

Apprentissage de dictionnaire faiblement cohérent par programmation quadratique mixte

Yuan LIU, Stéphane CANU, Paul HONEINE, Su RUAN

Normandie Univ, INSA Rouen, UNIROUEN, UNIHAVRE, LITIS;
Avenue de l'Université, 76801 Saint-Étienne-du-Rouvray Cedex, France
yuan.liu@insa-rouen.fr, stephane.canu@insa-rouen.fr
paul.honeine@univ-rouen.fr, su.ruan@univ-rouen.fr

Résumé – L'apprentissage de dictionnaire a permis des avancées considérables en représentations parcimonieuses. Il a été investi avec succès dans un large spectre d'applications en traitement du signal et des images, ainsi qu'en vision et reconnaissance des formes. Plusieurs études théoriques ont montré la pertinence de construire un dictionnaire à faible cohérence, c'est à dire faible corrélation entre ses éléments. Le problème d'optimisation associé est non convexe et non différentiable. Les méthodes qui s'y attaquent reposent sur la relaxation du problème, par exemple en ajoutant une étape de décorrélation à chaque itération. Dans cet article, nous proposons une méthode qui résout le problème avec les contraintes explicites. Pour cela, le sous-problème de codage parcimonieux est traité selon deux stratégies, par algorithme proximal ou par programme quadratique mixte en nombres entiers (MIQP). L'estimation du dictionnaire sous contraintes est abordée en combinant la méthode du lagrangien augmenté (ADMM) et la méthode *Extended Proximal Alternating Linearized Minimization* (EPALM), adaptée à des familles de problèmes non convexes. L'efficacité de la méthode MIQP+EPALM est démontrée en reconstruction d'image.

Abstract – Sparse representations with dictionary learning has been successfully explored in signal and image processing, as well as in vision and pattern recognition. Several theoretical studies have proven the importance of learning a low-coherence dictionary, *i.e.*, low correlation between its elements. The resulting optimization problem being non-convex and non-smooth, resolution methods rely on constraints relaxation, such as by adding a de-correlation step to each iteration. In this paper, we solve the problem with its explicit constraints. To this end, the sparse coding subproblem is addressed with two strategies, by proximal algorithm or by mixed integer quadratic program (MIQP). The dictionary update is addressed by combining the augmented Lagrangian method (ADMM) and the *Extended Proximal Alternative Linearized Minimization* (EPALM) method, which is suitable for non-convex problems. We show the relevance of the MIQP+EPALM method in image reconstruction.

1 Introduction

L'apprentissage de dictionnaire joue un rôle crucial dans les représentations parcimonieuses, permettant d'avoir une représentation adaptée aux données traitées, en traitement du signal et des images, ainsi qu'en vision, pour des applications de débruitage, d'*inpainting*, de classification et de reconnaissance des formes [12]. Les données étant représentées par une combinaison linéaire de quelques éléments d'un dictionnaire, l'apprentissage de dictionnaire repose sur l'estimation jointe de ses éléments et de leurs coefficients de pondération. Ce problème est abordé en utilisant une procédure de relaxation alternée en deux phases, une pour l'estimation du dictionnaire (dite adaptation du dictionnaire) et une pour l'estimation des coefficients de pondération (dite codage parcimonieux ou *sparse coding*). Les méthodes conventionnelles d'apprentissage de dictionnaire, telles que les méthodes de poursuite [8, 13], ne permettent pas de garantir la qualité du dictionnaire, ni celle de la représentation ou de la reconstruction.

Une mesure fondamentale pour caractériser un dictionnaire est la cohérence. Au delà de mesurer la plus grande corrélation entre les éléments du dictionnaire, cette mesure est intimement

liée au niveau de parcimonie et à la pertinence de la représentation résultante. En effet, plusieurs études théoriques ont montré l'importance d'avoir des dictionnaires à faible cohérence [9, 16]. Différentes méthodes ont été proposées pour apprendre de tels dictionnaires, au prix de relaxer le problème d'optimisation qui est non convexe et non différentiable. Les plus connues sont la méthode INK-SVD qui ajoute une étape de décorrélation après l'adaptation du dictionnaire à chaque itération [11] et la méthode itérative de projections et rotations (IPR) [4].

