4

Le PPA stochastique sous contraintes explicites

Comme au §3, on se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire $\boldsymbol{W}$ à valeurs dans l'espace $\mathbb{W}$ muni de sa tribu $\mathcal{W}$. On se donne un espace de Hilbert $\mathbb{U}$ ainsi qu'une partie non vide U^{ad} de $\mathbb{U}$, et une fonction j définie sur $\mathbb{U} \times \mathbb{W}$ à valeurs dans $\overline{\mathbb{R}}$. On note J l'espérance de j (supposée intégrable pour tout $u \in U^{\mathrm{ad}}$) :

$$J(u) = \mathbb{E}\big(j(u, \boldsymbol{W})\big) \ .$$

Pour exprimer les contraintes, on se donne un nouvel espace de Hilbert $\mathbb{V}$, un cône C inclus dans cet espace et une application Θ définie sur $\mathbb{U}$ à valeurs dans $\mathbb{V}$. On s'intéresse alors au problème suivant :

$$\min_{u \in U^{\mathrm{ad}}} J(u) \qquad \text{sous la contrainte} \qquad \Theta(u) \in -C \ . \tag{4.1}$$

Remarque 4.1. On s'intéresse ainsi au cas où le critère J s'exprime comme l'espérance d'une fonction aléatoire, alors que la contrainte Θ est prise sous forme déterministe uniquement. On présentera au §5 une extension du gradient stochastique au cas où la fonction Θ est l'espérance d'une fonction aléatoire θ définie sur $\mathbb{U} \times \mathbb{W}$ à valeurs dans $\mathbb{V}$ (voir Culioli et Cohen (1995)).

4.1 Rappels du cas déterministe

Notant $C^\star$ le cône dual [1] de C, le Lagrangien L du problème (4.1), défini sur $U^{\mathrm{ad}} \times C^\star$ à valeurs réelles, est donné par :

$$L(u, p) = J(u) + \langle p \,, \Theta(u) \rangle \ .$$

Sous les conditions classiques de convexité, de continuité et de qualification des contraintes, la résolution de (4.1) se fait en maxi-minimisant le Lagrangien.

1. On se reportera à l'annexe du §A pour toutes les notions relatives à la théorie de la dualité. On rappelle que le cône (positif) dual du cône C est défini par : $C^\star = \big\{p \in \mathbb{V}, \text{ tel que } \langle p \,, v \rangle \geq 0 \;\; \forall v \in C\big\}$.

4.1.1 Algorithmes d'Uzawa et d'Arrow-Hurwicz

La manière la plus classique d'effectuer la maxi-minimisation est de mettre en œuvre l'algorithme d'*Uzawa*, qui, à chaque itération, consiste à alterner une étape de *minimisation complète* du Lagrangien en u et une étape de type *pas de gradient* pour la maximisation en p. La k-ème itération de l'algorithme d'Uzawa consiste donc, connaissant le couple $(u^{(k)}, p^{(k)})$, à calculer :

$$u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} J(u) + \left\langle p^{(k)} , \Theta(u) \right\rangle , \tag{4.2a}$$

$$p^{(k+1)} = \mathrm{proj}_{C^\star}\big(p^{(k)} + \rho\, \Theta(u^{(k+1)})\big) . \tag{4.2b}$$

Une autre façon de procéder, connue sous le nom d'algorithme de *Arrow-Hurwicz*, consiste à chaque itération à alterner un pas de gradient pour la minimisation en u *et* un pas de gradient pour la maximisation en p. Une itération de l'algorithme d'Arrow-Hurwicz revient, connaissant le couple $(u^{(k)}, p^{(k)})$, à calculer :

$$u^{(k+1)} = \mathrm{proj}_{U^{\mathrm{ad}}}\Big(u^{(k)} - \epsilon\big(\nabla J(u^{(k)}) + \big(\Theta'(u^{(k)})\big)^\top \cdot p^{(k)}\big)\Big) , \tag{4.3a}$$

$$p^{(k+1)} = \mathrm{proj}_{C^\star}\big(p^{(k)} + \rho\, \Theta(u^{(k+1)})\big) . \tag{4.3b}$$

