

Apprentissage non-linéaire en ligne dans les réseaux de capteurs

Cédric RICHARD, Paul HONEINE, Mehdi ESSOLOH, Hichem SNOUSSI

Institut Charles Delaunay (FRE CNRS 2848) - LM2S - Université de Technologie de Troyes
12 rue Marie Curie, BP 2060, 10010 Troyes cedex - France

prenom.nom@utt.fr

Résumé – Cet article présente plusieurs stratégies d'apprentissage en ligne d'une fonctionnelle non-linéaire dans un réseau de capteurs sans fil. Le problème d'apprentissage est défini dans le cadre des espaces de Hilbert à noyau reproduisant. L'estimation de la fonctionnelle y est pratiquée de manière distribuée sur la base des mesures locales collectées en chacun des capteurs, selon un mode coopératif de type incrémental ou par diffusion de l'information entre les nœuds du réseau. Les approches proposées sont testées sur des données synthétiques de diffusion de chaleur. Elles démontrent une excellente capacité de suivi des évolutions du système, tout en affichant des coûts calculatoires réduits.

Abstract – This paper deals with nonlinear online learning strategies in wireless sensor networks. The learning problem is solved in a reproducing kernel Hilbert space. Functional estimation is performed distributively using local measurements collected by sensors. Each node communicates with all its neighbors as dictated by the incremental or diffusion modes of cooperation. The algorithms are validated with synthetic data governed by the heat conduction equation. They provide accurate tracking results with low computational cost.

1 Introduction

Le domaine des réseaux de capteurs sans fil fait actuellement l'objet d'un intérêt considérable de la part des communautés académique et industrielle. La dispersion d'une multitude de capteurs bon marché dans une région donnée, l'élaboration d'un protocole de routage adéquat, et une implémentation algorithmique efficace, ouvrent en effet de nombreuses perspectives d'applications civiles et militaires.

Dans cet article, on s'intéresse au problème d'identification distribuée d'un champ, par exemple de température ou de concentration d'une espèce chimique, et le suivi de son évolution au cours du temps. Le mode de calcul distribué est inhérent au caractère réparti des nœuds du réseau sur l'aire surveillée, dont la tâche est d'acquérir et traiter localement les mesures. L'efficacité de la procédure d'identification est conditionnée par les interactions des nœuds, dictées par la topologie du réseau. Deux principes de coopération ont été principalement considérés dans la littérature, illustrés par la figure 1. En mode incrémental, l'information transite de façon séquentielle et cyclique d'un nœud voisin à l'autre. Le coût énergétique, largement dicté par le volume des communications, tend à être minimum [6, 7]. En mode de diffusion, chaque nœud coopère avec l'ensemble de ses voisins pour une meilleure qualité d'estimation [5]. Il en résulte un nombre accru d'échanges d'informations, qu'il est possible de restreindre en limitant artificiellement le nombre de voisins concernés. On parle dans ce cas de diffusion probabiliste.

La complexité des applications considérées ici, telles que la diffusion de chaleur qui illustre la suite de cet article, nécessite de recourir à des méthodes d'identification adéquates. Fondées sur les travaux précurseurs d'Aronszajn [1], les récentes avancées de l'estimation fonctionnelle basée sur les espaces de Hil-

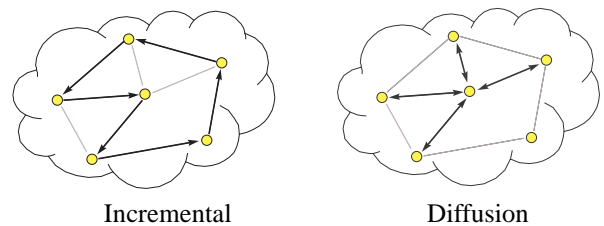


FIGURE 1 – Modes de coopération inter-capteurs

bert à noyau reproduisant, et de la théorie de la régularisation, ont apporté des réponses convaincantes aux problèmes impliquant des systèmes non-linéaires. Les deux stratégies d'apprentissage en ligne considérées, reposant sur un mode coopératif de type incrémental ou par diffusion, démontrent d'excellentes capacités de suivi des évolutions du système tout en affichant un coût calculatoire réduit. Ce travail fait suite à une étude portant uniquement sur le mode incrémental [3, 4]. Si plusieurs approches similaires existent dans la littérature, l'article présenté tire toute son originalité de l'usage du formalisme des espaces de Hilbert à noyau reproduisant. Il en résulte des modèles non-linéaires et des algorithmes fonctionnels adaptés à l'analyse de phénomènes complexes, pour un coût calculatoire réduit.

