

Algorithme de Douglas Rachford distribué

Adil Salim

En collaboration avec R. Mourya, P. Bianchi et C. Richard

Telecom ParisTech

21 novembre 2017

Sommaire

Optimisation distribuée

Algorithme de Douglas Rachford

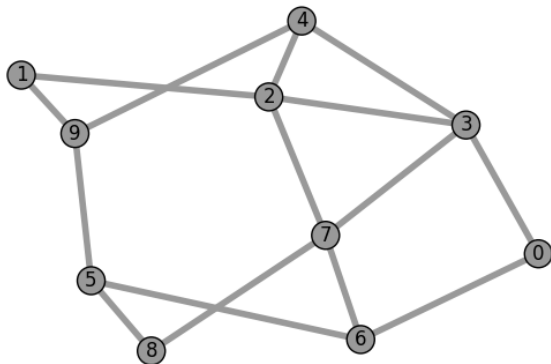
Application au problème de consensus

Analyse de l'algorithme de Douglas Rachford distribué

Contexte

Graphe $G = (V, E)$

$$\min_{x \in X^V} \sum_{i \in V} \mathbf{E}_{\theta} (\ell_i(x(i), \theta(i))) + \sum_{\{i,j\} \in E} \phi_{i,j}(x(i), x(j)) \quad (1)$$



Contexte

$$\min_{x \in X^V} \sum_{i \in V} \mathbf{E}_\theta (\ell_i(x(i), \theta(i))) + \sum_{\{i,j\} \in E} \phi_{i,j}(x(i), x(j))$$

- ▶ Variable locale $x(i)$, $i \in V$
- ▶ Variable aléatoire locale $\theta(i)$
- ▶ Perte locale convexe $\mathbf{E}_\theta (\ell_i(x(i), \theta(i)))$
- ▶ Terme de communication convexe $\phi_{i,j}(x(i), x(j))$, $\{i, j\} \in E$

Contexte

$$\min_{x \in X^V} \sum_{i \in V} \mathbf{E}_{\theta} (\ell_i(x(i), \theta(i))) + \sum_{\{i,j\} \in E} \phi_{i,j}(x(i), x(j))$$

Algorithme

1. Distribué : Information disponible pour $x(i)$
 - ▶ La perte locale
 - ▶ $x(j)$ si $j \sim i$.
2. Adaptatif : Loi de $\theta(i)$ inconnue. Réalisations $(\theta_n(i))_n$ i.i.d observées par i .
3. Asynchrone : Mise à jour à des instants aléatoires

Sommaire

Optimisation distribuée

Algorithme de Douglas Rachford

Application au problème de consensus

Analyse de l'algorithme de Douglas Rachford distribué

Algorithme de Douglas Rachford

Un grand nombre de problème en maths appliquées s'écrivent

$$\min_{x \in X} F(x) + G(x)$$

où $F, G : X \mapsto (-\infty, +\infty]$ sont **convexes** sur X , euclidien.

Exemple classique : Algorithme de **Douglas Rachford**

$$u_n = \text{prox}_{\gamma F}(x_{n-1})$$

$$z_n = \text{prox}_{\gamma G}(2u_n - x_{n-1})$$

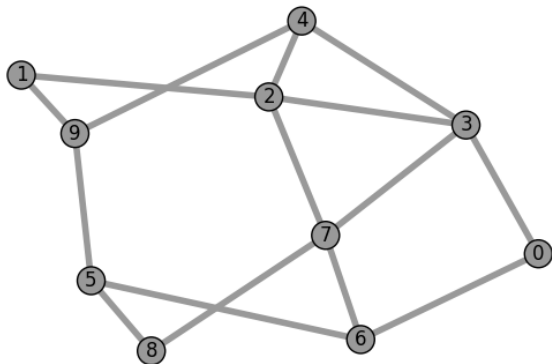
$$x_n = x_{n-1} + z_n - u_n.$$

où

$$\text{prox}_{\gamma F}(x) = \arg \min_{y \in X} \frac{1}{2} \|x - y\|^2 + \gamma F(y).$$

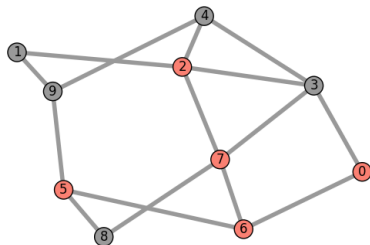
Algorithme de Douglas Rachford distribué

$x_{n-1} = (x_{n-1}(0), \dots, x_{n-1}(9))$. Un vecteur $x_{n-1}(i)$ pour chaque noeud i .



