Opérateurs monotones maximaux et approximation stochastique

Walid Hachem (CNRS / Télécom ParisTech)

Travail mené avec Pascal Bianchi et Adil Salim (Télécom ParisTech)

GdR ISIS, journée sur l'optimisation en milieu aléatoire, 8 novembre 2016.

Objectif

Problèmes déterministes :

- Problèmes d'optimisation convexe,
- Inégalités variationnelles,
- ...

Algorithmes déterministes :

- Point proximal, « splitting » de Douglas-Rachford, forward-backward, ADMM, algorithme primal-dual de Vũ-Condat, ...

Cadre théorique : **opérateurs monotones**.

Optimisation en milieu aléatoire :
Etudier des versions aléatoires des opérateurs monotones et des algorithmes associés.

Généralités

- Fct convexes, opérateurs monotones et algorithme du point proximal
- Le cas aléatoire : approximation stochastique

Présentation du problème

Le pas décroissant

Le pas constant

Applications et persectives

Recherche d'un minimum

$f : \mathbb{R}^N \to \overline{\mathbb{R}}$ fonction convexe semi continue inférieurement (*sci*)

- F est lisse : algorithme du **gradient** :

$$x_{n+1} = x_n - \gamma \nabla F(x_n)$$

Recherche d'un minimum

$f : \mathbb{R}^N \to \overline{\mathbb{R}}$ fonction convexe semi continue inférieurement (*sci*)

- Cas général : algorithme du **sous-gradient** :

$$x_{n+1} \in x_n - \gamma \partial F(x_n)$$

Recherche d'un minimum

$f : \mathbb{R}^N \to \overline{\mathbb{R}}$ fonction convexe semi continue inférieurement (*sci*)

- Cas général : algorithme du **sous-gradient** :

$$x_{n+1} \in x_n - \gamma \partial F(x_n)$$

- Algorithme du **point proximal** :

$$x_{n+1} = x_n - \gamma \partial F(x_{n+1})$$

Recherche d'un minimum

$f : \mathbb{R}^N \to \overline{\mathbb{R}}$ fonction convexe semi continue inférieurement (*sci*)

- Cas général : algorithme du **sous-gradient** :

$$x_{n+1} \in x_n - \gamma \partial F(x_n)$$

- Algorithme du **point proximal** :

$$\begin{aligned} x_{n+1} &= x_n - \gamma \partial F(x_{n+1}) \\ &= (I + \gamma \partial F)^{-1} x_n \end{aligned}$$

Recherche d'un minimum

$f : \mathbb{R}^N \to \overline{\mathbb{R}}$ fonction convexe semi continue inférieurement (*sci*)

- Cas général : algorithme du **sous-gradient** :

$$x_{n+1} \in x_n - \gamma \partial F(x_n)$$

- Algorithme du **point proximal** :

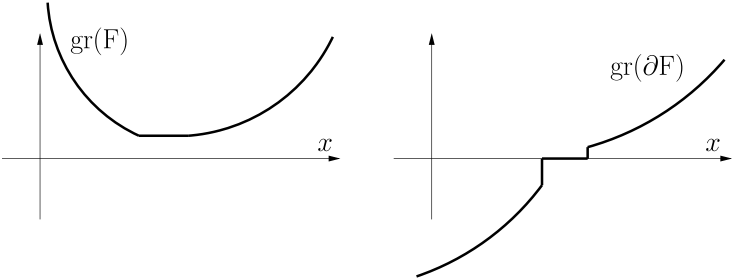
$$\begin{aligned} x_{n+1} &= x_n - \gamma \partial F(x_{n+1}) \\ &= (I + \gamma \partial F)^{-1} x_n \end{aligned}$$

Notation : $x_{n+1} = \mathrm{prox}_{\gamma F}(x_n)$.