Le présent article propose une méthode pour apprendre un dictionnaire faiblement cohérent, par la résolution du problème avec les contraintes explicites. Pour cela, le codage parcimonieux est traité selon deux stratégies, par algorithme proximal ou par programme quadratique mixte en nombres entiers (*mixed integer quadratic programming* (MIQP)) [10]. L'adaptation du dictionnaire sous contraintes est abordée en combinant la méthode du lagrangien augmenté, dite ADMM inexacte [5], et la méthode *Extended Proximal Alternating Linearized Minimization* (EPALM) [17]. La convergence de cette dernière a été démontrée pour une classe de problèmes non convexes [17]. L'efficacité de la méthode MIQP+EPALM est démontrée en reconstruction de l'image de Barbara.

2 Énoncé du problème

Soient $\mathbf{y}_1, \dots, \mathbf{y}_\ell$ des signaux de dimension n , et $Y = [\mathbf{y}_1, \dots, \mathbf{y}_\ell]$. Une représentation parcimonieuse de Y consiste à la décomposer selon $Y \approx DX$, où $X = [\mathbf{x}_1, \dots, \mathbf{x}_\ell] \in \mathbb{R}^{p \times \ell}$ est parcimonieuse (c'est à dire, avec seulement quelques termes non nuls), et $D = [\mathbf{d}_1, \dots, \mathbf{d}_p] \in \mathbb{R}^{n \times p}$ est le dictionnaire, ses colonnes étant appelées atomes. L'apprentissage de dictionnaire repose sur l'estimation de X et de D sous contrainte de parcimonie, par la résolution du problème d'optimisation :

$$\begin{aligned} \min_{\substack{D \in \mathbb{R}^{n \times p} \\ \mathbf{x}_i \in \mathbb{R}^p}} & \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{1}{2} \|\mathbf{y}_i - D\mathbf{x}_i\|_2^2 \right) \\ \text{avec} & \begin{cases} \|\mathbf{x}_i\|_0 \leq T, & \forall i = 1, \dots, \ell, \\ \mathbf{d}_k^\top \mathbf{d}_k = 1, & \forall k = 1, \dots, p, \end{cases} \end{aligned} \quad (1)$$

où la fonction objectif est l'erreur quadratique moyenne de reconstruction selon la norme euclidienne, et la parcimonie est mesurée par ℓ_0 qui compte le nombre de termes non nuls. Les atomes du dictionnaire D sont définis selon la contrainte de norme unité afin d'éviter que leurs normes euclidiennes soient trop grandes, ce qui conduirait à des coefficients de décomposition trop petits dans X .

Le problème (1) est abordé par relaxation alternée en deux phases. La première, dite de codage parcimonieux, consiste à estimer X en fixant D . A cause de la norme ℓ_0 , ce problème d'optimisation est non-convexe et NP-difficile. Pour surmonter cette difficulté, la plupart des travaux dans le domaine opèrent par relaxation, en remplaçant la norme ℓ_0 par un terme convexe comme la norme ℓ_1 [13], ou utilisent un algorithme approché comme celui de poursuite [8]. Nous avons récemment montré qu'il est possible d'avoir une résolution exacte, par la MIQP [10]. La seconde phase, dite d'adaptation de dictionnaire, consiste à estimer D en fixant X . Les algorithmes les plus utilisés sont ceux des moindres carrés et de descente de gradient stochastique [13]. Toutefois, le dictionnaire qui en résulte ne permet pas de garantir d'excellentes performances, parce que ses atomes peuvent être arbitrairement corrélés.

