

# Une Méthode à Noyaux pour l'Approximation Parcimonieuse des Représentations Temps-Fréquence Biliéaires

Manuel DAVY<sup>1</sup>, Patrick WOLFE<sup>2</sup>

<sup>1</sup>Laboratoire d'Automatique, de Génie Informatique et Signal  
BP 48, Cité Scientifique, 59651 Villeneuve d'Ascq cedex, France

<sup>2</sup>Harvard University, Cambridge, MA, USA

Manuel.Davy@ec-lille.fr, p.wolfe@ieee.org

**Résumé** – Nous présentons une méthode d'approximation des représentations temps-fréquence (RTF) reprenant le formalisme des méthodes à noyaux. Certains noyaux temps-fréquence étant définis positifs, la RTF est une fonction de l'espace de Hilbert à noyau reproduisant correspondant, qui est un espace de recherche naturel pour l'approximation. Cette dernière est recherchée avec une complexité minimale, et elle est parcimonieuse. Nous étendons le formalisme de la regression à vecteurs support pour résoudre le problème.

**Abstract** – We introduce a Time-Frequency Representation (TFR) approximation method, within the kernel machines framework. Some time-frequency kernels being positive definite, the TFR is a member of the related reproducing kernel Hilbert space which is the natural approximating function space. The approximation function is sought with minimum complexity and it is sparse. We solve the problem by an extension to the support vector regression formalism.

## 1 Introduction

Les représentations temps-fréquence (RTF) sont largement utilisées dans de multiples champs du traitement du signal, comme l'analyse des signaux audio, mécaniques, astrophysiques... Elle sont cependant largement redondantes, ce qui en fait des objets de grande dimension, et dont la manipulation intensive peut être coûteuse. Dans cet article, nous proposons une méthode d'approximation parcimonieuse des RTFs. L'objectif est de trouver une représentation aussi proche que possible de la représentation initiale, mais pouvant être résumée à peu de paramètres non nuls.

Depuis les travaux de Cohen [1], nous savons que toute représentation bilinéaire et covariante  $C_z^\phi(t, f)$  d'un signal peut s'écrire sous la forme d'une double convolution temps-fréquence de la représentation de Wigner-Ville, donnée par

$$\mathbb{W}V_z(t, f) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \quad (1)$$

par un noyau  $\phi(t, f)$ , comme suit :

$$C_z^\phi(t, f) = \iint \mathbb{W}V_z(\tau, \nu) \phi(t - \tau, f - \nu) d\tau d\nu \quad (2)$$

En pratique, nous utilisons des version discrétisées de ces expressions, selon une grille de  $N_t$  points temporels  $t_i, i = 1, \dots, N_t$  et  $N_f$  points fréquentiels  $f_i, i = 1, \dots, N_f$  (c'est-à-dire,  $C_z^\phi(t, f)$  est représenté sous la forme d'une matrice  $N_f \times N_t$ ).

L'idée de cet article est de construire une approximation de  $C_z^\phi(t, f)$ , recherchée dans la classe de fonctions de la forme

$$f(t, f) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_f} \alpha_{i,j} K((t, f), (t_i, f_j)) + b \quad (3)$$

où  $K(\cdot, \cdot)$  est un noyau défini positif défini sur  $\mathbb{R}^2 \times \mathbb{R}^2$ , et où chaque point  $(t_i, f_j)$  de la grille temps-fréquence porte un

coefficient  $\alpha_{i,j}, i = 1, \dots, N_t, j = 1, \dots, N_f$ . En outre, la solution recherchée doit être telle que les  $\alpha_{i,j}$  soient nuls pour la plupart – fournissant ainsi une approximation parcimonieuse – et positifs – conservant ainsi une interprétation en termes de représentation d'énergie.

Cet article est organisé de la façon suivante. La section 2 développe une méthode originale de régression à vecteurs support dans laquelle la contrainte de positivité des coefficients est imposée. La section 3 propose quelques commentaires et évoque le réglage des paramètres. La section 4 présente des résultats, et la dernière section propose quelques conclusions et des pistes de recherche future.

