OUTILS D'OPTIMISATION POUR LES SCIENCES DES DONNÉES ET DE LA DÉCISION

9 novembre 2023

Aujourd'hui : → Optimisation et décomposition 1/2
→ Fin TD 6 + début TD 7 (examen 2022-2023)

Demain : → Optimisation et décomposition 2/2
→ Fin TD 7

OPTIMISATION ET DÉCOMPOSITION

Motivation

- Difficulté des problèmes modernes en optimisation : la grande dimension
    - Très grand nombre de variables et/ou
    - Fonction objectif combinaison d'un très grand nombre de termes

Ex) Apprentissage profond, réseaux d'énergie, réseaux de télécommunication

- Opportunité : Tous ces problèmes sont fortement structurés
    - Relations entre les variables (typiquement des relations linéaires)
    - Possibilité de décomposer le problème en sous-problèmes de plus petite dimension

Q) Comment exploiter la structure dans les algorithmes d'optimisation ?

1) Optimisation sous contraintes linéaires

Problème : (P) $\underset{w \in \mathbb{R}^d}{\text{minimiser}} \quad f(w) \quad \text{s.c.} \quad Aw = b$

contrainte structurelle du problème (ex: contrainte physique)

$A \in \mathbb{R}^{m \times d}, \ b \in \mathbb{R}^m, \ f : \mathbb{R}^d \to \mathbb{R}$

Hypothèse : $\underbrace{\{w \in \mathbb{R}^d \mid Aw = b\}}_{\text{ensemble admissible}}$ est non vide

↳ Grâce à la théorie de la dualité, on peut remplacer le problème (P) par un problème sans contraintes basé sur une fonction appelée le lagrangien du problème.

Def : Le lagrangien associé à (P) est la fonction

$$\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^m \longrightarrow \mathbb{R}$$

$$(w, z) \longmapsto \mathcal{L}(w, z) = \underbrace{f(w)}_{\text{fonction objectif de (P)}} + \overset{\in \mathbb{R}^m}{z^T} \underbrace{(Aw - b)}_{\in \mathbb{R}^m}$$

w : variables de décision de (P)

$Aw-b \in \mathbb{R}^m$ représente la contrainte $Aw - b = 0$

z s'appelle le vecteur des variables duales / multiplicateurs de Lagrange

↳ Le lagrangien est une combinaison de la fonction objectif et des fonctions définissant les contraintes

↳ On peut montrer que (P) est équivalent au problème

"problème primal" $\left(\quad \underset{w \in \mathbb{R}^d}{\text{minimiser}} \quad \boxed{\max_{z \in \mathbb{R}^m} \mathcal{L}(w, z)} \right.$

Fonction $w \longmapsto \max_{z \in \mathbb{R}^m} \mathcal{L}(w, z)$

Définition Le problème dual de (P) est le problème

$$(D) \quad \underset{z \in \mathbb{R}^m}{\text{maximiser}} \; \underbrace{\min_{w \in \mathbb{R}^d} \mathcal{L}(w, z)}$$

$z \longmapsto \min_{w \in \mathbb{R}^d} \mathcal{L}(w, z)$: fonction duale du problème

Propriétés :

- La fonction duale est toujours concave (son opposée est convexe)

  $\Rightarrow$ (D) est équivalent à minimiser une fonction convexe sans contraintes !

- En règle générale, il y a dualité forte entre (P) et (D), càd que

1) $$\underbrace{\min_{w \in \mathbb{R}^d} \max_{z \in \mathbb{R}^m} \mathcal{L}(w, z)}_{\text{Valeur optimale de (P)}} = \underbrace{\max_{z \in \mathbb{R}^m} \min_{w \in \mathbb{R}^d} \mathcal{L}(w, z)}_{\text{Valeur optimale en (D)}}$$

2) $\exists (w^*, z^*) \in \mathbb{R}^d \times \mathbb{R}^m$ tels que w^* est solution de (P), z^* est solution de (D), et

$$\mathcal{L}(w^*, z^*) = \min_{w} \max_{z} \mathcal{L}(w, z) = \max_{z} \min_{w} \mathcal{L}(w, z)$$

$\hookrightarrow$ La théorie de la dualité lagrangienne permet de ramener la résolution d'un problème avec contraintes à la résolution d'un problème convexe sans contraintes

$\Rightarrow$ Pour résoudre le problème (P), on va donc maximiser la fonction duale $z \longmapsto \min_{w \in \mathbb{R}^d} \mathcal{L}(w, z)$

(2) Méthodes duales

Principe : Calculer une suite de variables duales $\{z_k\}$ via un algorithme d'optimisation et en déduire une suite de variables primales $\{w_k\}$

Méthode de montée duale (dual ascent)

Initialisation : $w_0 \in \mathbb{R}^d$, $z_0 \in \mathbb{R}^m$

Pour $k = 0, 1, \ldots$

- Calculer $w_{k+1} \in \underset{w \in \mathbb{R}^d}{\operatorname{argmin}}\, \mathcal{L}(w, z_k)$
- Poser $z_{k+1} = z_k + \alpha_k (A w_{k+1} - b)$ avec $\alpha_k > 0$