L'article est organisé ainsi. Dans une première partie, on rappelle les fondements de l'estimation fonctionnelle dans les espaces de Hilbert à noyau reproduisant. Dans une deuxième et troisième parties, on reformule et résout le problème sous forme distribuée dans le contexte des réseaux de capteurs coopérant en mode incrémental et par diffusion de l'information. Les algorithmes sont mis en œuvre sur des données synthétiques dans une quatrième et dernière partie.

2 Formulation du problème

Soit \mathcal{H} un espace de Hilbert à noyau reproduisant constitué de fonctions à valeurs réelles sur un compact $\mathcal{X} \subset \mathbb{R}^\ell$. On désigne par $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ son produit scalaire. Soit $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ le noyau de cet espace, qui vérifie donc la propriété reproductrice suivante : $\psi(\mathbf{x}_i) = \langle \psi, \kappa(\mathbf{x}_i, \cdot) \rangle_{\mathcal{H}}$ pour toute fonction ψ de \mathcal{H} et tout \mathbf{x}_i de \mathcal{X} . Cette propriété étant vérifiée par le noyau lui-même, on a évidemment

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \kappa(\mathbf{x}_i, \cdot), \kappa(\mathbf{x}_j, \cdot) \rangle_{\mathcal{H}}, \quad (1)$$

ce qui justifie l'appellation de noyau reproduisant. Le noyau gaussien, donné par $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ avec σ la largeur de bande, en est un exemple.

Étant donné un ensemble $\{(\mathbf{x}_n, \mathbf{d}_n)\}_{n=1, \dots, N}$ de données d'apprentissage i.i.d. régies par une loi de probabilité inconnue, on considère le problème de minimisation de l'erreur quadratique entre les sorties du modèle $\psi(\mathbf{x}_n)$ et les réponses souhaitées \mathbf{d}_n . Ce problème d'optimisation se formule et se réécrit :¹

$$\psi^o = \arg \min_{\psi \in \mathcal{H}} E|\psi(\mathbf{x}_n) - \mathbf{d}_n|^2 \quad (2)$$

$$= \arg \min_{\psi \in \mathcal{H}} E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2 \quad (3)$$

avec $\kappa_{\mathbf{x}_n} = \kappa(\mathbf{x}_n, \cdot)$. Le Théorème de Représentation [9] établit que la solution de ce problème peut être recherchée dans l'espace engendré par les N fonctions noyau $\kappa_{\mathbf{x}_i}$, soit

$$\psi^o = \sum_{i=1}^N \alpha_i^o \kappa_{\mathbf{x}_i}. \quad (4)$$

Il en résulte que l'optimum $\alpha^o = [\alpha_1^o \dots \alpha_N^o]^\top$ est solution des équations normales

$$R_x \alpha^o = r_{dx}, \quad (5)$$

avec $R_x = E \mathbf{k}_n \mathbf{k}_n^\top$ et $r_{dx} = E \mathbf{k}_n \mathbf{d}_n$. Dans ces expressions, le vecteur \mathbf{k}_n désigne $[\kappa(\mathbf{x}_n, \mathbf{x}_1) \dots \kappa(\mathbf{x}_n, \mathbf{x}_N)]^\top$. On suppose par ailleurs dans (5) que $E \mathbf{k}_n = 0$ et $E \mathbf{d}_n = 0$, ce qui peut nécessiter un centrage préalable des données concernées.

