Principe du Problème Auxiliaire et décomposition

15 décembre 2020

Rappel des cours précédents

On a jusqu'ici étudié des problèmes à structure additive :

- $u = (u_1, \ldots, u_N) \in U_1^{\mathrm{ad}} \times \cdots \times U_N^{\mathrm{ad}}$,
- $J(u_1, \ldots, u_N) = \sum_{i=1}^N J_i(u_i)$,
- $\Theta(u_1, \ldots, u_N) = \sum_{i=1}^N \Theta_i(u_i)$,

Pour ces problèmes, on a présenté 3 méthodes de décomposition :

1. par les prix, qui consiste à former le Lagrangien et à constater que le problème se décompose,
2. par allocation, où l'on introduit des variables supplémentaires permettant de décomposer,
3. par prédiction, pour laquelle on distribue les contraintes entre les sous-problèmes.

Dans tout cela, le fait que le problème soit à structure additive a joué un rôle essentiel !

Objectifs de la suite du cours I

Notre but est maintenant de se dispenser des hypothèses du modèle additif concernant J et Θ.[10] En effet,

- dans beaucoup de cas concrets, les fonctions de coût et de contrainte ne sont pas additives,
- les bonnes techniques de dualité, comme le Lagrangien augmenté, détruisent l'additivité de la contrainte.

Pour cela, on va introduire le Principe du Problème Auxiliaire, qui consiste à remplacer un problème par une suite de problèmes, ce qui permet de disposer d'un cadre général dans lequel on construit une grande variété d'algorithmes de résolution. Les avantages sont

- théoriques : étude de convergence unique,
- pratiques : décomposition, reconditionnement...

10. On continuera à supposer que l'ensemble U^{ad} se met sous la forme d'un produit cartésien, une contrainte n'ayant pas cette forme étant mise dans Θ.

Objectifs de la suite du cours II

Aujourd'hui, on va se limiter à l'étude des problèmes dans lesquels ne figurent pas de contrainte explicite :

$$\min_{u \in U^{\mathrm{ad}}} J(u) + J^{\Sigma}(u) \ ,$$

car c'est dans ce cadre que le Principe du Problème Auxilaire (**PPA**) se comprend le mieux et se met en œuvre simplement. Le terme supplémentaire $J^{\Sigma}(u)$ dans le critère représente la « partie qui ne pose pas de difficultés en décomposition », par exemple parce qu'elle est additive.

À la séance prochaine, on étudiera (rapidement) le cas général :

$$\min_{u \in U^{\mathrm{ad}}} J(u) + J^{\Sigma}(u) \quad \text{sous} \quad \Theta(u) + \Theta^{\Sigma}(u) \in -C \ .$$

Plan du cours

Cadre du problème étudié

Le problème étudié est :

$$\min_{u \in U^{\mathrm{ad}} \subset \mathcal{U}} J(u) + J^{\Sigma}(u) \ , \qquad (P_0)$$

avec

- U^{ad} convexe fermé non vide de $\mathcal{U}$,
- J et J^{Σ} convexes, s.c.i., propres,
- J différentiable,
- $J + J^{\Sigma}$ coercive.

Lorsqu'il sera question de décomposition, on supposera de plus que l'on a $U^{\mathrm{ad}} = U_1^{\mathrm{ad}} \times \cdots \times U_N^{\mathrm{ad}}$ et $J^{\Sigma}(u) = \sum_{i=1}^{N} J_i^{\Sigma}(u_i)$.

On rappelle qu'une solution optimale $u^{\sharp}$ du problème (P_0) est caractérisée par l'inéquation variationnelle :

$$\left\langle \nabla J(u^{\sharp}) \, , u - u^{\sharp} \right\rangle + J^{\Sigma}(u) - J^{\Sigma}(u^{\sharp}) \geq 0 \quad \forall u \in U^{\mathrm{ad}} \ .$$

Construction du problème auxiliaire I

La construction du problème auxiliaire associé à (P_0) autour d'un point $u^{(k)}$ repose sur les deux idées suivantes.

1. Remplacer $J(u)$ par son développement au premier ordre au point $u^{(k)}$: $J(u^{(k)}) + \langle \nabla J(u^{(k)}) \, , u - u^{(k)} \rangle$. En faisant cela,
   - on ne change pas trop la fonction J autour de $u^{(k)}$,
   - on ne perturbe pas la condition d'optimalité si $u^{(k)} = u^\sharp$,
   - on rend additive la partie du critère dépendant de J,

   mais on perd la coercivité du critère.
2. Choisir une fonction $K^{(k)}$ fortement convexe, et ajouter le terme $K^{(k)}(u)$ auquel on retranche son approximation à l'ordre 1 (pour ne pas perturber la condition d'optimalité) :

$$\frac{1}{\epsilon^{(k)}} \Big(K^{(k)}(u) - K^{(k)}(u^{(k)}) - \langle \nabla K^{(k)}(u^{(k)}) \, , u - u^{(k)} \rangle \Big) \, ,$$

où $\epsilon^{(k)} > 0$ est un coefficient de normalisation à choisir.

Construction du problème auxiliaire II

On forme alors le problème auxiliaire en appliquant ces deux idées. Éliminant les termes constants, on ajoute dans le problème (P_0) le terme $\big(K^{(k)}(u) - \big\langle \nabla K^{(k)}(u^{(k)}) \,, u \big\rangle\big)/\epsilon^{(k)}$, et on remplace le terme $J(u)$ par $\big\langle \nabla J(u^{(k)}) \,, u \big\rangle$, ce qui conduit, en multipliant le tout par $\epsilon^{(k)}$, au problème auxiliaire :

$$\min_{u \in U^{\mathrm{ad}}} K^{(k)}(u) + \big\langle \epsilon^{(k)} \nabla J(u^{(k)}) - \nabla K^{(k)}(u^{(k)}) \,, u \big\rangle + \epsilon^{(k)} J^{\Sigma}(u) \,. \quad (PA^{(k)})$$

L'unique solution du problème $(PA^{(k)})$ est notée $u^{(k+1)}$ et on l'utilise pour former le problème auxiliaire suivant $(PA^{(k+1)})$. Ainsi, la résolution itérée des problèmes auxiliaires engendre une suite $\{u^{(k)}\}_{k \in \mathbb{N}}$ dont on peut étudier la convergence.