## 2 Formulation proposée

Dans la suite, nous noterons par  $x_i, i = 1, \dots, m$  (avec  $m = N_f N_t$ ) les points en dimension 2 de la grille temps-fréquence définie par  $t_k, k = 1, \dots, N_t$  et  $f_{k'}, k' = 1, \dots, N_f$ . Ainsi, à chaque couple  $(k, k'), k = 1, \dots, N_t, k' = 1, \dots, N_f$ , correspond-il un indice  $i$ , tels que  $x_i = [t_k, f_{k'}]^T$ . L'équation (3) peut être réécrite sous la forme

$$f(x) = \sum_{i=1}^m \alpha_i K(x, x_i) + b \quad (4)$$

où  $x = [t, f]^T$ . Nous supposons pour l'instant que les  $\alpha_i$  et  $b$  sont dans  $\mathbb{R}$  et que  $K(\cdot, \cdot)$  est donné. Enfin, nous notons dans la suite  $y$  la valeur de la RTF en  $x$ , c'est-à-dire  $y = C_z^\phi((t, f) = x)$  pour  $z$  et  $\phi(\cdot, \cdot)$  donnés.

Approcher  $y = C_z^\phi(x)$  par une fonction  $f(x)$  telle que celle de l'équation (4), avec une contrainte de parcimonie revient à résoudre un problème variationnel. Avant de le développer, nous introduisons une fonction de coût et les risques correspondants.

## 2.1 Fonction de coût et risque

Reprenant le formalisme introduit dans le cadre de la régression par les méthodes à vecteurs support, nous utilisons dans la suite la fonction de coût  $c(\cdot, f(\cdot))$  de la figure 1, dite  $\epsilon$ -insensible [2]. Le coût  $c(\cdot, f(\cdot))$  indique l'importance que l'on accorde à un écart entre la fonction  $f(x)$  et la valeur  $y$  de la RTF.

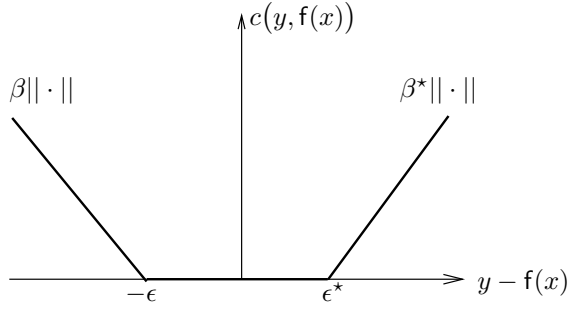


FIG. 1: Une fonction de coût de type  $\epsilon$ -insensible, non symétrique. Elle ne pénalise pas les erreurs se trouvant dans l'intervalle  $[-\epsilon, \epsilon^*]$ , et pénalise linéairement les erreurs plus importantes. Cela assure la parcimonie de la l'approximation trouvée.

Le risque empirique associé s'écrit comme la moyenne empirique de la fonction de coût

$$R_{\text{emp}}[f] = \frac{1}{m} \sum_{i=1}^m c(y_i, f(x_i)) \quad (5)$$

## 2.2 Problème variationnel

La fonction  $f(\cdot)$  de l'équation (4) appartient à un espace de fonctions de type *espace de Hilbert à noyau reproduisant*  $K(\cdot, \cdot)$  (car ce dernier est défini positif)  $\mathcal{H}$ , dans lequel il existe une norme induite notée  $\|\cdot\|_{\mathcal{H}}$ . A l'aide de cette norme, nous définissons le problème variationnel suivant, à résoudre numériquement :

$$\begin{aligned} \text{Minimiser} \quad & \frac{1}{2} \|f\|_{\mathcal{H}}^2 + CR_{\text{emp}}[f] \\ \text{Avec} \quad & \alpha_i \geq 0, \forall i = 1, \dots, m \end{aligned}$$

