Optimisation sous contraintes

Joon Kwon

Master 2 — MathSV

7 octobre 2021

Dans ce chapitre,

- $f : \mathbb{R}^d \to \mathbb{R}$
- $C \subset \mathbb{R}^d$ ensemble de points admissibles

on considère le problème :

$$\min_{x \in C} f(x),$$

qu'on écrit aussi :

$$\begin{array}{rl} \text{minimiser} & f(x) \\ \text{soumis à} & x \in C \end{array}$$

Methode du gradient projeté

Méthodes de pénalisation

Conditions KKT

Dualité lagrangienne

Methode du gradient projeté

Définition

On suppose que C est convexe et fermé.

Proposition

Soit $y \in \mathbb{R}^d$. $x \mapsto \|y - x\|^2$ admet un unique minimiseur sur C.

Definition

Cet unique minimiseur est appelé projection orthogonale de y sur C et est noté $\mathrm{proj}_C(y)$.

Definition (Gradient projeté)

Soit $x^{(1)} \in \mathbb{R}^d$ et $(\gamma^{(t)})_{t \geqslant 1}$ une suite strictement positive. On appelle descente de gradient projeté associée à la fonction objectif f, à l'ensemble convexe fermé C, au point initial $x^{(1)}$ et aux pas $(\gamma^{(t)})_{t \geqslant 1}$ la suite $(x^{(t)})_{t \geqslant 1}$ définie par :

$$x^{(t+1)} = \mathrm{proj}_C \left\{ x^{(t)} - \gamma^{(t)} \nabla f(x^{(t)}) \right\}, \qquad t \geqslant 1.$$

Pertinent quand la projection sur C est moins difficile à résoudre que le problème initial.

Garanties pour f convexe

Théorème

On suppose que f admet un minimiseur x^* sur C.

- *Si f est convexe et L-régulière, le choix $\gamma^{(t)} = 1/L$ garantit*

$$f(x^{(T+1)}) - f(x^*) \leqslant \frac{L \left\| x^{(1)} - x^* \right\|^2}{2T}.$$

- *Si f est L-régulière et K-fortement convexe, le choix $\gamma^{(t)} = 1/L$ garantit*

$$f(x^{(T+1)}) - f(x^*) \leqslant \frac{L}{2} \left\| x^{(1)} - x^* \right\|^2 \cdot \left(1 - \frac{K}{L}\right)^T.$$

Mêmes garanties que pour la descente de gradient pour la minimisation sans contraintes.

Garanties pour f non-convexe

On définit une application de «gradient alternatif» associée à f et C par :

$$G_\gamma(x) = \frac{x - \mathrm{Proj}_C(x - \gamma \nabla f(x))}{\gamma}, \quad x \in \mathbb{R}^d, \ \gamma > 0.$$

Ce qui assure $x^{(t+1)} = x^{(t)} - \gamma^{(t)} G_{\gamma^{(t)}}(x^{(t)})$.

Théorème

Si f est seulement L-régulière, le choix $\gamma^{(t)} = 1/L$ garantit

$$\min_{1 \leqslant t \leqslant T} \left\| G_{1/L}(x^{(t)}) \right\|^2 \leqslant \frac{4L(f(x^{(1)}) - f(x^*))}{T}.$$

Méthodes de pénalisation

Pénalisation extérieure

- On remplace la contrainte par un terme supplémentaire dans la fonction à minimiser.
- On se donne $\phi\colon \mathbb{R}^d \to \mathbb{R}$ une fonction telle que :

$$\phi(x) = 0 \text{ si } x \in C, \qquad \text{et} \qquad \phi(x) > 0 \text{ si } x \notin C.$$

- Soit $\varepsilon > 0$ et on considère le problème approché :

$$\begin{aligned} \text{minimiser} \quad & f(x) + \frac{1}{\varepsilon}\phi(x) \\ \text{soumis à} \quad & x \in \mathbb{R}^d. \end{aligned}$$

- Plus ε est petit, plus le problème est proche du problème initial.
- La solution x_ε^* peut ne pas appartenir à C.

Exemple de résolution par pénalisation

- Principe : on résout successivement des approximations de plus en plus proches du problème initial.
- On choisit $\varepsilon^{(1)} > 0$, et on résout approximativement :

$$\begin{array}{ll} \text{minimiser} & f(x) + \dfrac{1}{\varepsilon^{(1)}}\phi(x) \\ \text{soumis à} & x \in \mathbb{R}^d. \end{array}$$

  par une méthode de notre choix.
- À l'étape $t \geqslant 2$,
  - on choisit $\varepsilon^{(t)}$ tel que $0 < \varepsilon^{(t)} < \varepsilon^{(t-1)}$
  - on résout approximativement :

$$\begin{array}{ll} \text{minimiser} & f(x) + \dfrac{1}{\varepsilon^{(t)}}\phi(x) \\ \text{soumis à} & x \in \mathbb{R}^d. \end{array}$$

  en prenant $x^{(t-1)}$ pour point initial.

