

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

10 novembre 2023

Dernière séance:

→ Fin cours optimisation et décomposition

→ TD: Examen 2022-2023 (partie 2)

### (3) Décomposition

Problème: (P)  $\left\{ \begin{array}{l} \text{minimiser} \\ u \in \mathbb{R}^{d_1} \\ v \in \mathbb{R}^{d_2} \end{array} \right. f(u) + g(v) \quad \text{s.c.} \quad Au + Bv = c$

Problème séparable:

- Objectif: somme de deux fonctions de groupes de variables distincts
- Contraintes: définies via deux fonctions de ces mêmes groupes de variables

$$f: \mathbb{R}^{d_1} \rightarrow \mathbb{R}$$

$$g: \mathbb{R}^{d_2} \rightarrow \mathbb{R}$$

$$A \in \mathbb{R}^{m \times d_1}, \quad B \in \mathbb{R}^{m \times d_2} \\ c \in \mathbb{R}^m$$

### Lagrangien du problème

Rappel: Pour minimiser  $f(w)$  s.c.  $Aw=b$ , le lagrangien s'écrit  $\mathcal{L}(w, z) = f(w) + z^T(Aw-b)$

Le lagrangien associé à (P) est la fonction

$$\mathcal{L}: \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^m \rightarrow \mathbb{R}$$

$$(u, v, z) \mapsto \mathcal{L}(u, v, z) = f(u) + g(v) + z^T(Au + Bv - c)$$

↑  
Variable du problème (P)

↑  
variable duale /  
multiplicateur de Lagrange

- ↳ Comme on a défini un lagrangien, on peut également définir des méthodes d'optimisation duales pour résoudre (P)
- \* Fourche duale  $\Rightarrow$  Décomposition duale
- \* Lagrangien augmenté  $\Rightarrow$  ADMM

## Décomposition duale

• Initialisation:  $u_0 \in \mathbb{R}^{d_1}, v_0 \in \mathbb{R}^{d_2}, z_0 \in \mathbb{R}^m$

• Pour  $k=0, 1, \dots$

• Calculer  $u_{k+1} \in \underset{u \in \mathbb{R}^{d_1}}{\operatorname{argmin}} \mathcal{L}(u, v_k, z_k)$

} Optimisation selon  $u$

• Calculer  $v_{k+1} \in \underset{v \in \mathbb{R}^{d_2}}{\operatorname{argmin}} \mathcal{L}(u_k, v, z_k)$

} Optimisation selon  $v$   
( $u, z$  fixés)

• Calculer  $z_{k+1} = z_k + \alpha_k (A u_{k+1} + B v_{k+1} - c)$   $\alpha_k > 0$

↓  
Mise à jour de la variable duale  
(par de sous-gradient pour le problème dual)

## Intérêt pratique

• Les calculs de  $u_{k+1}$  et  $v_{k+1}$  sont indépendants et donc peuvent être effectués en parallèle

• Par rapport à l'approche duale générique, on résout deux problèmes de taille  $d_1$  et  $d_2$  au lieu d'un problème de taille  $d_1 + d_2$ , ce qui est souvent plus efficace en termes de calcul

## ADMM (Alternating Direction Method of Multipliers)

Méthode des multiplicateurs: autre nom de l'algorithme de Lagrangien augmenté

## Lagrangien augmenté associé à (P)

$\mathcal{L}^a: \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$

$(u, v, z, \lambda) \mapsto \mathcal{L}^a(u, v, z, \lambda)$

avec

$$\mathcal{L}^{\lambda}(u, v, z; \lambda) = \underbrace{f(u) + g(v) + z^T (Au + Bv - c)}_{\mathcal{L}(u, v, z)} + \frac{\lambda}{2} \|Au + Bv - c\|^2$$

↑  
Régularisation  
→ pénalisation des  
couples  $(u, v)$  qui ne  
vérifient pas la  
contrainte

### Algorithme

• Initialisation:  $u_0 \in \mathbb{R}^{d_1}$ ,  $v_0 \in \mathbb{R}^{d_2}$ ,  $z_0 \in \mathbb{R}^m$ ,  $\lambda > 0$ .