Algorithme de Douglas Rachford distribué

Certains noeuds (noeuds de calcul) se réveillent aléatoirement $\{0, 2, 5, 6, 7\}$.



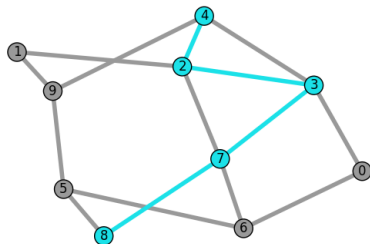
Ils observent $\theta_n(i), \forall i \in \{0, 2, 5, 6, 7\}$. Puis,

$$\forall i \in \{0, 2, 5, 6, 7\}, \quad u_n(i) = \text{prox}_{\frac{\gamma}{p_i} \ell_i(\cdot, \theta_n(i))}(x_{n-1}(i))$$

Pour les autres noeuds i , $u_n(i) = x_{n-1}(i)$

Algorithme de Douglas Rachford distribué

Certaines arêtes (arêtes de communication) se réveillent aléatoirement. Sous-graphe (ν_n, ε_n)



De l'information est échangée le long des arêtes de communication:

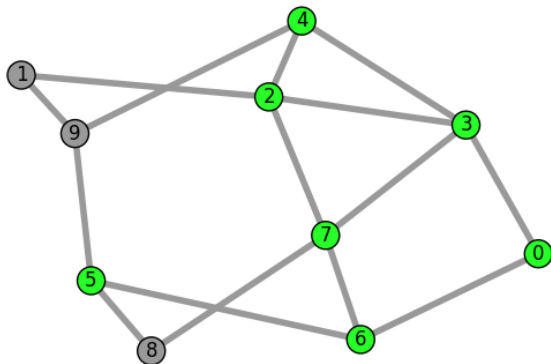
$$z_n|_{\nu_n} = \text{prox}_{\sum_{e \in \varepsilon_n} \frac{\gamma}{q_e} \phi_e} (2u_n|_{\nu_n} - x_{n-1}|_{\nu_n})$$

Pour les autres noeuds $i \notin \nu_n$, $z_n(i) = 2u_n(i) - x_{n-1}(i)$, ce qui revient à $z_n(i) = x_{n-1}(i)$ si $i \notin \nu_n$ n'est pas un noeud de calcul.

Algorithme de Douglas Rachford distribué

Finalement, pour tout $i \in V$,

$$x_n(i) = x_{n-1}(i) + z_n(i) - u_n(i)$$



L'algorithme est distribué, adaptatif, asynchrone.

Sommaire

Optimisation distribuée

Algorithme de Douglas Rachford

Application au problème de consensus

Analyse de l'algorithme de Douglas Rachford distribué

Problème du consensus

Si le graphe G est connexe,

$$\begin{aligned} & \min_{x \in X} \sum_{i \in V} \mathbf{E}_{\theta} (\ell_i(x, \theta(i))) \\ = & \min_{x \in X^V} \sum_{i \in V} \mathbf{E}_{\theta} (\ell_i(x(i), \theta(i))) \quad + \quad \sum_{\{i,j\} \in E} \phi_{i,j}(x(i), x(j)) \end{aligned}$$

$$\text{où } \phi_{i,j}(x(i), x(j)) = \begin{cases} 0 & \text{si } x(i) = x(j) \\ +\infty & \text{sinon} \end{cases}$$

On applique Douglas Rachford distribué au problème de consensus.
La perte locale ℓ_i est choisie quadratique.