La cohérence est une quantité fondamentale considérée pour la caractérisation de dictionnaires. Elle est définie par la plus grande corrélation, en valeur absolue, entre deux atomes d'un dictionnaire, c'est à dire pour des atomes de norme unité :

$$\mu = \max_{i \neq j} |\mathbf{d}_i^\top \mathbf{d}_j|. \quad (2)$$

La pertinence de cette mesure a été démontrée dans plusieurs travaux [9, 16]. L'importance d'apprendre des dictionnaires faiblement cohérents (e.g. dictionnaires à atomes orthogonaux si $\mu = 0$) est incontestable.

Deux stratégies ont été proposées pour apprendre un dictionnaire faiblement cohérent. La première consiste à ajouter, à chaque itération, une étape de décorrélacion après celle d'adaptation du dictionnaire, comme proposé par la méthode INK-SVD [11] et la méthode itérative de projections et rotations (IPR) [4]. La seconde stratégie repose sur l'ajout d'un terme de pénalisation dans le problème d'optimisation [1, 3, 15], ce

qui résulte en un problème d'optimisation non convexe et non différentiable à cause de la norme ℓ_0 sur la parcimonie et des contraintes sur la cohérence et la norme unité des atomes du dictionnaire. A notre connaissance, il n'existe aucun travail sur l'apprentissage de dictionnaire faiblement cohérent qui permet de résoudre le problème avec les contraintes explicites.

Le présent article propose de résoudre explicitement le problème d'apprentissage de dictionnaire faiblement cohérent, sans relaxation, c'est à dire selon le problème d'optimisation

$$\begin{aligned} \min_{\substack{D \in \mathbb{R}^{n \times p} \\ \mathbf{x}_i \in \mathbb{R}^p}} & \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{1}{2} \|\mathbf{y}_i - D\mathbf{x}_i\|_2^2 \right) \\ \text{avec} & \begin{cases} |\mathbf{d}_k^\top \mathbf{d}_h| \leq \mu_s, & \forall k, h \in \{1, \dots, p\}, k \neq h \\ \mathbf{d}_k^\top \mathbf{d}_k = 1, & \forall k = 1, \dots, p \\ \|\mathbf{x}_i\|_0 \leq T, & \forall i = 1, \dots, \ell, \end{cases} \end{aligned} \quad (3)$$

où μ_s est le niveau de cohérence prédéfini.

3 Méthode proposée

L'algorithme d'apprentissage de dictionnaire faiblement cohérent opère en optimisation alternée. Dans l'étape de codage parcimonieux, le problème est reformulé en un problème MIQP pour être résolu avec des techniques d'optimisation avancées [10]. Nous explorons également une résolution par la méthode du proximal [14]. Dans la seconde étape, le problème d'optimisation du dictionnaire avec contraintes de cohérence et de norme unité est un problème non convexe. Nous proposons de résoudre ce problème en combinant la méthode du lagrangien augmenté (ADMM) avec l'algorithme EPALM [17].

3.1 Codage parcimonieux

En supposant D connu, le problème d'optimisation (3) peut être divisé en ℓ problèmes indépendants plus petits, de la forme

$$\min_{\mathbf{x} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - D\mathbf{x}\|_2^2, \quad \text{avec } \|\mathbf{x}\|_0 \leq T, \quad (4)$$

où, par abus de notation, les \mathbf{y}_i et \mathbf{x}_i sont notés par \mathbf{y} et \mathbf{x} , respectivement. Pour résoudre ce problème, nous considérons deux méthodes plus performantes que les approches glouton [2, 7] : les méthodes de proximal [14, 6] et MIQP [10].

La méthode proximale repose sur une application itérative de l'opérateur proximal

$$\text{prox}_f(\mathbf{v}) = \arg \min_{\mathbf{x} \in \mathbb{R}^p} f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|^2. \quad (5)$$

Pour notre problème, $f(\mathbf{x}) = \delta_{\mathcal{S}_T}(\mathbf{x}) + \frac{1}{2} \|\mathbf{y} - D\mathbf{x}\|^2$, où $\delta_{\mathcal{S}_T}$ est la fonction indicatrice sur $\mathcal{S}_T = \{u \in \mathbb{R}^p ; \|u\|_0 \leq T\}$. Face à la difficulté de trouver une solution analytique à ce problème, on utilise une version linéarisée, qui permet d'avoir l'algorithme itératif suivant :

$$\mathbf{x}^{k+1} = \text{prox}_{\mathcal{S}_T}(\mathbf{x}^k - \lambda \nabla_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - D\mathbf{x}^k\|^2), \quad (6)$$

où λ désigne le pas décroissant. En pratique, \mathbf{x}^{k+1} est obtenue de l'argument ci-dessus en ne gardant que ses T plus grands termes, en valeur absolue.

