

Une approche convexe du partitionnement de données et de la segmentation d'image

Laurent CONDAT

Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble
Laurent.Condat@gipsa-lab.grenoble-inp.fr

Résumé – Une nouvelle formulation convexe du partitionnement de données et de la segmentation d'image est proposée, à nombre K de régions fixé, avec pénalisation du périmètre des régions. Ce problème est ainsi une version spatialement régularisée du problème des K -moyennes. L'approche proposée repose sur une discrétisation de l'espace de recherche, c'est-à-dire qu'il faut spécifier un nombre fini de candidats, parmi lesquels les K centroïdes sont déterminés. Après reformulation comme un problème d'affectation, une relaxation convexe est proposée, qui fait intervenir une boule de norme $l_{1,\infty}$. Un éclatement de celle-ci est proposé, afin d'éviter la projection coûteuse sur cette boule. Des exemples illustrent l'efficacité de l'approche.

Abstract – A new convex formulation to data clustering and image segmentation is proposed, with fixed number K of regions and penalization of the region perimeters. So, this problem is a spatially regularized version of the K -means problem. The proposed approach relies on a discretization of the search space; that is, a finite number of candidates must be specified, from which the K centroids are determined. After reformulation as an assignment problem, a convex relaxation is proposed, which involves a $l_{1,\infty}$ norm ball. A splitting of it is proposed, so as to avoid the costly projection on this ball. Some examples illustrate the efficiency of the approach.

1 Introduction

Le partitionnement (*clustering*) de données vise à décomposer un ensemble d'éléments en K groupes, de manière à minimiser un certain coût de dissimilarité intra-groupe. Ainsi, le problème classique des K -moyennes (*K-means*), en général NP-difficile, consiste à partitionner N points de \mathbb{R}^d en K groupes, en minimisant la somme des distances au carré de chaque point au centroïde (barycentre du groupe) le plus proche. Une application est la quantification d'image couleur [1] : on cherche la palette de K couleurs représentant au mieux une image donnée. Les points sont alors les valeurs de pixels dans \mathbb{R}^3 correspondant aux coordonnées dans un espace couleur.

Un problème fondamental en traitement d'image et vision, encore plus difficile, est celui de la segmentation d'image : on cherche à décomposer une image de N pixels en K régions correspondant aux objets qui composent la scène, sur la base, en plus d'une similarité intra-région, d'une certaine homogénéité spatiale [2]. Nous considérons dans cet article une approche variationnelle de type *partitionnement minimal* [3], c'est-à-dire que l'homogénéité spatiale est obtenue en pénalisant le périmètre des régions. Ce problème est parfois appelé problème de Potts [4] ou problème d'approximation constante par morceaux de Mumford-Shah [2].

De manière générale, le problème considéré peut être formalisé comme suit. Les données $y = (y_n)_{n \in \Omega}$ consistent en un signal de domaine $\Omega = \{1, \dots, N\}$ ou une image de domaine $\Omega = \{1, \dots, N_1\} \times \{1, \dots, N_2\}$ (ayant $N = N_1 N_2$ pixels), à valeurs y_n dans \mathbb{R}^d , muni de la norme euclidienne. Étant donné

un entier $K \geq 2$, on cherche à partitionner Ω en K régions¹ Ω_k (donc $\bigcup_{k=1}^K \Omega_k = \Omega$ et $\Omega_k \cap \Omega_{k'} = \emptyset$, pour tous $k \neq k'$), et à trouver les centroïdes $c_k \in \mathbb{R}^d$ associés, de telle sorte à

$$\underset{(\Omega_k)_{k=1}^K, (c_k)_{k=1}^K}{\text{minimiser}} \frac{1}{2} \sum_{k=1}^K \sum_{n \in \Omega_k} \|y_n - c_k\|^2 + \frac{\lambda}{2} \sum_{k=1}^K \text{per}(\Omega_k), \quad (1)$$

où per indique le périmètre et $\lambda \geq 0$ est un paramètre qui contrôle le niveau de régularisation spatiale. Lorsque $\lambda = 0$, on retrouve exactement le problème des K -moyennes. L'ordre de l'indexation ne joue alors plus aucun rôle et on peut raisonner sur le nuage de points $(y_n)_{n \in \Omega}$ dans \mathbb{R}^d , qu'il s'agit de partitionner en K groupes, dont les c_k sont les barycentres.