4.1.2 Principe du Problème Auxiliaire

La généralisation de la décomposition par les prix par le Principe du Problème Auxiliaire a été présentée au §B.2. Elle consiste à choisir une fonction K définie sur l'espace $\mathbb{U}$ à valeurs dans $\mathbb{R}$, et à remplacer la résolution du problème (4.1) par la résolution d'une suite de problèmes auxiliaires indexés par k. Le k-ème problème auxiliaire comporte deux étapes, l'une de minimisation en u et l'autre de maximisation en p. La résolution de ce problème auxiliaire constitue la k-ème itération de l'algorithme du PPA qui s'écrit :

$$u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} K(u) + \left\langle \epsilon\nabla J(u^{(k)}) - \nabla K(u^{(k)}) , u \right\rangle + \epsilon\left\langle p^{(k)} , \Theta'(u^{(k)}) \cdot u \right\rangle , \tag{4.4a}$$

$$p^{(k+1)} = \mathrm{proj}_{C^\star}\big(p^{(k)} + \rho\, \Theta(u^{(k+1)})\big) . \tag{4.4b}$$

Remarque 4.2. Dans la phase de minimisation en u de cet algorithme, on peut remplacer l'approximation du premier ordre $\left\langle p^{(k)}, \Theta'(u^{(k)}) \cdot u \right\rangle$ par le terme (a priori non linéaire) $\left\langle p^{(k)} , \Theta(u) \right\rangle$, ce qui conduit à une étape de minimisation en u de la forme :

$$u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} K(u) + \left\langle \epsilon\nabla J(u^{(k)}) - \nabla K(u^{(k)}) , u \right\rangle + \epsilon\left\langle p^{(k)} , \Theta(u) \right\rangle . \tag{4.4c}$$

On a vu au §B.2 que l'application du PPA dans ce cadre permettait, d'une part de retrouver les algorithmes d'Uzawa (problème (4.4c) avec les choix $K(u) = J(u)$ et $\epsilon = 1$), et d'Arrow-Hurwicz (problème (4.4a) avec le choix $K(u) = \|u\|^2/2$), et d'autre part de décomposer l'étape de minimisation en u de chaque problème auxiliaire en choisissant un noyau K additif par rapport à la décomposition de l'espace $\mathbb{U}$.

4.2 Extension stochastique de l'algorithme d'Uzawa ?

Une première tentative pour appliquer l'idée du gradient stochastique au problème (4.1) est, dans le cadre de l'algorithme d'Uzawa, de remplacer dans (4.2) la valeur $J(u)$ par la valeur $j(u, w^{(k+1)})$, où $w^{(k+1)}$ est un tirage de la variable aléatoire $\boldsymbol{W}$. Cette manière de faire conduit à l'algorithme suivant.

Algorithme 4.3. (Algorithme d'Uzawa stochastique)

1. Choisir un couple de points initiaux $(u^{(0)}, p^{(0)}) \in U^{\mathrm{ad}} \times C^{\star}$, ainsi qu'une suite $\{\rho^{(k)}\}_{k\in\mathbb{N}}$ de réels positifs.
2. À l'itération k, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire $\boldsymbol{W}$.
3. Calculer $u^{(k+1)} \in \arg\min_{u\in U^{\mathrm{ad}}} j(u, w^{(k+1)}) + \langle p^{(k)} , \Theta(u)\rangle$.
4. Calculer $p^{(k+1)} = \mathrm{proj}_{C^{\star}}\big(p^{(k)} + \rho^{(k)}\, \Theta(u^{(k+1)})\big)$.
5. Incrémenter l'indice k de 1 et retourner à l'étape 2.

Plutôt que de mener l'étude théorique de convergence de cet algorithme, on va montrer sur un exemple simple que l'algorithme proposé ne permet pas d'obtenir la solution du problème de départ. On s'appuiera pour cela sur le résultat (Kushner et Clark, 1978, Theorem 2.3.1) qui, dans le cadre de l'approximation stochastique étudié au §2.4, établit les liens entre la convergence de l'algorithme :

$$\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)} + \epsilon^{(k)}\Big(h(\boldsymbol{U}^{(k)}) + \boldsymbol{\xi}^{(k+1)}\Big) , \tag{4.5}$$

et le comportement asymptotique de l'équation différentielle ordinaire :

$$\overset{\bullet}{u} = h(u) . \tag{4.6}$$

Plus précisément, on utilisera le théorème suivant (Benveniste et collab., 1990, Chapter 2, Theorem 7).

Théorème 4.4. *Soit $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ la suite générée par l'algorithme (4.5). On suppose qu'il existe un point $u^{\sharp} \in \mathbb{U}$ tel que :*

$$\mathbb{P}\big(\lim_{k\to+\infty} \boldsymbol{U}^{(k)} = u^{\sharp}\big) > 0 .$$

Alors, le point $u^{\sharp}$ est un point d'équilibre stable de l'équation différentielle ordinaire (4.6).

En d'autres termes, ce résultat indique qu'un algorithme de type (4.5) ne peut converger (s'il converge...) que vers un point d'équilibre stable de l'équation différentielle ordinaire associée.

Contre-exemple.