La solution de ce problème assure à la fois une bonne approximation de la RTF par la fonction  $f$  (faible risque empirique), ainsi qu'une solution de faible complexité (norme faible de  $f$  dans  $\mathcal{H}$ ), évitant ainsi le sur-apprentissage, et favorisant la parcimonie. Le problème peut se réécrire comme suit (voir [2] pour une dérivation similaire) : g

$$\begin{aligned} \text{Minimiser} \quad & \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j K(x_i, x_j) + \frac{C}{m} \sum_{i=1}^m (\beta \xi_i + \beta^* \xi_i^*) \\ \text{par rapport à} \quad & \{\alpha_i, \xi_i, \xi_i^*\}_{i=1, \dots, m} \in \mathbb{R}^{3m} \\ \text{Avec} \quad & (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \epsilon + \xi_i, \\ & y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \epsilon^* + \xi_i^*, \\ & \xi_i \geq 0, \xi_i^* \geq 0, \alpha_i \geq 0 \text{ pour } i = 1, \dots, m \end{aligned}$$

Ce problème quadratique et convexe admet une formulation duale qui fournit les coefficients  $\alpha_i$ . Pour cela, nous introduisons les multiplicateurs de Lagrange  $\eta_i, \eta_i^*, \gamma_i, \gamma_i^*$  et  $\lambda_i$ , positifs, et nous formons le Lagrangien  $L$  qui doit être minimisé par

rapport aux variables primales  $b, \{\alpha_i, \xi_i, \xi_i^*\}_{i=1, \dots, m}$  et maximisé par rapport aux variables duales  $\eta_i, \eta_i^*, \gamma_i, \gamma_i^*$  et  $\lambda_i$

$$\begin{aligned} L = \quad & \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j K(x_i, x_j) + \frac{C}{m} \sum_{i=1}^m (\beta \xi_i + \beta^* \xi_i^*) \\ & - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^* + \lambda_i \alpha_i) \\ & - \sum_{i=1}^m \gamma_i (\epsilon + \xi_i + y_i - \sum_{j=1}^m \alpha_j K(x_i, x_j) - b) \\ & - \sum_{i=1}^m \gamma_i^* (\epsilon^* + \xi_i^* + \sum_{j=1}^m \alpha_j K(x_i, x_j) + b - y_i) \end{aligned} \quad (6)$$

Sous forme vectorielle, il s'écrit (où les notations en gras représentent les vecteurs et matrices, par exemple,  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^T$  et  $\mathbf{K} = [K(x_i, x_j)]_{(i,j=1, \dots, N)}$ )

$$\begin{aligned} L = \quad & \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + \frac{C}{m} (\boldsymbol{\beta} \cdot \mathbf{1}_m^T \boldsymbol{\xi} + \boldsymbol{\beta}^* \cdot \mathbf{1}_m^T \boldsymbol{\xi}^*) \\ & - \boldsymbol{\eta}^T \boldsymbol{\xi} - \boldsymbol{\eta}^{*T} \boldsymbol{\xi}^* - \boldsymbol{\alpha}^T \boldsymbol{\lambda} \\ & - \boldsymbol{\gamma}^T ((\epsilon - b) \mathbf{1}_m + \boldsymbol{\xi} + \mathbf{y} - \mathbf{K} \boldsymbol{\alpha}) \\ & - \boldsymbol{\gamma}^{*T} ((\epsilon^* + b) \mathbf{1}_m + \boldsymbol{\xi}^* + \mathbf{K} \boldsymbol{\alpha} - \mathbf{y}) \end{aligned} \quad (7)$$

