



# Apprentissage de Dictionnaires Multivariés et Décomposition Parcimonieuse Invariante par Translation et par Rotation 2D

Quentin BARTHÉLEMY<sup>1</sup>, Anthony LARUE<sup>1</sup>, Aurélien MAYOUE<sup>1</sup>, David MERCIER<sup>2</sup>, Jérôme I. MARS<sup>3</sup>

<sup>1</sup>CEA, LIST, Laboratoire d'Outils pour l'Analyse de Données

<sup>2</sup>CEA, LIST, Laboratoire Information, Modèle et Apprentissage  
Gif-sur-Yvette, F-91191, France

<sup>3</sup>GIPSA-lab, DIS, UMR 516 CNRS, Grenoble INP  
Grenoble, F-38402, France

prénom.nom@{cea.fr, gipsa-lab.grenoble-inp.fr}

**Résumé** – Cet article présente le Multivariate Dictionary Learning Algorithm qui apprend en ligne les structures élémentaires associées à un ensemble de signaux multivariés. Une fois apprises, le Multivariate Orthogonal Matching Pursuit décompose tous les signaux de cet ensemble avec parcimonie. Ces méthodes sont spécifiées dans le cas d'invariance par rotation qui produit un dictionnaire restreint de noyaux. Nos méthodes sont appliquées à des données d'écriture manuscrite, afin d'extraire les motifs caractéristiques de cette base de signaux.

**Abstract** – This article presents a new tool, Multivariate Dictionary Learning Algorithm, able to learn online the elementary structures associated to a multivariate signals set. Once learned, Multivariate Orthogonal Matching Pursuit codes sparsely all signals of this set. These methods are specified to the 2D rotation-invariant case which induces a small-sized kernels dictionary. Our methods are applied to 2D handwritten data to extract the characteristic patterns of this signals set.

## 1 Introduction

La décomposition d'un signal  $y \in \mathbb{C}^N$  sur un dictionnaire normé  $\Phi \in \mathbb{C}^{N \times M}$  s'écrit :  $y = \Phi x + \epsilon$ , avec  $x \in \mathbb{C}^M$  les coefficients et  $\epsilon \in \mathbb{C}^N$  le résidu. Le dictionnaire composé d'atomes  $\{\phi_m\}_{m=1}^M$  est dit redondant quand  $M > N$ . L'approximation parcimonieuse s'écrit :  $\min_x \|y - \Phi x\|^2$  t. q.  $\|x\|_0 \leq C(P)$ , avec  $C$  une constante, et  $\|x\|_0$  la pseudo-norme  $\ell_0$  définie comme le cardinal du support<sup>1</sup> de  $x$ . Afin de résoudre  $(P)$  au mieux, nous cherchons le dictionnaire  $\Phi$  adapté à la base des signaux étudiés : cela signifie qu'il permet de décomposer avec parcimonie tous les signaux de cette base. Les méthodes d'apprentissage de dictionnaire font émerger des données les motifs élémentaires caractéristiques des signaux considérés [1, 2].

Après avoir présenté les algorithmes de décomposition et d'apprentissage, nous nous placerons dans les cas multivarié et invariant par translation. Nous exposerons nos méthodes et leurs spécifications au cas d'invariance par rotation 2D. Elles seront appliquées à des signaux d'écriture manuscrite.

## 2 Etat de l'art

Les algorithmes de Poursuite [3], notamment l'Orthogonal Matching Pursuit (OMP) [4], résolvent séquentiellement  $(P)$ , mais cette optimisation n'est pas convexe. La solution peut être

un minimum local, surtout pour une forte cohérence<sup>2</sup>  $\mu_\Phi$ . Cependant, ces algorithmes sont très rapides quand on cherche juste quelques atomes. Une autre approche consiste à relaxer ce problème par une norme  $\ell_1$ . Différents algorithmes résolvent ce problème d'optimisation convexe et sont présentés dans [3]. Une forte cohérence  $\mu_\Phi$  ne garantit pas de retrouver le support de  $x$  optimal, et si c'est le cas, la convergence est longue.

