

# Sur un algorithme de Robbins-Monro distribué

Pascal BIANCHI, Gersende FORT, Walid HACHEM, Jérémie JAKUBOWICZ

Laboratoire Traitement et Communication de l'Information (LTCI)  
Télécom ParisTech & CNRS

[prenom].[nom]@telecom-paristech.fr

**Résumé** – Ce papier étudie la convergence d'un algorithme de Robbins-Monro distribué sur un réseau basé sur la technique de "gossip". On montre que, sous des hypothèses vérifiables en pratique, l'algorithme étudié converge vers un consensus entre les différents agents du réseau. Un théorème central limite est établi afin de préciser la vitesse de convergence. Enfin, le fonctionnement de l'algorithme est illustré sur un exemple d'estimation par maximum de vraisemblance distribué.

**Abstract** – This paper investigates the convergence of a distributed Robbins-Monro algorithm over a network, based on the "gossip" technique. It is shown that under practically verifiable assumptions, the algorithm under study converges towards a consensus between the agents of the network. In order to assess a speed of convergence, a central limit theorem is derived. Last, we illustrate this algorithm through a distributed maximum-likelihood estimation example.

## 1 Introduction

Chercher les racines d'une équation de la forme  $h(\theta) = 0$  est un problème récurrent que l'on rencontre par exemple dans les procédures d'optimisation de fonctions différentiables. Résoudre cette équation de façon itérative est une façon classique de procéder [1]. Lorsque  $h$  n'est pas connue de façon analytique mais que l'on ne dispose que d'une version bruitée  $H(\theta, X)$ , on peut utiliser l'algorithme de Robbins-Monro pour trouver les racines de  $h$  [2]. En traitement du signal, la maximisation de vraisemblance par descente de gradient en est une application typique. Dans ce cas particulier, l'objectif consiste à chercher les points stationnaires de la divergence de Kullback entre la vraie loi des observations et une famille paramétrique de lois. Dans ce cas le terme  $H(\theta, X)$  correspond au gradient (par rapport à  $\theta$ ) de la log-vraisemblance de l'observation  $X$  sous le modèle de paramètre  $\theta$ .

Dans ce papier, nous étudions des algorithmes de Robbins-Monro distribués (cf. par exemple [3]). Soit un graphe connexe : les sommets modélisent les agents et les arêtes définissent le voisinage de chaque agent i.e. l'ensemble des agents avec lesquels il peut communiquer. L'objectif est d'estimer  $\theta_* \in \mathbb{R}^d$  tel que  $h(\theta_*) = 0$  lorsque  $h(\theta) = \sum_{i=1}^N h_i(\theta)$ . Pour ce faire, chaque agent  $i$ ,  $i \in \{1, \dots, N\}$ , dispose d'approximations bruitées  $\{H_i(\cdot, X_{n,i})\}_n$  de la fonction  $h_i$ . De plus, les nœuds sont capables de communiquer avec leurs voisins à certains instants (éventuellement aléatoires). Lorsque deux nœuds communiquent, ils s'entendent sur une valeur commune de leurs estimées qui supplante alors leurs estimées individuelles d'avant la communication. Il s'agit donc d'un algorithme qui mêle une étape d'approximation stochastique (cf. [4, 1] par exemple) et une étape de "gossip" (cf. [5]). Sous des hypothèses liées no-

tamment à la procédure de "gossip", cet algorithme est un algorithme de consensus.

Les algorithmes de consensus ont fait l'objet de nombreux travaux dans la littérature récente de traitement du signal [6, 7]. Ils consistent à atteindre un état de consensus de façon distribuée et itérative dans le cas particulier où le consensus recherché est la valeur moyenne des quantités initiales figurant dans chaque nœud. Plus récemment, plusieurs auteurs ont formulé le problème d'approximation stochastique distribuée comme un problème de consensus [3, 6, 8].

La principale contribution de ce papier est de fournir une analyse complète de la convergence d'un algorithme de Robbins-Monro distribué : nous montrons que cet algorithme est *stable*, qu'il converge vers l'ensemble des zéros de  $h$  et nous étudions la vitesse de convergence. Ce papier généralise les travaux antérieurs de [3]. Tout d'abord, nos hypothèses sont moins restrictives ; en particulier, nous ne supposons pas que la fonction objectif  $h$  est bornée. Ensuite nous énonçons des conditions simples qui garantissent la stabilité de la procédure stochastique (sans recourir à des étapes de stabilisation comme par exemple, les techniques de reprojction (voir e.g. [1])). Enfin nous étudions le comportement limite de cet algorithme sous des hypothèses sur les perturbations  $H_i$  et les champs moyens  $h_i$  facilement vérifiables.