A l'optimum, dit *point selle*, les dérivées par rapport aux variables primales s'annulent,

$$\frac{\partial L}{\partial b} = (\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)^T \mathbf{1}_m = 0 \quad (8)$$

$$\frac{\partial L}{\partial \boldsymbol{\alpha}} = \mathbf{K}(\boldsymbol{\alpha} + \boldsymbol{\gamma} - \boldsymbol{\gamma}^*) - \boldsymbol{\lambda} = 0 \quad (9)$$

$$\frac{\partial L}{\partial \boldsymbol{\xi}} = \frac{C}{m} \boldsymbol{\beta} \cdot \mathbf{1}_m - \boldsymbol{\eta} - \boldsymbol{\gamma} = 0 \quad (10)$$

$$\frac{\partial L}{\partial \boldsymbol{\xi}^*} = \frac{C}{m} \boldsymbol{\beta}^* \cdot \mathbf{1}_m - \boldsymbol{\eta}^* - \boldsymbol{\gamma}^* = 0 \quad (11)$$

d'où l'on extrait ( $\mathbf{K}$  est inversible)

$$\boldsymbol{\alpha} = \mathbf{K}^{-1} \boldsymbol{\lambda} + \boldsymbol{\gamma}^* - \boldsymbol{\gamma} \quad (12)$$

Afin d'écrire ce problème sous une forme classique, nous introduisons  $\boldsymbol{\omega}^T = [\boldsymbol{\lambda}^T; \boldsymbol{\gamma}^T; \boldsymbol{\gamma}^{*T}]$  et

$$\mathbf{H} = \begin{bmatrix} \mathbf{K}^{-1} & -\mathbf{I}_m & \mathbf{I}_m \\ -\mathbf{I}_m & \mathbf{K} & -\mathbf{K} \\ \mathbf{I}_m & -\mathbf{K} & \mathbf{K} \end{bmatrix} \quad (13)$$

$$\mathbf{C} = [\mathbf{0}_m^T; \mathbf{y} + \epsilon \mathbf{1}_m^T; -\mathbf{y} + \epsilon^* \mathbf{1}_m^T]^T \quad (14)$$

$$\mathbf{A} = [\mathbf{0}_m^T; \mathbf{1}_m^T; -\mathbf{1}_m^T]^T \quad (15)$$

Enfin, soient  $\mathbf{l} = \mathbf{0}_{3m}$ ,  $\mathbf{u} = [+ \infty \mathbf{1}_m^T; \frac{C}{m} \boldsymbol{\beta} \mathbf{1}_m^T; \frac{C}{m} \boldsymbol{\beta}^* \mathbf{1}_m^T]^T$  et  $b = 0$ . Le problème d'optimisation à résoudre est quadratique sous contraintes linéaires, à résoudre par rapport à  $\boldsymbol{\omega}$

$$\begin{aligned} \text{minimiser} \quad & \frac{1}{2} \boldsymbol{\omega}^T \mathbf{H} \boldsymbol{\omega} + \mathbf{C}^T \boldsymbol{\omega} \\ \text{avec} \quad & \mathbf{A}^T \boldsymbol{\omega} = b \\ & \mathbf{l} \leq \boldsymbol{\omega} \leq \mathbf{u} \end{aligned} \quad (16)$$

Ce problème d'optimisation ne présente pas de difficultés numériques, et peut être traité par un algorithme de point intérieur [3] ou encore itérativement, en prenant en compte les points temps-fréquence l'un après l'autre.

### 3 Discussion

Le principe d'approximation proposé ici repose sur le choix d'un noyau défini positif. Ce choix peut être effectué de manière arbitraire, par exemple un noyau gaussien

$$K((t_1, f_1), (t_2, f_2)) = e^{-0.5 \left[ \frac{(t_1 - t_2)^2}{\sigma_t^2} + \frac{(f_1 - f_2)^2}{\sigma_f^2} \right]} \quad (17)$$