Pénalisation intérieure

Si on souhaite des solutions approchées appartenant strictement à C

- On se donne $\phi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ telle que :

$$\forall x \notin C, \quad \phi(x) = +\infty.$$

- Soit $\varepsilon > 0$ et on considère :

$$\begin{array}{ll} \text{minimiser} & f(x) + \varepsilon\phi(x) \\ \text{soumis à} & x \in \mathbb{R}^d \end{array}$$

- Exemple : $C = \mathbb{R}^d_+$. $\phi(x) = -\sum_{i=1}^d \log x_i$.

Conditions KKT

Problèmes de minimisation sous forme standard

$$\begin{array}{ll} \text{minimiser} & f(x) \\ \text{soumis à} & h_i(x) = 0 \quad (1 \leqslant i \leqslant p) \\ & g_j(x) \leqslant 0 \quad (1 \leqslant j \leqslant q) \end{array} \tag{P}$$

où les g_j et h_i sont des applications $\mathbb{R}^d \to \mathbb{R}$.

Autrement dit, il s'agit du problème

$$\min_{x \in C} f(x) \qquad \text{où} \qquad C = \left\{ x \in \mathbb{R}^d, \quad \begin{array}{ll} \forall i = 1, \ldots, p, & h_i(x) \leqslant 0 \\ \forall j = 1, \ldots, q, & g_j(x) = 0 \end{array} \right\}.$$

Definition (Contraintes actives et inactives)

Soit $x \in \mathbb{R}^d$. Une contrainte d'inégalité "$g_j(x) \leqslant 0$" est dite active (ou saturée) (resp. inactive) en x si :

$$g_j(x) = 0 \qquad (\text{resp.} \quad g_j(x) < 0).$$

Théorème de Fritz John

Théorème (Fritz John)

Soit $x^* \in C$ *une solution de* (P). *Si les fonctions* f, g_j, h_i *sont* $\mathcal{C}^1$ *en* x^*, *il existe* $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ *et* $(\mu_0, \mu_1, \ldots, \mu_q) \in \mathbb{R}_+^q \setminus \{0\}$ *tels que :*

$$\begin{cases} \mu_0 \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla g_j(x^*) = 0 \\ \forall j = 1, \ldots, q, \quad , \quad \mu_j g_j(x^*) = 0. \end{cases}$$

- Si $\mu_0 = 0$, la propriété ci-dessus ne donne aucune information sur f, seulement sur C.
- Les coefficients $(\lambda_i)_{1 \leqslant i \leqslant p}$ et $(\mu_j)_{1 \leqslant j \leqslant q}$ sont appelés multiplicateurs de Lagrange

Conditions KKT

Definition

Un point $x^* \in C$ vérifie les conditions KKT (Karush-Kuhn-Tucker) si les f, h_i, g_j sont $\mathcal{C}^1$ en x^* et s'il existe $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ et $\mu_1, \ldots, \mu_q \in \mathbb{R}_+$ tels que :

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla g_j(x^*) = 0 \\ \forall j = 1, \ldots, q, \quad \mu_j g_j(x^*) = 0. \end{cases}$$

- On appelle qualification des contraintes une condition nécessaire pour qu'une solution de (P) satisfasse les conditions KKT. Il en existe un grand nombre.
- On dit que les contraintes sont qualifiées en un point $x \in C$ si une telle condition nécessaire est satisfaite.

Qualification des contraintes

Proposition

Soit $x^* \in C$ une solution de (P). On suppose que les f, g_j, h_i sont $\mathcal{C}^1$ en x^*. Si de plus une des conditions suivantes est vérifiée, alors x^* satisfait les conditions KKT.

(i) *Les fonctions h_i et g_j sont affines*

(ii) *la famille composée des vecteurs $\nabla h_i(x^*)$ (pour $i = 1, \dots, p$) et des vecteurs $\nabla g_j(x^*)$ (pour j tel que $g_j(x^*) = 0$) est linéairement indépendante.*

(iii) *Les fonctions f, g_j sont convexes, les h_j affines, et il existe $x \in C$ tel que $g_j(x) < 0$ pour tout $j = 1, \dots, q$. (condition de Slater)*

Il existe de nombreuses autres conditions de qualification.

