la matrice non régularisée Chen et Ye (2008). Dans le contexte de l'alignement de séquences ou de séries temporelles, des approches de régularisation plus directes pour les NTE consistent à remplacer les opérateurs $\min$ ou $\max$ par un opérateur de sommation ($\sum$) dans les équations récursives qui définissent les distances élastiques. Il en résulte qu'au lieu de ne considérer que le meilleur chemin d'alignement possible entre deux séries temporelles, le noyau régularisé effectue la somme des *couts* (ou *gains*) de tous les chemins d'alignement possibles avec un mécanisme de pondération qui cherche à favoriser les *bons* alignements et à pénaliser les *mauvais* alignements. Ces principes ont été appliqués avec succès par Saigo et al. (2004) pour la mesure (non définie) de Smith et Waterman (1981) très utilisée pour la comparaison de séquences symboliques, et plus récemment pour la speudo distance Dynamic Time Warping (DTW, Velichko et Zagoruyko (1970), Sakoe et Chiba (1971)) Cuturi et al. (2007), Marteau et Gibet (2014).

Nous développons dans cet article, sur la base d'une analyse en composantes principales à noyau (K-PCA), une étude expérimentale permettant d'évaluer les noyaux listés en table 1 sur de tâches de classification (supervisée et non-supervisée) de séries temporelles dans des sous espaces de dimension réduite. En nous limitant à des ensembles de séries temporelles de taille fixe, nous proposons ainsi de comparer expérimentalement au travers d'une analyse K-PCA un noyau Gaussien construit à partir de la distance Euclidienne (noyau défini positif, non temporellement élastique, ce noyau servant de base de référence), un noyau Gaussien construit à partir de la pseudo distance DTW (noyau non défini en général, mais élastique), une version régularisée du noyau précédent basée sur la recherche de la matrice de corrélation la plus proche Higham (2002), et enfin le noyau DTW régularisé suivant la méthode proposée par Marteau et Gibet (2014), $K_{DTW}$, et une version normalisée, $K^t_{DTW}$.

# 2 Le noyau DTW régularisé $K_{DTW}$

Faisant suite aux travaux de Cuturi et al. (2007), la technique de régularisation développée dans Marteau et Gibet (2014) s'attache à transformer les équations récursives définissant la DTW (Dynamic Time Warping) de manière à produire une mesure de similarité notée $K_{DTW}$ constituant un noyau défini positif, c'est-à-dire s'apparentant à un produit scalaire dans un espace de Hilbert à noyau reproduisant. $K_{DTW}$ se distingue de l'approche proposée par Cuturi et al. en prenant la forme d'un noyau de convolution tel que défini par Haussler (1999) tout en imposant une condition sur les coûts locaux d'alignement moins restrictive. Pour rappel, un noyau défini sur $\mathbb{R}$ est est une fonction continue symétrique $K : A \times A \rightarrow \mathbb{R}$ telle que $K(x, y) = K(y, x)$. $K$ est dit défini positif si et seulement si :

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) c_i c_j \geq 0$$

pour toute séquence finie de points $(x_1, ..., x_n) \in A^n$ et toute séquence de nombres réels $[c_1, ..., c_n) \in \mathbb{R}^n$ associée.

## 2.1 Definition

La définition récursive du noyau $K_{DTW}$ est la suivante :

$$\mathcal{K}_{DTW}(X_p, Y_q) = K^{xy}_{DTW}(X_p, Y_q) + K^{xx}_{DTW}(X_p, Y_q)$$

où

$$K^{xy}_{DTW}(X_p, Y_q) = \frac{1}{3} e^{-d_E^2(x(p),y(q))/\sigma} \sum \begin{cases} h(p-1,q) K^{xy}_{DTW}(X_{p-1}, Y_q) \\ h(p-1,q-1) K^{xy}_{DTW}(X_{p-1}, Y_{q-1}) \\ h(p,q-1) K^{xy}_{DTW}(X_p, Y_{q-1}) \end{cases} \quad (1)$$

$$K^{xx}_{DTW}(X_p, Y_q) = \frac{1}{3} \sum \begin{cases} h(p-1,q) K^{xx}_{DTW}(X_{p-1}, Y_q) e^{-d_E^2(x(p),y(p))/\sigma} \\ \Delta_{p,q} h(p,q) K^{xx}_{DTW}(X_{p-1}, Y_{q-1}) e^{-d_E^2(x(p),y(q))/\sigma} \\ h(p,q-1) K^{xx}_{DTW}(X_p, Y_{q-1}) e^{-d_E^2(x(q),y(q))/\sigma} \end{cases}$$

avec

- $\Delta_{p,q}$ est le symbole de Kronecker,
- $h(.,.)$ est une fonction symétrique positive utilisée pour spécifier un corridor symétrique limitant le décompte des chemins d'alignement possibles,
- $\sigma \in \mathbb{R}^+$ est un paramètre d'ajustement qui permet de pondérer les contributions locales, i.e. les distances entre les positions localement alignées, et
- $d_E(.,.)$ est la distance euclidienne définie sur $\mathbb{R}^k$ (ou tout autre noyau négatif semi-défini sur $\mathbb{R}^k$).