Dans certains cas, le noyau temps-fréquence  $\phi$  lui-même est défini positif (par exemple, le noyau du spectrogramme à fenêtre gaussienne correspond au noyau gaussien ci-dessus), et il est naturel de choisir

$$K((t_1, f_1), (t_2, f_2)) = \phi(t_1 - t_2, f_1 - f_2) \quad (18)$$

Ainsi, la fonction  $f$  obtenue est proche de la représentation initiale, tout en comportant beaucoup de coefficients nuls. En d'autres termes, la fonction  $f$  appartient au même espace de Hilbert à noyau reproduisant que  $C_z^\phi(t, f)$ . Cela permet à la fois un débruitage, et une compression de la RTF, permettant par exemple la mise en œuvre d'algorithmes de classification ou de détection. Le coefficient de régularisation permet de fixer la proportion de coefficients nuls: pour  $C$  petit, beaucoup de coefficients  $\alpha_i$  sont nuls, mais l'approximation effectuée est moins précise que si  $C$  est grand, et que beaucoup de coefficients sont non nuls. De même, les paramètres  $\epsilon$  et  $\epsilon^*$  permettent de régler le niveau de précision souhaité pour l'approximation en fixant un «horizon» à partir duquel on pénalise l'écart entre la RTF originale et son approximation.

L'ensemble des points temps-fréquence pour lesquels  $-\epsilon^* \leq f(t, f) - C_z^\phi(t, f) \leq \epsilon$  ont un coefficient  $\alpha$  nul et ne contribuent pas à former l'approximation.

La procédure de calcul des  $\alpha_i$  est la suivante. Dans un premier temps, la RTF de noyau  $\phi$  est calculée, puis la matrice  $\mathbf{K}$  est calculée. Enfin, le problème d'optimisation 16 est résolu numériquement. Les  $\alpha_i$  significativement non nuls sont conservés et servent de base à l'approximation. Enfin, le paramètre  $b$  est calculé grâce aux conditions de Karush-Kuhn-Tucker, voir [2].

L'approche proposée ici montre de nombreux avantages. Tout d'abord, elle est d'une grande simplicité algorithmique. Ensuite, l'influence des paramètres de réglage est facile à interpréter, et ils sont faciles à régler. Enfin, le coût calculatoire peut être très réduit lorsqu'on utilise une procédure d'optimisation adaptée. Il faut noter toutefois que cette méthode ne prend pas en compte d'*a priori* structurels pouvant être définis dans certains problèmes, contrairement à l'approche bayésienne développée dans [4], où un champ de Markov modélise les relations entre coefficients voisins.

### 4 Résultats

Les résultats présentés concernent des signaux synthétiques comportant deux chirps se croisant. Le noyau  $\phi(t, f)$  est gaussien avec  $\sigma_t^2 = \sigma_f^2 = 2$ , et le signal traité comporte  $N = 40$  points (cette longueur est relativement courte car nous avons mis en œuvre un algorithme de résolution du problème (16) non itératif, simple mais coûteux; un algorithme itératif permettrait de traiter des signaux beaucoup plus long en un temps comparable). Nous avons réglé  $C\beta = C\beta^* = 5000$ ,  $\epsilon = \epsilon^* = 0.2$ . La figure 2 représente la RTF originale, et la figure 3 présente son approximation. On constate une excellente

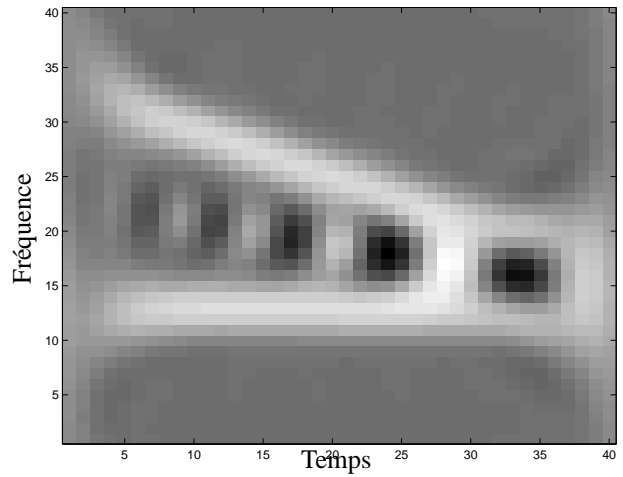


FIG. 2: Représentation temps-fréquence de deux chirps calculée pour un noyau gaussien.