Algorithme du PPA

La construction des problèmes auxiliaires suggère la mise en œuvre de l'algorithme suivant.

(0) **Initialisation**
Choisir une suite de noyaux $\{K^{(k)}\}_{k\in\mathbb{N}}$ et une suite de coefficients $\{\epsilon^{(k)}\}_{k\in\mathbb{N}}$. Choisir un point initial $u^{(0)} \in U^{\mathrm{ad}}$ et faire $k = 0$.

(1) **Itération** k
Résoudre le problème $(PA^{(k)})$.

(2) **Test de convergence**
Soit $u^{(k+1)}$ la solution de $(PA^{(k)})$:

- si $\|u^{(k+1)} - u^{(k)}\| > \sigma$, faire $k \leftarrow k+1$ et aller à l'étape (1), [11]
- sinon, fin de l'algorithme.

11. le test d'arrêt devrait être mieux choisi. . .

Bien fondé de l'algorithme du PPA

Le problème auxiliaire $(PA^{(k)})$ est donc :

$$\min_{u \in U^{\mathrm{ad}}} K^{(k)}(u) + \left\langle \epsilon^{(k)} \nabla J(u^{(k)}) - \nabla K^{(k)}(u^{(k)}) \, , u \right\rangle + \epsilon^{(k)} J^{\Sigma}(u) \; ,$$

et sa condition d'optimalité s'écrit :

$$\begin{aligned} &\left\langle \nabla K^{(k)}(u^{(k+1)}) - \nabla K^{(k)}(u^{(k)}) + \epsilon^{(k)} \nabla J(u^{(k)}) \, , u - u^{(k+1)} \right\rangle \\ &\qquad + \epsilon^{(k)} \big(J^{\Sigma}(u) - J^{\Sigma}(u^{(k+1)}) \big) \geq 0 \quad \forall u \in U^{\mathrm{ad}} \; . \end{aligned}$$

Supposons que la suite $\{u^{(k)}\}_{k \in \mathbb{N}}$ des solutions des problèmes auxiliaires $(PA^{(k)})$ converge vers un point $\overline{u}$, et que les gradients des fonctions J et $K^{(k)}$ sont suffisamment réguliers. Alors, le terme $\nabla K^{(k)}(u^{(k+1)}) - \nabla K^{(k)}(u^{(k)})$ tend vers zéro quand k tend vers l'infini, et la condition d'optimalité devient à la limite :

$$\left\langle \epsilon \nabla J(\overline{u}) \, , u - \overline{u} \right\rangle + \epsilon \big(J^{\Sigma}(u) - J^{\Sigma}(\overline{u}) \big) \geq 0 \quad \forall u \in U^{\mathrm{ad}} \; ,$$

qui est exactement celle du problème initial (P_0) !

Intérêt du PPA ?

On a donc remplacé la résolution du problème d'optimisation (P_0) par la résolution d'une suite de problèmes $\left\{(PA^{(k)})\right\}_{k\in\mathbb{N}}$. Ceci ne peut avoir d'intérêt que si les conditions suivantes sont remplies :

1. la suite des solutions $\{u^k\}_{k\in\mathbb{N}}$ converge, en un sens approprié,
2. la résolution de $(PA^{(k)})$ est plus facile que celle de (P_0),
   - en terme du nombre d'opérations à effectuer,
   - en terme du conditionnement du problème,
   - en terme de décomposition.

Pour réaliser ces conditions, on peut s'appuyer sur le choix des noyaux $K^{(k)}$, ce qui conduit à une grande diversité d'algorithmes.

On peut déjà noter que choisir des noyaux $K^{(k)}$ additifs permettra de décomposer le problème auxiliaire $(PA^{(k)})$:

$$\min_{u\in U^{\mathrm{ad}}} K^{(k)}(u) + \left\langle \epsilon^{(k)}\nabla J(u^{(k)}) - \nabla K^{(k)}(u^{(k)}) \, , u \right\rangle + \epsilon^{(k)} J^{\Sigma}(u) \, .$$

Énoncé du théorème I

Théorème – Hypothèses

On suppose que

- U^{ad} est un ensemble convexe fermé non vide d'un espace de Hilbert $\mathcal{U}$,
- J est convexe, s.c.i., propre, de gradient Lipschitzien de constante A,
- J^{Σ} est convexe, s.c.i., propre,
- $J + J^{\Sigma}$ est coercive,
- $K^{(k)}$ est fortement convexe de module $b^{(k)}$, s.c.i., propre, de gradient Lipschitzien de constante $B^{(k)}$, et on a :

$$\exists b > 0 \ \text{ et } \ B > 0 \,, \ \text{ tels que } \ b^{(k)} \geq b > 0 \ \text{ et } \ 0 < B^{(k)} \leq B \quad \forall k \,,$$

- $\epsilon^{(k)}$ est tel que

$$\exists \alpha > 0 \ \text{ et } \ \beta > 0 \,, \ \text{ tels que } \ \alpha \leq \epsilon^{(k)} \leq \frac{2b^{(k)}}{A + \beta} \quad \forall k \,.$$

Énoncé du théorème II

Théorème – Conclusions

Alors,

- le problème (P_0) admet au moins une solution $u^\sharp$,
- le problème $(PA^{(k+1)})$ admet une unique solution $u^{(k+1)}$,
- la suite $\{J(u^{(k)}) + J^\Sigma(u^{(k)})\}_{k\in\mathbb{N}}$ est décroissante et minorée,
- la suite $\{J(u^{(k)}) + J^\Sigma(u^{(k)})\}_{k\in\mathbb{N}}$ converge vers $J(u^\sharp) + J^\Sigma(u^\sharp)$,
- tout point d'accumulation de la suite $\{u^{(k)}\}_{k\in\mathbb{N}}$ est une solution de (P_0).