Ainsi, fondamentalement, l'opérateur *min* (ou *max*) est remplacé par un opérateur de sommation et une fonction de corridor symétrique (la fonction $h$ dans l'équation récursive ci-dessus) est introduite pour, éventuellement, limiter la sommation et et donc la complexité algorithmique. Enfin, un nouveau terme récursif nécessaire à la régularisation ($K^{xx}$) est ajouté,

de telle sorte que la preuve de la propriété de positivité du noyau peut être comprise comme une conséquence directe du théorème de convolution d' Haussler (1999).

## 2.2 Normalisation

Le noyau $K_{DTW}$ effectue la somme sur l'ensemble des chemins d'alignement possibles des produits des termes locaux d'alignement $e^{-d_E^2(x(p),y(p))/\sigma} \leq 1$. Pour les séries temporelles de grandes tailles, ces produits deviennent infimes et $K_{DTW}$ incalculable lorsque $\sigma$ est trop faible. Ainsi, le domaine de variation de $K_{DTW}$ s'amenuise en convergeant vers 0 lorsque $\sigma$ tend vers 0 sauf lorsque l'on compare deux séries temporelles identiques (la matrice de Gram correspondante souffre ainsi d'une *dominance diagonale*). Comme proposé dans Marteau et Gibet (2014), une manière de palier ce problème consiste à considérer le noyau *normalisé* :

$$\tilde{K}_{DTW}(.,.) = exp\left(\alpha \frac{log(K_{DTW}(.,.)) - log(K_{DTW_{Min}})}{log(K_{DTW_{Max}}) - log(K_{DTW_{Min}})}\right)$$

où $K_{DTW_{Max}}$ et $K_{DTW_{Min}}$ sont respectivement les valeurs maximale et minimale prises par le noyau sur des données d'apprentissage et $\alpha > 0$ une constante ($\alpha = 1$ par défaut). Si l'on oublie la constante de proportionnalité, cela revient simplement à élever le noyau $K_{DTW}$ à la puissance $t = \alpha/(log(K_{DTW_{Max}}) - log(K_{DTW_{Min}}))$ , ce qui montre que $\tilde{K}_{DTW}$ est lui aussi défini positif (Berg et al. (1984), Proposition 2.7). Une correction de dominance diagonale similaire (sous-polynomial, i.e. $t < 1$) a initialement été proposée dans Schölkopf et al. (2002).

L'effet de ce type de *normalisation*, sur le jeu de données SwedishLeaf (c.f. Table 2) est illustré en figure 1. La distribution des valeurs de la matrice de Gram associée au noyau non normalisé $K_{DTW}$ (évalué avec $\sigma = 1$) présente une très forte accumulation autour des valeurs très faibles (pour $\sigma = 1$, $K_{DTW_{Min}} = 2.1e-77$ et $K_{DTW_{Max}} = 1.9e-06$ sur le jeu de données testé), tandis que la distribution des valeurs de la matrice de Gram associée au noyau *normalisé* $\tilde{K}_{DTW} = K_{DTW}^t$ est plus diffuse. Les valeurs du noyau sont par ailleurs bornées : $\forall x, y, \tilde{K}_{DTW}(x, y) \in [0, e]$.

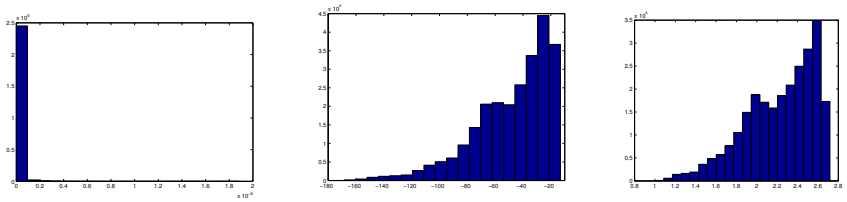

FIG. 1 – *Histogramme des valeurs des matrices de Gram : noyau $K_{DTW}$ à gauche, noyau $log(K_{DTW})$ au centre et noyau $K_{DTW}^t$ normalisé à droite, avec $t = 0.0061$.*

## 2.3 Complexité algorithmique

La définition récursive précédente permet de montrer que la complexité algorithmique liée au calcul du noyau $K_{DTW}$ est $O(n^2)$, où $n$ est la longueur des deux séries temporelles mises en correspondance, et lorsqu'aucun corridor n'est spécifié. Cette complexité est ramenée à $O(c.n)$ quand un corridor symétrique de taille $c$ est exploité par le biais de la fonction symmétrique $h$.

