Optimisation convexe – Méthodes itératives

Descentes de gradient

Bashar Dudin

May 17, 2019

EPITA

Optimisation sous contraintes d’égalités

Optimisation dans le cas général

Contraintes d'égalités | Cadre

On s'intéresse aux problèmes d'optimisation (P) de la forme

$$\begin{aligned} \text{minimiser} \quad & f(x) \\ \text{sujet à} \quad & \\ & Ax = b \end{aligned} \tag{P}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est convexe $\mathscr{C}^2$ et $A \in \mathscr{M}_{p,n}(\mathbb{R})$.

Contraintes d'égalités | Cadre

On s'intéresse aux problèmes d'optimisation (P) de la forme

$$\begin{aligned} &\text{minimiser} \quad f(x) \\ &\text{sujet à} \\ &\qquad\qquad Ax = b \end{aligned} \tag{P}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est convexe $\mathscr{C}^2$ et $A \in \mathscr{M}_{p,n}(\mathbb{R})$.

Hypothèses

- On suppose dans la suite que $\mathrm{rg}(A) < n$[a] ; chose qu'on peut en particulier garantir quand $p < n$.
- On suppose que notre point x_0 de départ est admissible[b].

[a] Quel est le sens de cette hypothèse?
[b] Qu'est-ce que cela implique?

Contraintes d'égalités | Cadre

On s'intéresse aux problèmes d'optimisation (P) de la forme

$$\begin{array}{ll} \text{minimiser} & f(x) \\ \text{sujet à} & \\ & Ax = b \end{array} \qquad (P)$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est convexe $\mathscr{C}^2$ et $A \in \mathscr{M}_{p,n}(\mathbb{R})$.

Hypothèses

- On suppose dans la suite que $\mathrm{rg}(A) < n$[a] ; chose qu'on peut en particulier garantir quand $p < n$. ***Dans ce cas la condition de Slater est satisfaite.***
- On suppose que notre point x_0 de départ est admissible[b].

[a] Quel est le sens de cette hypothèse?
[b] Qu'est-ce que cela implique?

Sous les hypothèses précédentes la résolution de (P) est équivalente à la résolution du système

$$\begin{cases} Ax & = & b \\ \nabla f(x) + A^T \nu & = & 0 \end{cases} \tag{KKT-P}$$

Cette formulation est au coeur d'une résolution de (P) basée sur l'algorithme de Newton qu'on abordera plus loin dans le cours.

Contraintes d’égalités | Attaque élémentaire

Sous les conditions de rang sur A et d’existence d’un point admissible le système linéaire $Ax = b$ a des solutions qu’on peut écrire paramétriquement.

Contraintes d'égalités | Attaque élémentaire

Sous les conditions de rang sur A et d'existence d'un point admissible le système linéaire $Ax = b$ a des solutions qu'on peut écrire paramétriquement.

L'algorithme de Gauss donne une matrice $F \in \mathscr{M}_{n,n-\mathrm{rg}(A)}(\mathbb{R})$ de rang maximal qui paramètre le lieu admissible de (P) par

$$x_0 + Fz$$

pour $z \in \mathbb{R}^{n-\mathrm{rg}(A)}$.

Le problème (P) est dans ce cas équivalent à (P_z) donné par

$$\min_z f(x_0 + Fz); \tag{P_z}$$

si z^* est un point optimal de (P_z) alors $x^* = x_0 + Fz^*$ est un point optimal de (P).

Contraintes d'égalités | Attaque élémentaire

Le problème (P) est dans ce cas équivalent à (P_z) donné par

$$\min_z f(x_0 + Fz); \tag{P_z}$$

si z^* est un point optimal de (P_z) alors $x^* = x_0 + Fz^*$ est un point optimal de (P).

Le problème (P_z) est un problème sans contraintes qu'on peut désormais résoudre par une descente de gradient.

Contraintes d’égalités | Attaque élémentaire

- L’avantage de la démarche précédente réside dans la facilité théorique de mise en place.

Contraintes d’égalités | Attaque élémentaire

- L’avantage de la démarche précédente réside dans la facilité théorique de mise en place.
- Son désavantage réside dans le fait qu’elle peut détruire une partie des propriétés du problème (P) de départ (comme le fait d’avoir une matrice A creuse).

