

Sur le problème de pré-image en reconnaissance des formes avec contraintes de non-négativité

Maya KALLAS^{1,2}, Paul HONEINE¹, Hassan AMOUD³, Clovis FRANCIS²

¹Institut Charles Delaunay (UMR STMR CNRS 6279), LM2S, Université de Technologie de Troyes, France

²Laboratoire d'analyse des systèmes (LASYS), Université Libanaise, Faculté de Génie, Liban

³Azm Center for Research in Biotechnology and its Applications, Université Libanaise, Liban

maya.kallas@utt.fr, paul.honeine@utt.fr, hassan.amoud@auf.org, cfrancis@ul.edu.lb

Résumé – Cet article traite le problème de la pré-image en méthodes à noyau pour la reconnaissance des formes. Il s'agit d'un passage obligé pour les applications dont le résultat du traitement doit pouvoir s'exprimer dans le même espace que les observations et non dans un espace fonctionnel difficile à appréhender. C'est le cas par exemple pour le débruitage de données au moyen de l'Analyse en Composantes Principales à noyau. On montre que le problème de la pré-image se prête à une résolution par des méthodes de descente. On profite alors de cette opportunité pour montrer qu'il est possible d'imposer des contraintes à la pré-image, telle que la non-négativité du résultat comme on peut la rencontrer en traitement d'images par exemple. Les algorithmes proposés sont illustrés sur le débruitage non-linéaire d'images et les performances atteintes montrent la pertinence de notre approche.

Abstract – This paper deals with the pre-image problem in kernel-based machine learning. This mapping back from a feature space into the input space is often crucial in many pattern recognition and denoising tasks. We study a gradient descent solution of the pre-image problem, and show the ease of imposing non-negativity constraints on the result. The relevance of the proposed approach is illustrated for denoising grayscale images.

1 Introduction

Au cours des deux dernières décennies, nous avons assisté à une prolifération des méthodes à noyau grâce à la diversité des traitements non-linéaires qu'elles autorisent avec un faible coût calculatoire [1]. Depuis les Machines à Support Vecteurs de Vapnik [2], ses performances ont été démontrées dans plusieurs domaines pour différentes finalités : détection, classification, estimation, régression, prédiction, débruitage, etc. Le principe clé réside dans l'interprétation d'un noyau défini positif comme un produit scalaire dans un espace transformé. Ainsi un tel noyau assure-t-il le passage des données de l'espace d'observation à l'espace dit de Hilbert à noyau reproduisant, sans la nécessité d'exhiber la fonction de transformation non-linéaire associée.

L'importance du passage de l'espace des observations à l'espace de Hilbert à noyau reproduisant est claire dans des problèmes décisionnels, tels que la classification et la régression. Cependant, la fonction réciproque, de l'espace transformé à l'espace d'observation, est souvent indispensable, surtout pour retrouver le résultat dans l'espace observable, *e.g.*, l'espace des signaux en traitement du signal. Or, les deux espaces ne sont pas en bijection, et très peu d'éléments de l'es-

pace transformé ont un antécédent dans l'espace initial. C'est un problème inverse, connu sous le nom du problème de la *pré-image*. La résolution de ce problème consiste à trouver une observation dont l'image, par la fonction noyau considérée, soit la plus proche possible de l'élément en question dans l'espace transformé. Plusieurs méthodes ont été proposées dans la littérature afin de résoudre ce problème mal-posé : une approche itérative par point-fixe [3], une technique issue du *multi-dimensional scaling* (MDS) [4], et une approche basée sur une transformation conforme [5]. Voir [6] pour une revue récente des différentes méthodes.

