Les inclusions différentielles et l'optimisation convexe

Azzouz Dermoune

Université de Sciences et Technologies de Lille.
Laboratoire Paul Painlevé-UMR-CNRS 8524, France.

22-24 septembre 2016 Kénitra

Plan d'exposé

1) Optimisation convexe et inclusion différentielle.
2) Oscillation autour des points d'équilibres.
3) Outils mathématiques pour résoudre les inclusions différentielles.
4) Schémas numériques.
5) Outils mathématiques pour résoudre les oscillations autours des points d'équilibres.

Optimisation convexe : Points d'équilibres

Soit

$$\varphi : \mathbb{R}^p \to \mathbb{R}$$

une fonction convexe. Les minimums de φ sont solutions de l'inclusion

$$0 \in \partial\varphi(x).$$

Se sont les points d'équilibre de l'inclusion différentielle

$$0 \in \frac{d\boldsymbol{x}(t)}{dt} + \partial\varphi(\boldsymbol{x}(t)).$$

Sous différentielle d'une fonction convexe

La sous différentielle de φ au point $\boldsymbol{x}$ est définie par

$$\partial\varphi(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^p : \quad \forall\, \boldsymbol{h} \in \mathbb{R}^p, \quad \varphi(\boldsymbol{x}+\boldsymbol{h}) - \varphi(\boldsymbol{x}) \geq \langle \boldsymbol{h}, \boldsymbol{y} \rangle\}.$$

Sous différentielle d'une fonction convexe

Si φ est différentiable en x, alors

$$\partial\varphi(x) = \{\nabla\varphi(x)\}.$$

Si φ n'est pas différentiable en x, alors $\partial\varphi(x)$ est une partie non vide, fermée et convexe de $\mathbb{R}^p$.

Exemple

$$\begin{aligned} \partial|x| &= \{sgn(x)\}, \quad x \neq 0, \\ \partial|0| &= [-1, 1]. \end{aligned}$$

Exemple : La loi de Coulomb

$$0 \in \frac{dv(t)}{dt} + \mu\partial\left|v(t)\right| \quad \text{inclusion différentielle}$$

La vitesse $v(t)$ converge vers le minimum de la foncton convexe $|v|$.

Points d'équilibres et inclusion différentielle

Les algorithmes pour trouver les minimums de φ sont liés à l'inclusion différentielle

$$0 \in \frac{dx(t)}{dt} + \partial \varphi(x(t)).$$

Sens exact d'une solution de l'inclusion différentielle

Une solution à l'inclusion différentielle

$$\boldsymbol{f}(t)dt \in d\boldsymbol{x}(t) + \partial\varphi(\boldsymbol{x}(t))dt$$

est un couple $(\boldsymbol{x}, \eta)$ avec $\boldsymbol{x}$ une fonction continue et η une fonction mesurable telle que

$$\begin{aligned} &\boldsymbol{x}(t) + \int_0^t \eta(s)ds = \int_0^t \boldsymbol{f}(s)ds + \boldsymbol{x}(0), \quad t \geq 0, \\ &dt\ p.p. \quad \eta(t) \in \partial\varphi(\boldsymbol{x}(t)). \end{aligned}$$

Oscillation autour des points d'équilibres

On suppose que

$$Z_\varepsilon = \int_{\mathbb{R}^p} \exp\{-\frac{2\varphi(x)}{\varepsilon}\}dx < +\infty.$$

Les oscillations autour de l'inclusion différentielle sont les solutions de l'inclusion différentielle stochastique suivante :

$$\varepsilon dw(t) \in dx(t) + \partial\varphi(x(t))dt,$$

où w est le mouvement brownien et $\varepsilon > 0$. A chaque instant t, $x^\varepsilon(t)$ est une variable aléatoire.

Oscillation autour des points d'équilibres

1)

$$\boldsymbol{P}(x^{\varepsilon}(t) \in A) \to \int_A \exp\{-\frac{2\varphi(x)}{\varepsilon}\}\frac{dx}{Z_{\varepsilon}}$$

lorsque $t \to +\infty$.
2) $\lim_{\varepsilon\to 0}\lim_{t\to+\infty} x^{\varepsilon}(t)$ sont les points d'équilibres (minimums de φ).

