

Estimation distribuée sur les réseaux multitâches en présence d'échanges d'informations bruitées

Roula NASSIF, Cédric RICHARD, André FERRARI

Laboratoire Lagrange, Université de Nice Sophia-Antipolis, France
{roula.nassif,cedric.richard,andre.ferrari}@unice.fr

Résumé – L'algorithme *diffusion LMS* est une stratégie efficace pour résoudre des problèmes d'estimation distribués sur des réseaux adaptatifs où les nœuds coopèrent en échangeant des informations pour estimer un même vecteur de paramètres. Pour certaines applications, le vecteur de paramètres à estimer peut, volontairement ou non, ne pas être identique pour tous les nœuds. De plus, les échanges d'informations peuvent se faire sur des liens bruités. Dans ce travail, nous étudions analytiquement l'impact de ces facteurs de nuisance sur les performances d'un réseau.

Abstract – The diffusion LMS algorithm is an efficient strategy for solving distributed optimization problems over adaptive networks where all nodes cooperate and exchange information to estimate the same parameter vector. In some applications, the optimum parameter vectors may not be the same for all nodes. Moreover, nodes usually exchange information over noisy links. In this work, we study analytically the impact on the network performance of such nuisance factors.

1 Introduction

On considère un réseau adaptatif constitué d'agents échangeant avec leurs voisins, à chaque instant, leurs observations et leurs estimés locaux afin d'optimiser une fonction de coût globale. Dans la littérature, différentes approches ont été proposées pour l'estimation distribuée d'un vecteur de paramètre commun à l'ensemble des agents. A titre d'exemple, on cite les stratégies de *consensus*, *incrementales* ou encore de *diffusion*. On s'intéresse dans cet article aux stratégies de diffusion [1–4] en raison de leur robustesse, de leur capacité d'adaptation, de leur performance et de leur stabilité.

Récemment, différents travaux ont été menés afin d'étudier le comportement de ces stratégies de diffusion sur des réseaux multitâches. Un réseau multitâche est composé de plusieurs groupes ou clusters, au sein desquels les nœuds cherchent à estimer un même vecteur de paramètres. L'estimation d'un vecteur de paramètres est appelée *tâche*. Deux scénarios multitâches ont été définis dans la littérature. Dans un premier cas, dit supervisé, la composition des clusters est connue. Des informations a priori sur les relations entre les tâches peuvent alors être exploitées afin d'améliorer les performances d'estimation [5–7]. Dans un second cas, dit non-supervisé, aucune information n'est connue sur la composition des clusters au sein desquels les nœuds réalisent une même tâche. On s'intéresse ici à ce scénario où chaque nœud n'a pas connaissance du fait que ses voisins partagent ou non la même tâche que lui. Il existe alors des algorithmes de clustering non-supervisés permettant de regrouper les nœuds réalisant une même tâche en clusters, et de restreindre les échanges d'information aux seuls nœuds dans un même cluster [8]. Les performances de l'algorithme diffusion LMS, simple-tâche par nature, ont par ailleurs été récemment étudiées lorsqu'il est appliqué à un réseau multitâche [8] sans que l'on recourt à un algorithme de clustering. On s'intéresse à une telle analyse dans cet article en y adjoignant une nuisance complémentaire. Celle-ci consiste à supposer que les échanges d'information entre les nœuds peuvent être entachés de bruit.

Dans cet article, on étudie théoriquement les performances de l'algorithme diffusion LMS lorsqu'il est mis en œuvre dans un environnement multitâche et que les communications entre les nœuds sont bruitées. Cet article est organisé ainsi. Le problème est modélisé en Section 2. Les performances de l'algorithme, en moyenne et en erreur quadratique moyenne, sont étudiées en Section 3. Des résultats expérimentaux présentés en Section 4 viennent conforter cette analyse. La Section 5 présente une conclusion et des perspectives.

2 Modélisation du problème

On considère un réseau constitué de N nœuds. À chaque instant i , chaque nœud k du réseau acquiert une observation scalaire de moyenne nulle $d_k(i)$ et un vecteur de régression $\mathbf{x}_k(i)$ de dimension M et de matrice de covariance $\mathbf{R}_{\mathbf{x},k} = \mathbb{E}\{\mathbf{x}_k(i)\mathbf{x}_k^\top(i)\} > 0$. On suppose que les couples de mesures $\{d_k(i), \mathbf{x}_k(i)\}$ vérifient le modèle linéaire suivant :

$$d_k(i) = \mathbf{x}_k^\top(i)\mathbf{w}_k^* + z_k(i), \quad (1)$$

avec \mathbf{w}_k^* un vecteur de paramètres inconnu recherché par le nœud k et $z_k(i)$ un bruit centré de variance $\sigma_{z,k}^2$. On suppose que le processus aléatoire $\{z_k(i)\}$ est temporellement blanc et spatialement indépendant.