On commencera par décrire précisément l'algorithme distribué étudié, ainsi que les hypothèses afférentes dans la section 2. Après quoi nous présenterons les deux principaux résultats, sans preuve, faute de place, toujours dans la section 2. Enfin nous illustrerons l'algorithme étudié sur un problème de localisation distribuée par un réseaux de capteurs dans la section 3. Ce problème de localisation sera abordé par la méthode du maximum de vraisemblance.

## 2 Algorithme : description et analyse de convergence

### 2.1 L'algorithme

Soit un graphe *connexe* composé de  $N$  nœuds,  $N \geq 1$ . Étant donnée une suite de variables aléatoires  $(X_{n,i})_{n \in \mathbb{N}}$  à valeur dans  $\mathbb{R}^{m_i}$  correspondant à l'observation au temps  $n$  par le nœud  $i$ , une fonction  $H_i : \mathbb{R}^d \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^d$ , et une famille de matrices  $N \times N$  aléatoires *doublement stochastiques*  $\{W_n\}_n$ , chaque nœud  $i$  construit itérativement un processus stochastique  $\theta_{n,i}$  à valeurs dans  $\mathbb{R}^d$ . Chaque itération comprend deux étapes :

[Étape locale] Au temps  $n$ , chaque nœud  $i$  calcule  $\tilde{\theta}_{n,i}$  par :

$$\tilde{\theta}_{n,i} = \theta_{n-1,i} + \gamma_n H_i(\theta_{n-1,i}; X_{n,i}), \quad (1)$$

où  $\gamma_n$  désigne un pas déterministe positif.

[Étape de "gossip"] Chaque nœud  $i$  calcule une nouvelle estimée comme la moyenne pondérée des estimées voisines :

$$\theta_{n,i} = \sum_{j=1}^N w_n(i, j) \tilde{\theta}_{n,j},$$

où  $W_n := [w_n(i, j)]_{i,j=1}^N$ . Lorsque  $w_n(i, i) = 1$ ,  $\theta_{n,i} = \tilde{\theta}_{n,i}$  et le nœud  $i$  n'échange avec aucun voisin. La loi des matrices  $W_n$  est telle que  $w_n(i, j) = 0$  si les nœuds  $i$  et  $j$  ne sont pas connectés.

**Formulation vectorielle :** Cet algorithme peut se résumer sous forme vectorielle par

$$\theta_n = (W_n \otimes I_d)(\theta_{n-1} + \gamma_n H(\theta_{n-1}; X_n)), \quad (2)$$

où  $\otimes$  désigne le produit de Kronecker,  $I_d$  la matrice identité  $d \times d$ , la fonction  $H : \mathbb{R}^{dN} \times \mathbb{R}^{\sum m_i} \rightarrow \mathbb{R}^{dN}$  est donnée par

$$H(\theta; x) := (H_1(\theta_1; x_1)^T, \dots, H_N(\theta_N; x_N)^T)^T$$

$X_n = (X_{n,1}, \dots, X_{n,N})$  et  $\theta = (\theta_1, \dots, \theta_N)$ .  $^T$  désigne la transposition.

### 2.2 Les hypothèses

La suite des communications effectuées par les nœuds du graphe au fil du temps est modélisée par une suite de matrices aléatoires  $(W_n)_{n \geq 1}$ . Pour  $n \geq 1$  on introduit la  $\sigma$ -algèbre  $\mathcal{F}_n = \sigma(\theta_0, X_{1:n}, W_{1:n})$ . On suppose que la loi du vecteur aléatoire  $X_{n+1}$  conditionnellement à  $\mathcal{F}_n$  vérifie l'hypothèse de Robbins-Monro :

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mu_{\theta_n}(A)$$

pour tout ensemble mesurable  $A \subset \mathbb{R}^{dN}$  et où  $(\mu_\theta)_{\theta \in \mathbb{R}^{dN}}$  est une famille donnée de mesure de probabilités sur  $\mathbb{R}^{dN}$ . Pour chaque  $\theta \in \mathbb{R}^{dN}$ , on note  $\mathbb{E}_\theta[g(X)] = \int g(x) \mu_\theta(dx)$ .