Si on suppose de plus que J est fortement convexe de module a, alors

- la suite $\{u^{(k)}\}_{k\in\mathbb{N}}$ converge vers l'unique solution $u^\sharp$ de (P_0),
- on dispose d'une majoration de l'erreur de convergence :

$$\|u^{(k+1)} - u^\sharp\| \leq \frac{\left(B^{(k)}/\epsilon^{(k)}\right) + A}{a} \|u^{(k+1)} - u^{(k)}\| \ .$$

Schéma de la preuve

Plutôt que de faire la preuve in extenso du théorème, on va en donner la structure, très générale, qui s'appuie sur la théorie du contrôle et l'étude de la stabilité des systèmes dynamiques.

1. On se donne une fonction réelle $\mathcal{L}$ opérant sur les variables de l'algorithme, telle que $\mathcal{L}$ soit bornée inférieurement et coercive. Une telle fonction est dite de Lyapunov.
2. La fonction $\mathcal{L}$ est choisie de telle sorte qu'elle décroit le long des trajectoires de l'algorithme : $\mathcal{L}(u^{(k+1)}) \leq \mathcal{L}(u^{(k)})$.
3. On a alors que la suite $\{\mathcal{L}(u^{(k)})\}_{k \in \mathbb{N}}$ est convergente et on calcule sa limite.
4. On a aussi que la suite $\{u^{(k)}\}_{k \in \mathbb{N}}$ est bornée et on caractérise les limites de ses sous-suites convergentes.

Résumé de la preuve

1. Un bon choix de fonction de Lyapunov $\mathcal{L}$ est le critère du problème $J + J^{\Sigma}$, qui est coercif et borné inférieurement par $J(u^{\sharp}) + J^{\Sigma}(u^{\sharp})$.
2. En écrivant la condition d'optimalité du problème $(PA^{(k)})$ au point $u^{(k)}$ et en utilisant les propriétés des fonctions J et $K^{(k)}$, on montre que
$$\mathcal{L}(u^{(k)}) - \mathcal{L}(u^{(k+1)}) \geq \left(\frac{b^{(k)}}{\epsilon^{(k)}} - \frac{A}{2}\right) \|u^{(k)} - u^{(k+1)}\|^2 ,$$
d'où la convergence de la suite $\left\{\mathcal{L}(u^{(k)})\right\}_{k\in\mathbb{N}}$ et le caractère borné de la suite $\{u^{(k)}\}_{k\in\mathbb{N}}$ par coercivité de $\mathcal{L}$.
3. Écrivant la condition d'optimalité de $(PA^{(k)})$ au point $u^{\sharp}$ et utilisant les propriétés de J et $K^{(k)}$, on montre que $\mathcal{L}(u^{\sharp}) - \mathcal{L}(u^{(k+1)}) \geq \xi^{(k+1)} \to 0$, d'où la limite $\mathcal{L}(u^{\sharp})$ de la suite $\left\{\mathcal{L}(u^{(k)})\right\}_{k\in\mathbb{N}}$.
4. Grâce au caractère s.c.i. de $J + J^{\Sigma}$, on obtient que toute limite $\overline{u}$ d'une sous-suite convergente de $\{u^{(k)}\}_{k\in\mathbb{N}}$ est solution du problème (P_0).

Un mot de dimension infinie

Un argument essentiel dans la preuve du théorème est que, la suite $\{u^{(k)}\}_{k \in \mathbb{N}}$ étant bornée, elle possède des sous-suites convergentes.

- Dans le cas où $\mathcal{U}$ est un espace de Hilbert de dimension finie muni de la topologie forte (celle induite par la norme), c'est une conséquence du fait que les ensembles fermés bornés de $\mathcal{U}$ sont compacts, d'où la conclusion par le théorème de Bolzano-Weierstrass. Si $\mathcal{U}$ est un Hilbert de dimension infinie, les fermés bornés ne sont plus compacts pour la topologie forte et l'argument ne tient plus...
- Pour s'en sortir, on change de point de vue et on considère sur l'espace $\mathcal{U}$ la topologie faible, qui est la topologie la plus grossière pour laquelle les éléments du dual $\mathcal{U}^\star$ de $\mathcal{U}$ sont continues. On sait que, dans la topologie faible, les ensembles fermés et bornés de $\mathcal{U}$ sont compacts. Il reste à prouver que les propriétés de fermeture et de semi-continuité inférieure, faites pour la topologie forte, restent vraies dans la topologie faible. Ceci est faux en toute généralité, mais est vérifié dans dans notre cas car on sait que les ensembles convexes fortement fermés sont faiblement fermés. De même, la fonction $J + J^\Sigma$ est convexe et fortement s.c.i., et est donc faiblement s.c.i..

1. Décomposition par le Principe du Problème Auxiliaire
   - Principe du Problème Auxiliaire
   - Théorème de convergence
   - Intérêts du PPA

2. Travaux dirigés sur la décomposition par prédiction
   - Le cas des contraintes inégalités
   - Étude d'une vallée hydraulique
   - Réseau de distribution d'eau

Une variante de l'algorithme du PPA

Il peut y avoir de l'intérêt à choisir dans l'algorithme du **PPA** des coefficients $\epsilon^{(k)}$ tous égaux à la valeur 1, par exemple pour éliminer des termes dans la différence $\nabla J(u^{(k)}) - \nabla K^{(k)}(u^{(k)})$. Comme on l'a vu dans le théorème de convergence, ceci implique :

$$\frac{2b^{(k)}}{A} > 1 \ ,$$

où $b^{(k)}$ est le module de forte convexité du noyau $K^{(k)}$. Si cette condition n'est pas vérifiée, on peut toujours ajouter au noyau $K^{(k)}$ un terme quadratique $\frac{\gamma^{(k)}}{2}\|\cdot\|^2$. Alors, le module de forte convexité du nouveau noyau $K^{(k)}(u) + \frac{\gamma^{(k)}}{2}\|u\|^2$ est égal à $b^{(k)} + \gamma^{(k)}$, et la condition sera remplie pourvu que $\gamma^{(k)}$ soit pris assez grand.

