Modèles de mélange de Gaussiennes temps-fréquence pour le débruitage de signaux

Emilie Villaron, Sandrine Anthoine, Bruno Torrésani

▶ **To cite this version:**

Emilie Villaron, Sandrine Anthoine, Bruno Torrésani. Modèles de mélange de Gaussiennes temps-fréquence pour le débruitage de signaux. GRETSI 2011, 2011, Bordeaux, France. pp.ID401. hal-00813604

HAL Id: hal-00813604
https://hal.science/hal-00813604

Submitted on 16 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèles de mélange de Gaussiennes temps-fréquence pour le débruitage de signaux

Emilie VILLARON, Sandrine ANTHOINE, Bruno TORRÉSANI

Université de Provence - CNRS
Laboratoire d'Analyse, Topologie et Probabilités, UMR CNRS 6632
39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
Emilie.Villaron@cmi.univ-mrs.fr, Sandrine.Anthoine@cmi.univ-mrs.fr,
Bruno.Torresani@cmi.univ-mrs.fr

Résumé – On considère le problème de débruitage d'un signal engendré par des coefficients (de synthèse) dans un dictionnaire temps-fréquence, modélisés par un mélange de Gaussiennes multivariées. L'existence de corrélations à la fois dans les coefficients de synthèse et entre les éléments du dictionnaire complique singulièrement le problème de débruitage. Un algorithme de type "Majoration-Minimisation" adapté au problème posé est développé et étudié, une preuve de convergence est donnée. Cet algorithme requiert en toute généralité l'inversion de matrices de grande dimension; des cas particuliers sont considérés (dictionnaires temps-fréquence invariants par translation, et covariances possédant une structure compatible) dans lesquels la complexité peut être considérablement réduite. L'inversion se ramène à une inversion de matrices bloc-circulantes, qui peut être effectuée efficacement. Les performances de l'algorithme sont illustrées par des simulations numériques.

Abstract – In this paper, we tackle the problem of denoising a signal generated by (synthesis) coefficients on a redundant time-frequency dictionary, in the particular case where the coefficients follow a multivariate mixture of Gaussians model. The crux of the problem is that we have to deal with correlations both between the synthesis coefficients and between the dictionary atoms. We propose a "Majorization-Minimization" algorithm adapted to this problem and prove its convergence. In principle, our algorithm requires the inversion of a matrix which dimensions may be large. To solve efficiently the problem, we exhibit particular cases where the dictionary and covariance matrix have a compatible structure. In these cases, we only need to invert bloc-circulant matrices, which can be done efficiently. Numerical simulations illustrate the performances of our algorithm.

1 Introduction

Les algorithmes de débruitage de signal basés sur des décompositions sur un dictionnaire supposent généralement la décorrélation des coefficients de la décomposition. Cette hypothèse, bien qu'elle fournisse des résultats souvent intéressants, n'est parfois pas réaliste dans certaines applications. On peut par exemple penser à des situations où les signaux observés sont le résultat de mesures après propagation dans un milieu complexe, pour lequel les corrélations induites par la propagation sont mal connues.

Motivés entre autres par des applications à l'analyse de signaux neurophysiologiques et aux interfaces cerveau-machine, nous nous intéressons ici à des modèles dans lesquels les coefficients de synthèse sont modélisés par des lois de mélange (modélisant par exemple différents "états cérébraux"), et nous considérons le problème de débruitage dans ce contexte. Ce problème peut être abordé via une formulation variationnelle, pour laquelle il est nécessaire de recourir à un algorithme adapté. Nous développons ici une approche basée sur le principe de *majoration-minimisation*, pour lequel nous donnons une preuve de convergence dans le cas de mélanges de Gaussiennes multivariées, de structure de covariance connue.

L'algorithme requiert à chaque itération une inversion matricielle, qui peut être de grande dimension. Nous montrons qu'en nous limitant à une classe spécifique de dictionnaires possédant des propriétés d'invariance par translation, les matrices considérées héritent d'une structure particulière (circulante par blocs) qui peut réduire énormément la complexité du problème. Parmi les exemples de dictionnaires de cette classe, on peut notamment mentionner les dictionnaires temps-fréquence (repères de Gabor) ainsi que des dictionnaires d'ondelettes invariants par translation. Notons que l'hypothèse d'invariance par translation est naturelle dans bon nombre de problèmes et d'applications, adaptée notamment aux situations où des hypothèses de stationnarité peuvent être faites sur le signal. Dans ce contexte nous considérons deux situations différentes. Dans la première, une distribution a priori globale est affectée à l'ensemble des coefficients, alors que dans la seconde, les coefficients sont regroupés en familles décorrélées.