• Pour  $k=0, 1, \dots$

• Calculer  $u_{k+1} \in \underset{u \in \mathbb{R}^{d_1}}{\operatorname{argmin}} \mathcal{L}^{\lambda}(u, v_k, z_k; \lambda)$

• Calculer  $v_{k+1} \in \underset{v \in \mathbb{R}^{d_2}}{\operatorname{argmin}} \mathcal{L}^{\lambda}(u_k, v, z_k; \lambda)$

• Calculer  $z_{k+1} = z_k + \lambda (Au_{k+1} + Bv_{k+1} - c)$

### Commentaires

→ On peut choisir d'autres valeurs de pas que  $\lambda$  dans le calcul de  $z_{k+1}$

→ On peut faire varier  $\lambda$  au cours des itérations  
 $\{\lambda_k\}$  (en général  $\lambda_{k+1} \geq \lambda_k$ )

Remarque: Toutes les idées ci-dessus s'adaptent à une décomposition en un nombre quelconque de vecteurs de variables

$$\begin{array}{l} \text{minimiser} \\ u_1 \in \mathbb{R}^{d_1} \\ \vdots \\ u_q \in \mathbb{R}^{d_q} \end{array} \sum_{i=1}^q f_i(u_i) \quad \text{s.c.} \quad \sum_{i=1}^q A_i u_i = b$$

$$\begin{array}{l} f_i: \mathbb{R}^{d_i} \rightarrow \mathbb{R} \quad \forall i \\ A_i \in \mathbb{R}^{m \times d_i} \quad \forall i \\ b \in \mathbb{R}^m \end{array}$$

Exemple: "Group LASSO"

$$\text{LASSO: } X \in \mathbb{R}^{m \times d}, y \in \mathbb{R}^m$$

$$\text{minimiser } w \in \mathbb{R}^d \quad \frac{1}{2} \|Xw - y\|^2 + \lambda \|w\|_1$$

"  $\sum_{j=1}^d |w_j|$  "

$\Rightarrow$  Solution est parcimonieuse (beaucoup de coeff. nuls)

$$\|w\|_1 = \sqrt{\sum v_j^2}$$

Group LASSO

$$\text{minimiser } w \in \mathbb{R}^d \quad \frac{1}{2} \|Xw - y\|^2 + \lambda \sum_{l=1}^L \|w^{[l]}\|$$

$$\text{avec } w = \begin{bmatrix} w^{[1]} \\ \vdots \\ w^{[L]} \end{bmatrix}$$

$L = d$  LASSO  
 $L < d \Rightarrow$  donne des solutions avec beaucoup de sous-vecteurs  $w^{[l]}$  nuls

① On introduit un vecteur de  $z \in \mathbb{R}^d$  pour rendre le problème séparable

② On force  $w$  et  $z$  à être égaux via une contrainte linéaire  
 $w - z = 0 \Leftrightarrow \sum_{l=1}^L (w^{[l]} - z^{[l]}) = 0$

$\Rightarrow$  On obtient

$$\begin{array}{l} \text{minimiser} \\ w \in \mathbb{R}^d \\ z \in \mathbb{R}^d \end{array} \frac{1}{2} \|Xw - y\|^2 + \lambda \sum_{l=1}^L \|z^{[l]}\|$$

$$\text{s.c.} \quad \sum_{l=1}^L (w^{[l]} - z^{[l]}) = 0$$

(-) On a ajouté des variables

(+) On a obtenu un problème séparable en  $w$  et  $z$  et sur ce problème les itérations d'ADMM en  $w$  et  $z$  sont plus simples à calculer qu'une itération de gradient proximal sur le problème de départ