Outils mathématiques pour étudier les inclusions différentielles

1) La fonction

$$\varphi_\lambda(\boldsymbol{x}) = \min\{\varphi(\boldsymbol{z}) + \frac{\|\boldsymbol{z}-\boldsymbol{x}\|_2^2}{2\lambda} : \quad \boldsymbol{z} \in \mathbb{R}^p\}$$

converge vers φ en croissant lorsque $\lambda \to 0$. C'est une fonction convexe de classe C^1.

2)

$$\nabla \varphi_\lambda(\boldsymbol{x}) = \frac{\boldsymbol{x} - \arg\min \varphi_\lambda(\boldsymbol{x})}{\lambda}.$$

3) Le vecteur $z^* = \arg\min \varphi_\lambda(\boldsymbol{x})$ est l'unique solution de l'inclusion

$$\boldsymbol{x} \in \boldsymbol{z} + \lambda\partial\,\varphi(\boldsymbol{z}), \quad \boldsymbol{x} \quad \text{fixé.}$$

En analyse convexe

$$\boldsymbol{z}^* = (\boldsymbol{I} + \lambda\partial\,\varphi)^{-1}(\boldsymbol{x}).$$

4) L'application

$$\boldsymbol{x} \in \mathbb{R}^p \to (\boldsymbol{I} + \lambda\partial\varphi)^{-1}(\boldsymbol{x})$$

est une contraction.

5) $\nabla\varphi_\lambda$ est lipschitzienne de constante de lipschitz égale à $\frac{1}{\lambda}$.

Schéma d'Euler

$$k = 0, \ldots, N, \quad \Delta t = (t_{k+1} - t_k) = \frac{1}{N},$$
$$\boldsymbol{f}(t_k)\Delta t = \boldsymbol{x}^N(t_{k+1}) - \boldsymbol{x}^N(t_k) + \Delta t \partial\varphi(\boldsymbol{x}^N(t_{k+1}))$$

est appelé schéma d'Euler implicite de l'inclusion différentielle

$$\boldsymbol{f}(t)dt \in d\boldsymbol{x}(t) + \partial\varphi(\boldsymbol{x}(t))dt.$$

Schéma d'Euler

Sachant $x^N(t_k)$, le vecteur $x^N(t_{k+1})$ est solution de

$$x^N(t_k) + f(t_k)\Delta t = x^N(t_{k+1}) + \Delta t \partial \varphi(x^N(t_{k+1})).$$

Donc

$$x^N(t_{k+1}) = (I + \Delta t \partial \varphi)^{-1}(x^N(t_k) + f(t_k)\Delta t).$$

En interpolant les points $(t_k, x^N(t_k))$: $\quad k = 0, \ldots, N$ on obtient la trajectoire absolument continue

$$x^N : [0,1] \to \mathbb{R}^p.$$

Elle vérifie

$$x^N(t) = x(0) + \int_0^t f^N(s)ds - \int_0^t \eta^N(s)ds,$$

où

$$f^N(s) = f(t_k), \quad \eta^N(s) = \partial\varphi(x^N(t_{k+1})) \quad s \in [t_k, t_{k+1}).$$

Elle est basée sur la propriété maximale monotone de $\partial\varphi$.

Propriété de monotonie $\partial\varphi$

$$\langle \boldsymbol{y}(1) - \boldsymbol{y}(2), \boldsymbol{x}(1) - \boldsymbol{x}(2) \rangle \geq 0, \quad \forall\, \boldsymbol{x}(1), \boldsymbol{x}(2) \in \mathbb{R}^{\boldsymbol{p}},$$
$$\forall\, \boldsymbol{y}(1) \in \partial\varphi(\boldsymbol{x}(1)), \quad \boldsymbol{y}(2) \in \partial\varphi(\boldsymbol{x}(2)).$$

Propriété maximal de $\partial\varphi$

Si $\boldsymbol{x}(n) \to \boldsymbol{x}$, et $\boldsymbol{y}(n) \to \boldsymbol{y}$ avec $\boldsymbol{y}(n) \in \partial\varphi(\boldsymbol{x}(n))$, alors $\boldsymbol{y} \in \partial\varphi(\boldsymbol{x})$.

Hypothèse de croissance linéaire : 1) Il existe $C > 0$ telle que pour tout $\boldsymbol{y} \in \partial\varphi(\boldsymbol{x})$

$$\|\boldsymbol{y}\| \leq C(1 + \|\boldsymbol{x}\|).$$

2) La fonction $\boldsymbol{f} \in L^{\infty}[0, 1]$.