On considère le problème défini de la manière suivante.
- $\mathbb{U} = \mathbb{R}^2$ et $U^{\mathrm{ad}} = \mathbb{U}$.
- $\mathbb{V} = \mathbb{R}$ et $C = \{0\}$ (contrainte égalité).
- $\mathbb{W} = \mathbb{R}^4$ et $w = (a_1, a_2, b_1, b_2)$.
- $j(u, w) = \frac{1}{2}\big(a_1 {u_1}^2 + a_2 {u_2}^2\big) + \big(b_1 u_1 + b_2 u_2\big)$.
- $\Theta(u) = \theta_1 u_1 + \theta_2 u_2$.

On suppose que les variables aléatoires $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{B}_1$ et $\boldsymbol{B}_2$ sont intégrables, que $\boldsymbol{A}_1$ et $\boldsymbol{A}_2$ sont strictement positives, et que θ_1 et θ_1 sont des constantes réelles non nulles. Le problème à résoudre est :

$$\min_{u \in \mathbb{R}^2} \mathbb{E}\left(j(u, \boldsymbol{W})\right) \qquad \text{sous} \qquad \Theta(u) = 0 \, .$$

1. *Résolution du problème déterministe.*
   Il n'y a aucune difficulté sur ce problème à calculer les espérances et résoudre le problème déterministe associé. Notant $(u_1^\sharp, u_2^\sharp)$ la solution primale et $p^\sharp$ le multiplicateur optimal associé à la contrainte, les conditions d'optimalité du problème s'écrivent :
   $$\begin{aligned} &\mathbb{E}\left(\boldsymbol{A}_1\right) u_1^\sharp + \mathbb{E}\left(\boldsymbol{B}_1\right) + \theta_1 p^\sharp = 0 \, , \\ &\mathbb{E}\left(\boldsymbol{A}_2\right) u_2^\sharp + \mathbb{E}\left(\boldsymbol{B}_2\right) + \theta_2 p^\sharp = 0 \, , \\ &\theta_1 u_1^\sharp + \theta_2 u_2^\sharp = 0 \, . \end{aligned}$$
   On en déduit que la valeur du multiplicateur optimale est :
   $$p^\sharp = - \frac{\theta_1 \dfrac{\mathbb{E}\left(\boldsymbol{B}_1\right)}{\mathbb{E}\left(\boldsymbol{A}_1\right)} + \theta_2 \dfrac{\mathbb{E}\left(\boldsymbol{B}_2\right)}{\mathbb{E}\left(\boldsymbol{A}_2\right)}}{\dfrac{\theta_1^2}{\mathbb{E}\left(\boldsymbol{A}_1\right)} + \dfrac{\theta_2^2}{\mathbb{E}\left(\boldsymbol{A}_2\right)}} \tag{4.7}$$
2. *Mise en œuvre de l'algorithme d'Uzawa stochastique.*
   La k-ème itération de l'algorithme 4.3 revient à résoudre le système :
   $$\begin{aligned} &a_1^{(k+1)} u_1^{(k+1)} + b_1^{(k+1)} + \theta_1 p^{(k)} = 0 \, , \\ &a_2^{(k+1)} u_2^{(k+1)} + b_2^{(k+1)} + \theta_2 p^{(k)} = 0 \, , \end{aligned}$$
   et à mettre à jour la valeur $p^{(k+1)}$ du multiplicateur par la relation :
   $$p^{(k+1)} = p^{(k)} + \rho^{(k)} \Big(\theta_1 u_1^{(k+1)} + \theta_2 u_2^{(k+1)}\Big) \, .$$

Les deux premières relations permettent de calculer $u_1^{(k+1)}$ et $u_2^{(k+1)}$ en fonction de $p^{(k)}$. Reportant ces valeurs dans la troisième relation, on en déduit l'équation de récurrence stochastique décrivant l'évolution du multiplicateur au cours des itérations :

$$\boldsymbol{P}^{(k+1)} = \boldsymbol{P}^{(k)} - \rho^{(k)}\bigg(\Big(\frac{\theta_1^2}{\boldsymbol{A}_1^{(k+1)}} + \frac{\theta_2^2}{\boldsymbol{A}_2^{(k+1)}}\Big)\boldsymbol{P}^{(k)} - \Big(\theta_1\frac{\boldsymbol{B}_1^{(k+1)}}{\boldsymbol{A}_1^{(k+1)}} + \theta_2\frac{\boldsymbol{B}_2^{(k+1)}}{\boldsymbol{A}_2^{(k+1)}}\Big)\bigg).$$