Cet article est organisé comme suit. La problématique et la modélisation sont décrites dans la section 2.1, et l'algorithme d'estimation est discuté en Section 2.2. Le cas particulier des dictionnaires invariants par translation est considéré en Section 3, et des résultats numériques sont discutés en Section 4.

2 Position du problème et algorithme

2.1 Débruitage avec *a priori* mélange de gaussiennes

Nous nous plaçons dans le cadre de signaux discrets de longueur finie R, et nous nous limitons pour simplifier au cas réel. Dans $\mathcal{H}=\mathbb{R}^R$, on considère un dictionnaire $\mathcal{D}=\{\phi_0,...\phi_{S-1}\}$ représenté par sa matrice $\Phi \in \mathbb{R}^{R\times S}$ (avec $S \geq R$). On suppose disposer d'observations $s \in \mathbb{R}^R$ de la forme

$$s = \Phi y + b \tag{1}$$

où $b \in \mathbb{R}^R$ est un bruit blanc Gaussien de variance σ_0^2, et où $y \in \mathbb{R}^S$ est un vecteur aléatoire, suivant une loi de mélange de Gaussiennes multivariées, dont on supposera les moyennes nulles pour simplifier, et dont on notera $\Sigma_1,\ldots,\Sigma_K \in \mathbb{R}^{S\times S}$ les matrices de covariance et $p_1,\ldots,p_K$ les probabilités d'appartenance. La loi *a priori* de y est donc donnée par

$$p(y) = \sum_{k=1}^{K} \frac{p_k}{\sqrt{(2\pi)^S \det \Sigma_k}} \exp\left(-\frac{1}{2} y^T \Sigma_k^{-1} y\right) , \tag{2}$$

où on note y^T la transposée du vecteur y. La maximisation de la log-vraisemblance associée à ce modèle et une observation s conduit à un problème d'optimisation

$$\hat{y} = \arg\min_{y\in\mathbb{R}^S} \left(\frac{1}{2\sigma_0^2}\|s-\Phi y\|_2^2 - \log(p(y))\right) = \arg\min_{y\in\mathbb{R}^S} L(y) , \tag{3}$$

qui doit être résolu numériquement. L'existence de corrélations entre coefficients de synthèse ne permet pas d'utiliser directement les approches usuelles pour ce type de problème.

Remarque 1 *Il est possible de traiter de façon tout à fait similaire le cas de bruits Gaussiens quelconques, notamment stationnaires en moyenne quadratique, en modifiant le terme d'attache aux données. On se limite ici au cas de bruit blanc pour simplifier les expressions.*

2.2 Algorithme d'estimation

Pour la résolution numérique du problème (3), nous développons une approche "MM" (pour Majoration-Minimisation, voir par exemple [3, 4]) basée sur la minimisation d'un majorant de la pénalité $L(y)$, obtenu comme suit. On note $\mathbf{I}_S$ la matrice identité de $\mathbb{R}^S$, et

$$C(y) = -\log(p(y)) , \qquad \mathbf{A} = \frac{1}{2}\sum_{k=1}^{K} p_k \Sigma_k^{-1} . \tag{4}$$

et on cherche un majorant quadratique f_t de C tangent à C au point d'estimation courante y_t. Une application directe de l'inégalité de Jensen nous donne :

$$\begin{aligned} C(y) &\leq -\sum_{k=1}^{K} p_k \log\left(\frac{\exp(-\frac{1}{2}y^T\Sigma_k^{-1}y)}{\sqrt{(2\pi)^S\det\Sigma_k}}\right) \\ &\leq y^T\mathbf{A}y + K. \end{aligned}$$

On en déduit la majoration suivante :

Proposition 1 *Pour tout $y_t \in \mathbb{R}^S$ donné, notons f_t la fonction définie par*

$$\begin{aligned} f_t(y) &= y^T(\mathbf{A}+\lambda\mathbf{I}_S)y + \langle V_t, y\rangle + c , \qquad \text{où} \\ V_t &= \nabla C(y_t) - 2(\mathbf{A}+\lambda\mathbf{I}_S)y_t , \end{aligned}$$

et où $\lambda \in \mathbb{R}^+$. La constante c peut être choisie de sorte que

$$\left\{\begin{array}{rcl} f_t(y) & \geq & y^T\mathbf{A}y + K_0 \quad \forall y \in \mathbb{R}^T \\ \nabla f_t(y_t) & = & \nabla C(y_t) \end{array}\right.$$

où K_0 est une constante.