Convergence du schéma d'Euler

Théorème Bernardin (2000). La suite $(\boldsymbol{x}^N, \eta^N)$ *converge vers une solution* $(\boldsymbol{x}, \eta)$ *de l'inclusion différentielle. De plus*

$$\sup_{t\in[0,1]} \|\boldsymbol{x}^N(t) - \boldsymbol{x}(t)\|_2 \leq \frac{C}{\sqrt{N}}.$$

Unicité de la solution

Soient (x^1, η^1), (x^2, η^2) deux solutions qui partent du même point $x(0)$. Alors dt p.p.

$$\frac{d}{dt}\|x^1(t) - x^2(t)\|^2 = 2\langle x^1(t) - x^2(t), \eta^2(t) - \eta^1(t)\rangle \leq 0.$$

Ceci implique $x^1(t) = x^2(t)$ pour tout t, et alors $\eta^1(t) = \eta^2(t)$ p.p.

Oscillation autour de l'équilibre

On perturbe l'inclusion différentielle

$$0 \in d\mathbf{x}(t) + \partial\varphi(\mathbf{x}(t))dt$$

par les oscillations browniennes

$$d\mathbf{w}(t) \in d\mathbf{x}(t) + \partial\varphi(\mathbf{x}(t))dt.$$

Oscillation autour de l'équilibre

Théorème (Pettersson 97) : Soit $\frac{1}{N} = t_{k+1} - t_k$ *une partition de* $[0,1]$*, et* $\boldsymbol{w}^N$ *l'interpolation linéaire des points* $(t_k, \boldsymbol{w}(t_k))$*. On considère la solution*

$$\boldsymbol{x}^N(t) = \boldsymbol{x}(0) + \boldsymbol{w}^N(t) - \int_0^t \eta^N(s)ds,$$

$$\eta^N(t) \in \partial\varphi(\boldsymbol{x}^N(t)) \quad dt \ p.p.$$

Oscillation autour de l'équilibre

Alors $x^N \to x$ *dans* $C[0,1]$ *et* $\eta^N \to \eta$ *faiblement dans* $L^1[0,1]$ *lorsque* $N \to +\infty$ *et*

$$x(t,w) = x(0) + w(t) - \int_0^t \eta(s,w)ds,$$
$$\eta(t,w) \in \partial\varphi(x(t,w)) \quad dt \ p.p.$$

De plus

$$\boldsymbol{E}[\sup_{t\in[0,1]} \|x^N(t) - x(t)\|_2^2] \leq C\frac{\sqrt{\ln(N)}}{\sqrt{N}}.$$

Continuité par rapport au temps et aux oscillations

Proposition (Dermoune, Ouknine, Sbai). L'application

$$(t, \boldsymbol{w}) \in [0,1] \times C[0,1] \to \boldsymbol{x}(t, \boldsymbol{w}) \in \mathbb{R}^p$$

est localement lipschitzienne.

Exemple : Oscillations autour de Lasso

Tibshirani 1996 : La régression linéaire

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} \in \mathbb{R}^n, \quad \boldsymbol{x} \in \mathbb{R}^p$$

lorsque $n < p$ se résout par l'optimisation

$$\arg\min\{\frac{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2}{2} + \mu\|\boldsymbol{x}\|_1 : \quad \boldsymbol{x} \in \mathbb{R}^p\} := \mathit{Lasso},$$

où $\mu > 0$, $\boldsymbol{A}$ une matrice $n \times p$ et $\boldsymbol{y} \in \mathbb{R}^n$ sont donnés,

$$\|\boldsymbol{x}\|_2^2 = \sum_{i=1}^{p} x_i^2, \quad \|\boldsymbol{x}\|_1 = \sum_{i=1}^{p} |x_i|.$$

Exemple : Oscillations autour de Lasso

Thèse Daoud Ounaissi (Lille 1) 2016 : Etude numérique des oscillations autour de lasso :

$$\varepsilon d\boldsymbol{w}(t) \in d\boldsymbol{x}^{\varepsilon}(t) + \partial\{\frac{\|\boldsymbol{A}\boldsymbol{x}^{\varepsilon} - \boldsymbol{y}\|_2^2}{2} + \mu\|\boldsymbol{x}^{\varepsilon}\|_1\}dt.$$

