

Modèle autorégressif non-linéaire à noyau : une première approche

Maya KALLAS^{1,2}, Paul HONEINE¹, Cédric RICHARD³⁺, Clovis FRANCIS², Hassan AMOUD⁴

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

³Laboratoire Fizeau (UMR CNRS 6525, OCA), Université de Nice Sophia-Antipolis, France

⁴Azm Center for Research in Biotechnology and its Applications, Université Libanaise, Liban

+Institut Universitaire de France

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

Résumé – L'analyse et la prédiction de séries temporelles par un modèle autorégressif ont été largement étudiées pour des systèmes linéaires. Toutefois, ce principe s'avère généralement inadapté pour l'analyse des systèmes non-linéaires. L'objectif de cette communication est de proposer un modèle autorégressif non-linéaire dans un espace de Hilbert à noyau reproduisant. On combine, d'une part le principe du coup du noyau qui permet d'estimer les paramètres du modèle, et d'autre part la résolution d'un problème de pré-image pour obtenir la valeur de la prédiction dans l'espace signal. L'approche proposée hérite de la simplicité algorithmique du modèle autorégressif classique, tout en étant non-linéaire par rapport aux échantillons d'entrée. Une comparaison avec différentes méthodes de prédiction non-linéaires illustre les performances du modèle autorégressif non-linéaire proposé sur des séries temporelles test de la littérature.

Abstract – This communication deals with the problem of analysis and prediction using an autoregressive model. Being linear by nature, such modeling is not adapted for many real-life applications, naturally non-linear. We present in this paper a nonlinear autoregressive model, using kernel machines. The proposed approach inherits the simplicity of autoregressive model, yet non-linear. By combining the principle of the kernel trick and the resolution of the pre-image problem required for the interpretation of the data, we derive a new nonlinear prediction scheme. A comparison with several nonlinear prediction methods illustrates the performance of the proposed nonlinear autoregressive model on chaotic time series.

1 Introduction

La modélisation autorégressive (AR) pour l'analyse et la prédiction de séries temporelles a toujours suscité un intérêt important en raison de sa simplicité conceptuelle [1]. Il s'agit d'un modèle de régression linéaire dans lequel le signal est expliqué par ses échantillons passés plutôt que par d'autres valeurs, permettant par là même de prédire des valeurs futures. Bien que plébiscitée dans de nombreuses applications pour son efficacité et sa grande simplicité de mise en œuvre, la modélisation AR n'est par essence pas adaptée à l'analyse de phénomènes non-linéaires.

Dans cet esprit, nous proposons dans cet article d'étendre le principe de ce modèle linéaire à un modèle autorégressif non-linéaire, au jour des méthodes à noyau en reconnaissance des formes [2]. Pour cela, on considère le modèle AR dans un espace de Hilbert à noyau reproduisant (RKHS), où chaque

échantillon est représenté par son image obtenue par une fonction non-linéaire. Dans ce cadre, on revisite le problème d'estimation des paramètres du modèle AR, en l'exprimant en terme de produit scalaire de ces images. En substituant le produit scalaire par une fonction noyau, le *coup du noyau* [3] permet d'identifier les paramètres du modèle non-linéaire, sans qu'il soit nécessaire d'explicitier ses images. Ainsi les données sont-elles implicitement transformées du domaine des échantillons au RKHS, sans un surcoût calculatoire important. Toutefois, le résultat de tout traitement, tel que la prédiction d'un future élément, demeure dans le RKHS.

Quoique le passage à l'espace transformé RKHS permet d'élaborer des techniques non-linéaires à partir de techniques linéaires classiques [4], la fonction inverse, du RKHS au domaine d'observation, est souvent d'intérêt primordial. Cette étape est nécessaire afin d'interpréter le résultat obtenu dans l'espace transformé. Le *problème de la pré-image* consiste à identifier l'élément dont l'image est la plus proche possible de la prédiction dans le RKHS. Différentes techniques ont été proposées pour résoudre le problème de la pré-image dans le

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.

cas général. Voir [5] pour une discussion approfondie avec des applications en traitement du signal. Ce principe, souvent négligé dans la littérature comme par exemple dans [6], est étudié ici pour l'approche proposée. En reportant le résultat dans le domaine des séries temporelles, unidimensionnelles aussi bien que multidimensionnelles, la résolution du problème de pré-image nous offre la possibilité de définir un schéma de prédiction. Cette approche est dénommée modélisation AR-à-noyau, pour l'analyse et la prédiction.