# 3 Analyse en composantes principals non linéaire (KPCA)

L'analyse en composantes principales non-linéaire, encore appelée ACP à noyau ou Kernel-PCA, Schölkopf et al. (1998) peut être vue comme une généralisation de l'ACP classique : elle permet d'engendrer une réduction de dimensionnalité non linéaire du point de vue de l'espace de représentation initial des données. Le principe consiste à projeter, par le biais d'une fonction non-linéaire $\Phi(.)$, les données initiales dans un espace en général de plus haute dimension de sorte que l'image de la variété (non linéaire) contenant les données initiales devienne plus facilement linéairement séparable dans le nouvel espace, appelé espace des caractéristiques. Il suffit alors d'effectuer une ACP classique dans cet espace linéaire pour obtenir une réduction de dimensionalité non linéaire dans l'espace des données initiales.
Si l'on exploite un noyau $K(.,.)$ défini positif, celui-ci induit de manière implicite une fonction nonlinéaire dite de *mapping* $\Phi(.)$ telle que $\forall x, y$, $K(x,y) =< \Phi(x)^T, \Phi(y) >$. Cette fonction $\Phi(.)$ n'a pas besoin d'être connue explicitement (on évoque ici l'astuce du noyau).

**Algorithm 1** ACP non linéaire

1: Choix du noyau (défini positif) $k$
2: Construction de la matrice de Gram à partir des données : $K = [k(x_i, x_j)]_{i,j=1,...,m}$
3: Centrage de la matrice de Gram (on retire la moyenne des données projetées dans l'espace des caractéristiques) :
$\tilde{K} = K - 2\mathbf{1}_{1/m} + \mathbf{1}_{1/m} K \mathbf{1}_{1/m}$,
où $\mathbf{1}_{1/m}$ est la matrice $m \times m$ dont tous les éléments sont égaux à $1/m$
4: Résolution du problème aux valeurs propres : $\tilde{K}\alpha_i = \lambda_i \alpha_i$, $j = 1, ..., m$
5: Projection d'un point $x$ quelconque dans l'espace des caractéristiques :
$y_j = \sum_{i=1}^{m} \alpha_{i,j} k(x, x_i)$, $j = 1, ..., d \leq m$

L'algorithme 1 présente succinctement les étapes de l'ACP non-linéaire, qui, à partir du choix d'un noyau défini positif, extrait les valeurs et vecteurs propres de la matrice de Gram centrée associée, puis projette toute donnée (initiale ou de test) dans un espace des caractéristiques de dimension réduite ($d$). Il est clair que l'ACP non-linéaire nécessite que le noyau utilisé soit défini positif.

# 4 Expérimentation

L'expérimentation proposée a pour objectif i) l'évaluation comparative de quelques approches de régularisation pour les noyaux temporellement élastiques type DTW sur des jeux de séries temporelles de natures diverses (en taille, volume et nombre de catégories) et ii) l'évaluation de leur apport par rapport à des noyaux non régularisés ou non élastiques. Cette évaluation porte à la fois sur la capacité des noyaux à classer (de manière supervisée ou non supervisée) les données et à les représenter dans des espaces de dimension réduite. Notons ici que le noyau de Fisher exploitable pour la mise en correspondance de modèles génératifs associés aux séries temporelles n'est pas considéré : il est réputé moins performant que les noyaux régularisés à base d'alignements locaux Vert et al. (2004) .

| Jeux de données | # séries temporelles | longueur des séries | # catégories |
|---|---|---|---|
| Adiac | 390 | 176 | 37 |
| CBF | 30 | 128 | 3 |
| ECG200 | 100 | 96 | 2 |
| FaceFour | 24 | 350 | 4 |
| FISH | 175 | 463 | 7 |
| Gun_Point | 50 | 150 | 2 |
| Lighting2 | 637 | 60 | 2 |
| Lighting7 | 319 | 70 | 7 |
| OSULeaf | 200 | 427 | 6 |
| SwedishLeaf | 500 | 128 | 15 |
| synthetic_control | 300 | 60 | 6 |
| yoga | 300 | 426 | 2 |
| 50words | 450 | 270 | 50 |

TAB. 2 – *Liste des jeux de données utilisés issus de Keogh et al. (2006)*

## 4.1 Contexte expérimental et expériences considérées

Les 13 jeux de données exploités dans le cadre de cette étude et listés en Table 2 sont issus de la banques de séries temporelles d'UCR Keogh et al. (2006). Ils sont de tailles modestes pour permettre (éventuellement) une visualisation la plus explicite possible en faible dimension. Le nombre de catégories varie de 2 à 50 et la longueurs des séries varie de 60 à 463.

Pour les 13 jeux de données listés en Table 2, et les 5 noyaux listés en Table 1, une ACP non linéaire est pratiquée puis les données sont projetées dans l'espace des caractéristiques obtenu en faisant varier le nombre de vecteurs propres, c'est à dire la dimension de l'espace réduit. Par exemple, en figure 2 les projections des séries temporelles du jeu de données Gun_Point sont présentées dans un espace des caractéristiques de dimension 3 pour les noyaux Gaussien-Euclidien, Gaussien-DTW régularisé par matrice de corrélation la plus proche (MCPP), $K_{DTW}$ et $K_{DTW}^t$. La valeur du paramètre $t$, exposant du noyau $K_{DTW}^t$, est estimé directement à partir des données d'apprentissage en évaluant les valeurs extrêmes prises par le noyau $K_{DTW}$ non normalisé. A titre d'exemple, on considère le jeu de données Gun_Point, pour lequel a matrice de Gram évaluée sur le noyau Gaussien DTW n'est pas définie. Comme le montre la figure 2, les projections dans le sous-espace des caractéristiques de dimension 3 sont très proches pour le noyau Gaussien DTW et sa version régularisée par matrice de corrélation la plus proche. La séparation des classes est, sur cet exemple, bien meilleure pour le noyau $KDTW$ et sa version normalisée $KDTW^t$.