On notera que ce terme additionnel dans le noyau $K^{(k)}$ introduit un terme de freinage $\frac{\gamma^{(k)}}{2}\|u - u^{(k)}\|^2$ dans le problème $(PA^{(k)})$.

Intérêt du PPA en décomposition I

On veut mettre en œuvre une méthode de décomposition sur un problème tel que :

$$U^{\mathrm{ad}} = U_1^{\mathrm{ad}} \times \cdots \times U_N^{\mathrm{ad}} \ ,$$

$$J^{\Sigma}(u) = \sum_{i=1}^{N} J_i^{\Sigma}(u_i) \ ,$$

mais pour lequel la fonction J n'est pas additive.

Pour cela, on choisit un noyau $K^{(k)}$ additif :

$$K^{(k)}(u) = \sum_{i=1}^{N} K_i^{(k)}(u_i) \ .$$

Intérêt du PPA en décomposition II

Le problème auxiliaire $(PA^{(k)})$:

$$\min_{u \in U^{\mathrm{ad}}} K^{(k)}(u) + \left\langle \epsilon^{(k)} \nabla J(u^{(k)}) - \nabla K^{(k)}(u^{(k)}) \, , u \right\rangle + \epsilon^{(k)} J^{\Sigma}(u) \, ,$$

s'écrit alors :

$$\min_{u \in U^{\mathrm{ad}}} \sum_{i=1}^{N} K_i^{(k)}(u_i) + \sum_{i=1}^{N} \left\langle \epsilon^{(k)} \nabla_{u_i} J(u^{(k)}) - \nabla K_i^{(k)}(u_i^{(k)}), u_i \right\rangle + \sum_{i=1}^{N} \epsilon^{(k)} J_i^{\Sigma}(u_i) \, ,$$

et la minimisation peut être faite indépendamment u_i par u_i :

$$\min_{u_i \in U_i^{\mathrm{ad}}} K_i^{(k)}(u_i) + \left\langle \epsilon^{(k)} \nabla_{u_i} J(u^{(k)}) - \nabla K_i^{(k)}(u_i^{(k)}) \, , u_i \right\rangle + \epsilon^{(k)} J_i^{\Sigma}(u_i) \, ,$$

Le choix d'un noyau $K^{(k)}$ additif a donc permis la décomposition.

Choix canonique de noyaux de décomposition

Dans le cas où la fonction J est fortement convexe, un choix possible, dit canonique de noyau de décomposition additif est de choisir les $K_i^{(k)}$ suivants :

$$K_i^{(k)}(u_i) = J(u_1^{(k)}, \ldots, u_{i-1}^{(k)}, u_i, u_{i+1}^{(k)}, \ldots, u_N^{(k)}) \; .$$

Avec ce choix, on a $\nabla K_i^{(k)}(u_i^{(k)}) = \nabla_{u_i} J(u^{(k)})$, et le sous-problème de minimisation en u_i issu de la décomposition s'écrit alors :

$$\min_{u_i \in U_i^{\mathrm{ad}}} K_i^{(k)}(u_i) + (\epsilon^{(k)} - 1)\big\langle \nabla_{u_i} J(u^{(k)}) \, , u_i \big\rangle + \epsilon^{(k)} J_i^{\Sigma}(u_i) \; .$$

Dans le cas $\epsilon^{(k)} = 1$, ce sous-problème se réduit à :

$$\min_{u_i \in U_i^{\mathrm{ad}}} J(u_1^{(k)}, \ldots, u_{i-1}^{(k)}, u_i, u_{i+1}^{(k)}, \ldots, u_N^{(k)}) \; + \; J_i^{\Sigma}(u_i) \; ,$$

et l'algorithme est donc la méthode de coordinate descent.

Algorithme du gradient et PPA

On considère la situation où $J^{\Sigma} \equiv 0$. Avec le choix du noyau quadratique $K(u) = \frac{1}{2}\|u\|^2$, le problème auxiliaire $(PA^{(k)})$ s'écrit

$$\min_{u \in U^{\mathrm{ad}}} \frac{1}{2}\|u\|^2 + \left\langle \epsilon^{(k)} \nabla J(u^{(k)}) - u^{(k)} \, , u \right\rangle .$$

Ce problème est équivalent à :

$$\min_{u \in U^{\mathrm{ad}}} \frac{1}{2}\left\| u - u^{(k)} + \epsilon^{(k)} \nabla J(u^{(k)}) \right\|^2 ,$$

dont la solution est donnée par :

$$u^{(k+1)} = \mathrm{proj}_{U^{\mathrm{ad}}}\left(u^{(k)} - \epsilon^{(k)} \nabla J(u^{(k)})\right) .$$

C'est l'algorithme du gradient projeté.

Dans cette situation, si l'on choisit un noyau K général, l'algorithme du PPA est identique à l'algorithme de Mirror Descent permettant de mettre en œuvre des méthodes de type gradient dans des espaces de Banach.