Cette communication exploite une nouvelle famille d'algorithmes pour la résolution du problème de pré-image. On montre que ce problème d'optimisation se prête facilement à une résolution par des algorithmes de descente du gradient. On en profite alors pour imposer des contraintes supplémentaires à la pré-image. On s'intéresse en particulier aux contraintes de non-négativité, comme c'est souvent le cas en traitement des images. On dérive alors un algorithme à gradient qui impose la non-négativité des résultats des observations dans l'expression de la pré-image. La non-négativité de la pré-image est retenue pour l'application de débruitage, à partir d'un algorithme d'Analyse en Composantes Principales à noyau (ACP-à-noyau). Pour ce problème, les algorithmes proposés ont des performances supérieures aux techniques classiques de pré-

Ce travail a été partiellement financé par les accords de coopération scientifique franco-libanais CEDRE sous le contrat No. 10 SCI F15/L5 et l'Université Libanaise.

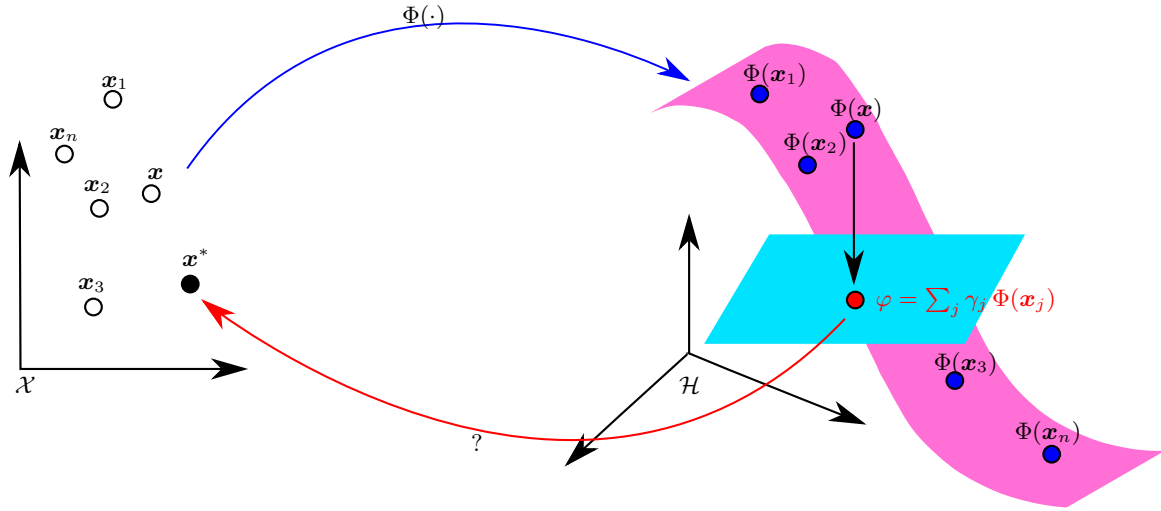


FIGURE 1 – Schéma du problème de pré-image. Le passage de l’espace des observations \mathcal{X} à l’espace transformé \mathcal{H} est accordé par le noyau employé, sans la nécessité d’exhiber la fonction de transfert $\Phi(\cdot)$. Suite à des opérations dans cet espace (illustrées ici par la projection dans un sous-espace), on a recours à résoudre le problème de pré-image, selon lequel on détermine la fonction inverse pour revenir à l’espace des observations, à savoir estimer \mathbf{x}^* à partir de φ telle que $\Phi(\mathbf{x}^*) \approx \varphi$.

image étudiées dans [3, 4]. Mais avant d’en arriver là, il est utile d’introduire les méthodes à noyau et le problème de pré-image.