3 Mode de coopération incrémental

Dans le contexte de l'apprentissage distribué dans un réseau de N capteurs, on s'intéresse ici à la modélisation d'un phénomène physique tel que la diffusion d'un champ de température en fonction des coordonnées du plan défini par \mathcal{X} . Chaque nœud n accède à des réalisations temporelles $\{x_n, d_{n,i}\}$ du couple $\{\mathbf{x}_n, \mathbf{d}_n\}$. Le critère de moindres carrés

$$J(\psi) = \sum_{n=1}^N E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2 \quad (6)$$

peut être décomposé en une somme de N critères $J_n(\psi)$ individuels, un en chaque nœud n ,

$$J(\psi) = \sum_{n=1}^N J_n(\psi) \quad \text{avec} \quad J_n(\psi) = E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2. \quad (7)$$

1. Les variables et vecteurs aléatoires sont notés en gras.

Pour déterminer la solution ψ^o , une méthode simple de descente de gradient peut être mise en œuvre. Elle s'exprime ainsi

$$\psi_i = \psi_{i-1} - \frac{\mu}{2} \sum_{n=1}^N \nabla J_n(\psi_{i-1}), \quad (8)$$

où $\nabla J_n(\psi_{i-1}) = 2E(\psi_{i-1}(\mathbf{x}_n) - \mathbf{d}_n) \kappa_{\mathbf{x}_n}$. En remplaçant l'espérance mathématique par la grandeur instantanée correspondante, il en résulte l'algorithme LMS *fonctionnel* suivant :

Stratégie incrémentale :

A chaque instant i , répéter

1. $\phi_{0,i} = \psi_{i-1}$
2. $\phi_{n,i} = \phi_{n-1,i} - \mu(\psi_{i-1}(\mathbf{x}_n) - \mathbf{d}_{n,i}) \kappa_{\mathbf{x}_n}$, $n = 1, \dots, N$
3. $\psi_i = \phi_{N,i}$

Le réseau est parcouru de manière répétée afin de suivre l'évolution du phénomène physique au cours temps. Cette approche nécessite, par l'instruction 2., que chaque nœud ait accès à l'information globale ψ_{i-1} . Afin d'éviter cet inconvénient, il est possible d'évaluer ∇J_n localement en $\phi_{n-1,i}$ reçu du précédent capteur. Ceci conduit à la mise-à-jour

$$\phi_{n,i} = \phi_{n-1,i} - \mu(\phi_{n-1,i}(\mathbf{x}_n) - \mathbf{d}_{n,i}) \kappa_{\mathbf{x}_n}. \quad (9)$$

On note que le modèle est malheureusement enrichi d'un terme à chaque rencontre d'un nouveau capteur, limitant en l'état son application à des réseaux peu denses. Il est toutefois possible d'élaborer des modèles d'ordre réduit en considérant un critère supplémentaire de parcimonie comme dans [3, 4, 8]. Aussi préconise-t-on, lors de la visite du capteur n à l'instant i , d'élaborer un modèle d'ordre réduit de la forme

$$\phi_{n,i} = \sum_{j \in \mathcal{D}} \alpha_{\omega_j} \kappa_{\mathbf{x}_{\omega_j}} + \alpha_n \kappa_{\mathbf{x}_n} \quad (10)$$

où \mathcal{D} est un sous-ensemble d'indices de $\{1, \dots, N\}$ appelé dictionnaire. Il convient à présent de préciser le critère de parcimonie utilisé, à-même de limiter l'ordre du modèle sans en compromettre les performances.

On a proposé, dans une précédente étude, un critère de parcimonie reposant uniquement sur la topologie du réseau [3]. Celui-ci garantit que les capteurs \mathbf{x}_{ω_j} sélectionnés reflètent une couverture raisonnable de la région scrutée. On déplore toutefois que la constitution de ce dictionnaire ne soit jamais remise en question, quelles que soient les évolutions du système et l'erreur de modélisation commise, étant donnée l'unique dépendance du critère vis-à-vis des entrées \mathbf{x}_i . Pour y remédier, on considère ici un critère angulaire dit de cohérence entre les fonctionnelles $\phi_{n,i}$ et $\phi_{n,i}^\perp$, où cette dernière désigne la projection orthogonale de $\phi_{n,i}$ sur le sous-espace $\mathcal{H}_{\mathcal{D}}$ engendré par les fonctions noyau désignées par \mathcal{D} . Ainsi propose-t-on de conserver la fonction candidate $\kappa_{\mathbf{x}_i}$ dans le modèle si