Contraintes d’égalités | Attaque élémentaire

- L’avantage de la démarche précédente réside dans la facilité théorique de mise en place.
- Son désavantage réside dans le fait qu’elle peut détruire une partie des propriétés du problème (P) de départ (comme le fait d’avoir une matrice A creuse).

La méthode de Newton s’étend naturellement au cas des contraintes d’égalités. C’est le point qu’on étudie dans la suite.

Le principe de cette extension prend racine dans le problème (quadratique) suivant

$$\begin{aligned} &\text{minimiser} \quad && \frac{1}{2}x^T P x + q^T x + r \\ &\text{sujet à} && \\ & && Ax = b \end{aligned} \tag{P_Q}$$

On rappelle que dans ce contexte P est une matrice symétrique positive. Les hypothèses imposées à nos problèmes sous contraintes étant toujours conservées.

Méthode de Newton | Principe

Le problème (P_Q) vérifie la condition de Slater, sa résolution est équivalente à

$$\left\{\begin{array}{ccc} Ax & = & b \\ Px + A^T \nu & = & -q \end{array}\right. \tag{KKT-P_Q}$$

Méthode de Newton | Principe

Équations qu'on écrit encore

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \nu \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}. \tag{KKT-P_Q}$$

Les hypothèses faites sur (P_Q) garantissent l'existence de solutions primal-dual.

Méthode de Newton | Principe

Équations qu'on écrit encore

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \nu \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}. \tag{KKT-P_Q}$$

Les hypothèses faites sur (P_Q) garantissent l'existence de solutions primal-dual. La méthode de Newton sous contraintes d'égalités est une méthode itérative qui applique la démarche précédente à l'approximation d'ordre 2 de (P) au voisinage de l'itéré courant.

Méthode de Newton

Soit x le point courant d'une méthode de Newton sous contraintes d'égalités. Il s'agit de définir le prochain itéré.

Méthode de Newton

Soit x le point courant d'une méthode de Newton sous contraintes d'égalités. Il s'agit de définir le prochain itéré. On note (P_x) le problème d'optimisation quadratique donné par

$$\begin{array}{ll} \text{minimiser} & f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{sujet à} & \\ & A(x+v) = b \end{array} \tag{P_x}$$

Méthode de Newton

Soit x le point courant d'une méthode de Newton sous contraintes d'égalités. Il s'agit de définir le prochain itéré. On note (P_x) le problème d'optimisation quadratique donné par

$$\begin{array}{ll} \text{minimiser} & f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{sujet à} & \end{array} \tag{P_x}$$

$$A(x+v) = b$$

Tout comme dans le cas sans contrainte, un point optimal Δx_N de (P_x) décrit une direction de descente en x.

Méthode de Newton

En dualisant, la direction Δx_N s'obtient comme première composante d'une solution de l'équation matricielle

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}. \qquad \text{(KKT-}P_Q\text{)}$$

Algorithme de Newton

Algorithm 1 Méthode de Newton

Input: f : a function, x_0 : an initial point in the domain of f, ε : tolerance.

Output: x^* : an optimal solution of (P) if bounded from below

function $\textsc{Newton\_Method}(f, x_0, \varepsilon)$
    $x \leftarrow x_0$
    $\Delta x_N \leftarrow -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$
    $\lambda^2(x) = -\nabla f(x)^T \Delta x_N$
    while $\frac{\lambda^2(x)}{2} > \varepsilon$ **do**
        $\Delta x_N \leftarrow -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$
        $\lambda^2(x) = -\nabla f(x)^T \Delta x_N$
        compute step $t > 0$ of descent
        $x \leftarrow x + t\Delta x_N$
    end while
    return x
end function

Optimisation sous contraintes d’égalités

Optimisation dans le cas général

Cas général | Cadre

On s'intéresse désormais aux problèmes d'optimisation (Q) de la forme

$$\begin{array}{lll} \text{minimiser} & f_0(x) & \\ \text{sujet à} & & \\ & f_i(x) \leq 0 & \forall i \in \{1, \ldots, m\} \\ & Ax = b & \end{array} \tag{Q}$$

où $f_0 : \mathbb{R}^n \to \mathbb{R}$ et les $f_i : \mathbb{R}^n \to \mathbb{R}$ sont convexes $\mathscr{C}^2$ et $A \in \mathscr{M}_{p,n}(\mathbb{R})$.