A l'issue de l'ACP non linéaire, nous proposons une expérience de classification supervisée basée sur la règle du plus proche voisin (1-PPV) et une expérience de clustering basée sur l'algorithme des $K$-moyennes (K correspondant ici au nombre effectif de catégories du jeux de données). Pour chaque jeux de données et pour chaque noyaux testés la classification supervisée et non supervisée sont effectuées dans le sous espace des caractéristiques défini par K-PCA en faisant varier la dimension du sous-espace de 1 à 20 (pour dix dimensions, cela correspond à une réduction dimensionnelle variant de 83% à 98% pour les jeux de données testés).

La qualité de la classification est ensuite évaluée à partir du taux d'erreur de classification estimé à partir d'une validation croisée en 5 sous-échantillons. La précision et l'information

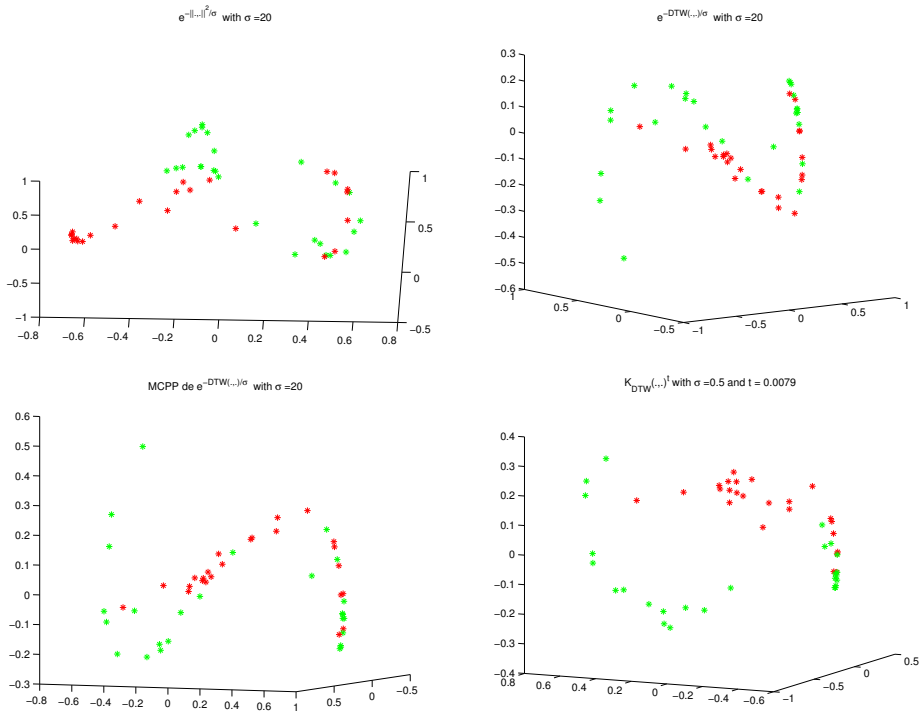


FIG. 2 – *Projections des séries du jeux de données Gun_Point dans le sous espace des caractéristiques de dimension 3 obtenu par K-PCA pour les noyaux : Gaussien Euclidien (en haut à gauche), Gaussien DTW (en haut À DROITE), MCPP du Gaussien DTW (en bas à gauche) et $K^t_{DTW}$ (en bas à droite).*

mutuelle normalisée ($IMN$) sont utilisées pour évaluer la qualité d'une classification non supervisée sur des données La précision est définie comme la fraction des individus correctement étiquetés, étant donné une correspondance 1-vers-1 entre les *vraies* classes et les clusters découverts. Si $p$ dénote une permutation quelconque des indices des clusters $\{\tilde{c}_i\}$ proposés (ou des *vraies* classes $\{c_j\}$), la précision est alors définie par : $P = \frac{1}{N} MAX_p \sum_{i=1\cdots K} n_{i,p(i)}$ où $n_{i,p(i)}$ représente le nombre d'individus partagés entre la classe $c_i$ et le cluster $\tilde{c}_{p(i)}$, $K$ est le nombre de clusters/classes, et $N$ est le nombre total d'individus.