Avec K_0 bien choisi, f_t est un majorant quadratique de C et est tangent à C en y_t.

Ainsi, pour $y_t \in \mathbb{R}^S$ donné, $L(y)$ est majorée par la forme quadratique Q_t tangente à L en y_t définie par

$$Q_t(y) = \frac{1}{2\sigma_0^2}\|s-\Phi y\|^2 + f_t(y) , \tag{5}$$

dont le point critique est obtenu en résolvant

$$\mathbf{B}y = \left(\frac{1}{\sigma_0^2}\Phi^T s - V_t\right) , \qquad \text{où} \tag{6}$$

$$\mathbf{B} = \left(\frac{1}{\sigma_0^2}\Phi^T\Phi + 2(\mathbf{A}+\lambda\mathbf{I}_S)\right) , \tag{7}$$

Φ^T étant la transposée de Φ. Comme $\mathbf{B}$ est symétrique définie positive, $\mathbf{B}$ est inversible et le point critique est un minimum de Q_t. L'algorithme de "Majoration-Minimisation" (MM) que nous proposons consiste en l'itération de ce processus de majoration par Q_t et de minimisation de la majorante Q_t.

Le résultat de la Proposition 1 permet d'utiliser les théorèmes généraux sur les algorithmes MM [4], ce qui fournit :

Corollaire 1 *L'itération définie par*

$$\begin{aligned} y_{t+1} = &\left(\frac{1}{\sigma_0^2}\Phi^T\Phi + 2(\mathbf{A}+\lambda\mathbf{I}_S)\right)^{-1} \\ &\times\left(\frac{1}{\sigma_0^2}\Phi^T s - \nabla C(y_t) + 2(\mathbf{A}+\lambda\mathbf{I}_S)y_t\right) \end{aligned}$$

converge vers un minimum (local) de L.

Il est à noter que cet algorithme demande, à chaque itération de résoudre un système linéaire impliquant la matrice $\mathbf{B} \in \mathbb{R}^{S\times S}$, qui peut être de grande dimension si S est grand ou si le dictionnaire est très redondant. L'adaptation à des situations réelles peut donc être problématique, sauf à faire des hypothèses sur la structure des matrices de covariance et du dictionnaire.

3 Invariance par translation

3.1 Dictionnaires stratifiés invariants

Nous nous focalisons maintenant sur le cas de dictionnaires spécifiques, stratifiés (c'est à dire dont les éléments sont indexés par un double indice $\lambda=(m,n)$) et invariants par translation dans le sens suivant :

Definition 1 *Un dictionnaire représenté par sa matrice Φ est invariant par translation (circulaire) si les colonnes ϕ_λ de Φ s'écrivent sous la forme*

$$\phi_\lambda[k] = \phi_{m,n}[k] = \phi_{0,n}[k-m] \ ,$$

où $m = 0, \ldots, M-1$ et $n = 0, \ldots, N-1$, $\quad S = MN$, les opérations étant à prendre modulo R.

Parmi les exemples usuels de dictionnaire invariants par translation, citons

• un dictionnaire de Gabor réel, où les atomes sont définis par

$$\phi_{m,n}[k] = \begin{cases} \cos\left(\frac{2\pi n\nu_0(k-mb_0)}{R}\right)\phi[k-mb_0] \ , & 0 \le n \le \lfloor \frac{N}{2} \rfloor \\ \sin\left(\frac{2\pi n\nu_0(k-mb_0)}{R}\right)\phi[k-mb_0] \ , & \lfloor \frac{N}{2} \rfloor < n < N, \end{cases}$$

b_0 et ν_0 définissant le réseau temps-fréquence considéré,

• plus généralement, un dictionnaire construit à partir d'un banc de filtre (par exemple ondelettes non décimées).