Algorithme prox et PPA

Dans le cas $J^{\Sigma} \neq 0$, avec le noyau quadratique $K(u) = \frac{1}{2}\|u\|^2$, le problème auxiliaire $(PA^{(k)})$ s'écrit

$$\min_{u \in U^{\mathrm{ad}}} \frac{1}{2}\|u\|^2 + \left\langle \epsilon^{(k)} \nabla J(u^{(k)}) - u^{(k)} \, , u \right\rangle + \epsilon^{(k)} J^{\Sigma}(u) \, .$$

Ce problème est équivalent à :

$$\min_{u \in U^{\mathrm{ad}}} \frac{1}{2}\left\| u - u^{(k)} + \epsilon^{(k)} \nabla J(u^{(k)}) \right\|^2 + \epsilon^{(k)} J^{\Sigma}(u) \, ,$$

et correspond au calcul de la régularisée de Moreau-Yosida de la fonction $\epsilon^{(k)} J^{\Sigma}$ au point $u^{(k)} - \epsilon^{(k)} \nabla J(u^{(k)})$, dont la solution est :

$$u^{(k+1)} = \mathrm{prox}_{\epsilon^{(k)} J^{\Sigma}} \left(u^{(k)} - \epsilon^{(k)} \nabla J(u^{(k)}) \right) \, .$$

C'est l'algorithme du gradient proximal.

Algorithme de Newton et PPA

On considère la situation où $J^\Sigma \equiv 0$ et où $U^{\mathrm{ad}} = \mathcal{U}$. On suppose que la fonction J est fortement convexe deux fois différentiable et on choisit le noyau $K^{(k)}(u) = \frac{1}{2}\langle u \,, \nabla^2 J(u^{(k)}) \cdot u\rangle$. Le problème auxiliaire $(PA^{(k)})$ s'écrit :

$$\min_{u \in \mathcal{U}} \frac{1}{2}\langle u \,, \nabla^2 J(u^{(k)}) \cdot u\rangle + \langle \epsilon^{(k)} \nabla J(u^{(k)}) - \nabla^2 J(u^{(k)}) \cdot u^{(k)} \,, u\rangle \,,$$

et sa condition d'optimalité :

$$\nabla^2 J(u^{(k)}) \cdot u^{(k+1)} + \epsilon^{(k)} \nabla J(u^{(k)}) - \nabla^2 J(u^{(k)}) \cdot u^{(k)} = 0 \,,$$

permet de calculer sa solution :

$$u^{(k+1)} = u^{(k)} - \epsilon^{(k)} \big[\nabla^2 J(u^{(k)})\big]^{-1} \cdot \nabla J(u^{(k)}) \,.$$

C'est l'algorithme de Newton.

Reconditionnement de l'algorithme de Newton par le PPA

On considère la situation précédente ($J^{\Sigma} \equiv 0$, $U^{\mathrm{ad}} = \mathcal{U}$ et J deux fois différentiable), mais on suppose que J est simplement convexe. Afin de disposer d'un noyau fortement convexe, on fait le choix de noyau $K^{(k)}(u) = \frac{1}{2}\langle u \, , \nabla^2 J(u^{(k)}) \cdot u \rangle + \frac{\gamma^{(k)}}{2}\|u\|^2$, avec $\gamma^{(k)} > 0$. Le problème auxiliaire $(PA^{(k)})$ s'écrit

$$\min_{u \in \mathcal{U}} \frac{1}{2}\big\langle u \, , \big(\nabla^2 J(u^{(k)}) + \gamma^{(k)} \mathrm{I_d}\big).u \big\rangle + \\ \big\langle \epsilon^{(k)} \nabla J(u^{(k)}) - \big(\nabla^2 J(u^{(k)}) + \gamma^{(k)} \mathrm{I_d}\big) \cdot u^{(k)} \, , u \big\rangle \; ,$$

et sa solution est :

$$u^{(k+1)} = u^{(k)} - \epsilon^{(k)} \big[\nabla^2 J(u^{(k)}) + \gamma^{(k)} \mathrm{I_d}\big]^{-1} \cdot \nabla J(u^{(k)}) \; .$$

C'est l'algorithme de Newton reconditionné.

Décomposition de l'algorithme de Newton par le PPA

Dans la situation précédente ($J^\Sigma \equiv 0$, $U^{\mathrm{ad}} = \mathcal{U}$), on extrait de la Hessienne par blocs de J au point $u^{(k)}$ sa partie diagonale :

$$\nabla^2 J(u(k)) = \begin{pmatrix} \nabla^2_{u_1 u_1} J(u^{(k)}) & \dots & \nabla^2_{u_1 u_N} J(u^{(k)}) \\ \vdots & \ddots & \vdots \\ \nabla^2_{u_N u_1} J(u^{(k)}) & \dots & \nabla^2_{u_N u_N} J(u^{(k)}) \end{pmatrix} \rightsquigarrow H_{\mathrm{D}}^{(k)} = \begin{pmatrix} \nabla^2_{u_1 u_1} J(u^{(k)}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \nabla^2_{u_N u_N} J(u^{(k)}) \end{pmatrix},$$

et on choisit $K^{(k)}(u) = \frac{1}{2}\langle u \,, H_{\mathrm{D}}^{(k)} \cdot u \rangle = \frac{1}{2} \sum_{i=1}^N \langle u_i \,, \nabla^2_{u_i u_i} J(u^{(k)}) \cdot u_i \rangle$.

Ce noyau additif permet dans le problème $(PA^{(k)})$ d'effectuer la minimisation indépendamment u_i par u_i :

$$\min_{u_i \in \mathcal{U}_i} \frac{1}{2}\langle u_i \,, \nabla^2_{u_i u_i} J(u^{(k)}) \cdot u_i \rangle + \langle \epsilon^{(k)} \nabla_{u_i} J(u^{(k)}) - \nabla^2_{u_i u_i} J(u^{(k)}) \cdot u_i^{(k)} \,, u_i \rangle ,$$

et la solution du sous-problème i est :

$$u_i^{(k+1)} = u_i^{(k)} - \epsilon^{(k)} \left[\nabla^2_{u_i u_i} J(u^{(k)}) \right]^{-1} \cdot \nabla_{u_i} J(u^{(k)}) .$$

C'est l'algorithme de Newton décomposé.