$$\text{coh}(\phi_{n,i}, \phi_{n,i}^\perp) = \frac{|\langle \phi_{n,i}, \phi_{n,i}^\perp \rangle_{\mathcal{H}}|}{\|\phi_{n,i}\|_{\mathcal{H}} \|\phi_{n,i}^\perp\|_{\mathcal{H}}} \leq \eta, \quad (11)$$

où η est un seuil positif déterminant le niveau de parcimonie souhaité. Il convient d'évaluer $\phi_{n,i}^\perp = \sum_{j \in \mathcal{D}} \beta_{\omega_j} \kappa(\mathbf{x}_{\omega_j}, \cdot)$. Ce dernier est défini par les β_{ω_j} minimisant $\|\phi_{n,i} - \phi_{n,i}^\perp\|_{\mathcal{H}}^2$, soit

$$\min_{\beta} \|\alpha_n \kappa_{\mathbf{x}_n} - \sum_{j \in \mathcal{D}} (\beta_{\omega_j} - \alpha_{\omega_j}) \kappa_{\mathbf{x}_{\omega_j}}\|_{\mathcal{H}}^2. \quad (12)$$

En annulant la dérivée de la forme quadratique ci-dessus par rapport aux variables β_{ω_j} , on aboutit enfin à

$$\beta = \alpha + \alpha_n \mathbf{K}_{n,i}^{-1} \mathbf{k}_{n,i} \quad (13)$$

avec α , β et $\mathbf{k}_{n,i}$ les vecteurs d'éléments α_{ω_j} , β_{ω_j} , et $\kappa(\mathbf{x}_{\omega_j}, \mathbf{x}_n)$. La matrice $\mathbf{K}_{n,i}$ est de terme général $\kappa(\mathbf{x}_{\omega_j}, \mathbf{x}_{\omega_k})$. L'algorithme proposé s'exprime finalement ainsi.

Stratégie incrémentale avec contrôle de l'ordre du modèle :

A chaque instant i , répéter

1. $\phi_{0,i} = \psi_{i-1}$
2. $\phi_{n,i} = \phi_{n-1,i} - \mu(\phi_{n-1,i}(\mathbf{x}_n) - d_{n,i}) \kappa_{\mathbf{x}_n}$, $n = 1, \dots, N$
si $\text{coh}(\phi_{n,i}, \phi_{n,i}^\perp) > \eta$, remplacer $\phi_{n,i}$ par $\phi_{n,i}^\perp$
3. $\psi_i = \phi_{N,i}$

Le test dans 2., lorsqu'il est activé, a pour effet de réduire l'ordre du modèle en retirant $\kappa_{\mathbf{x}_n}$ du développement. Avant de clore cette section, notons qu'il s'agissait ici d'un mode de coopération incrémental évoqué en introduction et illustré par la figure 1. La coopération par diffusion fait l'objet de la suite.

4 Mode de coopération par diffusion

Lorsque d'avantage de ressources en communication sont disponibles, il peut être intéressant d'exploiter un schéma de coopération plus sophistiqué. En mode de diffusion, chaque nœud interagit avec l'ensemble de ses voisins pour une meilleure qualité d'estimation [2, 5]. Plus précisément, à chaque instant i , chaque nœud n accomplit trois tâches. Il utilise les mesures acquises localement afin de procéder à une mise-à-jour provisoire. Il consulte également les nœuds voisins avec lesquels il collabore, noté \mathcal{N}_k , pour acquérir leurs propres estimées et/ou mesures brutes. Finalement, il agrège l'ensemble de ces informations afin de mettre à jour son estimée. Afin de mettre cette stratégie en équation, on considère comme précédemment le critère (6) dont on recherche l'optimum ψ^o . Cette fonction coût peut être décomposée comme suit, en chacun des nœuds k du réseau