2 Un modèle autorégressif à noyau

En modélisation autorégressive, le modèle de régression est basé sur la prédiction d'un échantillon à partir d'un certain nombre d'échantillons de son passé, selon une simple combinaison linéaire. Soit une série temporelle de n échantillons, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ d'un espace d'observation \mathcal{X} , décrite par un modèle AR d'ordre p , selon

$$\mathbf{x}_i = \sum_{j=1}^p \alpha_{p-j+1} \mathbf{x}_{i-j} + e_i \quad (1)$$

où les constantes α_k , indépendantes de l'instant i , sont les paramètres du modèle, et e_i est un bruit supposé blanc. Différentes méthodes ont été proposées dans la littérature pour le calcul des paramètres du modèle. Nous revisitons ici la méthode des moindres carrés, qui consiste à minimiser l'erreur quadratique de prédiction entre la vraie valeur de la série et la valeur prédite par le modèle, définie par

$$\sum_{i=p+1}^n \left(\mathbf{x}_i - \sum_{j=1}^p \alpha_{p-j+1} \mathbf{x}_{i-j} \right)^2.$$

Les valeurs optimales des paramètres α_i sont données par le passage par zéro de la dérivée de cette expression par rapport à $\alpha_1, \alpha_2, \dots, \alpha_p$. En définissant le vecteur des paramètres $\boldsymbol{\alpha} = [\alpha_p \alpha_{p-1} \dots \alpha_1]^\top$, la condition d'optimalité est alors donnée par

$$\boldsymbol{\alpha} = \left(\sum_{i=p+1}^n \mathbf{X}_i^\top \mathbf{X}_i \right)^{-1} \sum_{i=p+1}^n \mathbf{X}_i^\top \mathbf{x}_i,$$

où $\mathbf{X}_i = [\mathbf{x}_{i-1} \mathbf{x}_{i-2} \dots \mathbf{x}_{i-p}]$ comprend les p échantillons précédents à \mathbf{x}_i . Une fois ce calcul fait, il est possible de prédire des futurs échantillons à partir de leur passé, avec $\sum_{j=1}^p \alpha_{p-j+1} \mathbf{x}_{i-j}$ pour $i = n+1, n+2, \dots$.

Dans le même esprit et par le biais des méthodes à noyau, nous proposons d'étendre le modèle AR à une approche non-linéaire dans un RKHS, en appliquant à chaque échantillon une fonction non-linéaire. Considérons une fonction non-linéaire $\Phi(\cdot)$ de l'espace des observations \mathcal{X} à l'espace RKHS \mathcal{H} qui à chaque $\mathbf{x}_i \in \mathcal{X}$ fait correspondre son image $\Phi(\mathbf{x}_i) \in \mathcal{H}$. Le modèle AR décrit dans le RKHS est alors donné par

$$\Phi(\mathbf{x}_i) = \sum_{j=1}^p \alpha_{p-j+1} \Phi(\mathbf{x}_{i-j}) + \epsilon_i^\Phi. \quad (2)$$

En définissant $\boldsymbol{\varphi}_i = [\Phi(\mathbf{x}_{i-1}) \Phi(\mathbf{x}_{i-2}) \dots \Phi(\mathbf{x}_{i-p})]$, le vecteur regroupant les transformées par la fonction $\Phi(\cdot)$ des p échantillons précédents, on peut écrire notre modèle non-linéaire sous la forme matricielle

$$\Phi(\mathbf{x}_i) = \boldsymbol{\varphi}_i \boldsymbol{\alpha} + \epsilon_i^\Phi.$$

En utilisant aussi la méthode des moindres carrés, les paramètres sont alors déterminés en minimisant l'erreur quadratique moyenne entre la valeur prédite $\sum_{j=1}^p \alpha_{p-j+1} \Phi(\mathbf{x}_{i-j})$ dans le RKHS et la vraie valeur de l'image $\Phi(\mathbf{x}_i)$. Ainsi la fonction coût à minimiser est-elle définie par

$$\sum_{i=p+1}^n \|\Phi(\mathbf{x}_i) - \boldsymbol{\varphi}_i \boldsymbol{\alpha}\|^2,$$

où $\|\cdot\|$ désigne la norme de l'espace RKHS en question. En annulant la dérivée de cette expression par rapport à $\boldsymbol{\alpha}$, nous obtenons les valeurs optimales des paramètres avec

$$\boldsymbol{\alpha} = \left(\sum_{i=p+1}^n \boldsymbol{\varphi}_i^\top \boldsymbol{\varphi}_i \right)^{-1} \sum_{i=p+1}^n \boldsymbol{\varphi}_i^\top \Phi(\mathbf{x}_i).$$

Il est clair que cette expression n'implique que le produit scalaire entre les couples d'images $\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_n)$, ce qui permet de l'évaluer en utilisant simplement une fonction noyau $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}_j)$.