Selon le nombre de vecteur de paramètres, on distingue deux types de réseaux : les réseaux simple tâche et multitâches [5]. Dans un réseau simple-tâche, tous les nœuds doivent estimer le même vecteur de paramètres, i.e., $\mathbf{w}_k^* = \mathbf{w}^*, \forall k$. L'algorithme diffusion LMS vise à estimer \mathbf{w}^* de façon distribuée en minimisant le coût global quadratique [1–3] :

$$J^{\text{glob}}(\mathbf{w}) = \sum_{k=1}^N \mathbb{E}\{|d_k(i) - \mathbf{x}_k^\top(i)\mathbf{w}|^2\}. \quad (2)$$

Algorithm 1 : Adapt-then-Combine (ATC)

Initialisation :

Pour chaque nœud k , fixer un pas $\mu_k > 0$ et initialiser $\mathbf{w}_k(0)$. Définir les matrices stochastiques gauche \mathbf{A} et droite \mathbf{C} .

Algorithme : A chaque instant $i \geq 0$ et pour chaque nœud k :

$$\begin{aligned} \boldsymbol{\psi}_k(i+1) &= \mathbf{w}_k(i) \\ &+ \mu_k \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_\ell(i) (d_\ell(i) - \mathbf{x}_\ell^\top(i) \mathbf{w}_k(i)), \end{aligned} \quad (3)$$

$$\mathbf{w}_k(i+1) = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \boldsymbol{\psi}_\ell(i+1). \quad (4)$$

D'après le modèle (1), on note que le minimum de (2) est atteint pour \mathbf{w}^* . Dans l'Algorithme 1, on rappelle brièvement la stratégie ATC du diffusion LMS où $\mathbf{w}_k(i)$ représente l'estimé de $J^{\text{glob}}(\mathbf{w})$ au nœud k et à l'iteration i .

Les coefficients positifs $\{a_{\ell k}, c_{\ell k}\}$ associés aux liens vérifient :

$$\sum_{\ell=1}^N a_{\ell k} = 1, \quad a_{\ell k} = 0 \text{ si } \ell \notin \mathcal{N}_k, \quad (\text{gauche stoch.}) \quad (5)$$

$$\sum_{k=1}^N c_{\ell k} = 1, \quad c_{\ell k} = 0 \text{ si } k \notin \mathcal{N}_\ell. \quad (\text{droite stoch.}) \quad (6)$$

On note \mathcal{N}_k l'ensemble des voisins de k , incluant k . Dans l'étape d'adaptation (3), le nœud k combine ses données au travers des paramètres $\{c_{\ell k}\}$, ainsi que les données reçues de ses voisins ℓ afin d'adapter son estimé $\mathbf{w}_k(i)$. Puis, dans l'étape de combinaison (4), le nœud k combine son estimé intermédiaire $\boldsymbol{\psi}_k(i+1)$ avec ceux de son voisinage $\boldsymbol{\psi}_\ell(i+1)$ au travers des coefficients $\{a_{\ell k}\}$. Dans [9], les performances des stratégies de diffusion sont étudiées pour des réseaux simple-tâche lorsque les informations sont bruitées.

Pour modéliser les transmissions sur des liens bruités, on adopte le même principe que [9]. À chaque instant, le nœud ℓ envoie ses données $d_\ell(i)$, $\mathbf{x}_\ell(i)$, et $\boldsymbol{\psi}_\ell(i)$ à son voisin k . Ce dernier en reçoit des versions bruitées telles que :

$$d_{\ell k}(i) = d_\ell(i) + z_{\ell k}^{(d)}(i), \quad (7)$$

$$\mathbf{x}_{\ell k}(i) = \mathbf{x}_\ell(i) + \mathbf{z}_{\ell k}^{(\mathbf{x})}(i), \quad (8)$$

$$\boldsymbol{\psi}_{\ell k}(i) = \boldsymbol{\psi}_\ell(i) + \mathbf{z}_{\ell k}^{(\boldsymbol{\psi})}(i), \quad (9)$$

où $z_{\ell k}^{(d)}(i)$ est un bruit scalaire et $\mathbf{z}_{\ell k}^{(\mathbf{x})}(i)$, $\mathbf{z}_{\ell k}^{(\boldsymbol{\psi})}(i)$ sont des vecteurs de bruit de dimension $M \times 1$. Ainsi, l'algorithme de diffusion (3), (4) devient :

$$\begin{aligned} \boldsymbol{\psi}_k(i+1) &= \mathbf{w}_k(i) \\ &+ \mu_k \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell k}(i) [d_{\ell k}(i) - \mathbf{x}_{\ell k}^\top(i) \mathbf{w}_k(i)], \end{aligned} \quad (10)$$

$$\mathbf{w}_k(i+1) = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \boldsymbol{\psi}_{\ell k}(i+1). \quad (11)$$