L'information mutuelle normalisée, $IMN$, entre la *vraie* classification $\mathcal{C}$ et celle prédite $\tilde{\mathcal{C}}$ est définie par : $NMI(\tilde{\mathcal{C}}, \mathcal{C}) = I(\tilde{\mathcal{C}}, \mathcal{C})/(H(\tilde{\mathcal{C}}) + H(\mathcal{C}))/2$ où :
$I(\tilde{\mathcal{C}}, \mathcal{C}) = \sum_k \sum_j P(\tilde{c}_k \cap c_j) \log \frac{P(\tilde{c}_k \cap c_j)}{P(\tilde{c}_k) P(c_j)}$, $H(\tilde{\mathcal{C}}) = -\sum_k P(\tilde{c}_k) \log P(\tilde{c}_k)$ et
$H(\mathcal{C}) = -\sum_k P(c_k) \log P(c_k)$

## 4.2 Résultats et analyse

A titre d'exemple, nous présentons pour les jeux de données CBF et FISH, respectivement en figure 3 et 4, les taux d'erreur pour la classification supervisée 1-PPV et les mesures IMN

et Précision pour la classification non supervisée par k-moyennes lorsque la dimension de l'espace des caractéristiques varie de 1 à 20. Sur CBF la matrice de Gram évaluée sur le noyau Gaussien DTW est semi-définie positive, ce qui n'est pas le cas sur le jeu de données FISH.

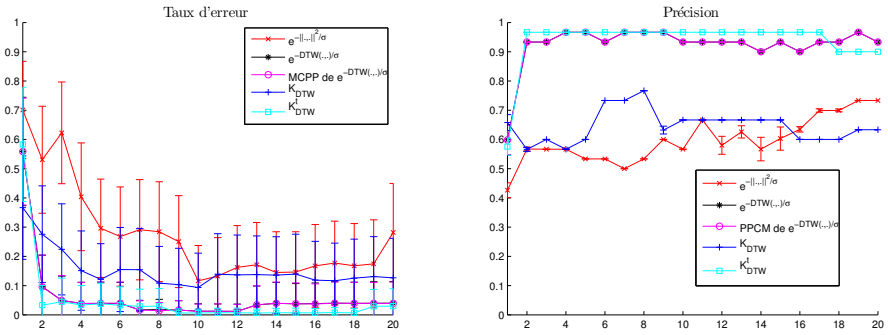


FIG. 3 – *Taux d'erreur de classification 1-PPV (à gauche) et précision (à droite) pour le clustering obtenu par les K-moyennes - pour les 5 noyaux testés (Gaussien Euclidien $\times$ rouge, Gaussien DTW $*$ noire, MCPP Gaussien DTW o rose, KDTW $+$ bleu, $KDTW^t$ $\square$ cyan) sur le jeu données CBF, lorsque la dimension du sous espace des caractéristiques varie de 1 à 20.*

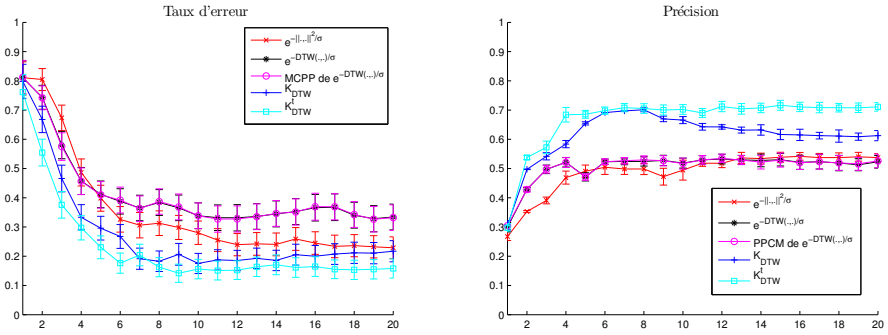


FIG. 4 – *Taux d'erreur de classification 1-PPV (à gauche)et précision (à droite) pour le clustering obtenu par les K-moyennes - pour les 5 noyaux testés (Gaussien Euclidien $\times$ rouge, Gaussien DTW $*$ noire, MCPP Gaussien DTW o rose, KDTW $+$ bleu, $KDTW^t$ $\square$ cyan) sur le jeu données FISH, lorsque la dimension du sous espace des caractéristiques varie de 1 à 20.*

Ces diagrammes permettent de comparer les noyaux sur la base des trois mesures (taux d'erreur, IMN et Précision) mais aussi sur leur capacité à offrir une réduction de dimension importante. Les figures 3 et 4 montrent que sur le jeux de données CBF, les noyaux Gaussien-DTW et $KDTW^t$ sont les plus efficaces, en offrant une très bonne séparation des classes dans un sous-espace de dimension 2. Sur le jeux de données FISH, ce sont les noyaux $KDTW$ et $KDTW^t$ qui sont les plus efficaces avec un optimum pour un sous-espace de dimension 10.