3 Prédiction utilisant une technique de pré-image

Une fois les observations ont été représentés dans le RKHS, et les paramètres déterminés, on peut alors prédire un élément à partir de son passé, avec

$$\boldsymbol{\psi}_i = \sum_{j=1}^p \alpha_{p-j+1} \Phi(\mathbf{x}_{i-j}), \quad (3)$$

pour $i = n+1, n+2, \dots$. Comme illustré dans la Figure 1, cet élément appartenant à l'espace engendré par les images des p précédents échantillons, il est alors nécessaire de revenir au domaine des échantillons. Alors que la fonction $\Phi(\cdot)$ permet de passer de ce domaine au RKHS, la fonction inverse n'existe pas dans le cas général. C'est le problème de la *pré-image* en méthodes d'apprentissage à noyau [5], qui consiste à déterminer l'échantillon \mathbf{x}_i^* dont l'image $\Phi(\mathbf{x}_i^*)$ est la plus proche possible de $\boldsymbol{\psi}_i$. En utilisant la norme définie dans le RKHS, le problème d'optimisation est donné par

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}} \left\| \sum_{j=1}^p \alpha_{p-j+1} \Phi(\mathbf{x}_{i-j}) - \Phi(\mathbf{x}) \right\|^2, \quad (4)$$

où les α_k sont les paramètres déjà calculés, ou encore

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}} \|\boldsymbol{\varphi}_i \boldsymbol{\alpha} - \Phi(\mathbf{x})\|^2.$$

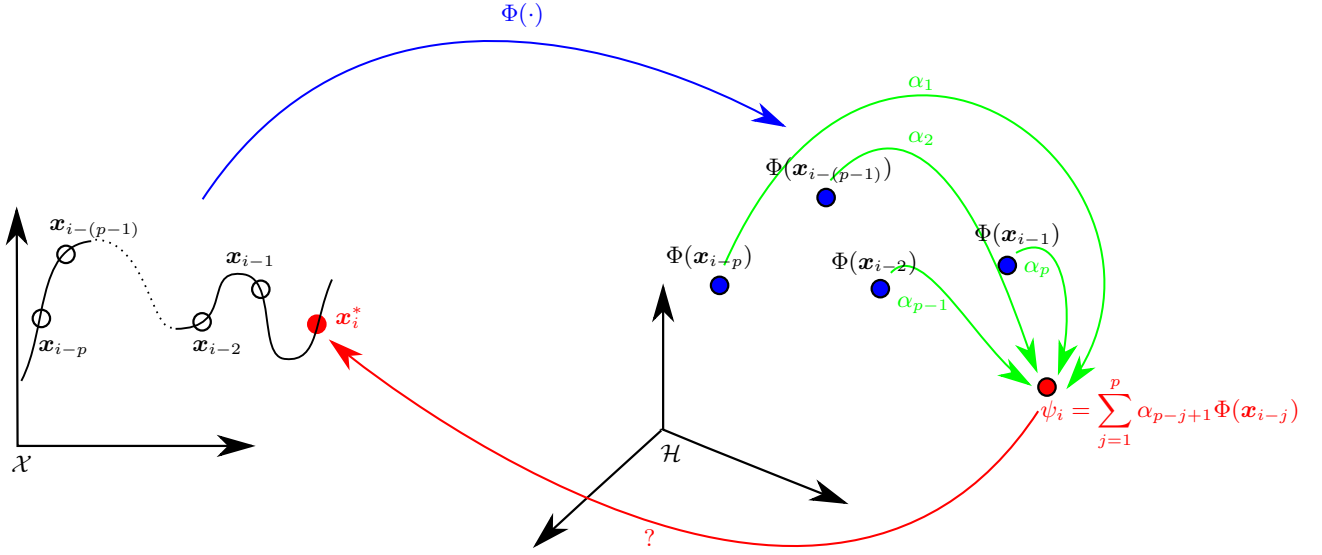


FIGURE 1 – Illustration du modèle AR-à-noyau. Les échantillons \mathbf{x}_i sont projetés de l'espace des observations vers l'espace RKHS, où un modèle AR est défini par l'estimation de ses paramètres α_k . Une technique de pré-image est alors nécessaire pour prédire un futur échantillon dans l'espace initial.