On peut définir le signal ou l'image quantifié / segmenté $x = (x_n)_{n \in \Omega}$ avec $x_n = c_k$ si $n \in \Omega_k$. Ainsi, x est une approximation de y constante par morceaux, prenant au plus K valeurs. Notons que dans le cas d'un signal 1-D, $\frac{1}{2} \sum_{k=1}^K \text{per}(\Omega_k) = |\{n : x_n \neq x_{n+1}\}|$, où $|\cdot|$ indique le cardinal d'un ensemble, est le nombre de sauts dans le signal². La définition en 2-D du périmètre basée sur des différences finies est beaucoup plus délicate [3, 5].

On remarque que, à régions Ω_k fixées, les centroïdes c_k , solutions du problème (1), sont les moyennes des éléments des régions : $c_k = \frac{1}{|\Omega_k|} \sum_{n \in \Omega_k} y_n$. À l'inverse, les centroïdes c_k étant fixés, 1) lorsque $\lambda = 0$, on obtient simplement les régions

1. Au sens strict, le nombre de régions est au plus K , et non exactement K , car certaines régions Ω_k pourraient être vides. Ce n'est jamais le cas dans les applications considérées.

2. On suppose des conditions de bord symétriques, donc le bord du domaine Ω n'est pas compté dans le périmètre.

Ω_k à partir des cellules de Voronoï associées : $\Omega_k = \{n \in \Omega : k = \operatorname{argmin}_{k'} \|y_n - c_{k'}\|\}$. 2) Lorsque $\lambda > 0$, il existe des méthodes efficaces pour résoudre le problème, par relaxation convexe [3, 5, 6] ou coupure de graphe [7], voir aussi [8]. Il s'ensuit qu'une stratégie alternée est possible, mettant à jour les Ω_k à c_k fixés, puis inversement. Dans le cas $\lambda = 0$, cela donne exactement l'algorithme classique des K -moyennes (à ne pas confondre avec le problème du même nom). Même s'il converge vers un minimum local du problème, il reste très dépendant de l'initialisation. Des méthodes globalement convexes, de forte complexité, ont été proposées pour les cas particuliers $K = 2$ ou $K = 4$ [9–11]. L'auteur n'a pas connaissance de méthode générique permettant d'approcher le minimum global du problème (1), réputé NP-difficile (l'approche par pénalisation des distances entre centroïdes [12] est inadaptée ici). Dans cet article, le problème est attaqué en **discrétisant** l'espace de recherche des centroïdes : on fixe un ensemble $\Gamma = \{a_m\}_{m=1}^M$ de M points (avec $K \ll M$) de \mathbb{R}^d , appelés *candidats*, et les centroïdes sont cherchés dans Γ plutôt que \mathbb{R}^d entier.

2 Approche proposée

2.1 Reformulation par *lifting* du problème

On a vu que dans le signal ou l'image segmenté x , chaque élément x_n prend comme valeur un des centroïdes c_k , qui est lui-même un des candidats a_m . On peut alors reformuler le problème (1) sous la forme d'un problème équivalent, dont l'inconnue est le tableau d'affectation z , qui a une dimension de plus que y , indexée par $m = 1, \dots, M$. Pour tout $n \in \Omega$ et $m = 1, \dots, M$, $z_{m,n}$ vaut 1 si $x_n = a_m$ et vaut 0 sinon. Chaque vecteur $z_{:,n} = (z_{m,n})_{m=1}^M$ appartient donc à l'ensemble \mathcal{A} des vecteurs d'affectation binaires, c'est-à-dire des vecteurs à éléments dans $\{0, 1\}$ dont la somme vaut 1.