Récemment, nous avons exploité une méthode de codage parcimonieux par MIQP [10], en réécrivant le problème d'optimisation sous une forme que l'on sait résoudre avec les logiciels d'optimisation modernes. L'idée repose sur l'introduction d'une variable binaire $\mathbf{z} \in \{0, 1\}^p$ qui indique si l'élément correspondant dans \mathbf{x} est nul, c'est à dire avec la relation logique

$$z_i = 0 \Leftrightarrow x_i = 0 \quad \text{pour tout } i = 1, 2, \dots, p. \quad (7)$$

La contrainte de parcimonie peut alors s'exprimer sous la forme $\mathbf{1}_p^\top \mathbf{z} \leq T$. Afin d'éliminer explicitement la relation logique (7) dans le problème d'optimisation, nous introduisons une contrainte de type *big M*, qui permet d'imposer les contraintes (7) en posant $-z_i M \leq x_i \leq z_i M$. La valeur de M est prise assez grande pour que la solution optimale $\hat{\mathbf{x}}$ soit dans la contrainte, c'est à dire $M > \|\hat{\mathbf{x}}\|_\infty$. La formulation *big M* du problème (4) est donc, pour $M > 0$ et $T > 0$:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^p, \mathbf{z} \in \{0,1\}^p} \quad & \frac{1}{2} \|\mathbf{y} - D\mathbf{x}\|_2^2 \\ \text{avec} \quad & \begin{cases} -\mathbf{z}M \leq \mathbf{x} \leq \mathbf{z}M \\ \mathbf{1}_p^\top \mathbf{z} \leq T. \end{cases} \end{aligned} \quad (8)$$

Toutes les contraintes sont désormais linéaires, et avec une fonction objective quadratique, ce problème d'optimisation mixte en nombre entiers (MIQP) peut être résolu en utilisant un logiciel d'optimisation comme CPLEX ou GUROBI.

3.2 Adaptation du dictionnaire

En supposant X connu, le dictionnaire est obtenu selon le problème d'optimisation suivant :

$$\begin{aligned} \min_{D \in \mathbb{R}^{n \times p}} \quad & \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{1}{2} \|\mathbf{y}_i - D\mathbf{x}_i\|_2^2 \right) \\ \text{avec} \quad & \begin{cases} |\mathbf{d}_k^\top \mathbf{d}_h| \leq \mu_s, & \forall k, h \in \{1, \dots, p\}, k \neq h \\ \mathbf{d}_k^\top \mathbf{d}_k = 1, & \forall k = 1, \dots, p. \end{cases} \end{aligned}$$

Nous proposons de le résoudre par la méthode ADMM, en introduisant une nouvelle variable $G \in \mathbb{R}^{p \times p}$ qui vérifie $G = D^\top D$. Le problème s'écrit alors sous la forme

$$\begin{aligned} \min_{D \in \mathbb{R}^{n \times p}, G \in \mathbb{R}^{p \times p}} \quad & \frac{1}{2} \|Y - DX\|_F^2 \\ \text{avec} \quad & \begin{cases} G = D^\top D, & G \in \mathcal{S}_G \\ \mathbf{d}_k^\top \mathbf{d}_k = 1, & \forall k = 1, \dots, p, \end{cases} \end{aligned}$$

où $\mathcal{S}_G = \{G \in \mathbb{R}^{p \times p} ; |G_{ij}| \leq \mu_s, \forall i \neq j\}$. Cette formulation semble permettre d'accélérer la convergence, comparée à l'ajout dans \mathcal{S}_G de la contrainte de norme unité sur les \mathbf{d}_k .