Sous-différentiel d'une fonction convexe *sci*

Propriétés de la fonction à valeurs ensemblistes ∂F :

- **Monotonie :** $\forall x, y, \ \ \forall u \in \partial F(x), v \in \partial F(y), \ \langle u - v, x - y \rangle \geq 0$,
- **Maximalité :** on ne peut ajouter aucun point au graphe $\{(x, y) \in \mathbb{R}^N \times \mathbb{R}^N \ : \ y \in \partial F(x)\}$ de ∂F sans violer la monotonie.
- Ensemble des **zéros** $\mathrm{zer}(\partial F) = \{x \ : \ 0 \in \partial F(x)\}$ coïncide avec l'ensemble des **minimiseurs** de F.

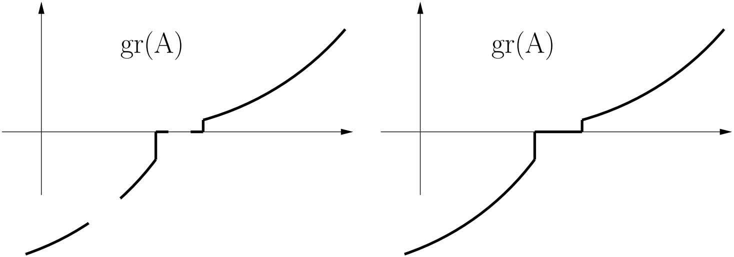


Généralisation : les opérateurs monotones maximaux

Opérateur **monotone** sur $\mathbb{R}^N$: fonction $A : \mathbb{R}^N \to 2^{\mathbb{R}^N}$ telle que

$$\forall x, y, \quad \forall u \in A(x), v \in A(y), \; \langle u - v, x - y \rangle \geq 0$$

Opérateur **monotone maximal** (MM) : on ne peut ajouter aucun point au graphe de A sans violer la monotonie.



Monotone *vs* monotone maximal

Maximalité

- Conséquences topologiques : le graphe est fermé, $A(x)$ est fermé convexe, $\overline{\mathrm{dom}(A)}$ est convexe où $\mathrm{dom}(A) = \{x \,:\, A(x) \neq \emptyset\}$, etc.
- Pour tout $\gamma > 0$, $(I + \gamma A)^{-1}$ est une **contraction définie sur $\mathbb{R}^N$ tout entier** ($\Leftrightarrow$ maximalité [Minty'61]).

Algorithme du point proximal

Pour tout $\gamma > 0$,

$$x_{n+1} = (I + \gamma A)^{-1} x_n$$

converge vers un point de $\mathrm{zer}(A)$ si $\mathrm{zer}(A) \neq \emptyset$.

- Une **brique de base** des algorithmes d'optimisation convexe, au delà du cas $A = \partial F$.

Comportement dynamique : ODE et inclusions différentielles

Soit F convexe dérivable. Considérons l'**ODE**

$$\begin{cases} \dot{z}(t) &= -\nabla F(z(t)), \quad t \in]0, \infty[\\ z(0) &= z_0 \in \mathrm{dom}(F). \end{cases}$$

Comportement dynamique : ODE et inclusions différentielles

Soit F convexe *sci*. Considérons l'**inclusion différentielle (ID)**

$$\begin{cases} \dot{z}(t) & \in & -\partial F(z(t)), \quad t \in]0, \infty[\\ z(0) & = & z_0 \in \mathrm{dom}(F). \end{cases}$$

Comportement dynamique : ODE et inclusions différentielles

Soit A un opérateur MM. Considérons l'**inclusion différentielle (ID)**

$$\begin{cases} \dot{z}(t) & \in & -A(z(t)), \quad t \in]0, \infty[\\ z(0) & = & z_0 \in \mathrm{dom}(A). \end{cases}$$

Comportement dynamique : ODE et inclusions différentielles

Soit A un opérateur MM. Considérons l'**inclusion différentielle (ID)**

$$\begin{cases} \dot{z}(t) & \in & -A(z(t)), \quad t \in]0, \infty[\\ z(0) & = & z_0 \in \mathrm{dom}(A). \end{cases}$$

Pour tout $z_0 \in \mathrm{dom}(A)$, cette ID admet une **solution unique**

Comportement dynamique : ODE et inclusions différentielles

Soit A un opérateur MM. Considérons l'**inclusion différentielle (ID)**

$$\begin{cases} \dot{z}(t) & \in & -A(z(t)), \quad t \in]0, \infty[\\ z(0) & = & z_0 \in \mathrm{dom}(A). \end{cases}$$

Pour tout $z_0 \in \mathrm{dom}(A)$, cette ID admet une **solution unique**

Une méthode de preuve :

- Construire des processus interpolés $z_\gamma(t)$ à partir de l'algorithme du point proximal. Faire $\gamma \to 0$, établir un résultat de compacité et montrer que $\lim_{\gamma \to 0} z_\gamma$ sur les intervalles de temps compacts est la solution de l'ID.