Les algorithmes d'apprentissage de dictionnaire (Dictionary Learning Algorithms : DLA) trouvent un dictionnaire adapté aux signaux que l'on veut décomposer. Les méthodes classiques alternent entre 2 étapes : i)  $\Phi$  est fixé et  $x$  est obtenu par décomposition parcimonieuse, ii)  $x$  est fixé et  $\Phi$  est mis à jour. Cette mise à jour est basée sur des critères tels que le Maximum de Vraisemblance (MV) [1]. Il existe aussi des méthodes d'apprentissage où  $x$  et  $\Phi$  sont mis à jour simultanément comme la K-SVD présentée en [2] dont la bibliographie résume bien l'état de l'art.

## 3 Cas multivarié et invariant par translation

Dans le cas multivarié, le signal étudié devient  $y \in \mathbb{C}^{N \times V}$ , avec  $V$  le nombre de composantes. Deux problèmes peuvent être envisagés en fonction des natures de  $\Phi$  et de  $x$  :

1. Le support de  $x$  est  $\text{support}(x) = \{m \in \mathbb{N}_M : x_m \neq 0\}$ .

2. La cohérence du dictionnaire  $\Phi$  est  $\mu_\Phi = \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$ .

- $\Phi \in \mathbb{C}^{N \times M}$  univarié et  $x \in \mathbb{C}^{M \times V}$  multivarié, le cas *multicanal* traité par les algorithmes [5, 6, 7];
- $\Phi \in \mathbb{C}^{N \times M \times V}$  multivarié et  $x \in \mathbb{C}^M$  univarié<sup>3</sup>, le cas *multivarié* seulement évoqué dans [8], mais avec une structure de dictionnaire particulière. Dans ce travail, nous étudions ce cas, avec  $\Phi$  multivarié et considéré normé.

Dans le cas invariant par translation, nous voulons décomposer le signal  $y$  comme une somme de quelques structures  $\{\psi_l\}_{l=1}^L$ , appelées noyaux, caractérisées indépendamment de leurs positions. Considérant le noyau  $\psi_l$ ,  $\sigma_l$  est un sous-ensemble des  $N$  indices  $t$ . Translatés à toutes les positions  $\tau \in \sigma_l$ , les noyaux  $\psi_l(t)$  engendrent tous les atomes  $\phi_m(t)$  :

$$y(t) = \sum_{m=1}^M x_m \phi_m(t) + \epsilon(t) = \sum_{l=1}^L \sum_{\tau \in \sigma_l} x_{l,\tau} \psi_l(t-\tau) + \epsilon(t) \quad (1)$$

Ainsi, le signal multivarié  $y$  est approché comme la somme pondérée de quelques noyaux multivariés  $\psi_l$ .

## 4 Présentation des méthodes

Nous exposons nos méthodes introduites dans [9] : le Multivariate OMP (M-OMP) pour la décomposition parcimonieuse, le Multivariate DLA (M-DLA) pour l'apprentissage de dictionnaire et leurs spécifications au cas d'invariance par rotation 2D.

### 4.1 Multivariate OMP

La décomposition parcimonieuse peut être effectuée par n'importe quel algorithme capable de surmonter la forte cohérence due au cas d'invariance par translation. Nous choisissons l'OMP pour sa rapidité ; une description plus précise est donnée dans [4]. Nous appelons Multivariate OMP (Algorithm 1) son extension au cas multivarié décrit précédemment (Section 3). A l'étape 4, le signal  $y$  (*resp.* dictionnaire  $D$ ) multivarié est déplié selon la dimension des composantes dans un vecteur  $y$  (*resp.* matrice  $D$ ) univarié. La projection orthogonale (étape 4) est souvent faite récursivement par différentes méthodes : nous choisissons celle basée sur l'inversion matricielle par bloc [4]. Notant  $k$  l'itération courante, le vecteur de coefficients obtenu  $x^k$  est réduit à ses éléments non nuls. Comparé à l'OMP original, la complexité du M-OMP est seulement multipliée par  $V$  le nombre de composantes.

### 4.2 Multivariate DLA

Notre méthode d'apprentissage, appelée Multivariate DLA (Algorithm 2), est l'alternance de deux étapes : une décomposition faite par M-OMP et une mise à jour du dictionnaire. Cette méthode est appliquée à un ensemble de signaux d'entraînement  $Y = \{y_p\}_{p=1}^P$ . Notre mise à jour est basée sur le

3. Ici,  $\Phi x$  est un produit élément par élément selon la dimension  $M$ .

4. La corrélation complexe entre les  $u$ èmes composantes des signaux multivariés  $a$  et  $b$  est notée  $\Gamma \{a[u], b[u]\}$ .