Ce problème d'optimisation sous contraintes est résolu par la minimisation du lagrangien augmenté :

$$\begin{aligned} L_{(c1,c2)}(D, G, \boldsymbol{\lambda}, H) = & \frac{1}{2} \|Y - DX\|_F^2 \\ & + \sum_{k=1}^p \lambda_k (\mathbf{d}_k^\top \mathbf{d}_k - 1) + \frac{c_1}{2} \sum_{k=1}^p (\mathbf{d}_k^\top \mathbf{d}_k - 1)^2 \\ & + \text{tr}(H(G - D^\top D)) + \frac{c_2}{2} \|G - D^\top D\|_F^2 + \delta_{\mathcal{S}_G}(G), \end{aligned}$$

Algorithme 1 L'algorithme ADMM pour résoudre (9)

pour tout $i = 0$ to $N_{iter} - 1$ **faire**
1. Déterminer la solution optimale (D^i, G^i) :
 $(D^i, G^i) = \arg \min L_{(c_1^i, c_2^i)}(D, G, \boldsymbol{\lambda}^i, H^i)$ (10)
2. Mettre à jour les multiplicateurs $(\boldsymbol{\lambda}^i, H^i)$:

$$\begin{cases} \boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i + c_1^i (\text{diag}((D^i)^\top D^i) - 1) \\ H^{i+1} = H^i + c_2^i (G^i - (D^i)^\top D^i) \end{cases}$$

3. Augmenter les paramètres de pénalité : $c_1^{i+1} = \rho_1 c_1^i$ et $c_2^{i+1} = \rho_2 c_2^i$
4. Sortir la solution si les critères d'arrêt sont atteints
si $\max_k |(\mathbf{d}_k^\top \mathbf{d}_k) - 1| \leq \epsilon$ et $\max |G^i - (D^i)^\top D^i| \leq \epsilon$ **alors** retour,
fin si
fin pour

Algorithme 2 L'algorithme EPALM pour résoudre (10)

Initialisation : $j = 0$, $G^{i,0} = (D^{i,0})^\top D^{i,0}$, $G^{i,0}(i_x, i_y) = \text{sign}(G^{i,0}(i_x, i_y)) \min(|G^{i,0}(i_x, i_y)|, \mu_s)$, $\Theta^i = \Theta^i(D^{i,0}, G^{i,0})$.
tant que $j < N_{iter}$ et $\Theta^i > \epsilon^i$ **faire**
1. Mettre à jour $D^{i,j}$ selon (11)
2. Calculer : $\tilde{G} = G^{i,j-1} - \frac{1}{t_2} \nabla_G (h(D^{i,j}, G^{i,j-1}))$
3. Obtenir $G^{i,j}$ par projection de \tilde{G} dans \mathcal{S}_G , selon (12)
4. Calculer le sous-différentiel $\Theta^i(D^{i,j}, G^{i,j})$
5. $j = j + 1$.
fin tant que

où $\boldsymbol{\lambda}$ et H sont le vecteur et la matrice de multiplicateurs de Lagrange, associés aux contraintes sur la diagonale de $D^\top D$ et l'égalité sur G , respectivement, c_1 et c_2 sont les paramètres positifs de pénalité (de l'augmentation), et $\delta_{\mathcal{S}_G}(G)$ est la fonction indicatrice sur l'ensemble \mathcal{S}_G . Le problème devient

$$\min_{D \in \mathbb{R}^{n \times p}, G \in \mathbb{R}^{p \times p}} L_{(c1,c2)}(D, G, \boldsymbol{\lambda}, H) \quad (9)$$

Ce n'est pas la méthode habituelle du lagrangien augmenté, à cause de sa non-convexité et sa non-différentiabilité. Cependant, il reste raisonnable de considérer l'approche dite ADMM inexacte [5], selon l'algorithme illustré dans Algorithme 1.