Rappel de la décomposition par prédiction I

$$\min_{(u_1,\dots,u_N)\in U_1^{\mathrm{ad}}\times\dots\times U_N^{\mathrm{ad}}} \sum_{i=1}^{N} J_i(u_i) \quad \text{sous} \quad \sum_{i=1}^{N} \Theta_i(u_i) - \theta = 0 \in \mathcal{V} \;.$$

Dans la méthode par prédiction, on commence par choisir à quel sous-problème i sera affectée chaque contrainte. Ce mécanisme de choix revient à partitionner l'espace d'arrivée des contraintes $\mathcal{V}$ en $\mathcal{V}_1 \times \dots \times \mathcal{V}_N$, où $\mathcal{V}_i$ est le sous-espace d'arrivée des contraintes affectées à l'unité i. Si aucune contrainte n'est affectée à i, l'espace $\mathcal{V}_i$ est vide.

Avec cette décomposition de l'espace d'arrivée $\mathcal{V}$, la contrainte du problème se met sous la forme de N contraintes, la contrainte affectée au sous- problème i s'écrivant :

$$\sum_{j=1}^{N} \Theta_j^i(u_j) - \theta^i = 0 \in \mathcal{V}_i \;.$$

Rappel de la décomposition par prédiction II

Chacune de ces N contraintes est écrite de manière équivalente en scindant la contrainte par ajout d'une variable $v^i \in \mathcal{V}_i$:

$$\Theta_i^i(u_i) - v^i = 0 \quad , \quad \sum_{j \neq i} \Theta_j^i(u_j) + v^i - \theta^i = 0 \; .$$

On reformule alors le problème en gardant telle quelle la première contrainte et en dualisant avec un multiplicateur p^i la seconde contrainte. Supposant l'existence d'un point-selle, le problème se met sous la forme :

$$\max_{(p^1,\dots,p^N)} \; \min_{(v^1,\dots,v^N)} \left\{ \min_{(u_1,\dots,u_N)} \sum_{i=1}^{N} \Big(J_i(u_i) + \big\langle p^i \, , \sum_{j \neq i} \Theta_j^i(u_j) + v^i - \theta^i \big\rangle \Big) \right.$$

$$\left. \text{sous} \quad \Theta_i^i(u_i) - v^i = 0 \, , \;\; i = 1, \dots, N \right\} .$$

On note $\mathcal{L}(v,p)$ la valeur du problème de minimisation en u.

Rappel de la décomposition par prédiction III

À $v = v^{(k)}$ et $p = p^{(k)}$ fixés, le problème de minimisation en u se décompose en N sous-problèmes indépendants :

$$\min_{u_i \in U_i^{\mathrm{ad}}} J_i(u_i) + \sum_{j \neq i} \langle p^{\,j,(k)} , \Theta_i^j(u_i) \rangle \quad \text{sous} \quad \Theta_i^i(u_i) - v^{\,i,(k)} = 0 \; ,$$

dont un point selle est notée : $(u_i^{(k+1)}, \lambda^{\,i,(k+1)})$.

La remise à jour des variables v et p se fait en résolvant les conditions d'optimalité du problème de maxi-minimisation :

$$\max_{p \in \mathcal{V}} \min_{v \in \mathcal{V}} \mathcal{L}(v, p) \; ,$$

à savoir $\nabla_v \mathcal{L}(v, p) = 0$ et $\nabla_p \mathcal{L}(v, p) = 0$ (car il n'y a aucune contrainte sur v et p), ce qui conduit par relaxation aux relations constituant l'étape de coordination :

$$p^{\,i,(k+1)} = \lambda^{\,i,(k+1)} \quad , \quad v^{\,i,(k+1)} = \theta^i - \sum_{j \neq i} \Theta_j^i(u_j^{(k+1)}) \; .$$

Décomposition par prédiction et contrainte inégalité E_1

On considère le cas d'une contrainte scalaire de type inégalité dans le problème d'optimisation à structure additive :

$$\min_{u_i \in U_i^{\mathrm{ad}} \subset \mathbb{R}^{n_i}} \sum_{i=1}^{N} J_i(u_i) \quad \text{sous} \quad \sum_{i=1}^{N} \Theta_i(u_i) - \theta \leq 0 \in \mathbb{R} ,$$

que l'on souhaite décomposer par prédiction. On suppose que la contrainte scalaire est affectée au sous-problème 1.

1. Quelles sont les deux possibilités pour écrire la méthode de décomposition par prédiction ?
2. Laquelle de ces possibilités permet la mise en œuvre effective de la prédiction ?

Décomposition par prédiction et contrainte inégalité $\quad R_1$

La prédiction consiste à introduire une variable v et à scinder la contrainte en deux parties. La première possibilité s'écrit :

$$\Theta_1(u_1) - v = 0 \quad , \quad \sum_{i=2}^{N} \Theta_i(u_i) + v - \theta \leq 0 \, .$$

La seconde partie de la contrainte, dualisée dans la méthode par prédiction, est une contrainte inégalité : son multiplicateur p doit donc être positif ou nul. La condition d'optimalité par rapport à p de l'opérateur $\mathcal{L}$ n'est pas $\nabla_p \mathcal{L}(v, p) = 0$! On en conclut que la méthode de prédiction ne s'applique pas directement.

Cette impossibilité disparaît si l'on utilise la seconde possibilité :

$$\Theta_1(u_1) - v \leq 0 \quad , \quad \sum_{i=2}^{N} \Theta_i(u_i) + v - \theta = 0 \, ,$$

car alors, le multiplicateur p est associé à une contrainte d'égalité et n'est donc soumis à aucune contrainte de signe.