---

#### Algorithm 1: $x = \text{Multivariate\_OMP}(y, \Psi)$

---

Initialisation :  $k = 1$ ,  $\epsilon^0 = y$ , dictionnaire  $D^0 = \emptyset$

**repeat**

1. Corrélations<sup>4</sup> : **for**  $l \leftarrow 1$  **to**  $L$  **do**

$C_l^k(\tau) \leftarrow \sum_{u=1}^V \Gamma \{ \epsilon^{k-1}[u], \psi_l[u] \}(\tau)$

2. Sélection :  $(l_{max}^k, \tau_{max}^k) \leftarrow \arg \max_{l,\tau} |C_l^k(\tau)|$

3. Dictionnaire actif :  $D^k \leftarrow D^{k-1} \cup \psi_{l_{max}^k}(t - \tau_{max}^k)$

4. Coefficients :  $x^k \leftarrow \arg \min_x \|y - D^k x\|^2$

5. Résidu :  $\epsilon^k \leftarrow y - D^k x^k$

6.  $k \leftarrow k + 1$

**until** *convergence*

---

critère MV [1], habituellement optimisé par Descente de Gradient. Nous préférons le Gradient du 2<sup>nd</sup> ordre de Levenberg-Marquardt [10] qui accélère la convergence, en couplant les méthodes du Gradient Stochastique et de Gauss-Newton. A l'itération  $i$ , la mise à jour de chaque noyau multivarié  $\psi_l$  est :

$$\psi_l^i(\underline{t}) = \psi_l^{i-1}(\underline{t}) + (H + \lambda I)^{-1} \cdot \sum_{\tau \in \sigma_l} x_{l,\tau}^{i*} \epsilon^{i-1}(\underline{t} + \tau) \quad (2)$$

avec  $\underline{t}$  les indices restreints au support temporel de  $\psi_l$ ,  $\lambda$  le pas adaptatif,  $*$  l'opérateur conjugué et  $H$  le hessien moyen calculé pour chaque noyau (et non pour chaque échantillon). Dans les cas de recouvrements de noyaux, la méthode d'apprentissage devient instable à cause de l'erreur faite sur l'estimation du gradient. Nous surestimons légèrement le hessien  $H$  pour compenser ce phénomène. L'étape de mise à jour, qui stabilise désormais l'apprentissage, est appelée LM\_modif (étape 2).

La convergence du M-DLA vers le minimum global n'est pas garantie à cause de l'optimisation alternée. Cependant le dictionnaire obtenu, minimum local ou global, assure la parcimonie des décompositions.

---

#### Algorithm 2: $\Psi = \text{Multivariate\_DLA}(\{y_p\}_{p=1}^P)$

---

Initialisation :  $i = 1$ ,  $\Psi^0 = \{L \text{ noyaux de bruit blanc}\}$

**repeat**

**for**  $p \leftarrow 1$  **to**  $P$  **do**

    1. Décomposition :  $x^i \leftarrow \text{M\_OMP}(y_p, \Psi^{i-1})$

    2. Mise à jour :  $\Psi^i \leftarrow \text{LM\_modif}(y_p, x^i, \Psi^{i-1})$

    3.  $i \leftarrow i + 1$

**until** *convergence*

---

### 4.3 Cas d'Invariance par Rotation 2D (2DRI)

Pour l'étude de signaux réels bivariés, des mouvements 2D par exemple, nous voulons les caractériser indépendamment de leurs orientations. L'invariance par rotation introduit une matrice  $R$  de rotation d'angle  $\theta_{l,\tau}$  pour chaque noyau bivarié  $\psi_l(t - \tau)$ . L'équation (1) devient :

$$y(t) = \sum_{l=1}^L \sum_{\tau \in \sigma_l} x_{l,\tau} R(\theta_{l,\tau}) \psi_l(t - \tau) + \epsilon(t) \quad (3)$$

Dans l'étape de sélection des noyaux (Algorithm 1, étape 2), le but est désormais de trouver l'angle  $\theta_{l^k, \tau_{max}^k}$  qui maximise les corrélations  $|C_l^k(\tau, \theta_{l, \tau})|$ . Une approche naïve consiste à échantillonner  $\theta_{l, \tau}$  en  $\Theta$  angles et d'ajouter un degré de liberté dans le calcul des corrélations (Algorithm 1, étape 1). La complexité est multipliée par  $\Theta$  par rapport au M-OMP utilisé dans le cas réel bivarié.