Utilisant le théorème 4.4, on sait que cette équation de récurrence stochastique ne peut converger (si elle converge. . .) que vers l'unique point $p^\star$ d'équilibre stable de l'équation différentielle ordinaire associée, à savoir :

$$p^\sharp = -\frac{\theta_1\mathbb{E}\left(\dfrac{\boldsymbol{B}_1}{\boldsymbol{A}_1}\right) + \theta_2\mathbb{E}\left(\dfrac{\boldsymbol{B}_2}{\boldsymbol{A}_2}\right)}{\dfrac{\theta_1^2}{\mathbb{E}\left(\boldsymbol{A}_1\right)} + \dfrac{\theta_2^2}{\mathbb{E}\left(\boldsymbol{A}_2\right)}}\ . \tag{4.8}$$

Si l'on compare alors les relations (4.7) et (4.8), on constate que,dès que les variables aléatoires $\boldsymbol{A}_1$ et $\boldsymbol{B}_1$ ne sont pas *indépendantes*, on peut avoir $\mathbb{E}\left(\frac{\boldsymbol{B}_1}{\boldsymbol{A}_1}\right) \neq \frac{\mathbb{E}(\boldsymbol{B}_1)}{\mathbb{E}(\boldsymbol{A}_1)}$, et donc $p^\sharp \neq p^\star$.

Conclusion.

Cet exemple montre que, même si l'algorithme 4.3 converge (ce que l'on n'a pas cherché à prouver. . .), il ne fournit de toute façon pas toujours la solution du problème initial. L'exemple choisi (linéaire-quadratique) est suffisamment simple et générique pour pouvoir exclure l'algorithme de la panoplie permettant de résoudre des problèmes de type (4.1), et donc pour tirer la conclusion suivante.

L'algorithme d'Uzawa ne peut pas être étendu au cas stochastique.

Si l'on compare l'algorithme de gradient stochastique standard à cette tentative d'extension de l'algorithme d'Uzawa au cadre stochastique, on constate que l'on a tenté de remplacer une évaluation de J par une évaluation de j, alors que le gradient stochastique consiste à remplacer une évaluation du *gradient* de J par une évaluation du *gradient* de j, et qu'on a donc utilisé des approximations différentes. De plus, alors que l'itérée $u^{(k+1)}$ de l'algorithme de gradient stochastique diffère de $u^{(k)}$ par une formule faisant intervenir un pas $\epsilon^{(k)}$ tendant vers zéro, l'itérée $u^{(k+1)}$ de l'algorithme d'Uzawa résulte d'une minimisation ne dépendant pas directement de $u^{(k)}$ et basée sur un tirage de l'aléa indépendant des tirages précédents. Contrairement au cas du gradient stochastique, il n'y a donc aucun *effet de moyenne* dans l'extension d'Uzawa, d'où l'échec.

4.3 Extension stochastique de l'algorithme issu du PPA

L'extension stochastique de l'algorithme issu du principe du problème auxiliaire consiste à remplacer dans les relations (4.4) le gradient $\nabla J(u^{(k)})$ (correspondant à une espérance portant sur la variable aléatoire $\boldsymbol{W}$) par le terme $\nabla_u j(u^{(k)}, w^{(k+1)})$ (correspondant à un tirage de la variable $\boldsymbol{W}$) ; on substitue alors à la résolution du problème (4.1) la résolution de la suite de problèmes auxiliaires :

$$u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} K(u) + \left\langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}) \, , u \right\rangle + \epsilon^{(k)} \left\langle p^{(k)} \, , \Theta(u) \right\rangle \, , \tag{4.9a}$$

$$p^{(k+1)} = \mathrm{proj}_{C^\star} \left(p^{(k)} + \epsilon^{(k)} \, \Theta(u^{(k+1)}) \right) . \tag{4.9b}$$

Remarque 4.5. Comme dans le cas déterministe, il est possible de remplacer dans la phase de minimisation en u le terme $\left\langle p^{(k)} \, , \Theta(u) \right\rangle$ par son approximation linéaire $\left\langle p^{(k)} \, , \Theta'(u^{(k)}) u \right\rangle$. L'étape de minimisation en u dans le problème (4.9) s'écrirait alors :

$$u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} K(u) + \left\langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}) \, , u \right\rangle + \epsilon^{(k)} \left\langle p^{(k)} \, , \Theta'(u^{(k)}) \cdot u \right\rangle . \tag{4.9c}$$

On propose alors l'algorithme suivant comme extension de l'algorithme issu du PPA au cadre stochastique sous contrainte déterministe.