Décomposition par prédiction et contrainte inégalité R_2

Avec cette seconde possibilité, l'application de la méthode de prédiction conduit à l'algorithme dont la $k$ème itération est :

- phase de décomposition :

$$\min_{u_1 \in U_1^{\mathrm{ad}}} J_1(u_1) \;\; \textit{sous} \;\; \Theta_1(u_1) - v^{(k)} \leq 0 \quad \rightsquigarrow \left(u_1^{(k+1)}, \lambda_1^{(k+1)}\right) ,$$

$$\min_{u_i \in U_i^{\mathrm{ad}}} J_i(u_i) + \left\langle p^{(k)} , \Theta_i(u_i) \right\rangle \quad \rightsquigarrow u_i^{(k+1)} , \quad i = 2, \ldots, N ,$$

- phase de coordination :

$$v^{(k+1)} = \theta - \sum_{i=2}^{N} \Theta_i(u_i^{(k+1)}) ,$$

$$p^{(k+1)} = \lambda_1^{(k+1)} .$$

La seule différence avec l'algorithme vu en cours est la présence d'une contrainte inégalité dans le sous-problème 1.

Optimisation d'une vallée hydraulique E_1

Le problème de l'optimisation d'une vallée hydraulique a été étudié lors de la séance de TP portant sur la décomposition par les prix et s'écrit :

$$\min_{x_{i,t},u_{i,t},w_{i,t}} \sum_i \sum_t L_{i,t}(x_{i,t}, u_{i,t}) ,$$

$$\text{sous} \quad f_{i,t}(x_{i,t}, u_{i,t}, w_{i,t}) - x_{i,t+1} = 0 \quad \forall i , \ \forall t ,$$

$$w_{i,t} - u_{i-1,t} = 0 \quad \forall i , \ \forall t .$$

On souhaite appliquer la décomposition par prédiction à ce problème, avec une décomposition barrage par barrage (i par i).

1. Parmi les contraintes du problème, quelles sont celles qui couplent les barrages et quelles sont celles qui sont locales ?
2. Appliquer la décomposition par prédiction en affectant à chaque barrage i les contraintes $w_{i,t} - u_{i-1,t} = 0 , \ \forall t$.

Optimisation d'une vallée hydraulique R_1

Les contraintes de dynamique $f_{i,t}(x_{i,t}, u_{i,t}, w_{i,t}) - x_{i,t+1} = 0$ font intervenir uniquement des variables du barrage i et peuvent donc être considérées comme des contraintes locales au sous-problème du barrage i.

Les contraintes $w_{i,t} - u_{i-1,t} = 0$ couplent les barrages i et $i-1$ à tous les instants t. Dans la méthode par prédiction, on introduit de nouvelles variables $v_{i,t}$ et on réécrit ces contraintes sous la forme :

$$w_{i,t} - v_{i,t} = 0 \quad , \quad v_{i,t} - u_{i-1,t} = 0 \, .$$

La première partie de cette contrainte est gardée telle quelle dans les sous-problèmes, alors que la seconde partie est dualisée à l'aide d'un multiplicateur $p_{i,t}$. À (v, p) fixé, le problème se décompose en N sous-problèmes (barrage par barrage). L'étape de coordination consiste à remettre à jour les variables (v, p).

Optimisation d'une vallée hydraulique $\quad R_2$

Les éléments à prendre en compte dans le sous-problème du barrage i à l'itération k de la méthode de prédiction sont alors :

- la fonction coût : $\sum_t L_{i,t}(x_{i,t}, u_{i,t})$,
- les contraintes locales : $f_{i,t}(x_{i,t}, u_{i,t}, w_{i,t}) - x_{i,t+1} = 0$,
- les contraintes affectées : $w_{i,t} - v_{i,t}^{(k)} = 0$,
- les termes de dualité : $-\langle p_{i+1,t}^{(k)} , u_{i,t} \rangle$,

d'où la forme du sous-problème associé au barrage i :

$$\min_{x_{i,t}, u_{i,t}, w_{i,t}} \sum_t L_{i,t}(x_{i,t}, u_{i,t}) - \sum_t \langle p_{i+1,t}^{(k)} , u_{i,t} \rangle \ ,$$

$$\text{sous} \quad f_{i,t}(x_{i,t}, u_{i,t}, w_{i,t}) - x_{i,t+1} = 0 \quad \forall t \ ,$$

$$w_{i,t} - v_{i,t}^{(k)} = 0 \quad \forall t \ ,$$

dont la solution est : $(x_{i,t}^{(k+1)}, u_{i,t}^{(k+1)}, w_{i,t}^{(k+1)}, \zeta_{i,t}^{(k+1)}, \lambda_{i,t}^{(k+1)})$.

Optimisation d'une vallée hydraulique R_3

Une fois les N sous-problèmes résolus et les solutions $u_{i,t}^{(k+1)}$ et $\lambda_{i,t}^{(k+1)}$ disponibles, l'étape de coordination s'écrit :

$$v_{i,t}^{(k+1)} = u_{i-1,t}^{(k+1)} \; ,$$
$$p_{i,t}^{(k+1)} = \lambda_{i,t}^{(k+1)} \; .$$

Remarque sur les sous-problèmes

En tenant compte de la coordination, le sous-problème du barrage i est :

$$\min_{x_{i,t}, u_{i,t}} \sum_t \Big(L_{i,t}(x_{i,t}, u_{i,t}) - \langle p_{i+1,t}^{(k)} \; , u_{i,t} \rangle \Big) \; ,$$

$$\text{sous} \quad f_{i,t}(x_{i,t}, u_{i,t}, u_{i-1,t}^{(k)}) - x_{i,t+1} = 0 \quad \forall t \; ,$$

qui est un problème de commande optimale en temps discret standard, où les apports $u_{i-1,t}$ en provenance du barrage en amont sont fixés et où les déversements $u_{i,t}$ vers le barrage aval sont valorisés au prix $p_{i+1,t}^{(k)}$.