Avant de présenter les résultats que l'on a établis sur les performances de l'algorithme (10), (11) lorsqu'il est appliqué à un réseau multitâche, on liste les hypothèses sur lesquelles on s'est basé dans notre analyse.

Hypothèse 1. Le vecteur de regression $\mathbf{x}_k(i)$ est issu d'un processus aléatoire de moyenne nulle, stationnaire dans le temps et spatialement indépendant.

Hypothèse 2. Les bruits $z_{\ell k}^{(d)}(i)$, $\mathbf{z}_{\ell k}^{(\mathbf{x})}(i)$ et $\mathbf{z}_{\ell k}^{(\boldsymbol{\psi})}(i)$ sont des variables aléatoires de moyenne nulle, stationnaires dans le temps et spatialement indépendantes avec des variances et des covariances respectives :

$$\sigma_{z, \ell k}^{(d)2} \quad \mathbf{R}_{z, \ell k}^{(\mathbf{x})} \quad \mathbf{R}_{z, \ell k}^{(\boldsymbol{\psi})}$$

Notons que ces moments existent uniquement si $\ell \in \mathcal{N}_k^-$ avec $\mathcal{N}_k^- = \mathcal{N}_k \setminus k$.

Hypothèse 3. On suppose $\{\mathbf{x}_k(i_1)\}$, $\{z_\ell(i_2)\}$, $\{z_{mn}^{(d)}(i_3)\}$, $\{z_{pq}^{(\mathbf{x})}(i_4)\}$, et $\{z_{st}^{(\boldsymbol{\psi})}(i_5)\}$ sont mutuellement indépendants.

3 Performance en moyenne et en erreur quadratique moyenne

3.1 Évolution du vecteur d'erreur

Les données bruitées $\mathbf{x}_{\ell k}(i)$, $d_{\ell k}(i)$ parvenant au nœud k de son voisin ℓ vérifient la relation linéaire (à comparer avec (1)) :

$$d_{\ell k}(i) = \mathbf{x}_{\ell k}^\top(i) \mathbf{w}_\ell^* + z_{\ell k}(i), \quad (12)$$

où nous introduisons le signal bruit de moyenne nulle :

$$z_{\ell k}(i) = z_\ell(i) + z_{\ell k}^{(d)}(i) - \mathbf{z}_{\ell k}^{(\mathbf{x})}(i)^\top \mathbf{w}_\ell^*, \quad \ell \in \mathcal{N}_k^-, \quad (13)$$

et de variance :

$$\sigma_{z, \ell k}^2 = \sigma_{z, \ell}^2 + \sigma_{z, \ell k}^{(d)2} + (\mathbf{w}_\ell^*)^\top \mathbf{R}_{z, \ell k}^{(\mathbf{x})} \mathbf{w}_\ell^*. \quad (14)$$

Soit $\tilde{\mathbf{w}}_k(i) = \mathbf{w}_k^* - \mathbf{w}_k(i)$ le vecteur d'erreur au nœud k et à l'instant i . En utilisant (12), l'erreur d'estimation qui apparaît dans l'étape d'adaptation (10) peut s'écrire sous la forme :

$$\begin{aligned} d_{\ell k}(i) - \mathbf{x}_{\ell k}^\top(i) \mathbf{w}_k(i) \\ = \mathbf{x}_{\ell k}^\top(i) \tilde{\mathbf{w}}_k(i) + \mathbf{x}_{\ell k}^\top(i) \mathbf{w}_{\ell k}^* + z_{\ell k}(i), \end{aligned} \quad (15)$$

avec $\mathbf{w}_{\ell k}^* = \mathbf{w}_\ell^* - \mathbf{w}_k^*$. Au niveau du réseau, on définit les vecteurs suivants :

$$\mathbf{w}^* = \text{col}\{\mathbf{w}_1^*, \dots, \mathbf{w}_N^*\}, \quad (16)$$