#### 2.2.1 Notations

On désigne par  $\mathbb{1}_N$  le vecteur  $N \times 1$  dont les composantes valent toutes 1. On suppose également que : On note  $J := (\mathbb{1}\mathbb{1}^T/N) \otimes I_d$  le projecteur orthogonal sur le sous-espace de

consensus  $\{\mathbb{1} \otimes \theta : \theta \in \mathbb{R}^d\}$  et  $J^\perp := I_{dN} - J$  le projecteur orthogonal sur le sous-espace orthogonal à l'espace de consensus. Pour tout  $\theta \in \mathbb{R}^{dN}$ , on a

$$\theta = \mathbb{1} \otimes \langle \theta \rangle + J^\perp \theta \quad \text{où} \quad \langle \theta \rangle := \frac{1}{N} (\mathbb{1}^T \otimes I_d) \theta. \quad (3)$$

Notons que par définition de  $J$ ,  $\langle \theta \rangle = (\theta_1 + \dots + \theta_N)/N$  est la moyenne des  $N$  sous-blocs de taille  $d$  du vecteur  $\theta$ .

Définissons le *champ moyen*  $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$  de l'algorithme de Robbins-Monro distribué par :

$$h(\theta) := \mathbb{E}_{\mathbb{1} \otimes \theta} [H(\mathbb{1} \otimes \theta; X)], \quad (4)$$

#### 2.2.2 Hypothèses

**Hypothèse A1** a) La matrice  $W_n$  est doublement stochastique, i.e.  $W_n \mathbb{1} = W_n^T \mathbb{1} = \mathbb{1}$ .

b) Les matrices  $(W_n)_{n \geq 1}$  sont i.i.d. et le rayon spectral de  $\mathbb{E}(W_1 W_1^T) - J$  est strictement inférieur à 1.

c) Pour tout couple de fonctions  $f, g$ ,

$$\mathbb{E}[f(W_{n+1})g(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[f(W_1)] \mathbb{E}[g(X_{n+1}) | \theta_n].$$

La condition **A1a**) est satisfaite pourvu que les nœuds coordonnent leurs poids. Des schémas de coordination sont étudiés dans [3, 5]. La condition **A1a**) est aussi satisfaite dans le cas d'un réseau asynchrone (cf. [5] pour les détails ainsi que la section 3 pour une brève discussion). La condition **A1b**) s'interprète comme suit. L'idée générale des algorithmes de type "Gossip" est que  $\mathbb{E}(W_n W_n^T)$  soit assez proche du projecteur orthogonal  $\mathbb{1}\mathbb{1}^T/N$  sur la droite  $\{t\mathbb{1} : t \in \mathbb{R}\}$ ; de sorte que l'algorithme (2) atteigne un consensus. La condition **A1b**) sur le rayon spectral assure que la quantité d'information échangée dans le réseau reste suffisante pour atteindre le consensus. L'hypothèse portant sur le caractère i.i.d. des matrices  $W_n$  peut être affaibli de sorte à couvrir le cas où le nombre moyen de communications entre les nœuds puisse varier dans le temps et même tendre vers 0 lorsque  $n$  tend vers l'infini. Dans ce cas il faut renforcer la condition sur le rayon spectral de façon adéquate. La condition **A1c**) implique que les variables aléatoires  $W_{n+1}$  et  $X_{n+1}$  soient indépendantes conditionnellement au passé. De plus, la loi conditionnelle de  $X_{n+1}$  ne dépend du passé que par l'intermédiaire de  $\theta_n$  : c'est l'hypothèse de Robbins-Monro.

On suppose également vérifiées les deux hypothèses suivantes qui portent sur les pas de l'algorithme.

**Hypothèse A2** a) La suite (déterministe)  $(\gamma_n)_{n \geq 1}$  est strictement positive et vérifie  $\sum_n \gamma_n = \infty$ .

b) Il existe  $\alpha > 1/2$  tel que :

$$\lim_{n \rightarrow \infty} n^\alpha \gamma_n = 0 \quad (5)$$

L'hypothèse **A2a**) est classique et nécessaire au bon déroulement des algorithmes d'approximation stochastique. Elle assure que les pas de temps permettent de discrétiser la demi-droite réelle positive dans son entier. L'hypothèse **A2b**) précise la vitesse de décroissance de ces pas de temps. Elle est vérifiée, par exemple, si  $\sum_n \gamma_n^2 < \infty$ .