Pour éviter ce surcoût, nous transformons le signal  $y$  de  $\mathbb{R}^{N \times 2}$  à  $\mathbb{C}^N$  (i.e.  $y \leftarrow y[1] + iy[2]$ ), puis nous appliquons le M-OMP : les noyaux et les coefficients de décomposition  $x$  sont maintenant complexes. Le module donne l'amplitude du coefficient et l'argument l'angle de rotation du noyau. Désormais susceptible de tourner, celui-ci n'est plus appris selon une orientation particulière comme dans l'approche précédente qualifiée d'*orientée* ( $V=2, y \in \mathbb{R}^{N \times 2}$ ). Ainsi, les noyaux sont invariants par rotation et par translation générant une décomposition *non-orientée* ( $V=1, y \in \mathbb{C}^N$ ). Cette spécification de la méthode de décomposition (resp. apprentissage) est appelée 2DRI-OMP (resp. 2DRI-DLA).

## 5 Données d'application

Nos méthodes sont appliquées aux signaux *Character Trajectories* disponibles sur la base UCI [11] et initialement traités avec un modèle probabiliste et une méthode d'apprentissage EM [12], mais sans parcimonie réelle dans les décompositions obtenues. Les données sont composées d'une centaine d'occurrences de 20 lettres écrites par la même personne. Les signaux sont les vitesses cartésiennes  $v_x$  et  $v_y$  de la pointe du stylo.

Nous souhaitons apprendre un dictionnaire adapté afin de décomposer avec parcimonie ces signaux de vitesse. Le dictionnaire est appris sur les 20 premières occurrences de chaque lettre et la décomposition est testée sur celles restantes. Les vitesses, sur lesquelles nos méthodes sont appliquées, sont intégrées seulement pour visualiser les trajectoires associées.

## 6 Résultats

Les résultats sont directement présentés en cas non-orienté. Le dictionnaire des noyaux intégrés (Fig 1) montre que le 2DRI-

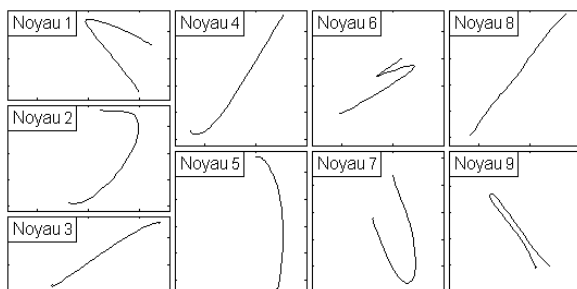


FIGURE 1 – Dictionnaire, appris par 2DRI-DLA, des trajectoires associées aux noyaux.

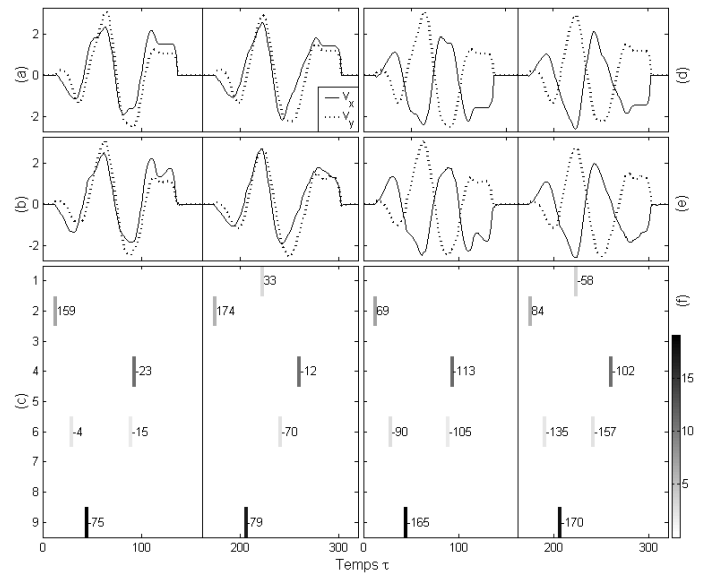


FIGURE 2 – Signaux de vitesse originaux (a) (resp. tournés de  $-90^\circ$ ) (d) et reconstruits (b) (resp. (e)) de 2 occurrences de la lettre *d*, ainsi que leurs spikegrammes associés (c) (resp. (f)).