Comparaison avec d'autres méthodes d'optimisation distribuée

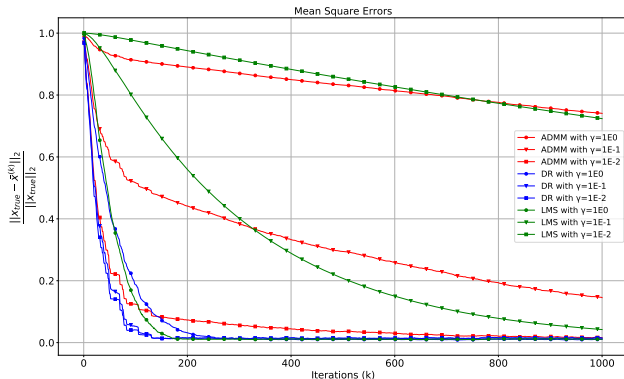


Figure 1: DR, ADMM asynchrone, Diffusion LMS asynchrone.

Comparaison avec d'autres méthodes d'optimisation distribuée

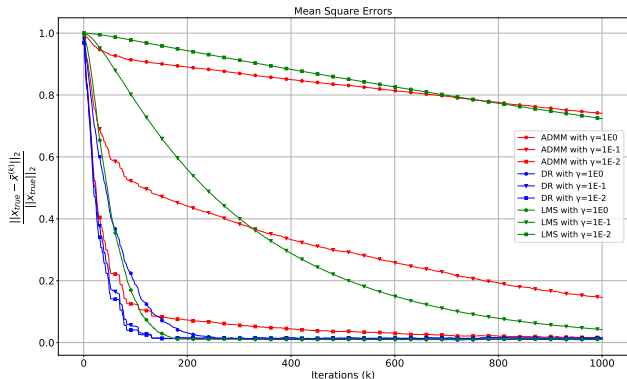


Figure 2: DR, ADMM asynchrone, Diffusion LMS asynchrone.

Application à la localisation de cible sous-marine

N récepteurs et M émetteurs dans le plan. Cible mouvante de coordonnées (x_t, y_t) à l'instant t .

Cas $N = 1$ récepteur. Estimateur de (x_t, y_t) :

$$\arg \min_{(x,y) \in \mathbf{R}^2} \|A(t)(x, y) - b(t)\|^2$$

où $A(t), b(t)$ encodent positions/mesures des émetteurs/ du récepteur

Application à la localisation de cible sous-marine

Cas N quelconque. Estimateur de (x_t, y_t) :

$$\arg \min_{(x,y) \in \mathbb{R}^2} \sum_{i \in V} \|A(t)_i(x, y) - b(t)_i\|^2$$

Problème du consensus.

Application à la localisation de cible sous-marine

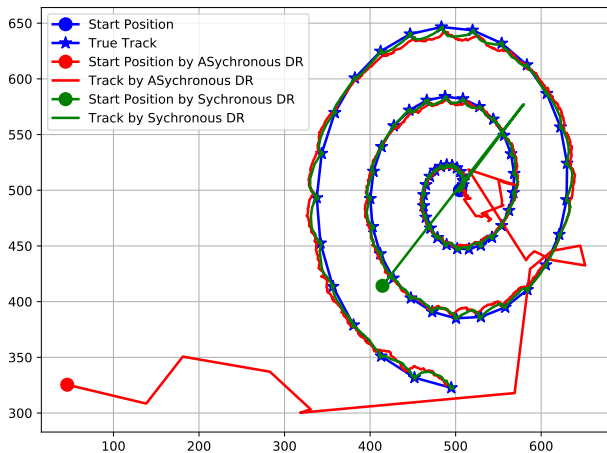


Figure 3: Trajectoire et trajectoire estimée de la cible.

Sommaire

Optimisation distribuée

Algorithme de Douglas Rachford

Application au problème de consensus

Analyse de l'algorithme de Douglas Rachford distribué

Optimisation stochastique

Après modélisation, le problème (1) est une instance de

Problème général :

$$\min_{x \in X} F(x) + G(x) \quad (2)$$

avec

$$F(x) = \mathbf{E}_{\xi}(f(x, \xi))$$

$$G(x) = \mathbf{E}_{\xi}(g(x, \xi))$$

où ξ est une variable aléatoire et $f(\cdot, s), g(\cdot, s) : X \rightarrow (-\infty, +\infty]$ est convexe pour tout s .