2 Méthodes à noyau et pré-image

Le concept des méthodes à noyau consiste à transformer des données \mathbf{x}_j , par le biais d’une fonction non-linéaire $\Phi(\cdot)$, de l’espace initial des observations \mathcal{X} vers un espace des caractéristiques \mathcal{H} , donnant ainsi leur image $\Phi(\mathbf{x}_j)$ par cette fonction $\Phi: \mathbf{x}_j \mapsto \Phi(\mathbf{x}_j)$. L’intérêt d’utiliser ces méthodes à noyau réside dans le fait qu’il n’est pas nécessaire d’exhiber cette fonction $\Phi(\cdot)$, permettant par là-même d’élaborer des méthodes de traitement non-linéaires pour une complexité calculatoire réduite. Pour ce faire, on exploite la propriété reproductrice du noyau $\kappa(\cdot, \cdot)$ associé à l’espace des caractéristiques \mathcal{H} . Cette propriété, dite *coup du noyau*, permet de transformer les méthodes linéaires de traitement de données en des méthodes non-linéaires, sous réserve qu’elles puissent s’exprimer en fonction de produits scalaires des observations. Ainsi la structure des algorithmes demeure-t-elle inchangée, et le surcoût calculatoire de la non-linéarité négligeable grâce à l’évaluation des noyaux.

C’est le cas de l’ACP-à-noyau, une extension non-linéaire de l’ACP classique. A l’instar de l’ACP, elle vise à identifier un sous-espace pertinent pour les données en maximisant leur variance projetée. Un tel espace permet d’assurer le débruitage des données, en y projetant chaque $\Phi(\mathbf{x})$. Ceci résulte en une signature de la forme

$$\varphi = \sum_{j=1}^n \gamma_j \Phi(\mathbf{x}_j), \quad (1)$$

en vertu du Théorème de Représentation [7]. Cependant, tout

le traitement étant réalisé dans l’espace des caractéristiques \mathcal{H} , il est nécessaire de faire le retour inverse vers l’espace des données, l’espace du traitement du signal, afin d’y interpréter le résultat et d’évaluer le débruitage comme illustré à la Figure 1.

A l’exception des éléments $\Phi(\mathbf{x}_j)$ de l’ensemble d’apprentissage pour lesquels les antécédents \mathbf{x}_j existent (et sont connus), la plupart des éléments de l’espace \mathcal{H} n’ont pas de pré-images exactes et uniques. Pour résoudre ce problème mal-posé, on a recours à trouver une solution approchée. Il s’agit alors de résoudre un problème d’optimisation, qui consiste à déterminer un antécédent \mathbf{x}^* qui, par la fonction Φ , nous donne l’image la plus proche possible de φ , soit $\mathbf{x}^* \in \mathcal{X}$ tel que $\Phi(\mathbf{x}^*) \approx \varphi$. On écrit alors le problème d’optimisation

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} J(\mathbf{x}),$$

où $J(\mathbf{x}) = \frac{1}{2} \|\varphi - \Phi(\mathbf{x})\|_{\mathcal{H}}^2$ est la fonction coût, et $\|\cdot\|_{\mathcal{H}}$ représente la norme dans l’espace des caractéristiques. En recourant au noyau reproductisant $\kappa(\cdot, \cdot)$ de \mathcal{H} , on peut écrire

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{2} \kappa(\mathbf{x}, \mathbf{x}) - \sum_{j=1}^n \gamma_j \kappa(\mathbf{x}_j, \mathbf{x}), \quad (2)$$

le terme indépendant de \mathbf{x} étant éliminé.

Par ailleurs, il est bien connu que la solution optimale est donnée par l’expression $\mathbf{x}^* = \sum_{i=1}^n \beta_i^* \mathbf{x}_i$ (voir par exemple [3, 4]). On peut alors écrire ce modèle sous forme matricielle, avec $\mathbf{x}^* = \boldsymbol{\beta}^{*T} \mathbf{X}$, où $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]^T$ et $\boldsymbol{\beta}^* = [\beta_1^* \ \beta_2^* \ \cdots \ \beta_n^*]^T$ est le vecteur des coefficients inconnus. Il s’agit alors d’estimer ces coefficients β_k^* optimaux au sens de la minimisation du coût $J(\mathbf{x})$. Bien que la fonction coût ne soit convexe, rien ne s’oppose à résoudre ce problème avec un algorithme de gradient.