DLA a extrait des primitives du mouvement. En effet, les traits et les courbes correspondent aux motifs élémentaires de l'ensemble des signaux d'écriture manuscrite.

Pour évaluer les qualités de ce dictionnaire, nous analysons les décompositions de quatre occurrences de la lettre *d* sur ce dictionnaire (Fig 2). Le signal de vitesse original (Fig 2a) (resp. tourné de  $-90^\circ$ ) (Fig 2d), composé des parties réelle  $v_x$  (trait plein) et imaginaire  $v_y$  (pointillé), est reconstruit (Fig 2b) (resp. (Fig 2e)). L'erreur quadratique moyenne (EQM) de reconstruction sur les vitesses est de l'ordre de 12% pour 5 noyaux. Les coefficients  $x_{l, \tau}$  sont visualisés par une représentation temps-noyau appelée spikegramme (Fig 2c) (resp. Fig 2f). Elle condense quatre informations : la position temporelle du noyau  $\tau$  (abscisse), l'indice du noyau  $l$  (ordonnée), l'amplitude du coefficient  $|x_{l, \tau}|$  (niveau de gris) et l'angle de rotation  $\theta_{l, \tau}$  (nombre à côté du spike, en degré). Le faible nombre de noyaux utilisés pour la reconstruction montre la parcimonie de la décomposition. Les noyaux primaires sont ceux de fortes amplitudes : ils concentrent l'information. Les noyaux secondaires codent les variabilités entre les différentes réalisations. Sur les spikegrammes (Fig 2c et 2f), la différence de  $90^\circ$  entre les angles des noyaux primaires montre l'invariance par rotation de la décomposition. Cette reproductibilité (valeurs d'amplitudes et d'angles des noyaux primaires) est la preuve d'un dictionnaire adapté.

La trajectoire de la lettre *d* originale (Fig 3a) (resp. tournée de  $-90^\circ$ ) (Fig 3d) est reconstruite avec ses noyaux primaires, en comparant les cas orienté (Fig 3b) (resp. (Fig 3e)) et non-orienté (Fig 3c) (resp. (Fig 3f)). Par exemple, la lettre (Fig 3c) est reconstruite comme la somme des noyaux 2, 4 et 9 (Fig 1) spécifiés par les amplitudes et les angles du spikegramme (Fig 2c). Quand les lettres sont tournées, dans le cas orienté, les

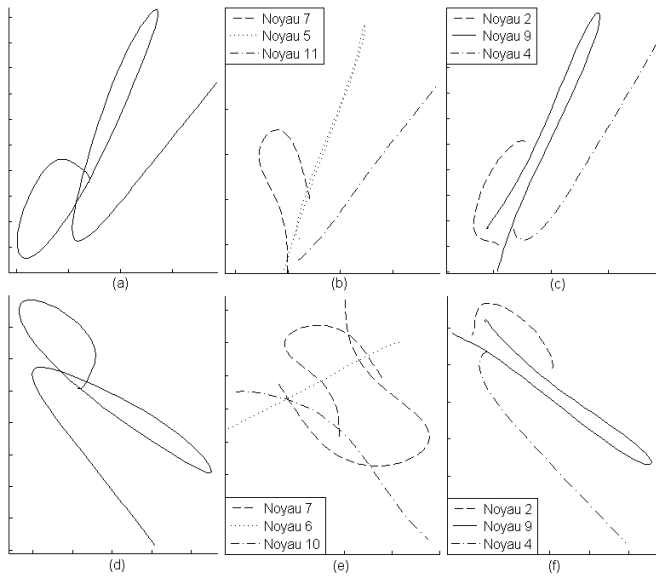


FIGURE 3 – Trajectoires de la lettre *d* originale (a) (*resp.* tournée de  $-90^\circ$  (d)), reconstruction orientée (b) (*resp.* (e)) et reconstruction non-orientée (c) (*resp.* (f)).

noyaux sélectionnés différent et l'EQM passe de 15% (Fig 3b) à 30% (Fig 3e). Dans le cas non-orienté, l'EQM est identique quel que soit l'angle de rotation (Fig 3c et 3f) et est généralement inférieure au cas orienté. La différence entre les reconstructions (Fig 3e et 3f) montre l'invariance par rotation de la décomposition non-orientée.