Algorithme 4.6. (Décomposition par les prix et PPA stochastique)

1. Choisir $(u^{(0)}, p^{(0)}) \in U^{\mathrm{ad}} \times C^\star$, et une suite $\left\{ \epsilon^{(k)} \right\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l'itération k, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire $\boldsymbol{W}$.
3. Calculer $u^{(k+1)}$ solution du problème auxiliaire (4.9a).
4. Calculer $p^{(k+1)}$ par la formule de mise à jour (4.9b).
5. Incrémenter l'indice k de 1 et retourner à l'étape 2.

On fait sur cet algorithme les commentaires suivants.

– Avec le choix de noyau $K(u) = \|u\|^2 / 2$, et utilisant la forme (4.9c) linéarisée en Θ, la minimisation en u dans le problème auxiliaire (4.9) se met sous la forme :

$$\min_{u \in U^{\mathrm{ad}}} \frac{1}{2} \|u\|^2 - \left\langle u^{(k)} - \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \epsilon^{(k)} \left(\Theta'(u^{(k)}) \right)^\top p^{(k)} \, , u \right\rangle ,$$

dont la solution $u^{(k+1)}$ se calcule explicitement :

$$u^{(k+1)} = \mathrm{proj}_{U^{\mathrm{ad}}} \left(u^{(k)} - \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \epsilon^{(k)} \left(\Theta'(u^{(k)}) \right)^\top p^{(k)} \right) ,$$

et correspond exactement à un pas de l'extension « naturelle » au cas stochastique de l'algorithme d'Arrow-Hurwicz déterministe (comparer avec les relations (4.3)).

– Cet algorithme ne peut *jamais* correspondre à celui d'Uzawa car, comme on le verra par la suite, les pas $\epsilon^{(k)}$ doivent tendre vers 0 de telle sorte que prendre un pas constant (égal à 1) comme dans l'algorithme d'Uzawa n'est pas envisageable.

On regarde enfin les propriétés de décomposition de cet algorithme. Utilisant toujours la forme linéarisée en Θ dans (4.9), supposant que l'espace $\mathbb{U}$ et l'ensemble admissible U^{ad} se mettent sous forme de produits cartésiens $\mathbb{U}_1 \times \ldots \times \mathbb{U}_N$ et $U_1^{\mathrm{ad}} \times \ldots \times U_N^{\mathrm{ad}}$, avec pour tout i l'inclusion $U_i^{\mathrm{ad}} \subset \mathbb{U}_i$, et en choisissant le noyau K sous forme additive par rapport à cette décomposition :

$$K(u) = \sum_{i=1}^{N} K_i(u_i) \ , \quad \text{avec} \quad u_i \in \mathbb{U}_i \ ,$$

l'étape de minimisation en u s'écrit :

$$\min_{(u_1,\ldots,u_N)\in U^{\mathrm{ad}}} \ \sum_{i=1}^{N} \Big(K_i(u_i) + \big\langle \epsilon^{(k)} \nabla_{u_i} j(u^{(k)}, w^{(k+1)}) - \nabla K_i(u_i^{(k)}) \ , u_i \big\rangle \\ + \epsilon^{(k)} \big\langle \big(\Theta'_{u_i}(u^{(k)})\big)^{\top} p^{(k)} \ , u_i \big\rangle \Big) \ ,$$

qui se décompose en N sous-problèmes indépendants dont la i-ème instance s'écrit :

$$\min_{u_i \in U_i^{\mathrm{ad}}} \ K_i(u_i) + \big\langle \epsilon^{(k)} \nabla_{u_i} j(u^{(k)}, w^{(k+1)}) - \nabla K_i(u_i^{(k)}) \ , u_i \big\rangle \\ + \epsilon^{(k)} \big\langle \big(\Theta'_{u_i}(u^{(k)})\big)^{\top} p^{(k)} \ , u_i \big\rangle \ .$$

Remarque 4.7. Dans le cas où la fonction Θ est aussi additive par rapport à la décomposition de l'espace :

$$\Theta(u) = \sum_{i=1}^{N} \Theta_i(u_i) \ , \quad \text{avec} \quad u_i \in \mathbb{U}_i \ ,$$

le produit scalaire $\big\langle \big(\Theta'_{u_i}(u^{(k)})\big)^{\top} p^{(k)} \ , u_i \big\rangle$ apparaissant dans le i-ème sous-problème peut être remplacé par le terme $\big\langle p^{(k)} \ , \Theta_i(u_i) \big\rangle$.

On va dans le paragraphe suivant donner un résultat de convergence de cet algorithme dans le cas où le Lagrangien du problème déterministe est *stable*[2]. On donnera ensuite un résultat adapté au cas où la fonction J est *fortement convexe*, ce qui permettra d'utiliser des « grands pas » dans l'étape de remise à jour du multiplicateur. On s'intéressera enfin au cas où la fonction J est *simplement convexe*, pour lequel on donnera un algorithme basé sur l'utilisation du Lagrangien augmenté.