Cas général | Cadre

On s'intéresse désormais aux problèmes d'optimisation (Q) de la forme

$$\begin{array}{lrcll} \text{minimiser} & & f_0(x) & & \\ \text{sujet à} & & & & \\ & f_i(x) & \leq & 0 & \forall i \in \{1, \ldots, m\} \\ & Ax & = & b & \end{array} \qquad (Q)$$

où $f_0 : \mathbb{R}^n \to \mathbb{R}$ et les $f_i : \mathbb{R}^n \to \mathbb{R}$ sont convexes $\mathscr{C}^2$ et $A \in \mathscr{M}_{p,n}(\mathbb{R})$.

Hypothèses

On suppose dans la suite que la condition de ***Slater*** est vérifiée. En particulier on peut supposer $\mathrm{rg}(A) < n$. On a dualité forte pour (Q) dans ce cas.

Cas général | Cadre

On rappelle dans ce cas que les conditions KKT pour (Q) s'écrivent

$$\begin{array}{rcll} Ax & = & b & \\ f_i(x) & \leq & 0 & \forall i \in \{1,\dots,n\} \\ \lambda_i f_i(x) & = & 0 & \forall i \in \{1,\dots,n\} \\ \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla_i f_i(x) + A^T \nu & = & 0 & \\ \lambda & = & 0 & \end{array} \tag{KKT-Q}$$

où $\lambda \in \mathbb{R}^m$ et $\nu \in \mathbb{R}^p$.

Remarque : On se limite à la méthode de la barrière logarithmique. Vous serez invités par la suite à vous attaquer par ce biais à une méthode de point intérieur.

Barrière logarithmique | Principe

On note I_- la fonction définie par

$$I_-(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ +\infty & \text{si } x > 0 \end{cases}$$

Le problème (Q) est désormais équivalent

$$\begin{array}{ll} \text{minimiser} & f_0(x) + \sum_{i=1}^{m} I_-\big(f_i(x)\big) \\ \text{sujet à} & \\ & Ax = b \end{array} \tag{Q}$$

Problème d'optimisation sans contraintes d'inégalités ; cas étudié précédemment.

Barrière logarithmique | Principe

On note I_- la fonction définie par

$$I_-(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ +\infty & \text{si } x > 0 \end{cases}$$

Le problème (Q) est désormais équivalent

$$\begin{array}{ll} \text{minimiser} & f_0(x) + \sum_{i=1}^{m} I_-\big(f_i(x)\big) \\ \text{sujet à} & \\ & Ax = b \end{array} \qquad (Q)$$

Problème d'optimisation sans contraintes d'inégalités ; cas étudié précédemment.

Problème!

La fonction objectif n'est pas différentiable en général.

Pour contourner le problème précédent on va approcher I_- par une famille de fonctions $\mathscr{C}^\infty$.

Définition

Pour tout $t > 0$ on note I_t la fonction $\mathscr{C}^\infty$ sur $\mathbb{R}_-^*$ définie par

$$I_t(x) = -\frac{1}{t}\ln(-x).$$

Pour contourner le problème précédent on va approcher I_- par une famille de fonctions $\mathscr{C}^\infty$.

Définition

Pour tout $t > 0$ on note I_t la fonction $\mathscr{C}^\infty$ sur $\mathbb{R}_-^*$ définie par

$$I_t(x) = -\frac{1}{t}\ln(-x).$$

La famille I_t converge uniformément quand $t \to +\infty$ vers I_-.

Barrière Logarithmique | Une famille de problèmes

Pour contourner le problème précédent on va approcher I_- par une famille de fonctions $\mathscr{C}^\infty$.