Nous proposons de résoudre le problème (9)-(10) en utilisant l'algorithme EPALM, dont la convergence a été démontrée pour une classe de problèmes non convexes [17]. Pour cela, nous identifions trois parties dans la fonction objective de (9) :

$$L_{(c_1^i, c_2^i)}(D, G, \boldsymbol{\lambda}^i, H^i) = f(D) + h(D, G) + \delta_{\mathcal{S}_G}(G),$$

où $f(D)$ et $h(D, G)$ sont définies respectivement par les 3 premiers termes et les deux suivants (ceux avec $G - D^\top D$) de (9). Par optimisation alternée, l'estimation de D est obtenue par une décente de gradient, selon

$$D^{i,j} = D^{i,j-1} - \frac{1}{t_1} \nabla_D f(D^{i,j-1}) + h(D^{i,j-1}, G^{i,j-1}), \quad (11)$$

et celle de G est obtenue par la méthode proximale, selon

$$G^{i,j}(i_x, i_y) = \begin{cases} \tilde{G}(i_x, i_y) & \text{si } |\tilde{G}(i_x, i_y)| \leq \mu_s \\ \text{sign}(\tilde{G}(i_x, i_y)) \mu_s & \text{sinon,} \end{cases} \quad (12)$$

où $\tilde{G} = G^{i,j-1} - \frac{1}{t_2} \nabla_G h(D^{i,j}, G^{i,j-1})$, et $i_x, i_y \in \{1, \dots, p\}$. L'algorithme résultant est présenté dans Algorithme 2. Les preuves de convergence et le calcul de sous-différentiel (pour le critère d'arrêt) sont omis de cet article par manque de place.

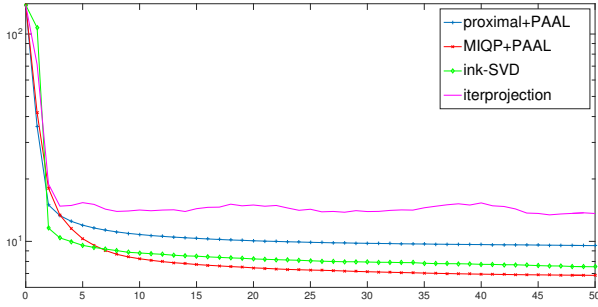


FIGURE 1 – La convergence des méthodes étudiées.

TABLE 1 – Résultats en terme de PSNR (en dB) pour différentes valeurs de cohérence $\mu_s = \cos(\theta_s)$

	Angle maximal θ_s entre deux atomes						
	5°	15°	30°	45°	60°	75°	83°
INK-SVD [11]	36.46	36.56	36.26	34.83	34.04	30.15	-
IPR [4]	36.82	36.57	35.72	31.51	30.60	27.84	-
Proximal+EPALM	27.40	27.42	28.06	29.80	29.31	29.75	22.97
MIQP+EPALM	37.60	37.26	38.89	38.55	36.52	35.31	33.97

4 Résultats expérimentaux

Dans cette section, nous étudions les performances de la méthode proposée d'apprentissage de dictionnaire, selon ses deux versions de codage parcimonieux : Proximal+EPALM et MIQP+EPALM. Nous avons comparé les performances avec INK-SVD [11] et IPR [4], leurs paramètres ont été fixés selon ces articles de référence. Nous avons considéré l'image de Barbara, avec des imagerie de taille 8×8 se chevauchant, pour définir la matrice Y de dimension $n = 64$ et $\ell > 3,5 \times 10^4$. Comme préconisé dans la littérature [8], le nombre d'atomes a été fixé à $p = 256$ et le niveau de parcimonie à $T = 20$.

La FIGURE 1 présente la convergence des algorithmes, obtenue en fixant le seuil de cohérence à $\mu_c = 0.6$ (53°). Ces résultats montrent que la méthode MIQP+EPALM, c'est à dire la combinaison de l'algorithme MIQP pour le codage parcimonieux et EPALM pour l'adaptation du dictionnaire, admet la plus grande vitesse de convergence. Dans la suite des expériences, le nombre maximal d'itérations a été fixé à 30.