2. Voir le commentaire A.2.2 page 113 pour la notion de stabilité du Lagrangien.

4.4 Théorème de convergence

Pour étudier la convergence de l'algorithme 4.6, on le formule en terme de variables aléatoires. Pour cela, on considère un échantillon $\{\boldsymbol{W}^{(k)}\}_{k\in\mathbb{N}}$ de taille infinie de la variable aléatoire $\boldsymbol{W}$, et les étapes 3 et 4 de l'algorithme 4.6 prennent la forme :

$$\begin{aligned}\min_{u\in U^{\mathrm{ad}}} K(u) + \left\langle \epsilon^{(k)}\nabla_u j(\boldsymbol{U}^{(k)}, \boldsymbol{W}^{(k+1)}) - \nabla K(\boldsymbol{U}^{(k)})\,, u\right\rangle \\ + \epsilon^{(k)}\left\langle \boldsymbol{P}^{(k)}\,, \Theta(u)\right\rangle\,, \qquad (4.10a)\end{aligned}$$

$$\boldsymbol{P}^{(k+1)} = \mathrm{proj}_{C^\star}\left(\boldsymbol{P}^{(k)} + \epsilon^{(k)}\,\Theta(\boldsymbol{U}^{(k+1)})\right)\,. \qquad (4.10b)$$

Il est sous-entendu que la minimisation dans (4.10a) ainsi que la projection dans (4.10b) sont effectuées ω par ω. Le résultat de la minimisation (4.10a) dépend de ω et est noté $\boldsymbol{U}^{(k+1)}$.

On se place dans le cas où le Lagrangien du problème (4.1) est stable. La question de la convergence de l'algorithme 4.6 est réglée par le résultat suivant.

Théorème 4.8.
On suppose que les hypothèses suivantes sont vérifiées.

1. *U^{ad} est une partie convexe fermée non vide d'un l'espace de Hilbert $\mathbb{U}$, et C est un cône convexe fermé saillant*[3] *d'un autre espace de Hilbert $\mathbb{V}$.*
2. *La fonction $j : \mathbb{U}\times\mathbb{W}\to\mathbb{R}$ est une intégrande normale, et l'espérance de $j(u, \boldsymbol{W})$ existe pour tout $u\in U^{\mathrm{ad}}$.*
3. *Pour tout $w\in\mathbb{W}$, la fonction $j(\cdot, w) : \mathbb{U}\to\mathbb{R}$ est propre, convexe, semi-continue inférieurement, et est différentiable sur un sous-ensemble ouvert contenant U^{ad} et son gradient partiel par rapport à u est noté $\nabla_u j(u, w)$.*
4. *La fonction $j(\cdot, w)$ est à gradient linéairement borné (GLB), uniformément en w :*

$$\exists c_1 > 0,\ \exists c_2 > 0,\ \forall w\in\mathbb{W},\ \forall u\in U^{\mathrm{ad}}\,,\quad \|\nabla_u j(u,w)\| \le c_1\,\|u\| + c_2\,.$$

5. *La fonction J est Lipschitzienne, coercive sur l'ensemble U^{ad}.*
6. *La fonction Θ est C-convexe, Lipschitzienne de rapport L_Θ.*
7. *Les contraintes sont qualifiées et le Lagrangien L est stable.*
8. *La fonction K est propre, fortement convexe de module b, semi-continue inférieurement, et est différentiable sur un sous-ensemble ouvert contenant U^{ad}.*
9. *La suite $\left\{\epsilon^{(k)}\right\}_{k\in\mathbb{N}}$ est une σ-suite.*

On a alors les conclusions suivantes.

3. ce qui signifie que $C\cap -C = \{0\}$

1. *Le problème (4.1) admet un ensemble de points selle $U^\sharp \times P^\sharp$ non vide.*
2. *Le problème (4.10a) admet une solution $\boldsymbol{U}^{(k+1)}$ unique.*
3. *Pour tout $p^\sharp \in P^\sharp$, la suite de variables aléatoires $\{L(\boldsymbol{U}^{(k)}, p^\sharp)\}_{k\in\mathbb{N}}$ converge presque sûrement vers $L(u^\sharp, p^\sharp)$, avec $u^\sharp \in U^\sharp$.*
4. *Les suites $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ et $\{\boldsymbol{P}^{(k)}\}_{k\in\mathbb{N}}$ engendrées par l'algorithme 4.6 sont bornées presque sûrement, et tout point d'accumulation d'une réalisation de la suite $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ appartient à $U^\sharp$, ensemble des solutions du problème.*

Remarque 4.9. L'hypothèse de stabilité du Lagrangien n'est en pratique pas aisée à vérifier. On peut la remplacer par la condition (plus forte) de stricte convexité de la fonction J. Il faut noter que sans l'hypothèse de stabilité, il y a peu de chances de pouvoir dire quoi que ce soit des limites de la suite de variables aléatoires $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$.