$$\tilde{\mathbf{w}}(i) = \text{col}\{\tilde{\mathbf{w}}_1(i), \dots, \tilde{\mathbf{w}}_N(i)\}, \quad (17)$$

$$\boldsymbol{\psi}(i) = \text{col}\{\boldsymbol{\psi}_1(i), \dots, \boldsymbol{\psi}_N(i)\}. \quad (18)$$

où $\text{col}\{\cdot\}$ est l'opérateur empilant les vecteurs en argument les uns sur les autres. En soustrayant \mathbf{w}_k^* des deux membres de l'équation (10) et en utilisant le résultat (15), on obtient :

$$\begin{aligned} \mathbf{w}^* - \boldsymbol{\psi}(i+1) \\ = (\mathbf{I}_{NM} - \mathcal{M}\mathcal{R}(i)) \tilde{\mathbf{w}}(i) - \mathcal{M}\mathbf{h}(i) - \mathcal{M}\mathbf{s}(i), \end{aligned} \quad (19)$$

avec \mathbf{I}_{NM} la matrice identité de taille $NM \times NM$ et

$$\mathcal{M} = \text{diag}\{\mu_k \mathbf{I}_M\}_{k=1}^N, \quad (20)$$

$$\mathcal{R}(i) = \text{diag}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell k}(i) \mathbf{x}_{\ell k}^\top(i)\right\}_{k=1}^N, \quad (21)$$

$$\mathbf{h}(i) = \text{col}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell k}(i) \mathbf{x}_{\ell k}^\top(i) \mathbf{w}_{\ell k}^*\right\}_{k=1}^N, \quad (22)$$

$$\mathbf{s}(i) = \text{col}\left\{\sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell k}(i) z_{\ell k}(i)\right\}_{k=1}^N. \quad (23)$$

Soit $\mathcal{A} = \mathbf{A} \otimes \mathbf{I}_M$ où \otimes désigne le produit de Kronecker. En utilisant (9), l'étape de combinaison (11) donne :

$$\mathbf{w}(i+1) = \mathcal{A}^\top \boldsymbol{\psi}(i+1) + \mathbf{z}^{(\psi)}(i+1), \quad (24)$$

avec

$$\mathbf{z}^{(\psi)}(i+1) = \text{col}\left\{ \sum_{\ell \in \mathcal{N}_k^-} a_{\ell k} \mathbf{z}_{\ell k}^{(\psi)}(i+1) \right\}_{k=1}^N. \quad (25)$$

En soustrayant \mathbf{w}^* des deux membres de l'équation (24) et en ajoutant la quantité nulle $\mathcal{A}^\top \mathbf{w}^* - \mathcal{A}^\top \mathbf{w}^*$, on trouve que le vecteur d'erreur au niveau du réseau $\tilde{\mathbf{w}}(i+1)$ évolue selon l'expression :

$$\tilde{\mathbf{w}}(i+1) = \mathcal{B}(i)\tilde{\mathbf{w}}(i) - \mathbf{g}(i) - \mathbf{r}(i) - \mathbf{z}^{(\psi)}(i+1), \quad (26)$$

avec

$$\mathcal{B}(i) = \mathcal{A}^\top (\mathbf{I}_{NM} - \mathcal{M}\mathcal{R}(i)), \quad (27)$$

$$\mathbf{g}(i) = \mathcal{A}^\top \mathcal{M}\mathbf{s}(i), \quad (28)$$

$$\mathbf{r}(i) = \mathcal{A}^\top \mathcal{M}\mathbf{h}(i) - (\mathbf{I}_{MN} - \mathcal{A}^\top) \mathbf{w}^*. \quad (29)$$

On note que les bruits sur les liens entraînent l'apparition de deux termes $\mathbf{g}(i)$ et $\mathbf{z}^{(\psi)}(i+1)$. Le scénario multitâche entraîne l'apparition du terme $\mathbf{r}(i)$.

Dans ce qui suit, compte tenu la limitation de l'espace disponible, nous présentons uniquement les résultats de l'analyse sans les arguments.