On retrouve x à partir de z par une simple sommation :

$$x_n = \sum_{m=1}^M z_{m,n} a_m, \quad \forall n \in \Omega. \quad (2)$$

De plus, on peut réécrire le terme d'attache aux données de (1) comme

$$\frac{1}{2} \sum_{n \in \Omega} \|x_n - y_n\|^2 = \frac{1}{2} \sum_{n \in \Omega} \sum_{m=1}^M z_{m,n} w_{m,n}, \quad (3)$$

ou, en abrégé, $\frac{1}{2} \|x - y\|_2^2 = \langle z, w \rangle$, où

$$w_{m,n} = \frac{1}{2} \|y_n - a_m\|^2. \quad (4)$$

En utilisant la formule de la co-aire [3], le terme de régularisation de (1), quant à lui, peut être écrit

$$\frac{\lambda}{2} \sum_{k=1}^K \operatorname{per}(\Omega_k) = \frac{\lambda}{2} \sum_{m=1}^M \operatorname{TV}(z_{m,:}), \quad (5)$$

où TV est une certaine forme discrète de la variation totale [5] et $z_{m,:} = (z_{m,n})_{n \in \Omega}$ est un signal ou une image scalaire de même domaine Ω que y .

Il reste donc juste à reformuler la contrainte que le nombre de régions Ω_k dans x est au plus K ou, de manière équivalente, que $|\{x_n\}| \leq K$, c'est-à-dire que x ne prend ses valeurs que dans K parmi les M candidats. On a la propriété suivante :

Proposition 1 *Le tableau d'affectation $z \in \mathcal{A}^\Omega$ correspond, via (2), à un signal ou une image $x \in \Gamma^\Omega$ qui prend au plus K valeurs distinctes, si et seulement si*

$$\|z\|_{1,\infty} \leq K, \quad (6)$$

$$\text{où}^3 \|z\|_{1,\infty} = \sum_{m=1}^M \max_{n \in \Omega} z_{m,n}.$$

En effet, comme z est à valeurs dans $\{0, 1\}$, un candidat a_m est affecté à au moins un élément x_n (et constitue donc un des centroïdes c_k) si et seulement si $z_{m,:}$ contient au moins un 1, soit si et seulement si $\max_{n \in \Omega} z_{m,n} = 1$.

On peut donc réécrire le problème (1), avec espace de recherche Γ des centroïdes, comme

$$\underset{z \in \mathcal{A}^\Omega}{\text{minimiser}} \langle z, w \rangle + \frac{\lambda}{2} \sum_{m=1}^M \operatorname{TV}(z_{m,:}) \quad \text{s.t.} \quad \|z\|_{1,\infty} \leq K. \quad (7)$$

2.2 Relaxation convexe du problème

Le problème (7) est non convexe ; plus précisément, les fonctions et ensembles sont tous convexes, hormis l'ensemble \mathcal{A} des vecteurs binaires d'affectation. On considère donc une relaxation convexe du problème, obtenue en remplaçant \mathcal{A} par son enveloppe convexe, qui est le simplexe Δ , soit l'ensemble des vecteurs à éléments positifs dont la somme vaut 1 [13]. Introduisons la boule $\mathcal{B} = \{s \in \mathbb{R}^{M \times \Omega} : \|s\|_{1,\infty} \leq K\}$, ainsi que la fonction indicatrice convexe $\iota_{\mathcal{E}}$ d'un ensemble \mathcal{E} qui vaut 0 si sa variable appartient à \mathcal{E} et $+\infty$ sinon. Le problème convexe proposé est donc

$$\underset{z \in \mathbb{R}^{M \times \Omega}}{\text{minimiser}} \langle z, w \rangle + \sum_{n \in \Omega} \iota_{\Delta}(z_{:,n}) + \frac{\lambda}{2} \sum_{m=1}^M \operatorname{TV}(z_{m,:}) + \iota_{\mathcal{B}}(z). \quad (8)$$