$$J(\psi) = J_k(\psi) + \sum_{\substack{n=1 \\ n \neq k}}^N J_n(\psi) \quad (14)$$

avec $J_k(\psi) = \sum_{n \in \mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2$. Les paramètres $c_{n,k}$ sont les composantes d'une matrice C , librement choisies par l'utilisateur sous les contraintes $c_{n,k} = 0$ si $n \notin \mathcal{N}_k$, $C\mathbb{1} = \mathbb{1}$ et $\mathbb{1}^\top C = \mathbb{1}^\top$ sous peine de modifier l'énoncé du problème de minimisation associé à (6). Soit $\psi_n^o = \arg \min_{\psi \in \mathcal{H}} J_n(\psi)$. On démontre que $J_k(\psi)$ peut

s'écrire $\sum_{n \in \mathcal{N}_k} \|\psi - \psi_n^o\|_{\mathcal{H}_n}^2 + \beta$ avec β une constante, après avoir complété un carré de sorte à aboutir à une identité remarquable et au prix d'une modification locale de la métrique repérée ici par $\|\cdot\|_{\mathcal{H}_n}$. Dans ces conditions, le critère (14) s'exprime également sous une forme locale

$$J_k^\ell(\psi) = \sum_{n \in \mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2 + \sum_{\substack{n=1 \\ n \neq k}}^N \|\psi - \psi_n^o\|_{\mathcal{H}_n}^2$$

Minimiser $J_k^\ell(\psi)$ nécessite que le nœud k ait accès à l'ensemble des optimums locaux ψ_n^o et des métriques $\|\cdot\|_{\mathcal{H}_n}$. On choisit de relaxer la fonction coût ainsi afin de limiter le volume des communications

$$J_k^r(\psi) = \sum_{n \in \mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\mathbf{x}_n} \rangle_{\mathcal{H}} - \mathbf{d}_n|^2 + \sum_{n \in \mathcal{N}_k / \{k\}} b_{n,k} \|\psi - \psi_n^o\|_{\mathcal{H}}^2 \quad (15)$$

où ψ_n^o désigne à présent la solution disponible au nœud n et à l'instant courant, dans le processus itératif d'optimisation qui va être présenté. Une méthode de descente de gradient peut être mise en œuvre pour minimiser (15). Elle s'exprime ainsi

$$\psi_{k,i} = \psi_{k,i-1} - \frac{\mu}{2} \nabla J_k^r(\psi_{k,i-1}) \quad (16)$$

avec

$$\begin{aligned} \nabla J_k^r(\psi_{k,i-1}) &= \sum_{n \in \mathcal{N}_k} 2 c_{n,k} E(\psi_{k,i-1}(\mathbf{x}_n) - \mathbf{d}_n) \kappa_{\mathbf{x}_n} \\ &+ \sum_{n \in \mathcal{N}_k / \{k\}} 2 b_{n,k} (\psi_{k,i-1} - \psi_n^o) \end{aligned} \quad (17)$$

Les méthodes d'optimisation itératives s'avèrent utiles pour minimiser une somme de fonctions coût convexes. Elles consistent à itérer selon chaque sous-gradient, dans un ordre pré-défini. S'agissant du gradient (17), on note qu'il est constitué d'une première composante visant à mettre le modèle local à jour à partir des mesures, et d'une seconde composante l'agrégant aux modèles locaux voisins. Dans [2], les auteurs suggèrent deux stratégies distinctes suivant l'ordre adopté pour l'adaptation et l'agrégation. On présente ci-dessous leurs homologues selon le formalisme des espaces de Hilbert à noyau reproduisant.