3.2 Analyse de la moyenne de l'erreur

En évaluant l'espérance de l'équation (26), on trouve :

$$\mathbb{E}\{\tilde{\mathbf{w}}(i+1)\} = \mathcal{B}\mathbb{E}\{\tilde{\mathbf{w}}(i)\} - \mathbf{g} - \mathbf{r}, \quad (30)$$

avec

$$\mathcal{B} = \mathcal{A}^\top (\mathbf{I}_{NM} - \mathcal{M}\mathcal{R}), \quad (31)$$

$$\mathcal{R} = \text{diag}\left\{ \sum_{\ell \in \mathcal{N}_k} c_{\ell k} (\mathbf{R}_{\mathbf{x},\ell} + \mathbf{R}_{\mathbf{z},\ell k}^{(\mathbf{x})}) \right\}_{k=1}^N, \quad (32)$$

$$\mathbf{g} = \mathcal{A}^\top \mathcal{M}\mathbf{s}, \quad (33)$$

$$\mathbf{s} = -\mathbf{R}_z^{(\mathbf{x})} \mathbf{w}^*, \quad (34)$$

$$\mathbf{r} = \mathcal{A}^\top \mathcal{M}\mathbf{h} - (\mathbf{I}_{MN} - \mathcal{A}^\top) \mathbf{w}^*, \quad (35)$$

$$\mathbf{h} = \text{col}\left\{ \sum_{\ell \in \mathcal{N}_k} c_{\ell k} (\mathbf{R}_{\mathbf{x},\ell} + \mathbf{R}_{\mathbf{z},\ell k}^{(\mathbf{x})}) \mathbf{u}_{\ell k}^* \right\}_{k=1}^N. \quad (36)$$

où $\mathbf{R}_z^{(\mathbf{x})}$ est une matrice de $N \times N$ blocs ayant pour bloc (ℓ, k) la matrice $c_{k\ell} \mathbf{R}_{z,k\ell}^{(\mathbf{x})}$ de dimension $M \times M$. Afin de garantir la convergence en moyenne de la récurrence (30), il suffit de choisir les pas de sorte que :

$$0 < \mu_k < \frac{2}{\lambda_{\max}\left\{ \sum_{\ell \in \mathcal{N}_k} c_{\ell k} (\mathbf{R}_{\mathbf{x},\ell} + \mathbf{R}_{\mathbf{z},\ell k}^{(\mathbf{x})}) \right\}}, \quad k = 1, \dots, N. \quad (37)$$

où $\lambda_{\max}\{\cdot\}$ désigne la plus grande valeur propre de son argument. Sous cette condition, la moyenne du vecteur d'erreur converge vers une valeur fixe lorsque $i \rightarrow \infty$ donnée par :

$$\mathbf{b} = \lim_{i \rightarrow \infty} \mathbb{E}\{\tilde{\mathbf{w}}(i)\} = -(\mathbf{I}_{MN} - \mathcal{B})^{-1}(\mathbf{g} + \mathbf{r}) \quad (38)$$

Le biais apparent à l'équation (38) est engendré par la présence de plusieurs tâches dans le réseau (résultant en \mathbf{h}), la coopération entre les nœuds (résultant en $(\mathbf{I}_{MN} - \mathcal{A}^\top) \neq 0$) et par l'échange des vecteurs de régression entre les nœuds voisins sur des liens bruités (résultant en \mathbf{s}). De plus, ces échanges bruités des vecteurs de régression interviennent dans la condition de stabilité de l'algorithme à travers les matrices de covariance des bruits $\{\mathbf{z}_{\ell k}^{(\mathbf{x})}(i)\}$.

3.3 Analyse de la variance de l'erreur

Dans cette section, on utilise le principe de conservation d'énergie proposé pour l'analyse des filtres adaptatifs. Nous avons analysé l'algorithme en étendant les analyses existantes dans la littérature [8, 9] à notre cas. De l'équation (26) et sous les Hypothèses 1, 2, 3, on établit la variance pondérée du vecteur d'erreur $\mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\}$, avec Σ une matrice quelconque définie positive. Soit $\boldsymbol{\sigma} = \text{vec}\{\Sigma\}$ où $\text{vec}\{\Sigma\}$ empile les colonnes de la matrice Σ les unes sur les autres. En développant les termes dans $\mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\}$ et en utilisant les propriétés suivantes sur les opérateurs de vectorisation et de trace

$$\text{vec}\{\mathbf{U}\Sigma\mathbf{W}\} = (\mathbf{W}^\top \otimes \mathbf{U})\boldsymbol{\sigma}, \quad \text{Tr}\{\Sigma\mathbf{W}\} = \text{vec}\{\mathbf{W}^\top\}^\top \boldsymbol{\sigma}, \quad (39)$$

l'espérance $\mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\}$ peut s'écrire sous la forme :

$$\begin{aligned} \mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\} &\approx \mathbb{E}\{\|\tilde{\mathbf{w}}(i)\|_{\mathcal{F}\boldsymbol{\sigma}}^2\} + [\text{vec}\{\mathbf{T}^\top\}]^\top \boldsymbol{\sigma} \\ &\quad - 2[\mathcal{B}\mathbb{E}\{\tilde{\mathbf{w}}(i-j)\} \otimes (\mathbf{g} + \mathbf{r})]^\top \boldsymbol{\sigma}, \end{aligned} \quad (40)$$