Généralités
Fct convexes, opérateurs monotones et algorithme du point proximal
Le cas aléatoire : approximation stochastique

Présentation du problème

Le pas décroissant

Le pas constant

Applications et persectives

L'algorithme du gradient stochastique

- $\{f_\xi(\cdot)\}$ famille de fonctions **indexée par une v.a.** ξ,
- Algorithme :

$$x_{n+1} = x_n - \gamma_{n+1} \nabla f_{\xi_{n+1}}(x_n),$$

  - (ξ_n) processus *iid*, $\xi_n \stackrel{\mathcal{L}}{=} \xi$,
  - (γ_n) suite de pas positifs
    - soit **décroissants**, en général $\sum \gamma_n = \infty$, $\sum \gamma_n^2 < \infty$ $(\in \ell^2 \setminus \ell^1)$,
    - soit **constants** : $\gamma_n \equiv \gamma > 0$.

- Nous voulons que la suite (x_n) se rapproche d'$\arg\min_x F(x)$ où $F(x) = \mathbb{E}_\xi f_\xi(x)$.

Cas plus général : algorithme de Robbins Monro

$$x_{n+1} = x_n - \gamma_{n+1} h_{\xi_{n+1}}(x_n) = x_n - \gamma_{n+1} H(x_n) - \gamma_{n+1} e_{n+1}$$

où

- $H(x) = \mathbb{E}_\xi h_\xi(x)$ est un **champ moyen**.
  Gradient stochastique : $H(x) = \nabla F(x)$ où $F(x) = \mathbb{E}_\xi f_\xi(x)$,
- $e_{n+1} = h_{\xi_{n+1}}(x_n) - H(x_n)$ est un **incrément de martingale**.

Méthode de l'ODE [Ljung'77] : la trajectoire interpolée des (x_n) se rapproche d'une solution de l'ODE $\dot{z}(t) = -H(z(t))$.

- Pas décroissant : convergence p.s. sur un compact glissant de l'axe des temps quand $n \to \infty$,
- Pas constant : cvg faible sur un compact fixe quand $\gamma \downarrow 0$.

Etude dynamique et comportement pour $n \to \infty$: [Kushner *et.al.*], [Benaïm *et.al.*], [Borkar *et.al.*], ...

Généralités

Présentation du problème
- Le prox stochastique et le forward-backward stochastique
- FB stochastique et inclusion différentielle

Le pas décroissant

Le pas constant

Applications et persectives

Le prox stochastique

- Analogue proximal du gradient stochastique :

$$\begin{aligned} x_{n+1} &= x_n - \gamma_{n+1} \partial f_{\xi_{n+1}}(x_{n+1}) \\ &= (I + \gamma_{n+1} \partial f_{\xi_{n+1}})^{-1} x_n = \mathrm{prox}_{\gamma_{n+1} f_{\xi_{n+1}}}(x_n). \end{aligned}$$

- Plus généralement, soit $\{A_\xi\}$ une **famille d'opérateurs MM indexée par une v.a.** ξ. On étudie

$$x_{n+1} = (I + \gamma_{n+1} A_{\xi_{n+1}})^{-1} x_n ,$$

où (ξ_n) est *iid*, $\xi_n \stackrel{\mathcal{L}}{=} \xi$.

Le « forward-backward » (FB) déterministe (fonctions)

- Objectif : trouver un minimum (supposé exister) de $G(x) + F(x)$ où G est convexe *sci* et F est convexe lisse.
- Algorithme FB : $x_{n+1} = (I + \gamma\partial G)^{-1}(x_n - \gamma\nabla F(x_n))$.
- Si $\gamma \leq 2/\|\nabla F\|_{\mathrm{Lip}}$, converge vers un zéro de $\partial G + \nabla F$.
- Exemple simple : le gradient projeté.