Nous avons étudié sept valeurs de cohérence $\mu_s = \cos(\theta_s)$, en fixant l'angle maximal entre deux atomes du dictionnaire à $\theta_s \in \{5^\circ; 15^\circ; 30^\circ; 45^\circ; 60^\circ; 75^\circ; 83^\circ\}$. Pour étudier la relation entre les performances en reconstruction et la cohérence du dictionnaire, nous avons examiné l'erreur de reconstruction en terme de rapport signal sur bruit (PSNR). Les résultats de reconstruction sont présentés dans la TABLE 1, pour différentes valeurs du paramètre de cohérence. Il est clair que les performances des méthodes INK-SVD et IPR diminuent quand la cohérence diminue, alors que la méthode Proximal+EPALM a la plus faible erreur pour $\mu_c = \cos(45^\circ)$, et la méthode MIQP+EPALM a la plus faible erreur pour $\mu_c = \cos(30^\circ)$. De plus, la méthode MIQP+EPALM surpasse les autres méthodes, quelle que soit la valeur du paramètre de cohérence. Le prix à payer est une complexité de calcul plus élevée.

5 Conclusion et perspectives

Nous avons proposé une nouvelle méthode d'apprentissage de dictionnaire faiblement cohérent. Le codage parcimonieux a été traité par un algorithme proximal ou par MIQP. Le dictionnaire a été obtenu en combinant la méthode du lagrangien augmenté (ADMM) avec l'algorithme EPALM. Nous avons montré l'efficacité de la méthode en reconstruction d'image. Les futurs travaux se concentreront sur l'apprentissage de dictionnaire pour des tâches spécifiques, notamment la classification.

Références

- [1] V. Abolghasemi, S. Ferdowsi, and S. Sanei. Fast and incoherent dictionary learning algorithms with application to fmri. *Signal, Image and Video Processing*, 9(1):147–158, 2015.
- [2] C. Bao, H. Ji, Y. Quan, and Z. Shen. ℓ_0 norm based dictionary learning by proximal methods with global convergence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3858–3865, 2014.
- [3] C. Bao, Y. Quan, and H. Ji. A convergent incoherent dictionary learning algorithm for sparse coding. In *European Conference on Computer Vision*, pages 302–316. Springer, 2014.
- [4] D. Barchiesi and M. D. Plumbley. Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. *IEEE Transactions on Signal Processing*, 61(8):2055–2065, April 2013.
- [5] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.
- [6] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier analysis and Applications*, 14(5-6):629–654, 2008.
- [7] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau. Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance. *IEEE Transactions on Signal Processing*, 64(6):1405–1419, 2016.
- [8] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.
- [9] P. Honeine. Analyzing sparse dictionaries for online learning with kernels. *IEEE Transactions on Signal Processing*, 63(23):6343–6353, 2015.
- [10] Y. Liu, S. Canu, P. Honeine, and S. Ruan. Mixed integer programming for sparse coding: Application to image denoising. *IEEE Transactions on Computational Imaging*, (in press) 2019.
- [11] B. Mailhé, D. Barchiesi, and M. D. Plumbley. INK-SVD: Learning incoherent dictionaries for sparse representations. In *Proc. ICASSP*, pages 3573–3576, March 2012.
- [12] J. Mairal, F. Bach, J. Ponce, et al. Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, 8(2-3):85–283, 2014.
- [13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696. ACM, 2009.
- [14] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.
- [15] I. Ramírez, F. Lecumberry, and G. Sapiro. Sparse modeling with universal priors and learned incoherent dictionaries. In *Proc. CAMSAP*, 2009.
- [16] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. on Information theory*, 50(10):2231–2242, 2004.
- [17] H. Zhu, X. Zhang, D. Chu, and L.-Z. Liao. Nonconvex and nonsmooth optimization with generalized orthogonality constraints: An approximate augmented lagrangian method. *Journal of Scientific Computing*, 72(1):331–372, 2017.