$\theta$	variable muette ( $\in \mathbb{R}^d$ )
$\boldsymbol{\theta}$	variable muette ( $\in \mathbb{R}^{dN}$ )
$\theta_{n,i}$	estimée du capteur $i$ au temps $n$ ( $\in \mathbb{R}^d$ )
$\boldsymbol{\theta}_n$	estimées des $N$ capteurs ( $\in \mathbb{R}^{dN}$ )
$\langle \boldsymbol{\theta}_n \rangle$	estimée moyenne ( $\in \mathbb{R}^d$ )
$J$	projecteur sur l'espace des consensus ( $\in \mathbb{R}^d$ )
$J^\perp \boldsymbol{\theta}_n$	vecteur des écarts au consensus ( $\in \mathbb{R}^{dN}$ )
$X_n$	observations concaténées au temps $n$ , ( $\in \mathbb{R}^{dN}$ )
$h(\theta)$	champ moyen de l'algorithme ( $\in \mathbb{R}^d \rightarrow \mathbb{R}^d$ )
$V(\theta)$	fonction de Lyapounov associée à $h$
$H(\boldsymbol{\theta}; X)$	champ perturbé ( $\in \mathbb{R}^{dN} \times \mathbb{R}^{dN} \rightarrow \mathbb{R}^{dN}$ )
$\langle H(\boldsymbol{\theta}; X) \rangle$	moyenne de $H$ ( $\in \mathbb{R}^{dN} \times \mathbb{R}^{dN} \rightarrow \mathbb{R}^d$ )
$\mathbb{1}_N$	vecteur $(1, \dots, 1)^T$ ( $\in \mathbb{R}^N$ )
$\gamma_n$	pas de l'algorithme ( $\in \mathbb{R}_+$ )

TABLE 1: Résumé des notations

**Hypothèse A3** Il existe une fonction  $V : \mathbb{R}^d \rightarrow \mathbb{R}^+$  telle que :

- $V$  soit différentiable et  $\nabla V$  soit Lipschitz.
- Pour tout  $\theta \in \mathbb{R}^d$ ,  $\nabla V(\theta)^T h(\theta) \leq 0$ .
- Il existe une constante  $C_1$ , telle que pour tout  $\theta \in \mathbb{R}^d$ ,  $|\nabla V(\theta)|^2 \leq C_1(1 + V(\theta))$ .
- Pour tout  $M > 0$ , l'ensemble  $\{\theta \in \mathbb{R}^d : V(\theta) \leq M\}$  soit compact.
- L'ensemble  $\mathcal{L} := \{\theta \in \mathbb{R}^d : \nabla V(\theta)^T h(\theta) = 0\}$  soit borné.
- $V(\mathcal{L})$  soit d'intérieur vide.

L'hypothèse A3b) signifie que  $V$  est une fonction de Lyapounov pour le champ moyen  $h$ . Quand  $h$  est connu (et continu), A3 combiné à la condition  $\sum_n \gamma_n = +\infty$  permet de prouver la convergence de la suite déterministe  $t_{n+1} = t_n + \gamma_{n+1} h(t_n)$  vers l'ensemble  $\mathcal{L}$ . Quand  $h$  est inconnu et remplacé par son approximation stochastique  $H$ , le comportement limite de l'algorithme bruité est le même pourvu que  $H$  satisfasse à des conditions de régularité et que  $(\gamma_n)$  vérifie  $\sum_n \gamma_n^2 < \infty$ .

Nous supposons que :

**Hypothèse A4** a) Il existe une constante  $C_2$  telle que pour tout  $\boldsymbol{\theta} \in \mathbb{R}^{dN}$ ,

$$\mathbb{E}_\theta [ |H(\boldsymbol{\theta}; X)|^2 ] \leq C_2 (1 + V(\langle \boldsymbol{\theta} \rangle) + |J^\perp \boldsymbol{\theta}|^2),$$

$$\mathbb{E}_\theta | \langle H(\boldsymbol{\theta}; X) \rangle - \langle H(J\boldsymbol{\theta}; X) \rangle | \leq C_3 |J^\perp \boldsymbol{\theta}|,$$

$$| \mathbb{E}_\theta \langle H(\boldsymbol{\theta}; X) \rangle - \mathbb{E}_{J\boldsymbol{\theta}} \langle H(J\boldsymbol{\theta}; X) \rangle | \leq C_4 |J^\perp \boldsymbol{\theta}|.$$

- La fonction  $h$  soit continue sur  $\mathbb{R}^d$ .