Le FB déterministe (opérateurs MM)

- Objectif : trouver un zéro (supposé exister) de $A + B$ où A est un opérateur MM et B est un opérateur MM dit cocoercif (généralisation de ∇F Lipschitz).

- Algorithme FB : $x_{n+1} = (I + \gamma A)^{-1}(x_n - \gamma B(x_n))$.

- Converge si γ est suffisamment petit vers un zéro de $A + B$.

- Exemples : gradient projeté, Chambolle-Pock, ADMM, Vũ-Condat qui les généralise, etc.

Le FB stochastique

$\{A_\xi\}$ et $\{B_\xi\}$ deux familles d'opérateurs MM indexées par ξ.
Nous étudions le processus (x_n) :

$$\begin{aligned} b_{n+1} &\in B_{\xi_{n+1}}(x_n), \\ x_{n+1} &= (I + \gamma_{n+1} A_{\xi_{n+1}})^{-1}(x_n - \gamma_{n+1} b_{n+1}), \end{aligned}$$

- dom $B_\xi = \mathbb{R}^N$,
- dom A_ξ dépend de ξ,
- (ξ_n) *iid*,
- Pas décroissant ou pas constant.

Littérature

Sous-différentiels : $x_{n+1} = \mathrm{prox}_{\gamma_{n+1} g_{\xi_{n+1}}}(x_n - \gamma_{n+1} \nabla f_{\xi_{n+1}}(x_n))$:

- $g_\xi \equiv G$ déterministe : [Nemirovski'09], ..., [Atchadé *et.al.*'15].
- f_ξ et g_ξ aléatoires : [Wang-Bertsekas'11 et 13], [Ryu-Boyd'15], hypothèses contraignantes.

Opérateurs monotones

- Point proximal, FB, etc. + erreurs additives déterministes sommables : riche littérature débutant par [Rockafellar'76], [Golstein-Tretyakov'79], ...
- Permutations aléatoires de résolvantes : [Passty'79].
- A déterministe et B_ξ aléatoire : [Rosasco *et.al.*'15].
- FB + erreurs additives aléatoires : [Combettes-Pesquet'16] (hypothèses de sommabilité des erreurs).

Généralités

Présentation du problème

Le prox stochastique et le forward-backward stochastique

FB stochastique et inclusion différentielle

Le pas décroissant

Le pas constant

Applications et persectives

Inclusion différentielle dans le cas aléatoire

Supposons $B_\xi = 0$. Nous suivons la méthode de l'ODE, en visant une ID de la forme

$$\dot{z}(t) \in -\mathcal{A}(z(t)) \quad \text{où} \quad \mathcal{A} = \mathbb{E}_\xi A_\xi .$$

Donner un sens à $\mathbb{E}_\xi A_\xi$, car A_ξ est à valeurs ensemblistes.
Intégrale d'Aumann :

$$\mathcal{A}(x) = \left\{ \mathbb{E}_\xi \varphi(\xi) \; : \; \varphi \text{ mesurable } \mathcal{L}^1, \; \varphi(\xi) \in A_\xi(x) \text{ p.s.} \right\} .$$

Propriétés de $\mathcal{A}$:

- dom $\mathcal{A}$ est l'intersection essentielle des dom A_ξ.
- $\mathcal{A}$ est trivialement monotone. Maximal quand $\overset{\circ}{D} \neq \emptyset$ + hypothèse de régularité sur les dom A_ξ [BH'16].
  Cas sous-différentiel étudié dans [Rockafellar'69] (échange $\mathbb{E} \leftrightarrow \partial$).