Il est immédiat de vérifier que la matrice de Gram associée à un dictionnaire invariant par translation est bloc-circulante :

$$\left(\Phi^T\Phi\right)_{\lambda\lambda'} = \langle \phi_{m,n}, \phi_{m',n'} \rangle = \langle \phi_{m-m',n}, \phi_{0,n'} \rangle,$$

Cette structure peut être exploitée pour peu que la matrice $\mathbf{A}$ définie en (4) possède une structure compatible. Il suffit par exemple que $\mathbf{A}$ soit bloc-diagonale, de blocs identiques et symétriques.

Proposition 2 *Supposons que le dictionnaire soit invariant par translation, et que le modèle* a priori *sur les coefficients de synthèse soit tel que, pour tout k, les matrices de covariances $\Sigma_k \in \mathbb{R}^{S\times S}$ soient bloc-diagonales, de la forme*

$$(\Sigma_k)_{\lambda,\lambda'} = (\Sigma_k)_{(m,n),(m',n')} = (S_k)_{n,n'}\,\delta_{m,m'} \ ,$$

pour certaines matrices de covariance réduites $S_k \in \mathbb{R}^{M\times M}$. Alors $\mathbf{A}$ et $\mathbf{B}$ (voir (7)*) sont bloc circulantes.*

$$\textit{d'où :} \quad \mathbf{B} = \begin{pmatrix} B_0 & B_1 & B_2 & \ldots & B_{N-1} \\ B_{N-1} & B_0 & B_1 & \ldots & B_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_1 & B_2 & \ldots & B_{N-1} & B_0 \end{pmatrix},$$

où chaque bloc B_i est une matrice de $\mathbb{R}^{M\times M}$.

Notons que cette hypothèse revient supposer les vecteurs de coefficients $\mathbf{y}_n = \{y_{m,n},\, m = 0, \ldots, M-1\}$ i.i.d.

Les matrices bloc circulantes peuvent être bloc-diagonalisées par transformation de Fourier [1, 5]. En notant F la matrice (unitaire) de Fourier et $\mathbf{F} = F \otimes \mathbf{I}_M$, $\mathbf{B}$ peut s'écrire sous la forme (voir [5] pour plus de détails)

$$\mathbf{B} = \mathbf{F}^*\mathbf{P}\mathbf{F} \ , \quad \text{où} \qquad \mathbf{P} = \mathrm{diag}(P_0, P_1, \ldots P_{N-1})$$

et où les matrices $P_k \in \mathbb{R}^{M\times M}$ ci-dessous sont inversibles

$$P_k = \sum_{n=1}^{N} \omega^{(n-1)\times k} B_k \ , \quad \omega = e^{\frac{2i\pi}{N}}$$

Ainsi la résolution du système (6) se ramène à des systèmes par blocs bien plus simples à résoudre numériquement :

$$\mathbf{P}z = x_t \ , \quad \text{où} \quad z = \mathbf{F}y \quad \text{et} \quad x_t = \mathbf{F}\left(\frac{1}{\sigma_0^2}\Phi^T s - V_t\right) \ . \quad (8)$$

Remarque 2 *Cette approche permet de gagner en efficacité, tant en termes de complexité qu'en termes de stockage. Par exemple, les matrices $\mathbf{B}$ et $\mathbf{A}$, de taille $MN\times MN$ sont stockées avec un gain d'un facteur N. De même, lors de l'itération, la mise à jour requiert N inversions de systèmes linéaires de taille M au lieu d'une inversion de taille $S = MN$.*

Remarque 3 *Nous avons déjà mentionné plus haut (voir Remarque 1) le cas de bruits Gaussiens corrélés, de matrice de covariance connue Σ_0. L'approche de la section 2.2 s'étend à cette situation sans difficulté, et conduit à l'itération*

$$\begin{aligned} y_{t+1} = \ & \left(\Phi^T\Sigma_0^{-1}\Phi + 2(\mathbf{A} + \lambda\mathbf{I}_S)\right)^{-1} \\ & \times \left(\Phi^T\Sigma_0^{-1}s - \nabla C(y_t) + 2(\mathbf{A} + \lambda\mathbf{I}_S)y_t\right) \end{aligned}$$

Cependant, des conditions supplémentaires (d'invariance par translation de nouveau) doivent être imposées au bruit pour que la structure utilisée dans la présente section soit conservée. Sans entrer dans les détails, notons que si b est stationnaire en moyenne quadratique, les résultats de cette section sont préservées, ainsi que la validité de l'algorithme correspondant.