Par ailleurs, le noyau Gaussien-DTW et sa variante régularisée à partir de la matrice de corrélation la plus proche conduisent à des résultats très similaires et sensiblement moins bons comparativement aux noyaux $KDTW$ et $KDTW^t$. La régularisation par matrice de corrélation la plus proche ne semble donc pas apporter de bénéfice significatif en terme de classification supervisée ou non supervisée sur ces jeux de données par rapport au noyau DTW non défini.

| **Données** | $e^{-\|\|.,.\|\|^2/\sigma}$ | $e^{-DTW(.,.)/\sigma}$ | $NC(e^{-DTW(.,.)/\sigma})$ | $K_{DTW}(.,.)$ | $K^t_{DTW}(.,.)$ |
|---|---|---|---|---|---|
| 50words | 0.3930 | 0.4547 | 0.4527 | 0.4183 | **0.3003** |
| Adiac | 0.5423 | 0.4158 | 0.4169 | 0.4170 | **0.3523** |
| CBF | 0.1188 | **0.0113** | 0.0116 | 0.0944 | 0.0347 |
| ECG200 | **0.1295** | 0.2589 | 0.2594 | 0.1680 | 0.1809 |
| FaceFour | 0.4408 | 0.2335 | 0.2315 | 0.2013 | **0.1601** |
| FISH | 0.2753 | 0.3371 | 0.3392 | 0.1759 | **0.1543** |
| Gun_Point | 0.1817 | 0.2443 | 0.2430 | 0.1137 | **0.0258** |
| Lighting2 | 0.3106 | 0.1155 | **0.1153** | 0.4034 | 0.1642 |
| Lighting7 | 0.4099 | 0.3925 | 0.3926 | 0.3949 | **0.2817** |
| OSULeaf | 0.4109 | 0.3436 | 0.3461 | 0.4043 | **0.3102** |
| SwedishLeaf | 0.2648 | 0.2643 | 0.2635 | 0.2627 | **0.1373** |
| synthetic_ctrl | 0.0150 | 0.0061 | 0.0060 | 0.0070 | **0.0045** |
| yoga | 0.3755 | 0.2608 | 0.2589 | 0.2840 | **0.2217** |

TAB. 3 – *Taux d'erreur de classification obtenus pour les 5 noyaux testés à partir d'un classifieur 1-PPV, d'une validation croisée à 5 sous-échantillons, et d'une projection sur les 10 axes principaux obtenus par K-PCA. Pour chaque test, les valeurs les plus faibles sont en gras.*

| **Données** | $e^{-\|\|.,.\|\|^2/\sigma}$ | $e^{-DTW(.,.)/\sigma}$ | $NC(e^{-DTW(.,.)/\sigma})$ | $K_{DTW}(.,.)$ | $K^t_{DTW}(.,.)$ |
|---|---|---|---|---|---|
| 50words | 0.3964 | 0.3838 | 0.3857 | 0.3865 | **0.4648** |
| Adiac | 0.3793 | 0.5041 | **0.5042** | 0.4785 | 0.4957 |
| CBF | 0.5667 | 0.9333 | 0.9333 | 0.6667 | **0.9667** |
| ECG200 | **0.8100** | 0.6800 | 0.6800 | 0.7500 | 0.7800 |
| FaceFour | 0.7500 | 0.7500 | 0.7500 | 0.7500 | **0.7917** |
| FISH | 0.4961 | 0.5155 | 0.5206 | 0.6689 | **0.6971** |
| Gun_Point | 0.5600 | 0.5600 | 0.5600 | **0.6200** | 0.5200 |
| Lighting2 | 0.5500 | **0.6000** | **0.6000** | 0.5333 | 0.5333 |
| Lighting7 | 0.4991 | 0.5536 | 0.5534 | 0.5571 | **0.6041** |
| OSULeaf | 0.4146 | 0.4410 | 0.4435 | 0.3945 | **0.4717** |
| SwedishLeaf | 0.4668 | 0.5486 | 0.5483 | 0.4917 | **0.6384** |
| synthetic_ctrl | 0.6767 | 0.9900 | 0.9900 | 0.8283 | **1** |
| yoga | 0.5233 | 0.5267 | 0.5267 | **0.5500** | 0.5033 |

TAB. 4 – *Précision associée au meilleur alignement entre les clusters obtenus par K-moyennes (effectué dans le sous espace défini par les 10 axes principaux obtenus par K-PCA) et les classes connues pour les 5 noyaux testés. Les meilleures valeurs pour chaque test sont en gras.*

Les tables 3, 4 et 5 présentent, pour les 13 jeux de données et pour les 5 noyaux testés, respectivement, les taux d'erreur moyens de classification 1-PPV obtenus par validation croi-