Stratégie de diffusion (adaptation/agrégation) :

A chaque instant i , répéter

1. $\phi_{k,i} = \psi_{k,i-1} - \mu \sum_{n \in \mathcal{N}_k} c_{n,k} (\psi_{k,i-1}(\mathbf{x}_n) - d_{n,i}) \kappa_{\mathbf{x}_n}$
2. $\psi_{k,i} = \sum_{n \in \mathcal{N}_k} b_{n,k} \phi_{n,i}$

Stratégie de diffusion (agrégation/adaptation) :

A chaque instant i , répéter

1. $\phi_{k,i-1} = \sum_{n \in \mathcal{N}_k} b_{n,k} \psi_{n,i-1}$
2. $\psi_{k,i} = \phi_{k,i-1} - \mu \sum_{n \in \mathcal{N}_k} c_{n,k} (\phi_{k,i-1}(\mathbf{x}_n) - d_{n,i}) \kappa_{\mathbf{x}_n}$

Avant de poursuivre, notons que l'ordre de chacun des modèles locaux est dicté par la taille des voisinages considérés

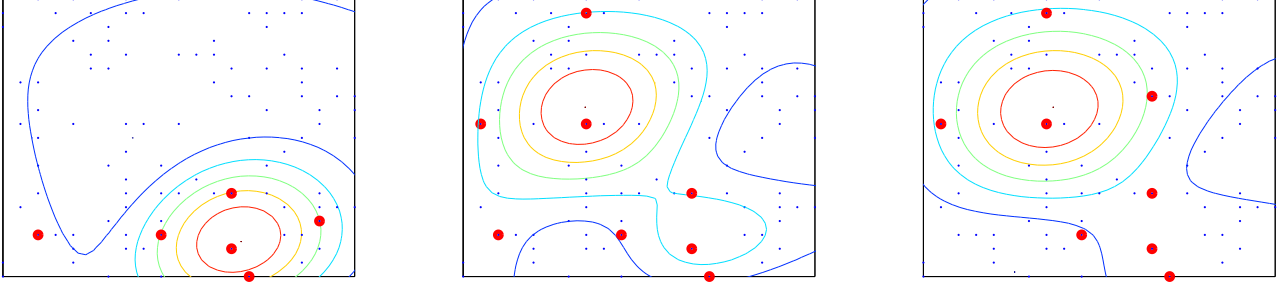


FIGURE 2 – Distribution de température estimée par la méthode LMS incrémentale à $t = 100$ (gauche), $t = 150$ (centre) et $t = 200$ (droite). Les capteurs appartenant au dictionnaire sont marqués d'un point rouge, les autres d'un point bleu.

pour chacun des nœuds. Aussi la nécessité d'un contrôle de celui-ci par un test de parcimonie n'est-il pas aussi pressant que dans le cas d'une stratégie incrémentale. La prochaine section est consacrée aux expérimentations, où l'on compare succinctement les modes de coopération incrémental et par diffusion.

5 Simulations

Pour illustrer la pertinence des algorithmes proposés, on s'intéresse à l'estimation d'un champ de température régi par l'équation différentielle suivante :

$$\frac{\partial \Theta(\mathbf{x}, t)}{\partial t} - c \nabla_{\mathbf{x}}^2 \Theta(\mathbf{x}, t) = Q(\mathbf{x}, t)$$

où $\Theta(\mathbf{x}, t)$ est la température dépendant de la position et du temps, $\nabla_{\mathbf{x}}^2 \Theta(\mathbf{x}, t)$ est l'opérateur spatial de Laplace et $Q(\mathbf{x}, t)$ correspond à la chaleur ajoutée. On considère une surface, de conductivité thermique $c = 0.1$, sur laquelle 100 capteurs sont uniformément tirés sur une grille de taille 21×21 . Deux sources de chaleur de 200 W sont activées successivement, la première des instants $t = 1$ à $t = 100$ (située dans le quart-plan inférieur-droit), la seconde de $t = 100$ à $t = 200$ (située dans le quart-plan supérieur-gauche). L'objectif de la simulation est, étant données les mesures

$$d_{n,i} = \Theta(\mathbf{x}_n, t_i) + \varepsilon_{n,i} \quad (18)$$

avec $\varepsilon_{n,i}$ un bruit blanc gaussien centré de variance 10^{-4} , d'estimer $\Theta(\mathbf{x}, t_i)$ via $\psi_i(\mathbf{x})$. Les simulations ont été réalisées avec un noyau gaussien de largeur de bande $\sigma = 0.5$, et un pas de gradient $\mu = 0.5$. Pour la méthode incrémentale, le seuil de cohérence η a été fixé à 0.985. Pour la méthode de diffusion, la stratégie adaptation/agrégation a été ici retenue en accordant un même poids à l'ensemble des voisins de chaque nœud, c'est-à-dire $b_{n,k} = c_{n,k} = 1/\text{card}(\mathcal{N}_k)$. La notion de voisinage a été définie par seuillage de la distance séparant les nœuds.