## 2.3 Analyse de la convergence

**Résultat 1 (Convergence)** Sous les hypothèses A1, A2, A3, A4 l'algorithme (2) converge presque-sûrement,

$$\lim_{n \rightarrow \infty} |\boldsymbol{\theta}_n - \mathbb{1} \otimes \langle \boldsymbol{\theta}_n \rangle| = 0, \quad \lim_{n \rightarrow \infty} d(\langle \boldsymbol{\theta}_n \rangle, \mathcal{L}) = 0.$$

De plus, presque-sûrement,  $(\langle \boldsymbol{\theta}_n \rangle)_{n \geq 1}$  converge vers une composante connexe de  $\mathcal{L}$ .

Le théorème 1 affirme que, presque-sûrement, le vecteur  $\boldsymbol{\theta}_n$  converge vers l'espace de consensus lorsque  $n \rightarrow \infty$ . Et que

de plus, la valeur moyenne des estimées sur le graphe converge vers une composante connexe de  $\mathcal{L}$ . Quand  $\mathcal{L}$  est un ensemble fini, le théorème 1 entraîne, que  $\langle \boldsymbol{\theta}_n \rangle$  converge presque-sûrement vers un point de  $\mathcal{L}$ .

Le résultat suivant précise la vitesse de convergence sous des hypothèses supplémentaires.

On note  $\theta_*$  un point de  $\mathbb{R}^d$  vérifiant :

**Hypothèse A5** a)  $\theta_* \in \mathcal{L}$ .

b) Le champ moyen  $h$  est différentiable au point  $\theta_*$  et  $h(\theta) = \nabla h(\theta_*)(\theta - \theta_*) + O(|\theta - \theta_*|^2)$  pour tout  $\theta$  dans un voisinage de  $\theta_*$ , où  $\nabla h(\theta_*)$  désigne la matrice Jacobienne  $d \times d$  de  $h$  au point  $\theta_*$ .

c)  $\nabla h(\theta_*)$  une matrice stable : la plus grande partie réelle  $-L$  de ses valeurs propres est strictement négative.

d) Il existe  $\delta > 0$  tel que la fonction :

$$\theta \mapsto \mathbb{E}_\theta [ |H(\boldsymbol{\theta}; X)|^{2+\delta} ]$$

soit localement bornée en  $\mathbb{1} \otimes \theta_*$ .

e) La fonction  $Q : \mathbb{R}^{dN} \rightarrow \mathbb{R}^{d \times d}$  définie par :

$$Q(\boldsymbol{\theta}) = \mathbb{E}_\theta [ (\langle H(\boldsymbol{\theta}; X) \rangle - \mathbb{E}_\theta \langle H(\boldsymbol{\theta}; X) \rangle) \cdot (\langle H(\boldsymbol{\theta}; X) \rangle - \mathbb{E}_\theta \langle H(\boldsymbol{\theta}; X) \rangle)^T ]$$

à valeurs matricielles est continue au point  $\mathbb{1} \otimes \theta_*$ .

f)  $Q(\mathbb{1} \otimes \theta_*)$  est définie positive.

**Résultat 2 (Théorème Central Limite)** En supposant A1, A3, A4 et  $\gamma_n = \gamma_0 n^{-a}$  avec  $a \in (1/2, 1]$ ; en supposant qu'il existe un point  $\theta_*$  assurant A5; alors, conditionnellement à l'évènement  $\{\langle \boldsymbol{\theta}_n \rangle \rightarrow \theta_*\}$ , on a :

$$\gamma_n^{-1/2} (\boldsymbol{\theta}_n - \mathbb{1} \otimes \theta_*) \rightarrow \mathbb{1} \otimes Z$$

où  $Z$  est un vecteur gaussien  $d \times 1$  centré de matrice de covariance  $\Sigma$  est l'unique solution de :

$$(\nabla h(\theta_*)^T + \zeta I_d) \Sigma + \Sigma (\nabla h(\theta_*) + \zeta I_d) = -Q(\theta^*)$$

où  $\zeta = 0$  si  $a \in (1/2, 1)$  et  $\zeta = 1/(2\gamma_0)$  si  $a = 1$ . Si  $a = 1$ , il faut  $2L\gamma_0 < 1$  où  $L = \max \Re \lambda$  pour  $\lambda$  valeur propre de  $\nabla h(\theta_*)$ .