F, G inconnues mais révélées *online* à travers des réalisations $(f(\cdot, \xi_n))_n$ et $(g(\cdot, \xi_n))_n$

Douglas Rachford stochastique

L'algorithme de Douglas Rachford distribué est une instance de

Algorithme de **Douglas Rachford stochastique** :

$$u_n = \text{prox}_{\gamma f(\cdot, \xi_n)}(x_{n-1}^\gamma)$$

$$z_n = \text{prox}_{\gamma g(\cdot, \xi_n)}(2u_n - x_{n-1}^\gamma)$$

$$x_n^\gamma = x_{n-1}^\gamma + z_n - u_n.$$

- ▶ pas $\gamma > 0$ constant
- ▶ $\text{dom}(g(\cdot, \xi)) = \{x \in X, g(x, \xi) < +\infty\}$ dépend de ξ
- ▶ (ξ_n) famille de variables aléatoires *i.i.d*

Comportement Dynamique

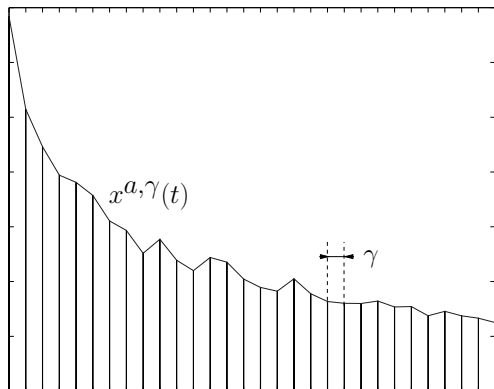


Figure 4: Processus continu interpolé : $x^{a, \gamma}(t)$ démarrnant en $x(0) = a$.

Comportement Dynamique

Supposons que F est différentiable et notons ∇F son gradient.

Notons ∂G la sous-différentielle de G .

Considérons l'inclusion différentielle (ID)

$$\dot{x}(t) \in -(\nabla F + \partial G)(x(t)).$$

Il existe une unique fonction absolument continue x^a qui satisfait cette inclusion p.p et telle que $x^a(0) = a \in \text{Dom}(G)$.

[MBSR'17]

$$x^{a,\gamma} \xrightarrow{\gamma \rightarrow 0} x^a$$

au sens de la convergence de processus aléatoires.

Stabilité

On cherche à étudier le comportement de (x_n^γ) lorsque $n \rightarrow \infty$ et $\gamma \rightarrow 0$.

Pour cela, l'étude dynamique ne suffit pas.

Les hypothèses suivantes donnent un résultat de stabilité de la chaîne de Markov $(x_n^\gamma)_n$ qui suffit.

- ▶ $F + G \xrightarrow{\infty} +\infty$
- ▶ $\exists L(\xi) > 0, x_\star \in \arg \min F + G$ tels que $L(\xi) \|\nabla f(x_\star, \xi)\|^2$ est intégrable et pour tout $x \in E$,

$$\|\nabla f(x, \xi) - \nabla f(x_\star, \xi)\| \leq L(\xi) \|x - x_\star\|$$

Comportement asymptotique des itérées

[MBSR'17]

$$\text{Soit } \bar{x}_n^\gamma = \frac{1}{n} \sum_{k=0}^{n-1} x_k^\gamma$$

$$\forall \varepsilon > 0, \quad \limsup_{n \rightarrow \infty} \mathbf{P} [d(\bar{x}_n^\gamma, \arg \min F + G) \geq \varepsilon] \xrightarrow{\gamma \rightarrow 0} 0.$$

$$\limsup_{n \rightarrow \infty} d(\mathbf{E} \bar{x}_n^\gamma, \arg \min F + G) \xrightarrow{\gamma \rightarrow 0} 0.$$

$$\forall \varepsilon > 0, \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{P} [d(x_k^\gamma, \arg \min F + G) \geq \varepsilon] \xrightarrow{\gamma \rightarrow 0} 0.$$