Comme l'indique la figure 2, l'algorithme incrémental suit parfaitement les évolutions de la distribution de chaleur par un choix approprié des capteurs représentés dans le modèle et cerclés ici de rouge. On rappelle que leur nombre correspond à l'ordre du modèle. La convergence de l'erreur quadratique moyenne normalisée de prédiction est illustrée par la figure 3.

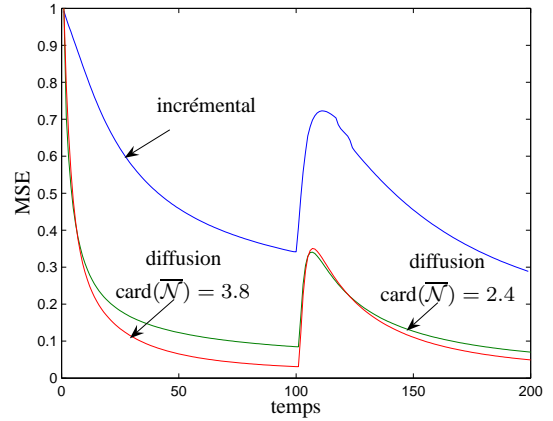


FIGURE 3 – Evolution de l'erreur quadratique moyenne normalisée de prédiction, pour les stratégies incrémentale et de diffusion. La taille moyenne des voisinages est indiquée pour cette dernière.

Il est à noter la croissance de celle-ci à l'instant $t = 100$, provoquée par l'extinction de la première source et l'allumage de la seconde. Elle est comparée à celle obtenue avec la méthode procédant par diffusion d'information. Cette dernière s'avère plus performante, au prix d'un volume de communications plus élevé même si la taille moyenne des voisinages reste faible.

Références

- [1] N. Aronszajn, *Theory of reproducing kernels*. Trans. Amer. Math. Soc., vol. 68, pp. 337-404, 1950.
- [2] F. Cattivelli and A. Sayed, *Diffusion LMS algorithms with information exchange*. Proc. Asilomar'08, Pacific Grove, USA, 2008.
- [3] P. Honeine, M. Essoloh, C. Richard, and H. Snoussi, *Distributed regression in sensor networks with a reduced-order kernel model*. Proc. IEEE Globecom'08, New Orleans, USA, 2008.
- [4] P. Honeine, C. Richard, J. C. M. Bermudez, M. Essoloh, H. Snoussi, and F. Vincent, *Functional estimation in Hilbert space for distributed learning in wireless sensor networks*. Proc. IEEE ICASSP'09, Taipei, Taiwan, 2009.
- [5] C. Lopes and A. Sayed, *Diffusion least-mean squares over adaptive networks*. Proc. IEEE ICASSP'07, Hawaii, USA, 2007.
- [6] C. Lopes and A. Sayed, *Incremental adaptive strategies over distributed networks*. IEEE Transactions Signal Processing, Vol. 55(8), pp. 4064-4077, 2007.
- [7] M. Rabbat and R. Nowak, *Distributed optimization in sensor networks*. Proc. IPSN'04, Berkeley, USA, 2004.
- [8] C. Richard, J. C. M. Bermudez, and P. Honeine, *Online prediction of time series data with kernels*. IEEE Transactions on Signal Processing, Vol. 57(3), pp. 1058-1067, 2009.
- [9] B. Schölkopf, R. Herbrich, and R. Williamson, *A generalized Representer Theorem*. NeuroCOLT, Royal Holloway College, Univ. London, UK, Tech. Rep. NC2-TR-2000-81, 2000.