## 3 Maximum de Vraisemblance distribué

On dispose d'observations i.i.d.  $\{(X_{n,1}, \dots, X_{n,N})\}_n$  de loi de densité  $f_*$  sur  $\mathbb{R}^{\sum m_i}$ , et d'une famille de densités  $\{f(x, \theta)\}_{\theta \in \mathbb{R}^d}$  sur  $\mathbb{R}^{\sum m_i}$  de la forme  $f(x, \theta) = \prod_{i=1}^N f_i(x_i; \theta)$  où  $f_i(\cdot; \theta)$  est une densité sur  $\mathbb{R}^{m_i}$ . Il s'agit d'estimer l'estimateur de maximum de vraisemblance distribué.

Dans cette application,  $H_i(\theta, x) = \nabla_\theta \log f_i(x; \theta)$ ;  $\mu_\theta(dx) = f_*(x) dx$  pour tout  $\boldsymbol{\theta} \in \mathbb{R}^{dN}$  et

$$h(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \int \nabla_\theta \log f_i(x; \theta) f_*(x) dx, \quad \boldsymbol{\theta} \in \mathbb{R}^d.$$

Il est facile de vérifier que la fonction  $V(\boldsymbol{\theta}) := \int \log f_*(x) f_*(x) dx - \int \log f(x; \boldsymbol{\theta}) f_*(x) dx$  est une fonction de Lyapounov. Puisque  $\nabla V = -h$ , on a  $\mathcal{L} = \{\boldsymbol{\theta} \in \mathbb{R}^d : \nabla V(\boldsymbol{\theta}) = 0\}$  et les résultats énoncés en Section 2 assurent la convergence de l'algorithme de

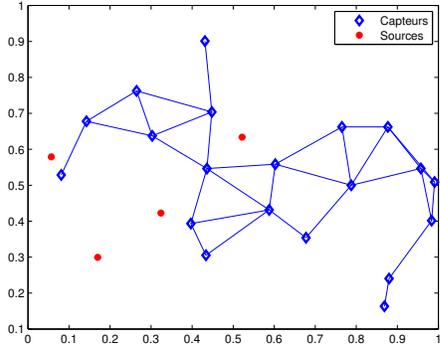


FIGURE 1: Réseau de 20 capteurs et 4 sources

Maximum de Vraisemblance distribué vers les mêmes points limites que l'algorithme de Maximum de Vraisemblance standard (i.e. centralisé). En observant que  $V$  est la distance de Kullback entre  $f_*$  et la famille  $\{f(x, \theta)\}_{\theta \in \mathbb{R}^d}$ , on retrouve l'interprétation de l'estimateur du maximum de vraisemblance comme recherche de la densité dans la famille  $\{f(\cdot, \theta)\}_{\theta \in \mathbb{R}^d}$  qui minimise la distance de Kullback à la vraie loi  $f_*$ .

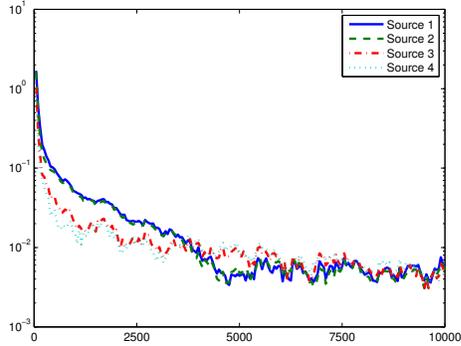


FIGURE 2: Pour chacune des sources  $k$  (ordonnées par abscisse croissante), tracé de  $n \mapsto \Delta_n(k)$