Réciproque dans le cas convexe

Proposition

On suppose que les f, g_j, h_i sont $\mathcal{C}^1$, que les f, g_j sont convexes, et les h_j affines. Soit $x^* \in C$ vérifiant les conditions KKT. Alors, x^* est une solution de (P).

Dualité lagrangienne

Lagrangien

Definition

Le lagrangien de (P) est la fonction $L\colon \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}_+^q \to \mathbb{R}$ définie par :

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x).$$

- On voit que : $$\sup_{\substack{\lambda\in\mathbb{R}^p \\ \mu\in\mathbb{R}_+^q}} L(x,\lambda,\mu) = \begin{cases} f(x) & \text{si } x \in C \\ +\infty & \text{sinon.} \end{cases}$$
- Donc (P) se réécrit : $$\min_{x\in C} f(x) = \min_{x\in\mathbb{R}^d} \sup_{\substack{\lambda\in\mathbb{R}^p \\ \mu\in\mathbb{R}_+^q}} L(x,\lambda,\mu)$$

Dualité faible

On définit le problème dual de (P) par

$$\begin{aligned} &\text{maximiser} \quad && \inf_{x\in\mathbb{R}^d} L(x,\lambda,\mu) \\ &\text{soumis à} \quad && \lambda \in \mathbb{R}^p \\ & && \mu \in \mathbb{R}^q_+. \end{aligned} \tag{P*}$$

Proposition

$$\sup_{\substack{\lambda\in\mathbb{R}^p\\ \mu\in\mathbb{R}^q_+}} \inf_{x\in\mathbb{R}^d} L(x,\lambda,\mu) \leqslant \inf_{x\in\mathbb{R}^d} \sup_{\substack{\lambda\in\mathbb{R}^p\\ \mu\in\mathbb{R}^q_+}} L(x,\lambda,\mu).$$

Dualité forte

Definition

$(x^*, \lambda^*, \mu^*) \in \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}_+^q$ est un point-selle du lagrangien si :

$$\forall (x, \lambda, \mu) \in \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}_+^q, \quad L(x^*, \lambda, \mu) \leqslant L(x^*, \lambda^*, \mu^*) \leqslant L(x, \lambda^*, \mu^*).$$

Proposition

S'il existe un point-selle (x^*, λ^*, μ^*) de L, alors,

$$\sup_{\substack{\lambda \in \mathbb{R}^p \\ \mu \in \mathbb{R}_+^q}} \inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu) = \inf_{x \in \mathbb{R}^d} \sup_{\substack{\lambda \in \mathbb{R}^p \\ \mu \in \mathbb{R}_+^q}} L(x, \lambda, \mu) = L(x^*, \lambda^*, \mu^*).$$

Si de plus, les fonctions f, h_i, g_j sont $\mathcal{C}^1$, x^* vérifie les conditions KKT avec (λ^*, μ^*) pour multiplicateurs de Lagrange.

Alors,

- On dit qu'il y a dualité forte.
- x^* est solution du problème primal.
- (λ^*, μ^*) est solution du problème dual.

Réciproque dans le cas convexe

Proposition

On suppose que :

- *les fonctions f, h_i, g_j sont $\mathcal{C}^1$*
- *les fonctions f, g_j sont convexes*
- *les fonctions h_i affines.*
- *$x^* \in C$ une solution de (P) vérifiant les conditions KKT (on note (λ^*, μ^*) les multiplicateurs de Lagrange).*

Alors (x^*, λ^*, μ^*) est un point-selle de L.

Et il y a donc dualité forte.

Algorithme d'Uzawa

L'algorithme d'Uzawa, consiste à appliquer la méthode du gradient projeté au problème dual (P^*).

- Initialisation : $\lambda^{(1)} \in \mathbb{R}^p, \mu^{(1)} \in \mathbb{R}_+^q$ quelconques.
- Pour $t \geqslant 1$,

$$
\begin{aligned}
x^{(t)} &= \operatorname*{arg\,min}_{x \in \mathbb{R}^d} L(x, \lambda^{(t)}, \mu^{(t)}) \\
\lambda^{(t+1)} &= \lambda^{(t)} + \gamma^{(t)} h_i(x^{(t)}), \qquad \forall i = 1, \ldots, p, \\
\mu^{(t+1)} &= \max\left(0,\ \mu^{(t)} + \gamma^{(t)} g_j(x^{(t)})\right), \qquad \forall j = 1, \ldots, q,
\end{aligned}
$$