## 7 Discussion

L'apprentissage de dictionnaire permet de retrouver les primitives des signaux étudiés : les décompositions basées sur un tel dictionnaire sont faites sur les structures caractéristiques de la base de ces signaux. Considérant l'erreur de reconstruction, le faible nombre de noyaux utilisés montre l'efficacité de cette méthode d'approximation parcimonieuse.

L'approche non-orientée réduit la taille du dictionnaire de noyaux dans deux cas :

- quand les signaux étudiés peuvent tourner, comme l'application présentée (la tablette d'acquisition est tournée par exemple). Pour générer un dictionnaire adapté, l'approche orientée a besoin d'apprendre des primitives du mouvement pour chacune des directions possibles, contrairement au cas non-orienté. Ceci entraîne une réduction remarquable de la taille du dictionnaire.
- quand les signaux ne sont pas tournés. L'approche non-orientée détecte les invariants rotationnels ce qui réduit aussi la taille du dictionnaire [9].

Ainsi, l'invariance par translation et par rotation génère un dictionnaire  $\Psi$  de petite taille (Fig 1). De plus, elle permet d'être robuste à la direction de l'écriture (rotation de la tablette) et à son inclinaison (variabilités intra et inter utilisateurs).

## 8 Conclusion

Nous avons présenté des nouveaux outils pour les signaux multivariés. Le M-DLA apprend en ligne un dictionnaire composé des structures élémentaires associées à une base de signaux multivariés. Une fois le dictionnaire appris, le M-OMP décompose avec parcimonie tous les signaux de cette base. Ces algorithmes sont spécifiés au cas d'invariance par rotation 2D, respectivement nommé 2DRI-DLA et 2DRI-OMP. L'invariance par translation et par rotation produit un dictionnaire de noyaux de petite taille. Les applications sont la réduction de dimension, la compression, le débruitage, la représentation et l'analyse de gestes et tout autre traitement basé sur l'extraction de caractéristiques multivariées.

Les perspectives envisagées sont de se comparer aux autres algorithmes d'apprentissage et d'étendre les méthodes présentées à l'invariance par rotation 3D.

## Références

- [1] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set : a strategy employed by V1 ? *Vision Research*, 37 :3311–3325, 1997.
- [2] M. Aharon, M. Elad, and A. Bruckstein. K-SVD : an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 54 :4311–4322, 2006.
- [3] J.A. Tropp and S.J. Wright. Computational methods for sparse solution of linear inverse problems. *Proc. of the IEEE*, 98 :948–958, 2010.
- [4] Y.C. Pati, R. Rezaifar, and P.S. Krishnaprasad. Orthogonal Matching Pursuit : recursive function approximation with applications to wavelet decomposition. In *Asilomar Conf. on Signals, Systems and Comput.*, 1993.
- [5] A. Lutoborski and V.N. Temlyakov. Vector greedy algorithms. *J. Complex.*, 19 :458–473, 2003.
- [6] J.A. Tropp, A.C. Gilbert, and M.J. Strauss. Algorithms for simultaneous sparse approximation ; Part I : Greedy pursuit. *Signal Proc.*, 86 :572–588, 2006.
- [7] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. on Signal Processing*, 54 :4634–4643, 2006.
- [8] R. Gribonval and M. Nielsen. Beyond sparsity : Recovering structured representations by  $\ell_1$ -minimization and greedy algorithms. Technical Report PI-1684, 2005.
- [9] Q. Barthélemy, A. Larue, A. Mayoue, D. Mercier, and J.I. Mars. Multivariate Dictionary Learning and Shift & 2D Rotation Invariant Sparse Coding. In *IEEE Workshop on Statistical Signal Processing*, 2011, Accepted.
- [10] K. Madsen, H.B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems, 2nd ed. Technical report, 2004.
- [11] A. Frank and A. Asuncion. UCI machine learning repository. <http://archive.ics.uci.edu/ml>, 2010.
- [12] B.H. Williams, M. Toussaint, and A.J. Storkey. A primitive based generative model to infer timing information in unpartitioned handwriting data. In *Int. Joint Conf. on Artificial Intelligence IJCAI*, pages 1119–1124, 2007.