On propose de résoudre ce problème par une technique de descente de gradient. Afin de simplifier cette présentation, on considère le noyau le plus utilisé dans la littérature qui est le noyau Gaussien, défini par

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{1}{2\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2},$$

où le paramètre $\sigma > 0$ désigne sa largeur de bande. Dans ce cas, le problème d'optimisation devient

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}} - \sum_{j=1}^p \alpha_{p-j+1} e^{-\frac{1}{2\sigma^2} \|\mathbf{x}_{i-j} - \mathbf{x}\|^2}.$$

En désignant par $J_i(\mathbf{x})$ la fonctionnelle ci-dessus, son gradient par rapport à \mathbf{x} est donné par

$$\nabla_{\mathbf{x}} J_i(\mathbf{x}) = \frac{1}{\sigma^2} \sum_{j=1}^p \alpha_{p-j+1} (\mathbf{x}_{i-j} - \mathbf{x}) e^{-\frac{1}{2\sigma^2} \|\mathbf{x}_{i-j} - \mathbf{x}\|^2}.$$

En annulant ce gradient à l'optimum \mathbf{x}_i^* , on obtient l'expression suivante qui peut être résolue par une méthode de point-fixe :

$$\mathbf{x}_i^* = \left(\sum_{k=1}^p \alpha_{p-k+1} e^{-\frac{1}{2\sigma^2} \|\mathbf{x}_{i-k} - \mathbf{x}_i^*\|^2} \right)^{-1} \times \sum_{j=1}^p \alpha_{p-j+1} e^{-\frac{1}{2\sigma^2} \|\mathbf{x}_{i-j} - \mathbf{x}_i^*\|^2} \mathbf{x}_{i-j}.$$

Le modèle AR-à-noyau se prête à une interprétation dans le cas du noyau Gaussien, et en général pour les fonctions de base radiale (noyaux RBF). Dans ce cas, la fonction $\Phi(\cdot)$ correspondante consiste en une transformation de chaque échantillon en une fonction Gaussienne centrée en ce point. Le modèle AR défini dans (3) est alors une combinaison linéaire de fonctions Gaussiennes. Or, une somme (pondérée) de

Gaussiennes ne peut pas s'exprimer sous forme d'une unique fonction gaussienne. Comme défini dans (4), le problème de pré-image consiste alors à déterminer le vecteur moyenne de la fonction gaussienne la plus proche de cette combinaison de Gaussiennes.

4 Applications

Pour montrer l'intérêt du modèle AR-à-noyau, nous avons étudié ses performances en prédiction sur 4 séries temporelles classiques [7, 8] : deux séries unidimensionnelles, celle du *Laser* proposée par la compétition Santa Fe et la *MG30* qui modélise l'évolution de la production des cellules sanguines, et deux séries chaotiques multidimensionnelles, la carte *Ikeda* liée à une dynamique du laser et l'attracteur de *Lorenz* défini par la solution d'un système d'équations différentielles.

Deux expériences ont été réalisées sur chacune des séries, avec $n = 50$ et $n = 300$. Nous avons utilisé n échantillons pour l'apprentissage et autant de données pour évaluer les performances du modèle. La phase d'apprentissage permet d'estimer les paramètres du modèle, l'ordre optimal (p), ainsi que la largeur de bande du noyau Gaussien (σ). Le Tableau 1 présente les valeurs de la largeur de bande σ choisies pour chacune des séries temporelles et utilisées pour chaque expérience. La valeur optimale de l'ordre a été $p = 3$ pour l'attracteur de *Lorenz* et $p = 6$ pour toutes les autres séries. Pour étudier les performances du modèle obtenu, nous avons calculé l'erreur quadratique moyenne de prédiction qui est estimée sur les échantillons $n + 1, n + 2, \dots, 2n$, et est donnée par l'expression

$$\frac{1}{n} \sum_{t=n+1}^{2n} \|\mathbf{x}_t^* - \mathbf{x}_t\|^2,$$

TABLE 1 – Valeurs de la largeur de bande σ du noyau Gaussien pour chaque série temporelle pour les deux expériences.