avec

$$\mathcal{F} = \mathbb{E}\{\mathcal{B}^\top(i) \otimes \mathcal{B}^\top(i)\} \approx \mathcal{B}^\top \otimes \mathcal{B}^\top, \quad (41)$$

$$\mathbf{T} = \mathbf{G} + \mathbf{R}_z^{(\psi)} + \mathbf{R}_r + 2\mathbf{G}_r, \quad (42)$$

$$\mathbf{G} = \mathcal{A}^\top \mathcal{M}\mathcal{C}^\top \mathcal{S}\mathcal{C}\mathcal{M}\mathcal{A}, \quad (43)$$

$$\mathbf{S} = \text{diag}\{\sigma_{z,k}^2 \mathbf{R}_{\mathbf{x},k}\}_{k=1}^N + \mathbf{s}\mathbf{s}^\top + \text{diag}\{\mathbf{D}_k\}_{k=1}^N, \quad (44)$$

$$\mathbf{R}_r = \mathbf{r}\mathbf{r}^\top + \mathcal{A}^\top \mathcal{M}\text{diag}\{\mathbf{H}_k\}_{k=1}^N \mathcal{M}\mathcal{A}, \quad (45)$$

$$\mathbf{G}_r = \mathbf{g}\mathbf{r}^\top - \mathcal{A}^\top \mathcal{M}\text{diag}\{\mathbf{J}_k\}_{k=1}^N \mathcal{M}\mathcal{A}, \quad (46)$$

$$\mathbf{R}_z^{(\psi)} = \text{diag}\left\{ \sum_{\ell \in \mathcal{N}_k^-} a_{\ell k}^2 \mathbf{R}_{z,\ell k}^{(\psi)} \right\}_{k=1}^N. \quad (47)$$

Notons que $\mathcal{C} = \mathbf{C} \otimes \mathbf{I}_M$ et que :

$$\mathbf{D}_k = \sum_{\ell \in \mathcal{N}_k} c_{\ell k}^2 ((\sigma_{z,\ell k}^{(d)2} + (\mathbf{w}_\ell^*)^\top \mathbf{R}_{z,\ell k}^{(\mathbf{x})} \mathbf{w}_\ell^*) \mathbf{R}_{\mathbf{x},\ell} + (\sigma_{z,\ell k}^{(d)2} + \sigma_{z,\ell}^2) \mathbf{R}_{z,\ell k}^{(\mathbf{x})}),$$

$$\mathbf{H}_k = \sum_{\ell \in \mathcal{N}_k} c_{\ell k}^2 (\mathbf{R}_{\mathbf{x},\ell} \mathbf{u}_{\ell k}^* (\mathbf{u}_{\ell k}^*)^\top \mathbf{R}_{z,\ell k}^{(\mathbf{x})} + (\mathbf{u}_{\ell k}^*)^\top \mathbf{R}_{z,\ell k}^{(\mathbf{x})} \mathbf{u}_{\ell k}^* \mathbf{R}_{\mathbf{x},\ell} +$$

$$(\mathbf{u}_{\ell k}^*)^\top \mathbf{R}_{\mathbf{x},\ell} \mathbf{u}_{\ell k}^* \mathbf{R}_{z,\ell k}^{(\mathbf{x})} + \mathbf{R}_{z,\ell k}^{(\mathbf{x})} \mathbf{u}_{\ell k}^* (\mathbf{u}_{\ell k}^*)^\top \mathbf{R}_{\mathbf{x},\ell}),$$

$$\mathbf{J}_k = \sum_{\ell \in \mathcal{N}_k} c_{\ell k}^2 (\mathbf{R}_{\mathbf{x},\ell} \mathbf{u}_{\ell k}^* (\mathbf{w}_\ell^*)^\top \mathbf{R}_{z,\ell k}^{(\mathbf{x})} + (\mathbf{w}_\ell^*)^\top \mathbf{R}_{z,\ell k}^{(\mathbf{x})} \mathbf{u}_{\ell k}^* \mathbf{R}_{\mathbf{x},\ell}).$$

L'approximation apparaissant dans l'équation (40) provient de deux approximations. Dans le second terme du second membre de l'équation, l'évaluation de certaines espérances nécessite la connaissance des moments d'ordre 4 des variables, qui sont inconnus. Ceux-ci sont approximatés par le produit des variances et covariances. De plus, les calculs conduisant au dernier terme de l'équation (40) et de l'équation (41) négligent parfois l'influence des termes contenant des puissances d'ordre 2 des pas, sous hypothèse de pas μ_k assez petits. Les approximations effectuées ici sont classiques dans l'analyse des algorithmes de type diffusion LMS. Voir [4, 9] pour plus d'informations.