Preuve. La démonstration des 2 premières conclusions du théorème découle des théorèmes généraux relatifs à l'optimisation convexe sous contraintes. Le fait que la solution $\boldsymbol{U}^{(k+1)}$ du problème (4.10a) soit une variable aléatoire, et donc une fonction mesurable, provient de ce que l'on a supposé que j était une intégrande normale (voir le théorème 3.13 de l'annexe 3.5, ainsi que le raisonnement fait au début de la preuve du théorème 3.3). La démonstration des 2 dernières conclusions se fait en suivant le déroulement « classique » en quatre étapes d'un algorithme d'optimisation, à savoir :

1. choix d'une fonction de Lyapunov opérant sur $(u^{(k)}, p^{(k)})$,
2. majoration de sa variation d'une itération de l'algorithme sur l'autre,
3. convergence de l'algorithme, à l'aide d'un argument de type martingale,
4. analyse des limites des suites permettant de caractériser la solution.

Pour alléger les écritures, on utilisera la notation :

$$g^{(k)} = \nabla_u j\big(u^{(k)}, w^{(k+1)}\big) \ .$$

Le fait que $u^{(k+1)}$ soit solution du problème (4.9a) est caractérisé par la condition d'optimalité suivante :

$$\begin{aligned}\forall u \in U^{\mathrm{ad}},\ \big\langle \nabla K(u^{(k+1)}) - \nabla K(u^{(k)}) + \epsilon^{(k)} g^{(k)} \,, u - u^{(k+1)} \big\rangle \\ + \epsilon^{(k)} \big\langle p^{(k)} \,, \Theta(u) - \Theta(u^{(k+1)}) \big\rangle \geq 0 \ . \quad (4.11)\end{aligned}$$

1. **Choix de la fonction de Lyapunov**. Soit $(u^\sharp, p^\sharp) \in U^\sharp \times P^\sharp$ un point selle du problème (4.1). On choisit la fonction de Lyapunov Λ de la forme :

$$\Lambda(u,p) = K(u^\sharp) - K(u) - \big\langle \nabla K(u) \,, u^\sharp - u \big\rangle + \frac{1}{2}\big\|p - p^\sharp\big\|^2 \ ,$$

et l'on note :

$$\psi^{(k)} = \Lambda(u^{(k)}, p^{(k)}) \ .$$

De la forte convexité de K et de la définition de Λ, on déduit que Λ est bornée inférieurement et coercive :

$$\left\|u^{(k)} - u^\sharp\right\|^2 \leq \frac{2}{b}\,\psi^{(k)}\,, \tag{4.12a}$$

$$\left\|p^{(k)} - p^\sharp\right\|^2 \leq 2\,\psi^{(k)}\,. \tag{4.12b}$$

2. **Majorations**.
   a. On majore pour commencer la variation $\left\|u^{(k+1)} - u^{(k)}\right\|$. Évaluant la condition d'optimalité (4.11) au point $u = u^{(k)}$:

$$\begin{aligned}\left\langle \nabla K(u^{(k)}) - \nabla K(u^{(k+1)})\,, u^{(k)} - u^{(k+1)}\right\rangle &\leq \left\langle \epsilon^{(k)} g^{(k)}\,, u^{(k)} - u^{(k+1)}\right\rangle \\ &\quad + \epsilon^{(k)}\left\langle p^{(k)}\,, \Theta(u^{(k)}) - \Theta(u^{(k+1)})\right\rangle .\end{aligned}$$

   Utilisant la forte convexité de K, l'inégalité de Schwartz et le caractère lipschitzien de Θ, on obtient :

$$\begin{aligned}b\left\|u^{(k+1)} - u^{(k)}\right\|^2 &\leq \epsilon^{(k)}\left\|g^{(k)}\right\|\left\|u^{(k)} - u^{(k+1)}\right\| \\ &\quad + \epsilon^{(k)}\left\|p^{(k)}\right\| L_\Theta \left\|u^{(k)} - u^{(k+1)}\right\|\,,\end{aligned}$$

   et donc :

$$b\left\|u^{(k+1)} - u^{(k)}\right\| \leq \epsilon^{(k)}\Big(\left\|g^{(k)}\right\| + L_\Theta\left\|p^{(k)}\right\|\Big)\,.$$