Généralités

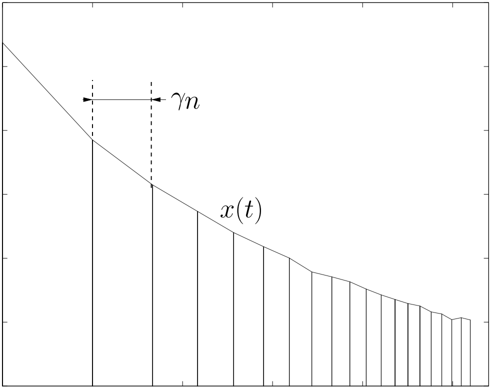
Présentation du problème

Le pas décroissant

Le pas constant

Applications et persectives

Pseudo trajectoire asymptotique (APT)



$x(t)$ est l'interpolation continue des (x_n)

Pseudo trajectoire asymptotique (APT)

Soit $\Phi_t(z_0)$ la solution de l'ID $\dot{z}(t) \in -(\mathcal{A}+\mathcal{B})(z(t))$ initialisée à z_0.

Résultat : presque sûrement, pour tout $T > 0$,

$$\sup_{s\in[0,T]} \|x(t+s) - \Phi_s(x(t))\| \xrightarrow[t\to\infty]{} 0,$$

($x(t)$ est une **APT** [Benaïm-Hirsch'96] du flot Φ).

Hypothèses :

- Existence d'un zéro de $\mathcal{A}+\mathcal{B}$,
- Régularité des dom A_ξ,
- Croissance linéaire de $B_\xi(x)$,
- Hypothèses de moments.

Preuve

$B_\xi = 0$ par simplicité.

La **régularisée de Yosida** d'un opérateur MM A pour $\gamma > 0$:

$$A^\gamma(x) = \frac{I - (I + \gamma A)^{-1}}{\gamma}(x)$$

« Approximation » univoque de A pour γ petit.

Yosida :

$$\begin{aligned} x_{n+1} &= (I + \gamma_{n+1} A_{\xi_{n+1}})^{-1} x_n = x_n - \gamma_{n+1} \frac{x_n - (I + \gamma_{n+1} A_{\xi_{n+1}})^{-1} x_n}{\gamma_{n+1}} \\ &= x_n - \gamma_{n+1} A_{\xi_{n+1}}^{\gamma_{n+1}}(x_n) \\ &= x_n - \gamma_{n+1} h^{\gamma_{n+1}}(x_n) - \gamma_{n+1} e_{n+1} \end{aligned}$$

où $h^{\gamma_{n+1}}(x_n) = \mathbb{E}_\xi A_\xi^{\gamma_{n+1}}(x)$ est un « champ moyen » et e_{n+1} est un incrément de martingale.

Preuve

Stabilité :

- Pour $x_\star \in \mathrm{zer}(\mathcal{A})$, la suite $(\|x_n - x_\star\|)$ est p.s. convergente,

Convergence rapide vers le domaine commun

- $x_n \to \mathrm{dom}\,\mathcal{A}$ p.s. au taux γ_n,

Arzelà-Ascoli

- La famille $(x(t + \,.\,))_{t\geq 0}$ est p.s. équicontinue,

Caractérisation des points d'accumulation

- Solutions de l'ID.

Corollaires : comportement asymptotique des itérées

- Pour $n \to \infty$, nous étudions la convergence de (x_n) et des moyennes empiriques
$$\bar{x}_n = \frac{\sum_{k=1}^{n} \gamma_k x_k}{\sum_{k=1}^{n} \gamma_k}$$

- La convergence de (x_n) vers $\mathrm{zer}(\mathcal{A})$ est **non garantie**, même dans le cas déterministe, quand $(\gamma_n) \in \ell^2$.

- Idée : certaines propriétés asymptotiques de l'ID se transposent à son APT [Benaïm'99]. Etudier les solutions de l'ID.

Comportement asymptotique des solutions de l'ID

- [Brézis'73] Pour tout A MM, les solutions de l'ID $\dot{z} \in -A(z)$ **convergent ergodiquement** :

$$\exists z_\star \in \mathrm{zer}(A), \quad \frac{1}{t}\int_0^t z(s)\,ds \xrightarrow[t\to\infty]{} z_\star,$$

- [Bruck'77] Si A est **demipositif** [1], alors

$$\exists x_\star \in \mathrm{zer}(A), \quad z(t) \to x_\star.$$

La demipositivité est satisfaite dans la plupart des cas, notamment le cas d'un sous-différentiel.