| **Données** | $e^{-\|\|.,.\|\|^2/\sigma}$ | $e^{-DTW(.,.)/\sigma}$ | $NC(e^{-DTW(.,.)/\sigma})$ | $K_{DTW}(.,.)$ | $K^t_{DTW}(.,.)$ |
|---|---|---|---|---|---|
| 50words | 0.6608 | 0.6554 | 0.6566 | 0.6530 | **0.7277** |
| Adiac | 0.6059 | 0.7019 | **0.7029** | 0.6910 | 0.6931 |
| CBF | 0.1970 | 0.7986 | 0.7986 | 0.3743 | **0.9020** |
| ECG200 | **0.3845** | 0.1576 | 0.1576 | 0.3589 | 0.1922 |
| FaceFour | 0.6002 | 0.6944 | 0.6944 | 0.6225 | **0.7823** |
| FISH | 0.3826 | 0.4338 | 0.4341 | 0.5844 | **0.5874** |
| Gun_Point | 0.0126 | 0.0126 | 0.0126 | **0.0413** | 0.0015 |
| Lighting2 | 0.0067 | **0.0850** | **0.0850** | 0.0234 | 0.0038 |
| Lighting7 | 0.4743 | 0.5575 | 0.5574 | 0.5325 | **0.5802** |
| OSULeaf | 0.3189 | 0.2550 | 0.2638 | 0.2720 | **0.3453** |
| SwedishLeaf | 0.5642 | 0.5925 | 0.5943 | 0.6230 | **0.7081** |
| synthetic_ctrl | 0.7281 | 0.9726 | 0.9726 | 0.7766 | **1** |
| yoga | 0.0000 | 0.0026 | 0.0026 | **0.0034** | 0.0003 |

TAB. 5 – *Information mutuelle associée entre les clusters obtenus par K-moyenne (effectué dans le sous espace défini par les 10 axes principaux obtenus par K-PCA) et les classes connues pour les 5 noyaux testés. Les meilleurs valeurs pour chaque test sont en gras.*

sée à 5 sous-échantillons (Tab.4), la Précision de la classification prédite par un clustering K-moyennes (Tab.3), et l'information mutuelle normalisée entre cette même classification prédite et la *vraie* classification (Tab.5). Ici, les classifications non supervisée et supervisée ont toutes été effectuées dans les sous-espaces des caractéristiques de dimension 10 obtenus par K-PCA à partir des matrices de Gram centrées associées aux noyaux. Pour les autres dimensions (en particulier 3, 5, 15 et 20) les noyaux se classent de manière comparable au classement obtenus pour la dimension 10, les mesures d'évaluations s'améliorant en général avec l'accroissement de dimension comme illustré sur les figures 3 et 4. Il ressort de cette étude que le noyau régularisé normalisé $K^t_{DTW}$ est de loin le plus robuste. Il obtientles taux d'erreur de classification les plus faibles sur 10 des 13 jeux de données (CBF est mieux classé par le noyau Gaussien-DTW, ECG200 par le noyau Gaussien-Euclidien et Lighting2 par le noyau Gaussien-DTW régularisé par matrice de corrélation la plus proche). Pour la classification non supervisée, $K^t_{DTW}$ obtient également les meilleurs résultats d'après les mesures Précision et IMN pour 8 des 13 jeux de données. Pour cette tâche, $K_{DTW}$ est meilleur sur les jeux de données yoga et Gun_Point, tandis que le noyau Gaussien-Euclidien se distingue sur ECG200 et les noyaux Gaussien-DTW sur Lighting2. Sur les jeux de données pour lesquels $K^t_{DTW}$ n'arrive pas en tête, ce noyau se positionne entre le noyau Gaussien-Euclidien et les noyaux Gaussien-DTW. Contrairement à la régularisation par matrice de corrélation la plus proche, le principe de régularisation mise en oeuvre dans $K_{DTW}$ et $K^t_{DTW}$ modifie en profondeur la nature même de la fonction de similarité sous-jacente à la mesure DTW en apportant en général une meilleure capacité à séparer ou regrouper les séries temporelles dans des espaces de dimension réduite.

# 5 Conclusion

Nous avons évalué expérimentalement la capacité de quelques noyaux (élastiques, non-élastiques, définis, non-définis) à classer des séries temporelles de manière supervisée ou non

dans des espaces de caractéristiques à dimension réduite obtenus par ACP non linéaire. Les résultats présentés montrent que les approches récentes de régularisation de noyaux élastiques offrent des alternatives bien meilleures que les principes classiques de régularisation basés sur des approches spectrales portant directement sur les valeurs propres des matrices de Gram construites à partir des noyaux non définis. Le noyau DTW régularisé $K_{DTW}$ exploité dans cet article dans sa version *normalisé* $K^t_{DTW}$ offre un bon compromis entre les noyaux définis non-élastiques (tel que le noyau Gaussien-Euclidien) et les noyaux non définis élastiques (tel que le noyau Gaussien-DTW). Non seulement celui-ci conserve une caractéristique d'élasticité temporelle tout en étant défini positif, mais il se marie également bien avec les approches à noyau tel que l'ACP non linéaire. Sa capacité à proposer des espaces de caractéristiques discriminantes en dimension réduite en font un outil en général efficace pour l'analyse exploratoire d'ensembles de séries temporelles. Ces résultats confirment et complètent ainsi ceux présentés par Marteau et Gibet (2014) et Marteau et al. (2014) dans le cadre d'une classification supervisée par machine à support vectoriel en apportant un éclairage sur la *normalisation* de ce type de noyau.

# Références


Berg, C., J. P. R. Christensen, et P. Ressel (1984). *Harmonic Analysis on Semigroups : Theory of Positive Definite and Related Functions*, Volume 100 of *Graduate Texts in Mathematics*. New York : Springer-Verlag.