Propriété de la solution

$$\bigcup_{i=1}^{p} \{t \in [0,1] : \quad x_i^\varepsilon(t) = 0\}$$

et de mesure de Lebesgue nulle. L'inclusion différentielle stochastique devient une équation différentielle stochastique avec dérive discontinue :

$$d\mathbf{x}^\varepsilon(t) = -\{\mathbf{A}^\top(\mathbf{A}\mathbf{x}^\varepsilon(t) - \mathbf{y}) + \mu sgn(\mathbf{x}^\varepsilon(t))\}dt + \varepsilon d\mathbf{w}(t).$$

Dynamique de P.G. Degennes

En dimension 1, cette équation a été utilisée par P.G. Degennes 1990 (Prix Nobel de Chimie). Dans ce cas $\boldsymbol{x}$ représente la vitesse de la dynamique de Coulomb soumise à la force extérieure

$$f(\boldsymbol{x}) = -\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}).$$

Nous allons considérer le cas général

$$dv^{\varepsilon}(t) = \{f(t, v^{\varepsilon}(t)) - \mu sgn(v^{\varepsilon}(t))\}dt + \varepsilon dw(t),$$

où f est continue et à croissance linéaire.

Astuces probabilistes

$$
\begin{aligned}
\nu^{\varepsilon,-}(ds,t)dt &= \mathbf{1}_{[v^{\varepsilon}(t)\leq 0]}\delta_{sgn(v^{\varepsilon}(t))}(ds)dt,\\
\nu^{\varepsilon,+}(ds,t)dt &= \mathbf{1}_{[v^{\varepsilon}(t)>0]}\delta_{sgn(v^{\varepsilon}(t))}(ds)dt.
\end{aligned}
$$

Astuces probabilistes

Les familles v^ε, $\nu^{\varepsilon,-}$, $\nu^{\varepsilon,+}$, sont tendues. Par extraction d'une sous suite les limites v, ν^-, ν^+ sont solutions de

$$dv(t) = \{f(t, v(t)) - \mu(\nu^+(\mathbb{R}, t) - \nu^-(\mathbb{R}, t))\}dt$$

Les quantités $\nu^+(\mathbb{R}, t)$, $\nu^-(\mathbb{R}, t)$ indiquent dans quel sens le solide glisse :

$$v(t) < 0 \quad \text{implique} \quad \nu^-(\mathbb{R}, t) = 1, \quad \nu^+(\mathbb{R}, t) = 0,$$

le solide glisse vers la gauche.

$$v(t) > 0 \quad \text{implique} \quad \nu^-(\mathbb{R}, t) = 0, \quad \nu^+(\mathbb{R}, t) = 1,$$

le solide glisse vers la droite.

Astuces probabilistes

$$v(t) = 0 \quad \text{implique} \quad 0 < \nu^-(\mathbb{R}, t) < 1, \quad 0 < \nu^+(\mathbb{R}, t) < 1,$$
$$\nu^-(\mathbb{R}, t) + \nu^+(\mathbb{R}, t) = 1.$$

Le solide s'arrête.

Bilan

La solution limite v est la solution de la dynamique de Coulomb soumise à la force extérieure f :

$$dv(t) \in f(t, v(t))dt - \mu\partial\,|v(t)|dt.$$

Les grandes déviations

Elles consistent à contrôler les probabilités

$$\limsup_{\varepsilon \to 0} \ln\{P(v^\varepsilon \in F)\}, \quad \liminf_{\varepsilon \to 0} \ln\{P(v^\varepsilon \in O)\}$$

pour les fermés F et les ouverts O de $C([0,1])$.

Freidlin-Wentzell 1991

Si

$$dv^{\varepsilon}(t) = b(t, v^{\varepsilon}(t))dt + \varepsilon d\boldsymbol{w}(t)$$

avec b lipschitzienne, alors

$$\limsup_{\varepsilon\to} \varepsilon^2 \ln\{\boldsymbol{P}(v^{\varepsilon} \in F)\} \leq -I(F),$$
$$\liminf_{\varepsilon\to} \varepsilon^2 \ln\{\boldsymbol{P}(v^{\varepsilon} \in O)\} \geq -I(O),$$

avec la fonction de taux

$$I(A) := \inf\{\int_0^1 \|\boldsymbol{b}(\varphi(t)) - \varphi'(t)\|^2 dt : \quad \varphi \in A\}.$$

Freidlin 1980

Intuitivement

$$\boldsymbol{P}(v^{\varepsilon} \approx \varphi) \approx \exp\{-\frac{1}{\varepsilon^2} \int_0^1 \|\boldsymbol{b}(\varphi(t)) - \varphi'(t)\|^2 dt\}$$

est maximale lorsque $\varphi'(t) = b(\varphi(t))$.