1. $\exists w \in \mathrm{zer}(A), \forall (u_n) \to u, \forall (v_n)$ bornée , $v_n \in A(u_n),\ \langle v_n, u_n - w\rangle \to 0 \Longrightarrow u \in \mathrm{zer}(A)$

Corollaires : comportement asymptotique des itérées

APT + stabilité :

Presque sûrement, il existe $x_\star \in \mathrm{zer}(\mathcal{A}+\mathcal{B})$ tel que

$$\overline{x}_n \to x_\star,$$

Si $\mathcal{A}+\mathcal{B}$ est demipositif, alors, p.s., $\exists x_\star \in \mathrm{zer}(\mathcal{A}+\mathcal{B})$ tel que

$$x_n \to x_\star.$$

Généralités

Présentation du problème

Le pas décroissant

Le pas constant

Applications et persectives

FB aléatoire à pas constant

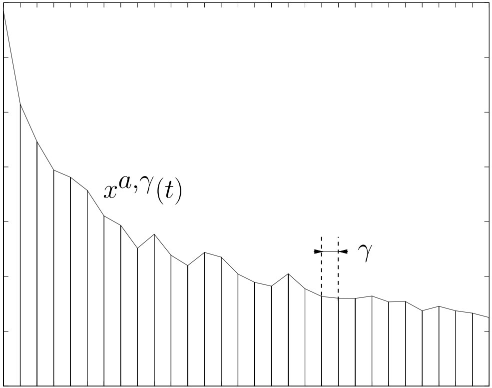
On suppose que B_ξ est univoque, $\operatorname{dom} B_\xi = \mathbb{R}^N$.

Algorithme :

$$x_{n+1}^\gamma = (I + \gamma A_{\xi_{n+1}})^{-1}(x_n^\gamma - \gamma B_{\xi_{n+1}}(x_n^\gamma))$$

- $\operatorname{dom} A_\xi$ dépend de ξ,
- (ξ_n) *iid*.

1er résultat : « weak APT »



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

1er résultat : « weak APT »

Convergence en probabilité de $x^{a,\gamma}(\cdot)$ vers la solution de l'ID, quand $\gamma \to 0$, au sens de la convergence uniforme sur les compacts.

En supposant $\mathrm{dom}(\mathcal{A}+\mathcal{B}) = \mathbb{R}^N$ par simplicité :

w-APT :

$$\forall \varepsilon > 0, \forall T > 0, \forall K \text{ compact },$$
$$\sup_{a \in K} \mathbb{P}\Big[\sup_{t \in [0,T]} \|x^{a,\gamma}(t) - \Phi_t(a)\| > \varepsilon \Big] \xrightarrow[\gamma \to 0]{} 0.$$

Hypothèses légères de moments et de régularité des domaines. Stabilité non nécessaire.

2e résultat : tension

- Comportement asymptotique des itérées : un résultat de stabilité est nécessaire.
- On regarde (x_n^γ) comme une **chaîne de Markov fellerienne** dont on étudie la **tension des lois invariantes quand** $\gamma \to 0$.

Soit I_γ l'ensemble des lois invariantes de cette chaîne.

Stabilité :
l'ensemble de lois de probabilité $I = \bigcup_{\gamma \in (0,\gamma_0]} I_\gamma$ est **tendu**.

Hypothèses :

- $\mathrm{zer}(\mathcal{A} + \mathcal{B}) \neq \emptyset$,
- Satisfaction d'un **critère de tension de type Pakes**.

Comportement asymptotique des itérées

w-APT + Tension :
Chaque point d'accumulation μ de l'ensemble I est une **mesure invariante** pour le flot Φ, *i.e.*, $\Phi_t(\mu) = \Phi(\mu)$ pour tout $t \geq 0$.