La projection sur le simplexe peut être réalisée de manière efficace [13]. Par contre, la projection sur \mathcal{B} , même si elle peut être réalisée exactement en temps fini [14], reste très coûteuse. C'est pourquoi nous proposons un éclatement (*splitting* [15]) de la fonction maximum :

Proposition 2 *La fonction maximum d'un vecteur, ou plus généralement d'un tableau à N éléments, s peut s'écrire comme une convolution infimale :*

$$\max_{n=1,\dots,N} s_n = \min_{q \in \mathbb{R}} q / \sqrt{\mu N} + \iota_{\mathcal{C}}(s - S^* q / \sqrt{\mu N}), \quad (9)$$

où $\mu > 0$ est un paramètre dont la valeur est fixée, \mathcal{C} est le cône des vecteurs à éléments négatifs, S est l'opérateur linéaire qui renvoie la somme des éléments d'un tableau et son adjoint, S^* , duplique un scalaire en un tableau de même taille que s à N éléments identiques.

3. On abuse dans cet article des termes de norme et de boule $l_{1,\infty}$: les éléments de z étant positifs, on n'a pas besoin d'en prendre la valeur absolue.

Notons que, dans la Proposition 2, la norme de l'opérateur linéaire $(s, q) \mapsto s - S^*q/\sqrt{\mu N}$ est $1 + 1/\mu$.

Introduisons $\mathcal{G} = \{s \in \mathbb{R}^M : \sum_{m=1}^M s_m \leq K\sqrt{\mu N}\}$. Le problème convexe que nous proposons de résoudre est ainsi :

$$\begin{aligned} & \underset{z \in \mathbb{R}^{M \times \Omega}, q \in \mathbb{R}^M}{\text{minimiser}} \quad \langle z, w \rangle + \sum_{n \in \Omega} \iota_{\Delta}(z_{:,n}) + \iota_{\mathcal{G}}(q) + \quad (10) \\ & \frac{\lambda}{2} \sum_{m=1}^M \text{TV}(z_{m,:}) + \sum_{m=1}^M \iota_{\mathcal{C}}(z_{m,:} - S^*q_m/\sqrt{\mu N}). \end{aligned}$$

Faute de place, nous ne détaillons pas l'algorithme développé, qui est la forme sur-relaxée ($\rho = 1.9$) de l'algorithme de Chambolle-Pock [15]. Nous utilisons dans la section suivante la forme isotrope de la variation totale discrète [5]. z ainsi que les variables duales nécessaires à représenter les deux derniers termes de (10) ont une taille en $O(NM)$. La complexité de chaque itération est en $O(NM \log M)$ (voire en $O(NM)$ avec une projection sur Δ en $O(M)$ plutôt que par tri [13]).

3 Validation expérimentale

On considère le problème des K -moyennes ($\lambda = 0$, $K = 20$) visant à partitionner le nuage y de $N = 3000$ points en dimension $d = 2$ du jeu de données A1 disponible à l'adresse <https://cs.joensuu.fi/sipu/datasets/>. Dans une première expérience, l'ensemble Γ forme une grille uniforme de 40×80 points dont on n'a gardé que les $M = 2280$ points dans l'enveloppe convexe de y . L'algorithme converge en un nombre fini d'itérations vers la solution exacte du problème (10), représentée en Fig. 2, qui se trouve être binaire (ce que l'on teste aisément en vérifiant que $\min_{n \in \Omega} \max_{m=1}^M z_{m,n} = 1$). Ainsi, et cela est remarquable, on a obtenu la solution du problème non convexe (1) (pour ce choix de Γ). Dans une seconde expérience, les candidats coïncident avec les données : $M = N$ et $\Gamma = y$. On obtient là aussi en temps fini la solution, identique, des problèmes (10) et (1).