Calcul de z_{k+1} : Pas de sous-gradient pour la fonction

$$d : z \longmapsto -\min_{w \in \mathbb{R}^d} \mathcal{L}(w, z)$$

en z_k

$$(-(A w_{k+1} - b)) \in \partial d(z_k))$$

Calcul de w_{k+1} : Résolution d'un problème d'optimisation sans contraintes

$$\mathcal{L}(w, z_k) = f(w) + z_k^T (Aw - b)$$

- Solution n'est pas forcément unique
- Contraintes ont une importance dans la minimisation

Si w_{k+1} est admissible pour (P) (càd $Aw_{k+1} = b$),

alors $z_{k+1} = z_k + \alpha_k \underbrace{(Aw_{k+1} - b)}_{=0} = z_k$

et z_k est une solution de (D)

$\Rightarrow w_k$ est une solution de (P)

↳ La méthode de montée duale possède des garanties théoriques proches de celles des méthodes de sous-gradient, et en particulier elle converge lentement vers une solution (et donc lentement vers un w_k qui soit admissible)

$\Rightarrow$ On peut obtenir de meilleures garanties en régularisant le problème

Déf : Le Lagrangien augmenté associé à (P) est la fonction

$$\mathcal{L}^a : \mathbb{R}^d \times \mathbb{R}^m \times \underbrace{\mathbb{R}_{++}}_{(0,+\infty)} \longrightarrow \mathbb{R}$$

$$(w, z, \lambda) \longmapsto \mathcal{L}^a(w, z; \lambda) = f(w) + z^T (Aw - b) + \frac{\lambda}{2} \|Aw - b\|^2$$

$$\mathcal{L}^a(w, z; \lambda) = \underbrace{f(w) + z^T(Aw - b)}_{\mathcal{L}(w,z)} + \frac{\lambda}{2}\|Aw - b\|^2$$

Plus la valeur de λ est élevée, plus on force la solution à satisfaire les contraintes

terme de régularisation $\lambda > 0$ qui pénalise les vecteurs w non admissibles pour (P)

Méthode du lagrangien augmenté (ou méthode des multiplicateurs)

<u>Initialisation</u> : $w_0 \in \mathbb{R}^d$, $z_0 \in \mathbb{R}^m$

<u>Itération k</u> :

- Calculer $w_{k+1} \in \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \; \mathcal{L}^a(w, z_k, \lambda)$
- Poser $z_{k+1} = z_k + \lambda (A w_{k+1} - b)$

<u>Commentaires</u>

→ <u>Version de base</u> : λ utilisé comme paramètre de régularisation <u>et</u> comme taille de pas constante.

→ <u>Versions avancées</u> : • Remplacer la taille de pas λ par $\alpha_k > 0$

• Faire varier $\lambda \to \{\lambda_k\}$ (en assurant que $\lambda_{k+1} \geq \lambda_k \; \forall k$)

→ <u>Avantage du lagrangien augmenté</u> : Garantit une convergence plus rapide vers un point vérifiant les contraintes

- Pour λ suffisamment grand, le problème $\underset{w \in \mathbb{R}^d}{\text{minimiser}}\ \mathcal{L}^a(w, z_k; \lambda)$ aura une unique solution

(Vrai $\forall \lambda > 0$ si f est convexe)

3 Décomposition

Cadre :
→ minimisation sous contraintes d'égalité linéaires
→ Structure séparable

$$\underset{\substack{u \in \mathbb{R}^{d_1} \\ v \in \mathbb{R}^{d_2}}}{\text{minimiser}}\ f(u) + g(v) \quad \text{s.c.} \quad Au + Bv = c$$

$f : \mathbb{R}^{d_1} \to \mathbb{R}$ $\quad A \in \mathbb{R}^{m \times d_1}$

$g : \mathbb{R}^{d_2} \to \mathbb{R}$ $\quad B \in \mathbb{R}^{m \times d_2}$

$c \in \mathbb{R}^m$

Séparable :
- La fonction objectif est une somme d'une fonction de u et d'une fonction de v
- Les fonctions définissant les contraintes sont sommes d'une fonction de u et d'une fonction de v

↳ Approche naïve : Poser $w = \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^{d_1+d_2}$ $\quad d = d_1 + d_2$

$$\hat{f}(w) = f(u) + g(v) \qquad \hat{A} = [A \ \ B] \qquad \hat{b} = c$$

$$\Rightarrow \underset{w \in \mathbb{R}^d}{\text{minimiser}}\ \hat{f}(w) \ \text{s.c.} \ \hat{A}w = b$$

$\ominus$ Cette approche ne tient pas compte de la structure du problème et conduit (via les méthodes duales notamment) à résoudre des problèmes en dimension d_1+d_2 alors que les variables ont des rôles indépendants dans la fonction objectif

$\Rightarrow$ Problématique lorsque $d \gg 1$, on préfère donc les techniques de décomposition

Remarque : La propriété de séparabilité s'étend à plus de deux groupes de variables

$$\underset{\substack{u^{(1)} \in \mathbb{R}^{d_1} \\ \vdots \\ u^{(m)} \in \mathbb{R}^{d_m}}}{\text{minimiser}} \sum_{i=1}^{m} f_i(u^{(i)}) \quad \text{s.c.} \quad \sum_{i=1}^{m} A_i u^{(i)} = b$$