3.2 *A priori* mélange de gaussiennes structurées

Le modèle défini ci-dessus suppose une valeur unique $k \in \{1, \ldots K\}$ pour la variable latente. Les dictionnaires stratifiés permettent de traiter de façon similaire un modèle modifié, dans lequel une valeur de variable latente est associée à chaque groupe de coefficients. Le vecteur de coefficients y est décomposé en N sous-vecteurs de taille M, $y = (\underline{y}_1 \quad \underline{y}_2, \ldots, \underline{y}_N)'$, chaque $\underline{y}_i$ représentant les coefficients de synthèse correspondant au i-ème groupe (pour fixer les idées, la i-ème fenêtre temporelle pour un dictionnaire de Gabor). Ces vecteurs sont supposés i.i.d et distribués selon un mélange de K Gaussiennes centrées multivariées de matrices de covariances $\tilde{\Sigma}_1$ et $\tilde{\Sigma}_2$ dans $\mathbb{R}^{M\times M}$. Cet *a priori* permet de prendre en compte l'évolution temporelle des signaux, ce qui n'est pas possible avec une matrice de covariance globale. On peut alors envisager de détecter des changements d'états dans le signal. Une approche identique à celle développée ci-dessus conduit à une nouvelle majoration pour la pénalité $\tilde{C}(y)$, de la forme

$$\begin{aligned} \tilde{C}(y) \ &= \ -\sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \frac{p_k \exp\left(-\frac{1}{2}\underline{y}_n^T\tilde{\Sigma}_k^{-1}\underline{y}_n\right)}{\sqrt{(2\pi)^M \det\tilde{\Sigma}_k}}\right) \\ &\le \ y^T(\mathbf{I}_N \otimes \tilde{\mathbf{A}})y + K_2 \end{aligned}$$

où la nouvelle matrice $\tilde{\mathbf{A}}$, définie par

$$\tilde{\mathbf{A}} = \mathbf{I}_G \otimes \left(\frac{1}{2}\sum_{k=1}^{K} p_k\tilde{\Sigma}_k^{-1}\right) \quad (9)$$

est une matrice bloc-diagonale et bloc-circulante. On peut alors utiliser la méthode proposée en Section 3 pour résoudre le système (6). L'algorithme décrit en Section 2.2 s'applique alors sans autre modification.

4 Résultats numériques

L'approche proposée a été implémentée dans l'environnement logiciel MATLAB. Nous illustrons ici l'approche proposée sur des signaux simulés de longueur $R = 1000$, générés sur un dictionnaire de Gabor réel avec pour paramètres $N = 100$ valeurs d'échantillonnage temporel et $M = 20$ valeurs de la fréquence par fenêtre temporelle (ainsi $S = 2000$). Nous considérons le cas du modèle structuré (défini en section 3.2), avec $K = 2$ états, associés chacun à une matrice de covariance diagonale et $[p_1, p_2] = [0.85, 0.15]$. Quelques résultats obtenus par l'algorithme décrit en Section 3.1 (utilisant la structure bloc-diagonale de la matrice $\mathbf{B}$) sont illustrés en FIGURE 1. L'évolution du critère L (FIGURE 2) montre une décroissance rapide du critère jusqu'à saturation (le test d'arrêt porte sur la différence relative entre deux estimations successives).

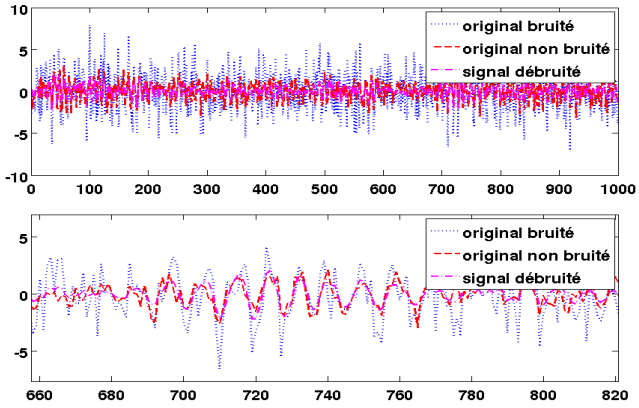

FIGURE 1 – Haut : Signal original (s), signal original non bruité (Φy) et signal débruité. Bas : zoom sur une petite partie du signal.