En conclusion, la variance pondérée $\mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\}$ converge lorsque $i \rightarrow \infty$ si et seulement si la matrice \mathcal{F} est stable. Sous l'approximation introduite dans (41), la condition (37) assure encore la stabilité de la matrice \mathcal{F} .

L'équation (40) permet d'établir une récurrence quantifiant la variance pondérée au cours du temps. Ainsi, on obtient [8] :

$$\begin{aligned} \mathbb{E}\{\|\tilde{\mathbf{w}}(i+1)\|_{\Sigma}^2\} &= \mathbb{E}\{\|\tilde{\mathbf{w}}(i)\|_{\Sigma}^2\} - \mathbb{E}\{\|\tilde{\mathbf{w}}(0)\|_{(\mathbf{I}-\mathcal{F})\boldsymbol{\sigma}}^2\} \\ &\quad + [\text{vec}\{\mathbf{T}^\top\}]^\top \mathcal{F}^i \boldsymbol{\sigma} - 2[(\mathcal{B}\mathbb{E}\{\tilde{\mathbf{w}}(i)\} \otimes (\mathbf{g} + \mathbf{r}))^\top + \boldsymbol{\Gamma}(i)] \boldsymbol{\sigma}, \end{aligned} \quad (48)$$

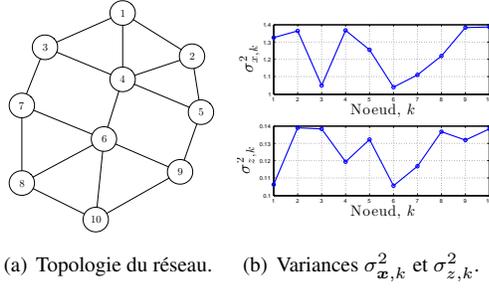


FIGURE 1 – Paramètres expérimentaux.

avec

$$\begin{aligned} \Gamma(i) &= \Gamma(i-1)\mathcal{F} \\ &+ (\mathcal{B}\mathbb{E}\{\tilde{\mathbf{w}}(i-1)\} \otimes (\mathbf{g} + \mathbf{r}))^\top (\mathcal{F} - \mathbf{I}_{(NM)^2}), \end{aligned} \quad (49)$$

et $\Gamma(0) = \mathbf{0}_{1 \times (MN)^2}$. De plus, sous la condition (37), en régime permanent, on a :

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E}\{\|\tilde{\mathbf{w}}(i)\|_{(\mathbf{I}-\mathcal{F})\sigma'}^2\} \\ = [\text{vec}\{\mathbf{T}^\top\}]^\top \sigma' - 2[\mathcal{B}\mathbf{b} \otimes (\mathbf{g} + \mathbf{r})]^\top \sigma'. \end{aligned} \quad (50)$$

Dans les simulations suivantes, on choisit $\Sigma = \frac{1}{N}\mathbf{I}_{NL}$, i.e., $\sigma = \frac{1}{N}\text{vec}\{\mathbf{I}_{NL}\}$ et $\sigma' = \frac{1}{N}(\mathbf{I} - \mathcal{F})^{-1}\text{vec}\{\mathbf{I}_{NL}\}$. Ce choix de Σ permet d'évaluer le MSD (mean-square-deviation) du réseau donnée par :

$$\text{MSD}(i) = \frac{1}{N}\mathbb{E}\{\|\tilde{\mathbf{w}}(i)\|^2\} = \frac{1}{N}\sum_{k=1}^N \mathbb{E}\{\|\tilde{\mathbf{w}}_k(i)\|^2\}. \quad (51)$$