↳ On peut même introduire encore plus de séparabilité

$$\underset{w \in \mathbb{R}^d}{\text{minimiser}} \quad -\frac{1}{2} \|Xw - y\|^2 + \lambda \sum_{l=1}^L \|v_l\|$$

$$\underbrace{v_1, \dots, v_L}_{\substack{d_1 \\ \vdots \\ d_L}} \quad \text{s.c.} \quad w[l] - v_l = 0 \quad l=1..L$$

$$d_1 + \dots + d_L = d$$

NB: Très grand nombre de variables sur ADMM:

- stochastique

- proximal

- accélération (sur certains problèmes convexes)

## (4) Optimisation par consensus et décentralisée

↳ Contexte:  $m$  entités / agents qui possèdent chacun leur fonction objectif  $f_i: \mathbb{R}^d \rightarrow \mathbb{R}$  qui n'est pas visible par les autres agents

$$\text{But:} \quad \underset{w \in \mathbb{R}^d}{\text{minimiser}} \quad \sum_{i=1}^m f_i(w)$$

↳ Décentralisation: On considère que chaque agent dispose d'une copie des variables  $w^{(i)} \in \mathbb{R}^d$  et on optimise donc la somme des  $f_i(w^{(i)})$

$\Rightarrow$  On impose (via des contraintes) le fait que les agents doivent arriver à un consensus, c'est à dire que  $w^{(i)} = w^{(j)} \quad \forall (i, j)$

Deux problèmes d'optimisation possibles :

① minimiser  $\sum_{i=1}^m f_i(w^{(i)})$  s.c.  $w - w^{(i)} = 0 \quad \forall i \in \{1, \dots, m\}$

$w \in \mathbb{R}^d$   
 $w^{(1)} \in \mathbb{R}^d$   
 $\vdots$   
 $w^{(m)} \in \mathbb{R}^d$

↓  
 Fonctions optimiser

↓  
 Contrainte de consensus

Optimisation par consensus en distribué avec une variable "maître"  $w$  partagée par tous les agents

$\Rightarrow$  On peut appliquer une technique de décomposition avec  $u = \begin{bmatrix} w^{(1)} \\ \vdots \\ w^{(m)} \end{bmatrix}$  et  $v = w$

(on peut décomposer en  $u^{(1)} = w^{(1)}, \dots, u^{(m)} = w^{(m)}, u^{(m+1)} = w$ )

$\Rightarrow$  le problème à résoudre pour mettre à jour  $w$  est très simple

② minimiser  $\sum_{i=1}^m f_i(w^{(i)})$  s.c.  $w^{(i)} - w^{(j)} = 0 \quad \forall 1 \leq i < j \leq m$

$w^{(1)} \in \mathbb{R}^d$   
 $\vdots$   
 $w^{(m)} \in \mathbb{R}^d$

Problème d'optimisation décentralisé (pas de variable maître partagée entre les agents)  $\Rightarrow$  les techniques duales s'appliquent encore

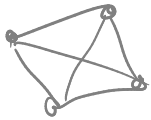
↳ Ce concept d'optimisation décentralisée se généralise  
 encore plus. à un réseau de  $m$  agents qui communiquent  
 via un graphe  $V = \{1, \dots, m\}$  sommets  
 $E \subseteq \{1, \dots, m\}^2$  arêtes

(3) minimiser  $w^{(i)} \in \mathbb{R}^d$   
 $w^{(m)} \in \mathbb{R}^d$

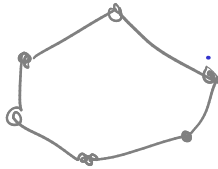
$$\sum_{i=1}^m f_i(w^{(i)}) \quad \text{s.c.} \quad w^{(i)} - w^{(j)} = 0$$

$$\forall (i,j) \in E$$

"Consensus entre voisins"



(2)



(3)