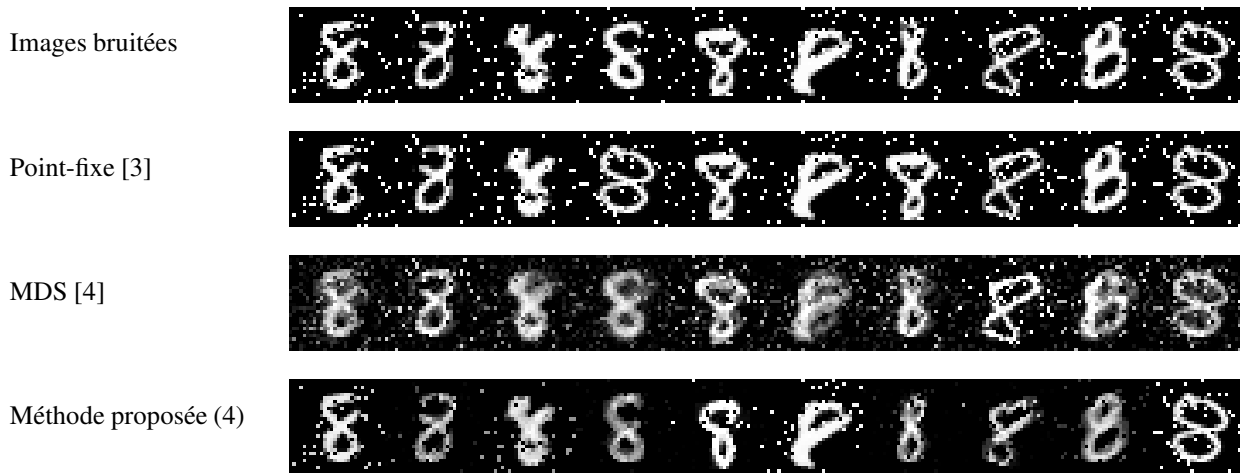


FIGURE 2 – Un ensemble de chiffres “8” bruités (première ligne), sur lequel on a appliqué l’ACP-à-noyau pour le débruitage, la pré-image étant obtenue par l’algorithme point-fixe [3] (seconde ligne), l’approche MDS [4] (troisième ligne), et l’algorithme avec contrainte de non-négativité de la pré-image (dernière ligne), selon la règle itérative (4).

TABLE 1 – Expressions des noyaux les plus utilisés ainsi que les expressions de leur gradient correspondant.

| Noyau | Expression | $\nabla_{\mathbf{x}} J(\mathbf{x})$ |
|------------|---|---|
| Polynomial | $\kappa_p(\mathbf{x}_i, \mathbf{x}_j) = (c + \mathbf{x}_i \cdot \mathbf{x}_j)^p$ | $\sum_{i=1}^n \gamma_i p \kappa_{p-1}(\mathbf{x}_i, \mathbf{x}) \mathbf{x}_i + p \kappa_{p-1}(\mathbf{x}, \mathbf{x}) \mathbf{x}$ |
| Gaussien | $\kappa_G(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{2\sigma^2} \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$ | $\frac{1}{\sigma^2} \sum_{i=1}^n \gamma_i \kappa_G(\mathbf{x}_i, \mathbf{x}) (\mathbf{x}_i - \mathbf{x})$ |

3 Pré-image avec contraintes de non-négativité

Nous proposons de résoudre le problème de pré-image au moyen d’un algorithme de gradient descendant, selon la règle itérative

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \eta_t \nabla_{\mathbf{x}} J(\mathbf{x}(t)), \quad (3)$$

où $\nabla_{\mathbf{x}} J(\mathbf{x})$ est le gradient de $J(\mathbf{x})$ par rapport à \mathbf{x} et η_t est le pas d’adaptation de l’algorithme. La présence de ce paramètre permet de contrôler la convergence de l’algorithme, contrairement à l’approche point-fixe proposée dans [3]. Pour illustrer cette expression, on considère le noyau Gaussien et le noyau Polynomial. Le Tableau 1 résume les expressions de ces deux noyaux, ainsi que les expressions du gradient correspondant.