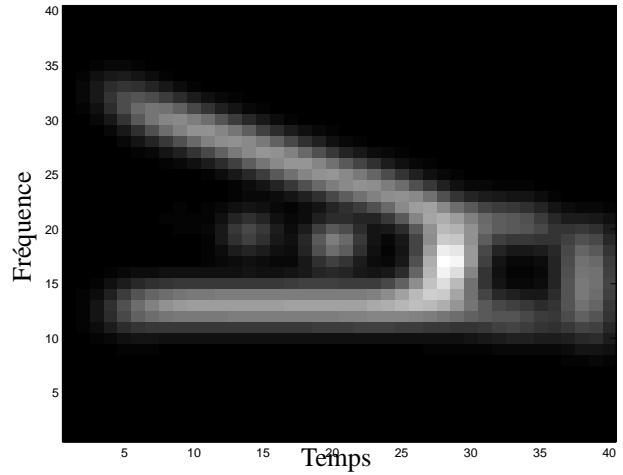


FIG. 3: Représentation temps-fréquence approchée de la RTF de la figure 2 obtenue par la méthode proposée.

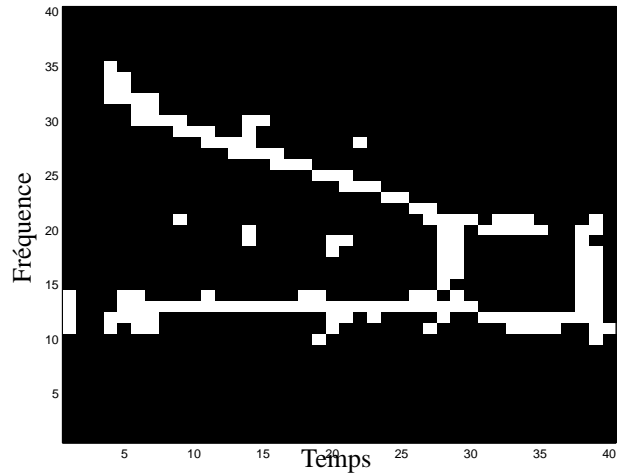


FIG. 4: Positions temps-fréquence des coefficients  $\alpha_i$  significativement non nuls et utilisés pour calculer la RTF approchée de la figure 3 obtenue par la méthode proposée.

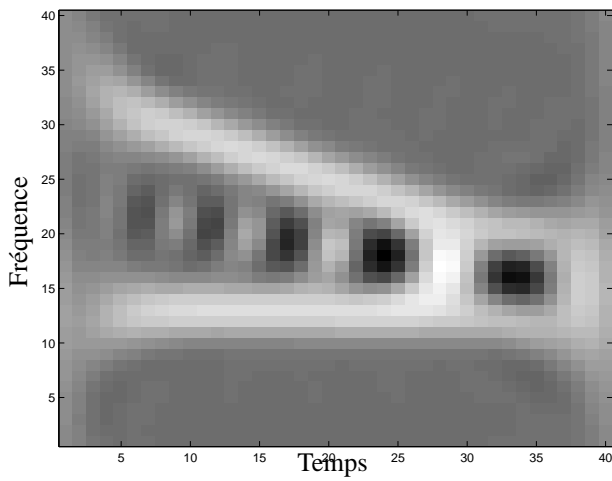


FIG. 5: Représentation temps-fréquence de deux chirps calculée pour un noyau gaussien.