On considère ensuite le problème de quantification d'image couleur [1], qui se ramène au problème des K -moyennes visant à partitionner les valeurs de pixels y_n dans l'espace couleur CIELAB ($d = 3$). En effet, la distance euclidienne dans cet espace est une bonne approximation de la distance perceptuelle. On construit une palette de $M = 279$ couleurs, représentée en Fig. 1. L'algorithme converge à nouveau en un nombre fini d'itérations vers la solution exacte binaire des problèmes (10) et (1), dans les six cas illustrés en Fig. 3.

Enfin, on considère le problème de segmentation, avec $K = 5$, $\lambda = 500$ et la palette de la Fig. 1, pour les deux images en haut de la Fig. 3. On voit bien apparaître 5 couleurs dans les images segmentées, voir Fig. 4, mais z n'est pas binaire et des mélanges de couleurs apparaissent au niveau des contours. Notons que l'orange des tournesols est différent de celui obtenu par quantification en Fig. 3 : cela justifie l'approche globalement convexe proposée, plutôt que d'estimer d'abord les K couleurs c_k puis de résoudre (1) à c_k fixées.

Références

- [1] L. Brun and A. Trémeau, *Digital Color Imaging Handbook*, CRC Press, 2012, Chapter "Color quantization", pp. 589–638.
- [2] L. Bar, T. F. Chan, G. Chung, M. Jung, N. Kiryati, N. Sochen, and L. A. Vese, "Mumford and Shah model and its applications to image segmentation and image restoration," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed. 2015, Springer.
- [3] A. Chambolle, D. Cremers, and T. Pock, "A convex approach to minimal partitions," *SIAM J. Imaging Sci.*, vol. 5, no. 4, pp. 1113–1158, 2012.
- [4] R. B. Potts, "Some generalized order-disorder transformations," *Proc. Cambridge Philos. Soc.*, vol. 48, pp. 106–109, 1952.
- [5] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imaging Sciences*, 2017, à paraître.
- [6] L. Condat and N. Pustelnik, "Segmentation d'image par optimisation proximale," in *Actes du GRETSI*, Lyon, France, Sept. 2015.
- [7] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, "A continuous max-flow approach to Potts model," in *Proc. of ECCV*, 2010, pp. 379–392.
- [8] C. Zach, C. Häne, and M. Pollefeys, "What is optimized in convex relaxations for multilabel problems: Connecting discrete and continuously inspired MAP inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 157–170, Jan. 2014.
- [9] E. S. Brown, T. F. Chan, and X. Bresson, "Completely convex formulation of the Chan–Vese image segmentation model," *Int. J. Comput. Vis.*, vol. 98, no. 1, pp. 103–121, 2012.
- [10] E. Bae, J. Yuan, and X.-C. Tai, "Simultaneous convex optimization of regions and region parameters in image segmentation models," in *Innovations for Shape Analysis: Models and Algorithms*, M. Breuß, A. Bruckstein, and P. Maragos, Eds. 2013, pp. 421–438, Springer Berlin Heidelberg.
- [11] E. Bae and X.-C. Tai, "Efficient global minimization methods for image segmentation models with four regions," *J. Math. Imaging and Vision*, vol. 51, no. 1, pp. 71–97, Jan. 2015.
- [12] C. Zhu, H. Xu, C. Leng, and S. Yan, "Convex optimization procedure for clustering: theoretical revisit," in *Proc. of NIPS*, Montreal, Canada, Dec. 2014, pp. 1619–1627.
- [13] L. Condat, "Fast projection onto the simplex and the 11 ball," *Math. Program. Series A*, vol. 158, no. 1, pp. 575–585, July 2016.
- [14] A. Quattoni, X. Carreras, M. Collins, and T. Darrell, "An efficient projection for $l_{1,\infty}$ regularization," in *Proc. of ICML*, Montreal, Canada, June 2009, pp. 857–864.
- [15] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.