Pour simplifier la présentation, nous considérons dans un premier temps les contraintes de non-négativité sur la pré-image, avec $\mathbf{x}^* \geq 0$. Le problème d’optimisation avec contraintes est alors défini par

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \geq 0} J(\mathbf{x}).$$

Nous proposons alors une résolution itérative selon

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \eta_t(\mathbf{x}(t)) \nabla_{\mathbf{x}} J(\mathbf{x}(t)), \quad (4)$$

où le pas $\eta_t(\mathbf{x}(t))$, une fonction de $\mathbf{x}(t)$, est optimisé afin de ne pas sortir du domaine admissible de la non-négativité. Comme

préconisé dans [8], on écrit $\eta_t(\mathbf{x}(t)) = \eta_t \text{diag}[\mathbf{x}(t)]$, où $\text{diag}[\mathbf{x}(t)]$ est la matrice diagonale formée par les éléments de $\mathbf{x}(t)$. Une telle formulation du pas permet une convergence rapide vers les valeurs nulles, comme illustré dans la Figure 3 avec des pas plus importants pour les grandes valeurs de \mathbf{x}_t , donnant ainsi en un résultat parcimonieux. En outre, une telle pondération du pas η_t par $\text{diag}[\mathbf{x}(t)]$ permet de simplifier la contrainte imposée sur le pas, où η_t doit satisfaire l’inégalité suivante :

$$\eta_t \leq \min_i \frac{1}{[\nabla_{\mathbf{x}} J(\mathbf{x}(t))]_i}.$$

Utilisant le modèle $\mathbf{x}^* = \sum_{i=1}^n \beta_i^* \mathbf{x}_i$, nous pouvons étendre notre approche à une méthode qui impose la non-négativité des coefficients β_i^* . Le problème d’optimisation avec contraintes s’écrit dans ce cas

$$\beta^* = \arg \max_{\beta \geq 0} J(\beta^T \mathbf{X}).$$

En tenant compte de la linéarité entre \mathbf{x}^* et β^* , on obtient la relation entre les gradients associés, avec

$$\nabla_{\beta} J(\beta^T \mathbf{X}) = \nabla_{\mathbf{x}} J(\mathbf{x}) \mathbf{X}^T.$$

Ainsi obtient-on à partir de (4) la règle suivante

$$\beta(t+1) = \beta(t) - \eta_t \text{diag}[\beta(t)] \nabla_{\mathbf{x}} J(\mathbf{x}(t)) \mathbf{X}^T.$$

On retrouve ici l’algorithme suggéré dans notre précédent travail [9], et dont les performances ont été montrées en traitement de signaux biomédicaux EEG.

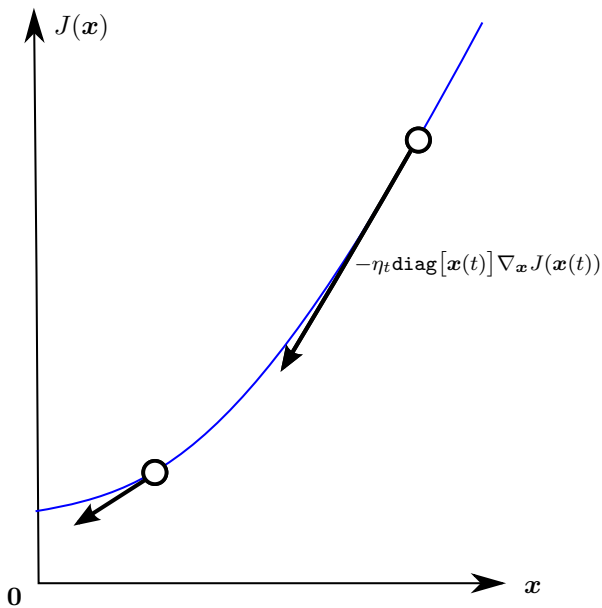


FIGURE 3 – Illustration de la pondération du pas par la valeur de x correspondante.