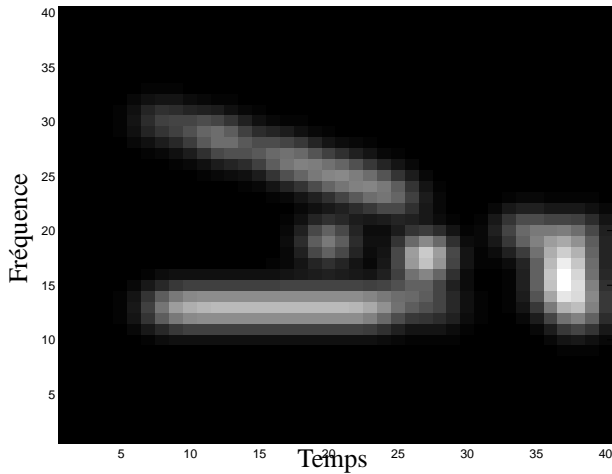


FIG. 6: Représentation temps-fréquence approchée de la RTF de la figure 2 obtenue par la méthode proposée. La régularisation est plus forte qu'à la figure 3.

adéquation. Enfin, la figure 4 montre les points temps-fréquence où les coefficients  $\alpha_i$  sont non nuls. Ils représentent 9.75% du nombre total des points temps-fréquence.

Les figures 5, 6 et 7 ci-dessous montrent le résultat obtenu lorsque  $C\beta = C\beta^* = 500$ , c'est-à-dire avec une régularisation plus importante, les autres paramètres étant inchangée. Les coefficients  $\alpha_i$  non nuls ne représentent plus que 2.62% du nombre total de coefficients temps-fréquence.

## 5 Conclusions

La méthode présentée permet de réduire significativement l'information temps-fréquence par un problème d'optimisation bien posé et sans difficultés numériques. Elle permet une compression de l'information pouvant atteindre plus de 97% sans trop dénaturer l'information temps-fréquence. Par ailleurs, les coefficients  $\alpha_i$  étant positifs, ils peuvent être interprétés comme une RTF épurée de l'énergie du signal qui pourra être employée dans des problèmes de décision.

Un algorithme itératif spécifique va maintenant être développé

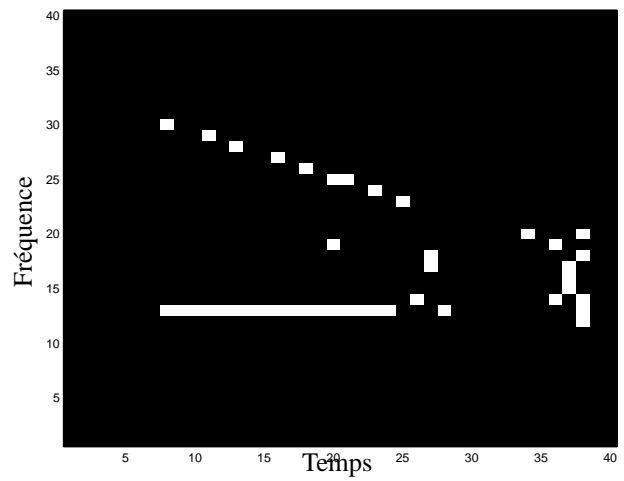


FIG. 7: Positions temps-fréquence des coefficients  $\alpha_i$  significativement non nuls et utilisés pour calculer la RTF approchée de la figure 6 obtenue par la méthode proposée. La régularisation est plus forte qu'à la figure 4

et la méthode pourra être appliquée à la classification temps-fréquence de signaux et à la compression audio.

## Références

- [1] P. Flandrin, *Temps-Fréquence*, Série Traitement du signal. Hermès, 1993.
- [2] B. Schölkopf and A. Smola, *Learning with Kernels*, MIT press, Cambridge, MA, 2002.
- [3] R. J. Vanderbei, "Loqo: An interior point code for quadratic programming", Tech. Rep. TR SOR-94-15, Department of Civil Engineering and Operations Research, Princeton University, 1995.
- [4] P. J. Wolfe, S. J. Godsill, and W.J. Ng, "Bayesian variable selection and regularisation for time-frequency surface estimation", *Journal of the Royal Statistical Society: Series B*, vol. 66, no. 3, pp. 575–589, Aug. 2004.