Conséquences :

$\mathcal{A} + \mathcal{B}$ MM quelconque :

$$\forall \varepsilon > 0, \quad \limsup_{n\to\infty} \mathbb{P}\left[d\left(\frac{1}{n}\sum_{k=0}^{n-1} x_k^\gamma, \mathrm{zer}(\mathcal{A}+\mathcal{B})\right) \geq \varepsilon\right] \xrightarrow[\gamma\to 0]{} 0.$$

$\mathcal{A} + \mathcal{B}$ demipositif :

$$\forall \varepsilon > 0, \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}\left[d\left(x_k^\gamma, \mathrm{zer}(\mathcal{A}+\mathcal{B})\right) \geq \varepsilon\right] \xrightarrow[\gamma\to 0]{} 0\,.$$

Satisfaction du critère de tension

- Le critère de tension découle de la coercivité d'une fonction associée à A_ξ et B_ξ.
- **Conditions suffisantes** obtenues dans des cas génériques.
  Exemple : sous-différentiels $A_\xi(x) = \partial g_\xi(x)$ et $B_\xi(x) = \nabla f_\xi(x)$: tension si l'ensemble $\text{zer}(\mathcal{A} + \mathcal{B})$ des minima de $G(x) + F(x)$ est un compact non vide et $\|\nabla f_\xi\|_{\text{Lip}}$ bornée p.s.
- De manière générale, la compacité de $\text{zer}(\mathcal{A} + \mathcal{B})$ est nécessaire.

Généralités

Présentation du problème

Le pas décroissant

Le pas constant

Applications et persectives

Minimisation contrainte

Problème : $$\min_x \mathbb{E}_\xi g_\xi(x) + \mathbb{E}_\xi f_\xi(x) \ .$$

Algorithme : $$x_{n+1} = (I + \gamma_{n+1}\partial g_{\xi_{\mathbf{n+1}}})^{-1}(x_n - \gamma_{n+1}\nabla f_{\xi_{\mathbf{n+1}}}(x_n))$$

Instance de cet algorithme : $$\min_x \mathbb{E}_\chi f_\chi(x), \quad x \in \bigcap_i C_i$$

où la projection sur $\bigcap_i C_i$ est **compliquée** alors que la projection sur le convexe C_i est **simple**.

Avec $G(x) = \sum_i \iota_{C_i}(x)$, l'algorithme s'écrit

$$x_{n+1} = \mathrm{proj}_{C_i}(x_n - \gamma_{n+1}\nabla f_{\chi_{\mathbf{n+1}}}(x_n)), \quad i \ \textbf{aléatoire}$$

Régularisation TV sur un graphe

Notation : Sur un graphe $G = (V, E)$, la variation totale de $x \in \mathbb{R}^{|V|}$ est

$$\mathrm{TV}(x) = \sum_{\{i,j\} \in E} |x_i - x_j|$$

Problème :
$$\min_{x \in \mathbb{R}^{|V|}} F(x) + \mathrm{TV}(x)$$

Exemples : « trend filtering » [R. Tibshirani'14], « graph inpainting » [Chen *et.al.*'14], ...

Algorithme FB déterministe :

$$x_{n+1} = \mathrm{prox}_{\gamma \mathrm{TV}}(x_n - \gamma \nabla F(x_n))$$

- Calcul de $\mathrm{prox}_{\mathrm{TV}}$ difficile sur les graphes non structurés,
- Mais $\exists$ un algorithme simple pour les graphes 1D : le « taut string ».

Régularisation TV sur un graphe

Une solution : écrire TV comme une espérance : soit ξ une marche aléatoire simple sur G. Alors, $TV_G(x) \propto \mathbb{E}\, TV_\xi(x)$.

Algorithme : [SBHJ'16] au moment $n+1$,

- Tirer une « marche aléatoire sans boucles » ξ_{n+1},
- Calculer $x_{n+1} = \mathrm{prox}_{\gamma_{n+1} \mathrm{TV}_{\xi_{n+1}}}(x_n - \gamma_{n+1} \nabla F(x_n))$.
  Facile car la marche est un graphe 1D.

Quelques perspectives

- Bruits (ξ_n) markoviens,
- Etude poussée des algorithmes primaux-duaux aléatoires,
- Généralisation à des inclusions différentielles non forcément issues d'opérateurs MM, [Benaïm, Sorin, Hofbauer'03], [Borkar *et.al.*], [Roth'13], ...
- Analyse au second ordre,
- Passage de $\mathbb{R}^N$ à des espaces de Hilbert.