4 Résultats expérimentaux

Dans cette section, on présente des simulations validant les modèles théoriques présentés précédemment. Pour ceci, on considère un réseau formé de 10 nœuds présenté dans la Figure 1(a). Les vecteurs de paramètres sont uniformément distribués sur un cercle de rayon $r = 0.03$ et de centre $\mathbf{w}_c = [1, 0.5]^\top$ (voir la section V-A de [8] pour plus de détails sur ce modèle). Les vecteurs de regressions sont i.i.d. Gaussiens de moyenne nulle et de matrice de covariance $\mathbf{R}_{\mathbf{x},k} = \sigma_{\mathbf{x},k}^2 \mathbf{I}_M$. Les signaux bruit $z_k(i)$ sont Gaussiens de moyenne nulle et de variance $\sigma_{z,k}^2$. Les variances $\{\sigma_{\mathbf{x},k}^2\}$ et $\{\sigma_{z,k}^2\}$ sont visualisées dans la Figure 1(b). Les signaux de bruit sur les liens $z_{\ell k}^{(d)}(i)$, $z_{\ell k}^{(\mathbf{x})}(i)$ et $z_{\ell k}^{(\psi)}(i)$ sont encore supposés i.i.d. Gaussiens de moyenne nulle, de variance et matrices de covariance $\sigma_{z,\ell k}^{(d)2} = \sigma_{z,\ell k}^2$ et $\mathbf{R}_{z,\ell k}^{(\mathbf{x})} = \mathbf{R}_{z,\ell k}^{(\psi)} = \sigma_{z,\ell k}^2 \mathbf{I}_M$, respectivement, avec $\sigma_{z,\ell k}^2 = 10^{-4} \forall \ell \in \mathcal{N}_k^-$ et $\forall k$. Nous fixons les pas $\mu_k = 0.01$ pour tous les nœuds du réseau. Nous considérons la stratégie de diffusion ATC avec une matrice \mathbf{C} uniforme, c.à.d., nous fixons $c_{\ell k} = (\text{card}\{\mathcal{N}_\ell\})^{-1}$ pour $k \in \mathcal{N}_\ell$. De même, nous choisissons une matrice de combinaison \mathbf{A} uniforme en fixant $a_{\ell k} = (\text{card}\{\mathcal{N}_k\})^{-1}$ pour $\ell \in \mathcal{N}_k$. Nous considérons 100 expériences de Monte-Carlo dans nos simulations.

On considère l'algorithme diffusion LMS selon trois configurations : le LMS non coopératif dans lequel chaque nœud cherche son propre optimum sans échanges d'informations avec ses voisins ($\mathbf{A} = \mathbf{C} = \mathbf{I}_N$), le ATC sans bruit sur les liens obtenu de l'Algorithme 1 et le ATC avec liens bruités obtenu des équations (10), (11). Comme on peut le remarquer sur la Figure 2, les résultats théoriques se superposent convenablement aux résultats expérimentaux. Pour un certain degré de similitude entre les tâches, la stratégie de diffusion est

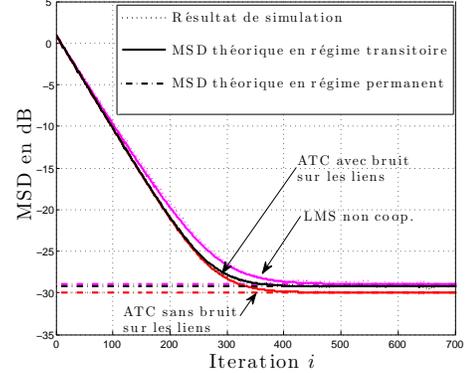


FIGURE 2 – Performances du réseau.

capable d'améliorer la MSD comparé au LMS non coopératif malgré la présence du biais dû au scénario multitâche. La présence du bruit sur les liens entraîne une dégradation des performances.

5 Conclusion et perspectives

Dans ce travail, nous avons considéré la stratégie de diffusion pour l'estimation distribuée sur des réseaux multitâches. Nous avons analysé le comportement du réseau en présence de bruit sur les liens. Nous envisageons dans un futur travail de proposer des stratégies permettant aux nœuds d'adapter leur voisinage de sorte à profiter de la coopération en réduisant le biais introduit par le scénario multitâche et le bruit sur les liens.

Références

- [1] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks : Formulation and performance analysis," *IEEE Trans. Sig. Proc.*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [2] F. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Sig. Proc.*, vol. 58, no. 3, pp. 1035–1048, March 2010.
- [3] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, Elsevier, Ed., vol. 3, pp. 322–454. Elsevier, 2014.
- [4] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, April 2014.
- [5] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Sig. Proc.*, vol. 62, no. 16, pp. 4129–4144, August 2014.
- [6] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," arXiv:1412.1798 (Submitted), Dec. 2014.
- [7] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion LMS with sparsity-based regularization," in *Proc. IEEE ICASSP*, April 2015.
- [8] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Sig. Proc.*, vol. 63, no. 11, pp. 2733–2748, June 2015.
- [9] X. Zhao, S.-Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. Sig. Proc.*, vol. 60, no. 7, pp. 3460–3475, July 2012.