FIGURE 1 – Palette de $M = 279$ couleurs obtenue en échantillonnant l'espace CIELAB sur un treillis cubique centré (les deux dernières couleurs orange et rouge ont été ajoutées manuellement).

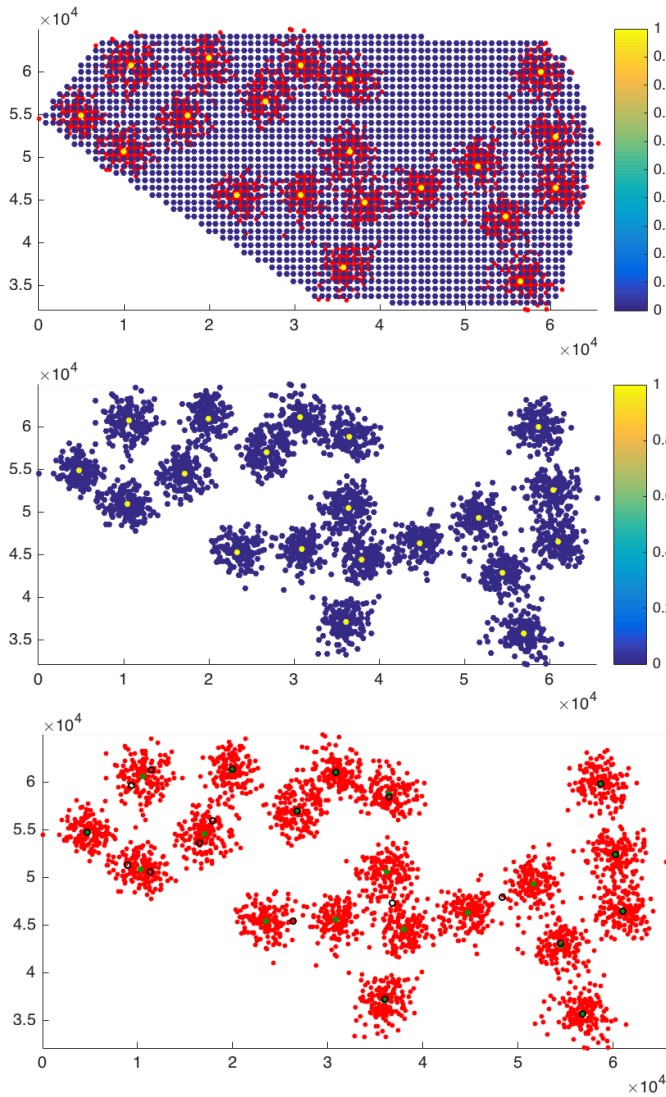


FIGURE 2 – Partitionnement de $N = 3000$ points (en rouge) en $K = 20$ groupes. Les M candidats a_m sont représentés avec une couleur correspondant à $\max_{n \in \Omega} z_{m,n}$, où z est la solution du problème (10). En haut, les candidats sont sur une grille uniforme. Au milieu, les candidats coïncident avec les données. Dans les deux cas, z est binaire et donc solution de (7). En bas, centroïdes trouvés par l'algorithme kmeans de Matlab, en noir, et même chose mais en initialisant kmeans avec les centroïdes trouvés par la méthode proposée (en jaune sur la figure du milieu), en vert.



FIGURE 3 – Résultats x de quantification couleur de deux images y de taille 168×254 et 228×298 , ligne du haut, sur $K = 6$, $K = 5$, $K = 4$ couleurs parmi la palette Γ représentée en Fig. 1, de haut en bas.



FIGURE 4 – Résultats x de segmentation des deux images y en haut de la Fig. 3, pour $K = 5$, $\lambda = 500$, et la palette Γ des couleurs candidates en Fig. 1.