	<i>Laser</i>	<i>MG₃₀</i>	<i>Ikeda</i>	<i>Lorenz</i>
1 ^{ère} expérience : $n = 300$	0.3	0.015	0.0046	0.04
2 ^{nde} expérience : $n = 50$	0.4	0.035	0.0255	0.03

TABLE 2 – Erreur quadratique moyenne pour différentes méthodes de prédiction et plusieurs séries temporelles

	n	<i>Laser</i>	<i>MG₃₀</i>	<i>Ikeda</i>	<i>Lorenz</i>
perceptron multicouche	300	1.4326	0.0461	0.00071	0.2837
régression à vecteurs de support	300	0.2595	0.0313	0.00081	0.1811
filtre de Kalman non-linéaire	300	0.2325	0.0307	0.00077	0.3133
modèle AR-à-noyau	300	0.0702	0.0008	0.00088	0.1792
modèle AR-à-noyau	50	0.1813	0.0049	0.00530	0.2625

où x_t^* est la valeur prédite par notre approche à l'instant t , et x_t est la vraie valeur de la série au même instant.

Nous avons comparé les résultats obtenus par notre méthode avec d'autres méthodes de prédiction non-linéaires, pour $n = 300$ échantillons d'apprentissage : le perceptron multicouche, la régression à vecteurs de support et le filtre de Kalman non-linéaire. Ces méthodes sont décrites en détail dans [8]. Une comparaison avec ces différentes méthodes est donnée dans le tableau 2. Même avec une faible taille de l'ensemble d'apprentissage, avec $n = 50$ échantillons, l'approche proposée présente de très bons résultats, à l'exception des données *Ikeda*. Il est clair que pour le même nombre d'échantillons d'apprentissage, notre méthode donne de meilleurs résultats, et l'amélioration des performances est considérable par rapport aux autres méthodes. On note aussi la simplicité algorithmique de l'AR-à-noyau, propriété héritée du modèle linéaire, où l'estimation des paramètres du modèle ne nécessite qu'une inversion de matrice de taille $p \times p$. Les autres méthodes non-linéaires souffrent d'une charge calculatoire conséquente.

5 Conclusion et perspectives

Dans cette communication, une première approche non-linéaire du modèle autorégressif a été proposée, en se basant sur les méthodes à noyau. Pour ce faire, un noyau reproduisant a été utilisé pour projeter les observations dans l'espace RKHS. Nous avons montré que les paramètres du modèle peuvent être estimés grâce au coup du noyau. Une technique de pré-image a été proposée pour prédire les futurs échantillons dans l'espace initial des séries temporelles. Une comparaison avec d'autres méthodes non-linéaires présentes dans la littérature sur plusieurs séries temporelles et chaotiques montre la pertinence de l'approche proposée.

Nos recherches futures porteront sur divers aspects, tout d'abord l'implémentation d'une technique pour le choix de l'ordre optimal du modèle AR-à-noyau, à l'instar du critère

d'information d'Akaike, ensuite la mise en œuvre d'autres méthodes pour l'estimation des paramètres telles que la méthode de Yule-Walker et la technique de Levinson-Durbin, et enfin une comparaison de performance entre les différentes techniques existantes de pré-image.

Références

- [1] R. S. Tsay, *Analysis of Financial Time Series (Wiley Series in Probability and Statistics)*, Wiley-Interscience, 2nd edition, Aug. 2005.
- [2] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [3] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," in *Automation and Remote Control*, 1964, number 25, pp. 821–837.
- [4] V. N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, September 1998.
- [5] P. Honeine and C. Richard, "The pre-image problem in kernel-based machine learning," *IEEE Signal Processing Magazine, special issue on "dimensionality reduction via subspace and manifold learnin"*, vol. 28 (2), March 2011.
- [6] R. Kumar and C. V. Jawahar, "Kernel approach to autoregressive modeling," in *The 13th National Conference on Communications (NCC)*, Kanpur, India, January 2007.
- [7] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines," in *IEEE Workshop on Neural Networks for Signal Processing VII*, 1997, pp. 511–519, IEEE Press.
- [8] L. Ralaivola and F. D'alche-Buc, "Time series filtering, smoothing and learning using the kernel kalman filter," in *Proc. IEEE International Joint Conference on Neural Networks*, 2005, vol. 3, pp. 1449–1454.