Définition

Pour tout $t > 0$ on note I_t la fonction $\mathscr{C}^\infty$ sur $\mathbb{R}_-^*$ définie par

$$I_t(x) = -\frac{1}{t}\ln(-x).$$

La famille I_t converge uniformément quand $t \to +\infty$ vers I_-. On remplace désormais l'étude de (Q) par celles des problèmes (Q_t)

$$\begin{array}{ll} \text{minimiser} & f_0(x) + \sum_{i=1}^m -\frac{1}{t}\ln\left(f_i(x)\right) \\ \text{sujet à} & \\ & Ax = b \end{array} \qquad (Q_t)$$

Barrière Logarithmique | Premières remarques

$$\begin{array}{ll} \text{minimiser} & f_0(x) + \sum_{i=1}^{m} -\frac{1}{t} \ln\left(-f_i(x)\right) \\ \text{sujet à} & \\ & Ax = b \end{array} \tag{Q_t}$$

- Si l'on prend t assez grand on devrait avoir une bonne approximation d'un point optimal de (Q).

Barrière Logarithmique | Premières remarques

$$\begin{array}{ll} \text{minimiser} & f_0(x) + \sum_{i=1}^{m} -\frac{1}{t}\ln\big(-f_i(x)\big) \\ \text{sujet à} & \\ & Ax = b \end{array} \qquad (Q_t)$$

- Si l'on prend t assez grand on devrait avoir une bonne approximation d'un point optimal de (Q).
- Pour t grand le calcul de la hessienne de la fonction objectif devrait introduire des difficultés numériques importantes.

Barrière Logarithmique | Comment itérer?

La difficulté liée au calcul de la hessienne à chaque itération nous impose de réféchir avec un peu de finesse à la manière de procédér. On note ϕ la fonction donnée par

$$\phi(x) = -\sum_{i=1}^{m} \ln\big(-f_i(x)\big)$$

fonction qu'on qualifie de ***barrière logarithmique***.

Barrière Logarithmique | Comment itérer?

La difficulté liée au calcul de la hessienne à chaque itération nous impose de réféchir avec un peu de finesse à la manière de procédér. On note ϕ la fonction donnée par

$$\phi(x) = -\sum_{i=1}^{m} \ln\big(-f_i(x)\big)$$

fonction qu'on qualifie de ***barrière logarithmique***. Le problème (Q_t) est équivalent à

$$\begin{array}{ll} \text{minimiser} & tf_0(x) + \phi(x) \\ \text{sujet à} & \\ & Ax = b \end{array} \tag{Q_t}$$

Barrière Logarithmique | Comment itérer?

La difficulté liée au calcul de la hessienne à chaque itération nous impose de réféchir avec un peu de finesse à la manière de procédér. On note ϕ la fonction donnée par

$$\phi(x) = -\sum_{i=1}^{m} \ln\big(-f_i(x)\big)$$

fonction qu'on qualifie de ***barrière logarithmique***. Le problème (Q_t) est équivalent à

$$\begin{array}{ll} \text{minimiser} & tf_0(x) + \phi(x) \\ \text{sujet à} & \\ & Ax = b \end{array} \tag{Q_t}$$

Au lieu d'attaquer le problème avec t grand, on procède séquentiellement.

- On se donne une suite (t_n) qui tend vers $+\infty$.
- Un point initial $t_{0,0}$ (strictement admissible) pour résoudre (Q_{t_0}).
- Le point initial servant à résoudre $(P_{t_{n+1}})$ est le point optimal de (P_{t_n}).

Pour l’instant on a rien montré!

Méthode de la barrière logarithmique

Algorithm 2 Méthode de la barrière logarithmique

Input: f : a function, x_0 : a strictly feasible point, $t_0 > 0$, $\mu > 1$, ε : tolerance.
Output: x^* : an optimal solution of (Q) if bounded from below

```
function LOG_BARRIER_METHOD(f, x_0, t_0, μ, ε)
    x ← x_0
    t ← t_0
    while m/t > ε do
        x ← optimal point of (Q_t) with starting point x
        t ← μt
    end while
    return x
end function
```

Méthode de la barrière logarithmique

- L'étape d'optimisation interne ligne 4 est appelée ***étape de centrage***. La liste des points optimaux des problèmes intermédiaire est appelée ***chemin central***.

Méthode de la barrière logarithmique

- L'étape d'optimisation interne ligne 4 est appelée ***étape de centrage***. La liste des points optimaux des problèmes intermédiaire est appelée ***chemin central***.
- On a *tradeoff* dans le choix de μ entre μ proche de 1 pour beaucoup d'itérations externes mais petit nombre d'itération interne et un comportement inverse pour petit μ.