Dans le cas

$$\boldsymbol{b}(\boldsymbol{v}) = \boldsymbol{f}(\boldsymbol{v}) - \mu sgn(\boldsymbol{v}),$$

on pose

$$\begin{aligned}
b_i^-(\boldsymbol{v}) &:= f_i(\boldsymbol{v}) + \mu, \\
b_i^+(\boldsymbol{v}) &:= f_i(\boldsymbol{v}) - \mu, \\
L_i^-(\boldsymbol{v}, \beta_i) &= |b_i^-(\boldsymbol{v}) - \beta_i|^2, \\
L_i^+(\boldsymbol{v}, \beta_i) &= |b_i^+(\boldsymbol{v}) - \beta_i|^2, \\
L_i^0(\boldsymbol{v}, \beta_i) &= \inf\{\alpha|b_i^-(\boldsymbol{v}) - \beta_i^-|^2 + (1-\alpha)|b_i^+(\boldsymbol{v}) - \beta^+|^2\}.
\end{aligned}$$

L'inf est pris sur $\alpha \in (0,1)$ et $\beta_i^- \geq 0$, $\beta_i^+ \leq 0$ tels que

$$\alpha\beta_i^- + (1-\alpha)\beta_i^+ = \beta_i.$$

Proposition (Dermoune, Ouknine, Sbai). *Nous avons les grandes déviations avec la fonction de taux*

$$I(\varphi) = \int_0^1 \tilde{L}(\varphi(t), \varphi'(t))dt,$$

où

$$\begin{array}{rcl} \tilde{L}(\boldsymbol{v}, \beta) & = & \displaystyle\sum_{i=1}^{p} \tilde{L}_i(\boldsymbol{v}, \beta_i), \\ \tilde{L}_i(\boldsymbol{v}, \beta_i) & = & L_i^-(\boldsymbol{v}, \beta_i), \quad si \quad v_i < 0, \\ \tilde{L}_i(\boldsymbol{v}, \beta_i) & = & L_i^+(\boldsymbol{v}, \beta_i), \quad si \quad v_i > 0, \\ \tilde{L}_i(\boldsymbol{v}, \beta_i) & = & L_i^0(\boldsymbol{v}, \beta_i), \quad si \quad v_i = 0. \end{array}$$

Interprétation

$$\boldsymbol{P}(v^\varepsilon \in A) \approx \exp(-\frac{I(A)}{\varepsilon^2})$$

En particulier

$$\boldsymbol{P}(v^\varepsilon \approx \boldsymbol{v}) \approx \exp(-\frac{1}{\varepsilon^2}\int_0^1 \tilde{L}(\boldsymbol{v}(t), \boldsymbol{v}'(t))dt).$$

Elle est donc maximale lorsque $\boldsymbol{v}$ est solution de la dynamique de Coulomb.
Ceci est une autre explication de la limite $\boldsymbol{v}^\varepsilon \to \boldsymbol{v}$.

Quelques références

1) P.G. Degennes 2005, Brownien motion with dry friction, J. Stat. Phys. 119, 953-962.
2) R. Tibshirani, Regression shrinkage and selection via LASSO, J. R. Stat. Soc. Ser. B Stat. Methodol. vol. 58 no. 1 267–288 (1996).
3) R. Tibshirani, The LASSO problem and uniqueness, Electron, J. Stat. vol. 7 1456–1490 (2013).
4) A. Dermoune, D. Ounaissi, N. Rahmania, MCMC convergence diagnosis using geometry of Bayesian LASSO, arXiv :1512.01366v1 [math.ST] (2015).

Quelques références

5) A. Dermoune, D. Ounaissi, N. Rahmania, Multilevel Monte Carlo simulation of a diffusion with non-smooth drift, arXiv :1504.06441v1 [math.ST] (2015).
6) A. Dermoune, D. Ounaissi, N. Rahmania, Oscillation of Metropolis-Hastings and simulated annealing algorithms around LASSO estimator, Math. Comput. Simulation (2015).
7) A. Beck, M. Teboulle, A Fast Iterative Shrinkage-thresholding algorithm for linear inverse problem, SIAM J. Imaging Sci. 183–202 (2009).
8) D. Ounaissi, Thèse Lille 1 (Juin 2016).