Décomposition par prédiction d'un réseau d'eau E_1

On rappelle la formulation compacte du problème d'optimisation du grand réseau d'eau connecté :

$$\min_{(u_{i,1},u_{i,2})_{i=1,\dots,N+1}} \sum_{i=1}^{N+1} J_i(u_{i,1},u_{i,2}) \quad \text{sous} \quad \sum_{i=1}^{N} u_{i,t} - u_{N+1,t} = 0 \;, \quad t = 1,2 \;,$$

avec les contraintes de bornes :

$$(u_{i,1},u_{i,2}) \in U_i^{\mathrm{ad}} = [0,\bar{v}_{i,1}] \times [0,\bar{v}_{i,2} - \bar{v}_{i,1}] \;, \quad i = 1,\dots,N \;.$$

Pour $i = 1,\dots,N$, l'expression détaillée de la fonction J_i est :

$$J_i(u_{i,1},u_{i,2}) = \min_{(v_{i,1},v_{i,2})} \frac{1}{2}\left(a_{i,1}v_{i,1}^2 + a_{i,2}v_{i,2}^2\right) \;,$$

$$\text{sous} \quad u_{i,1} + v_{i,1} \geq \bar{v}_{i,1} \;,$$

$$u_{i,1} + u_{i,2} + v_{i,1} + v_{i,2} - \bar{v}_{i,2} = 0 \;,$$

ainsi que : $J_{N+1}(u_{N+1,1},u_{N+1,2}) = \frac{1}{2}\left(a_{N+1,1}u_{N+1,1}^2 + a_{N+1,2}u_{N+1,2}^2\right)$.

Décomposition par prédiction d'un réseau d'eau E_2

1. Écrire sur la formulation compacte du problème l'algorithme de décomposition par prédiction, en affectant les deux contraintes couplantes au sous-problème $N+1$.
2. Donner la formulation détaillée de chaque sous-problème i, $i = 1, \ldots, N+1$.
3. Discuter de la résolution des conditions de KKT de ces sous-problèmes.

Décomposition par prédiction d'un réseau d'eau $\quad R_1$

Les contraintes couplantes $\sum_{i=1}^{N} u_{i,t} - u_{N+1,t} = 0$ sont réécrites à l'aide de nouvelles variables w_t sous la forme :

$$-u_{N+1,t} - w_t^{(k)} = 0 \quad , \quad \sum_{i=1}^{N} u_{i,t} + w_t^{(k)} = 0 \; , \;\; t = 1, 2 \; ,$$

et l'étape de décomposition de l'algorithme de la méthode par prédiction à l'itération k s'écrit :

$$\min_{(u_{i,1},u_{i,2}) \in U_i^{\mathrm{ad}}} J_i(u_{i,1}, u_{i,2}) + p_1^{(k)} u_{i,1} + p_2^{(k)} u_{i,2} \; , \;\; i = 1 \dots N \; ,$$

$$\min_{(u_{N+1,1},u_{N+1,2}) \in \mathbb{R}^2} J_{N+1}(u_{N+1,1}, u_{N+1,2}) \;\; \text{sous} \;\; \begin{cases} -u_{N+1,1} - w_1^{(k)} = 0 \\ -u_{N+1,2} - w_2^{(k)} = 0 \end{cases} \; .$$

Décomposition par prédiction d'un réseau d'eau R_2

Notant respectivement $(u_{i,1}^{(k+1)}, u_{i,2}^{(k+1)})$ et $(p_{N+1,1}^{(k+1)}, p_{N+1,2}^{(k+1)})$ la solution des sous-problèmes i et les multiplicateurs optimaux associés aux contraintes du sous-problème $N+1$, l'étape de coordination s'écrit, pour $t = 1, 2$:

$$w_t^{(k+1)} = -\sum_{i=1}^{N} u_{i,t}^{(k+1)} \quad , \quad p_t^{(k+1)} = p_{N+1,t}^{(k+1)} \; .$$

Le sous-problème $N+1$ est identique à celui de la méthode par les quantités. Utilisant la forme explicite de la fonction de coût J_{N+1}, on obtient directement l'expression des multiplicateurs :

$$p_{N+1,1}^{(k+1)} = a_{N+1,1}\Big(\sum_{i=1}^{N} u_{i,1}^{(k+1)}\Big) \quad , \quad p_{N+1,2}^{(k+1)} = a_{N+1,2}\Big(\sum_{i=1}^{N} u_{i,2}^{(k+1)}\Big) \; .$$

Décomposition par prédiction d'un réseau d'eau R_3

Pour $i = 1, \ldots, N$, utilisant la forme détaillée de la fonction J_i, on obtient l'expression du i-ème sous-problème, qui est identique à celle de la méthode par les prix :

$$\begin{aligned}
\min_{(v_{i,1}, v_{i,2}, u_{i,1}, u_{i,2}) \in \mathbb{R}^4} \quad & \frac{1}{2}\left(a_{i,1} v_{i,1}^2 + a_{i,2} v_{i,2}^2\right) + p_1^{(k)} u_{i,1} + p_2^{(k)} u_{i,2} \ , \\
\text{sous} \quad & \bar{v}_{i,1} - v_{i,1} - u_{i,1} \leq 0 \ , \\
& v_{i,1} + v_{i,2} + u_{i,1} + u_{i,2} - \bar{v}_{i,2} = 0 \ , \\
& 0 \leq u_{i,1} \leq \bar{v}_{i,1} \ , \\
& 0 \leq u_{i,2} \leq \bar{v}_{i,2} - \bar{v}_{i,1} \ .
\end{aligned}$$

Ce sous-problème comporte 5 contraintes inégalité : résoudre les conditions de KKT est donc faisable...