Chen, J. et J. Ye (2008). Training svm with indefinite kernels. In W. W. Cohen, A. McCallum, et S. T. Roweis (Eds.), *ICML*, Volume 307 of *ACM International Conference Proceeding Series*, pp. 136–143. ACM.

Chen, Y., E. K. Garcia, M. R. Gupta, A. Rahimi, et L. Cazzanti (2009). Similarity-based classification : Concepts and algorithms. *J. Mach. Learn. Res. 10*, 747–776.

Cuturi, M., J.-P. Vert, Ã. Birkenes, et T. Matsui (2007). A kernel for time series based on global alignments. In *IEEE ICASSP 2007*, Volume 2, pp. II–413–II–416.

Graepel, T., R. Herbrich, P. Bollmann-Sdorra, et K. Obermayer (1998). Classification on pairwise proximity data. In *NIPS*, pp. 438–444. The MIT Press.

Haasdonk, B. (2005). Feature space interpretation of svms with indefinite kernels. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 27*(4), 482–492.

Haussler, D. (1999). Convolution kernels on discrete structures. Technical Report UCS-CRL-99-10, University of California at Santa Cruz, Santa Cruz, CA, USA.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis 22*(3), 329–343.

Keogh, E. J., X. Xi, L. Wei, et C. Ratanamahatana (2006). The UCR time series classification-clustering datasets. http ://wwwcs.ucr.edu/ eamonn/time_series_data/.

Marteau, P.-F. et S. Gibet (2014). On Recursive Edit Distance Kernels with Application to Time Series Classification. *IEEE Trans. on Neural Networks and Learning Systems*, 1–14.

Marteau, P.-F., S. Gibet, et C. Reverdy (2014). Down-Sampling coupled to Elastic Kernel Machines for Efficient Recognition of Isolated Gestures. In *ICPR 2014*, Stockholm, Sweden.

Narita, H., Y. Sawamura, et A. Hayashi (2007). Learning a kernel matrix for time series data from dtw distances. In M. Ishikawa, K. Doya, H. Miyamoto, et T. Yamakawa (Eds.), *ICONIP (2)*, Volume 4985 of *Lecture Notes in Computer Science*, pp. 336–345. Springer.

Saigo, H., J.-P. Vert, N. Ueda, et T. Akutsu (2004). Protein homology detection using string alignment kernels. *Bioinformatics 20*(11), 1682–1689.

Sakoe, H. et S. Chiba (1971). A dynamic programming approach to continuous speech recognition. In *Proceedings of the Seventh International Congress on Acoustics, Budapest*, Volume 3, Budapest, pp. 65–69. Akadémiai Kiadó.

Schölkopf, B., A. Smola, et K.-R. Müller (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput. 10*(5), 1299–1319.

Schölkopf, B., J. Weston, E. Eskin, C. S. Leslie, et W. S. Noble (2002). A kernel approach for learning from almost orthogonal patterns. In T. Elomaa, H. Mannila, et H. Toivonen (Eds.), *ECML*, Volume 2430 of *LNCS*, pp. 511–528. Springer.

Smith, T. et M. Waterman (1981). Identification of common molecular subsequences. *Journal of Molecular Biology 147*, 195–197.

Velichko, V. M. et N. G. Zagoruyko (1970). Automatic recognition of 200 words. *International Journal of Man-Machine Studies 2*, 223–234.

Vert, J.-P., H. Saigo, et T. Akutsu (2004). Local alignment kernels for biological sequences. In K. T. B. Schölkopf et J.-P. Vert (Eds.), *Kernel Methods in Computational Biology*, pp. 131–154. MIT Press.

Wu, G., E. Y. Chang, et Z. Zhang (2005). Learning with non-metric proximity matrices. In *MULTIMEDIA '05 : Proceedings of the 13th annual ACM international conference on Multimedia*, New York, NY, USA, pp. 411–414. ACM.

Zhang, D., W. Zuo, D. Zhang, et H. Zhang (2010). Time series classification using support vector machine with gaussian elastic metric kernel. In *Intern. Conf. on Pattern Recognition*, ICPR '10, Washington, DC, USA, pp. 29–32. IEEE.

# Summary

In the context of time series data mining, recent studies exploit kernels constructed from elastic distances such as Dynamic Time Warping within kernel based methods. Yet matrix, related to Gram matrices, constructed from these kernels do not always have the required definiteness property which can make them unsuitable for such use. Emerging approaches dedicated to the regularization of time elastic kernels can be used in place of classical ones such as direct spectral approaches. We present in this paper a recent regularization method ($K_{DTW}$) for the DTW kernel and propose an experimental study exploiting a kernel principal component analysis for evaluating the ability of some kernels (elastic v.s. non elastic, definite v.s. not definite) to provide good classifications of the analyzed data, while providing an important reduction of dimensionality. This study shows the effectiveness of the regularization technique for time elastic kernels that is behind $K_{DTW}$.