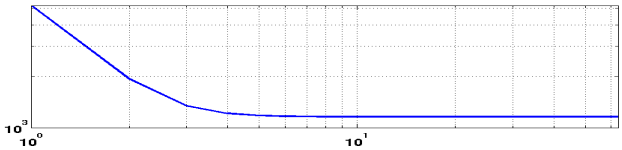

FIGURE 2 – Evolution de la fonctionnelle à minimiser en fonction du nombre d'itérations de l'algorithme (plot en log-log)

Afin de comparer les temps de calcul de la méthode globale (inversion de la matrice $\mathbf{B}$ et produit matriciel pour la mise à jour de y) et de celle exploitant la structure bloc-diagonale (8), on applique les deux méthodes sur 20 jeux de données générés avec les mêmes paramètres que précédemment. Les statistiques des deux familles de temps sont présentées dans la TABLE 1.

	Moyenne	Ecart-type
Méthode globale	351.95 s	25.54
Méthode structurée	92.45 s	1.25

TABLE 1 – Statistiques des séries de temps de calcul

La significativité de ces résultats est cependant à relativiser : on impose ici un maximum de 100 itérations. Si on autorise un nombre maximum d'itérations de 500, les temps de calcul des deux algorithmes deviennent comparables. Ceci est dû au langage de programmation utilisé (MATLAB), qui est optimisé pour le calcul matriciel et au fait que la dimension de $\mathbf{B}$ est ici raisonnable. Néammoins la méthode proposée se révélera d'autant plus intéressante dans le cas de données de grande taille. En effet (voir la remarque 2), la complexité de la mise à jour diminue d'un facteur N dans la méthode structurée. Par exemple, dans le cas d'un signal de longueur $R = 1000$ généré à partir d'un dictionnaire de Gabor avec un échantillonnage temporel de pas $a = 5$, un gain de facteur 200 est obtenu pour la mise à jour de l'estimateur de y. Par ailleurs, un gain équivalent est également obtenu en termes d'espace mémoire.

5 Conclusions

Cette contribution montre que les problèmes de débruitage de signaux dans un dictionnaire peuvent être abordés dans le cas de coefficients corrélés, quand le dictionnaire possède certaines structures spécifiques, et les corrélations des coefficients de synthèse sont compatibles avec ces structures. Le modèle considéré ici est un modèle de mélange de Gaussiennes multivariées, des extensions à d'autres modèles exhibant des corrélations peuvent être envisagées. Ce travail se place dans le cadre d'applications à la détection de caractéristiques dans des signaux EEG. Les développements à venir concernent l'extension de cette approche aux signaux multicapteurs, pour lesquels les gains en complexité et stockage seront importants, et son intégration dans un cadre global intégrant estimation de paramètres et décision.

Remerciements : Ce travail a bénéficié du soutien de l'ANR, dans le cadre du projet CoAdapt (*Co-adaptation cerveau-machine pour de meilleures interfaces*, ANR-09-EMER-002-02).

Références

[1] T. De Mazancourt et D. Gerlic, The inverse of a block-circulant matrix, *IEEE Transactions on Antennas and Propagation* **31** :5 (1983), pp. 808-810.

[2] P. J. Davis, *Circulant Matrices*, Wiley, New York (1979).

[3] M. Figueiredo, J. Bioucas-Dias, R. Nowak, Majorization-minimization algorithms for wavelet-based image restoration, *IEEE Transactions on Image Processing*, **16** :12 (2007), pp. 2980-2991.

[4] D.R. Hunter et K. Lange, A Tutorial on MM algorithms, *The American Statistician* **58** (2004), pp. 30-37.

[5] A. Kaveh et H. Rahami, Block circulant matrices and applications in free vibration analysis of cyclically repetitive structures, *Acta Mechanica*, **217** (2010), pp. 51-62.