4 Applications

Dans cette section, l'efficacité de l'approche proposée est illustrée en débruitage d'une collection d'images du chiffre manuscrit "8", de la base de données MNIST¹. Chaque image est de taille 28×28 pixels, codées en niveau de gris, donc avec des valeurs comprises entre 0 et 255. Les images ont été bruitées en ajoutant un bruit de type "Sel-et-Poivre" de densité 0.1. Une collection de 500 images a été exploitée pour l'apprentissage de l'ACP-à-noyau, avec un noyau Gaussien de largeur de bande $\sigma = 500$. Pour construire le sous-espace optimal (illustré par le plan dans la Figure 1), 50 vecteurs propres ont été retenus. Une autre collection de dix images a été utilisée pour tester les différents algorithmes de pré-image.

Nous avons comparé notre méthode de non-négativité sur la pré-image, avec la technique du point fixe [3] définie par

$$\mathbf{x}(t+1) = \frac{\sum_{i=1}^n \gamma_i \kappa_G(\mathbf{x}_i, \mathbf{x}(t)) \mathbf{x}_i}{\sum_{i=1}^n \gamma_i \kappa_G(\mathbf{x}_i, \mathbf{x}(t))},$$

et l'approche MDS [4]. Dans la Figure 2, on trouve les résultats obtenus en appliquant les méthodes du point-fixe (seconde ligne), la MDS (troisième ligne), et l'approche proposée (dernière ligne) sur la série de dix images corrompues par le bruit (première ligne). La méthode proposée donne de meilleurs résultats de débruitage par rapport aux autres techniques. Puisque nous utilisons des méthodes itératives, il est nécessaire d'indiquer le nombre d'itérations pour aboutir aux résultats souhaités. Nous avons appliqué 100 itérations pour la technique du point-fixe, alors que seulement 5 itérations ont été nécessaires pour notre approche.

1. La base de donnée MNIST est accessible via <http://yann.lecun.com/exdb/mnist/>.

5 Conclusion et perspectives

Dans cette communication, nous avons étudié le problème de pré-image sous contraintes de non-négativité. Nous avons proposé une approche itérative pour résoudre ce problème. La pertinence de l'approche proposée est illustrée sur une application de débruitage des images, en la comparant avec des techniques existantes. Nos recherches futures porteront sur une étude de convergence des algorithmes proposés, ainsi qu'une extension de notre approche à des contraintes de type boîte.

Références

- [1] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [2] V. N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, September 1998.
- [3] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Proceedings of the 1998 conference on Advances in neural information processing systems II*, Cambridge, MA, USA, 1999, pp. 536–542, MIT Press.
- [4] J. T. Kwok and I. W. Tsang, "The pre-image problem in kernel methods," in *ICML*, T. Fawcett and N. Mishra, Eds. 2003, pp. 408–415, AAAI Press.
- [5] P. Honeine and C. Richard, "Solving the pre-image problem in kernel machines : a direct method," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Grenoble, France, September 2009.
- [6] P. Honeine and C. Richard, "Pre-image problem in kernel-based machine learning," in *IEEE Signal Processing Magazine*, March 2011, to appear, vol. 28 (2).
- [7] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory*, London, UK, 2001, COLT '01/EuroCOLT '01, pp. 416–426, Springer-Verlag.
- [8] H. Lantéri, C. Theys, F. Benvenuto, and D. Mary, "Méthode algorithmique de minimisation de fonctions d'écart entre champs de données. application à la reconstruction d'images astrophysiques," in *Colloque GRETSI'2009*, Dijon, France, 8-11 septembre 2009.
- [9] M. Kallas, P. Honeine, C. Richard, H. Amoud, and C. Francis, "Nonlinear feature extraction using kernel principal component analysis with non-negative pre-image," in *Proc. 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Buenos Aires, Argentina, 31 Aug. - 4 Sept. 2010.