Nous illustrons numériquement le comportement asymptotique de cet algorithme. Pour ce faire, on considère le réseau décrit Figure 1 : on dispose  $N = 20$  capteurs sur le carré  $[0, 1] \times [0, 1]$  (représentés par un cercle), qui veulent estimer la position de  $D = 4$  sources (représentées par une étoile) ; les traits pleins représentent le système de voisinage de chaque capteur. La position des capteurs est connue (et notée  $z_i, z_i \in \mathbb{C}$ ). On veut estimer  $\theta_* = (\theta_*(1), \dots, \theta_*(D)) \in \mathbb{C}^D$  - ou, de façon équivalente, un vecteur de  $\mathbb{R}^d$  avec  $d = 2D$ , le vecteur collectant les coordonnées des  $D$  sources. Les observations sont simulées en prenant  $f_* \sim CN(\theta_*, \text{Diag}(\sigma_{i,1}^2 \dots \sigma_{i,D}^2))$  où les variances  $\sigma_{i,1}^2 \dots \sigma_{i,D}^2$  sont supposées connues et données par  $\sigma_{i,k}^2 = 10|\theta_*(k) - z_i|^2 + 0.1$ . On implémente l'algorithme (2) en prenant  $\gamma_n \sim 0.1 n^{-0.6}$ . Pour l'étape de "gossip", on sélectionne aléatoirement un capteur  $i$  et un de ses voisins  $j$ , et on pose  $w_n(i, j) = w_n(j, i) = 1/2$ ,  $w_n(k, k) = 1$  pour  $k \neq \{i, j\}$ .

Sur la Figure 2, on visualise la convergence vers l'espace de consensus. On définit  $\Delta_n(k) := |\theta_n(k) - \mathbb{1} \otimes \theta_*(k)|$ , où  $\theta_n(k) \in \mathbb{C}^N$

est le vecteur extrait de  $\theta_n$  qui collecte les  $N$  estimées de la position de la source  $k$ . Pour chacune des sources  $k \in \{1, \dots, D\}$ , on trace l'évolution de  $\Delta_n(k)$  en fonction du nombre d'itérations  $n$ . Cette quantité tend vers zéro, ce qui illustre la convergence vers l'espace de consensus et plus précisément, vers le maximum de vraisemblance  $\theta_*$ .

Sur la Figure 3, on visualise la vitesse de convergence donnée par le TCL : on trace  $n \mapsto \gamma_n^{-1} \mathcal{E}_n$  où  $\mathcal{E}_n$  est une estimation par une méthode de Monte Carlo basée sur 500 répliques, de l'erreur quadratique  $\mathbb{E}[|\theta_n - \mathbb{1} \otimes \theta_*|^2]$ . On peut observer la convergence de cette quantité, ce qui confirme l'expression de la vitesse de convergence de l'erreur quadratique d'estimation.

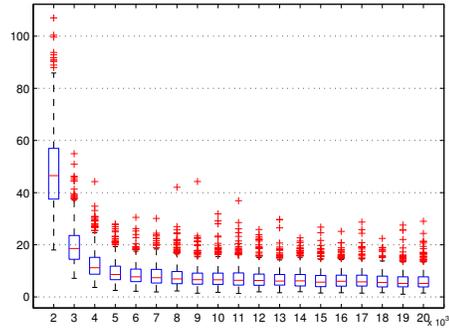


FIGURE 3: Tracé de  $n \mapsto \gamma_n^{-1} \mathcal{E}_n$

## Références

- [1] H.J. Kushner and G.G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, 2003.
- [2] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. of Mathem. Statist.*, vol. 22, no. 3, pp. 400–407, 1951.
- [3] S.S. Ram, A. Nedic, and V.V. Veeravalli, "Distributed stochastic sub-gradient projection algorithms for convex optimization," *Arxiv preprint arXiv:0811.2595*, 2008.
- [4] A. Benveniste, M. Metivier, and Priouret P., *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, 1987.
- [5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. on Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
- [6] I.D. Schizas, A. Ribeiro, and G.B. Giannakis, "Consensus in ad hoc WSNs with noisy links-Part I : Distributed estimation of deterministic signals," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.
- [7] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, July 2007.
- [8] S. Kar and J.M.F. Moura, "Distributed consensus algorithms in sensor networks : Quantized data and random link failures," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, 2010.
- [9] C. Andrieu, E. Moulines, and P. Priouret, "Stability of stochastic approximation under verifiable conditions," *SIAM J. Control Optim.*, vol. 44, no. 1, pp. 283–312, 2005.
- [10] M. Pelletier, "Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing," *Annals of Applied Probability*, vol. 8, no. 1, pp. 10–44, 1998.