   L'hypothèse GLB et les majorations (4.12) impliquent l'existence de constantes c_3 et c_4 telles que :

$$\left\|u^{(k+1)} - u^{(k)}\right\| \leq \epsilon^{(k)}\Big(c_3\sqrt{\psi^{(k)}} + c_4\Big)\,. \tag{4.13}$$

   b. On majore ensuite la variation $\left\|p^{(k+1)} - p^\sharp\right\|^2$. À l'aide de la relation $p^\sharp = \mathrm{proj}_{C^\star}\left(p^\sharp + \epsilon^{(k)}\Theta(u^\sharp)\right)$ [4], de la relation (4.9b) donnant $p^{(k+1)}$, et comme l'opérateur de projection sur $C^\star$ est contractant, on obtient :

$$\left\|p^{(k+1)} - p^\sharp\right\| \leq \left\|p^{(k)} - p^\sharp + \epsilon^{(k)}\big(\Theta(u^{(k+1)}) - \Theta(u^\sharp)\big)\right\|\,.$$

   Élevant cette inégalité au carré, développant le terme de droite et utilisant le fait que Θ est Lipschitzienne, il vient :

$$\begin{aligned}\left\|p^{(k+1)} - p^\sharp\right\|^2 &\leq \left\|p^{(k)} - p^\sharp\right\|^2 + \left(\epsilon^{(k)} L_\Theta\right)^2\left\|u^{(k+1)} - u^\sharp\right\|^2 \\ &\quad + 2\epsilon^{(k)}\left\langle p^{(k)} - p^\sharp\,, \Theta(u^{(k+1)}) - \Theta(u^\sharp)\right\rangle .\end{aligned} \tag{4.14}$$

4. Cette relation, vraie pour tout $\epsilon^{(k)} > 0$, s'obtient en écrivant la condition d'optimalité du problème de projection sur le cône $C^\star$ et en constatant que cette condition est identique à l'inégalité de gauche du point selle associée au problème (4.1).

c. On majore pour finir la variation $\psi^{(k+1)} - \psi^{(k)}$. Formant cette différence, utilisant le fait que la fonction K est convexe et que donc $K(u^{(k)}) - K(u^{(k+1)}) + \big\langle \nabla K(u^{(k)}) \,, u^{(k+1)} - u^{(k)} \big\rangle \leq 0$, ainsi que la relation (4.14), on obtient :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} \leq\ & \big\langle \nabla K(u^{(k)}) - \nabla K(u^{(k+1)}) \,, u^\sharp - u^{(k+1)} \big\rangle \\ & + \epsilon^{(k)} \big\langle p^{(k)} - p^\sharp \,, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \big\rangle \\ & + \frac{\big(\epsilon^{(k)} L_\Theta\big)^2}{2} \big\| u^{(k+1)} - u^\sharp \big\|^2 \,. \end{aligned}$$

Par la condition d'optimalité (4.11) évaluée au point $u = u^\sharp$, on obtient :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} \leq & +\epsilon^{(k)} \big\langle g^{(k)} \,, u^\sharp - u^{(k+1)} \big\rangle \\ & + \epsilon^{(k)} \big\langle p^\sharp \,, \Theta(u^\sharp) - \Theta(u^{(k+1)}) \big\rangle \\ & + \frac{\big(\epsilon^{(k)} L_\Theta\big)^2}{2} \big\| u^{(k+1)} - u^\sharp \big\|^2 \,. \end{aligned}$$

Cette dernière inégalité se met sous la forme :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} \leq\ & \epsilon^{(k)} \big\langle g^{(k)} \,, u^\sharp - u^{(k)} \big\rangle + \epsilon^{(k)} \big\langle p^\sharp \,, \Theta(u^\sharp) - \Theta(u^{(k)}) \big\rangle \\ & + \epsilon^{(k)} \big\| u^{(k+1)} - u^{(k)} \big\| \big(\big\| g^{(k)} \big\| + L_\Theta \big\| p^\sharp \big\| \big) \\ & + \frac{\big(\epsilon^{(k)} L_\Theta\big)^2}{2} \big\| u^{(k+1)} - u^\sharp \big\|^2 \,. \end{aligned}$$

Utilisant (4.12), (4.13) et l'hypothèse GLB, on en déduit l'existence de constantes c_5, c_6 et c_7 telles que :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} \leq & \ \epsilon^{(k)} \big\langle g^{(k)} \,, u^\sharp - u^{(k)} \big\rangle + \epsilon^{(k)} \big\langle p^\sharp \,, \Theta(u^\sharp) - \Theta(u^{(k)}) \big\rangle \\ & + (\epsilon^{(k)})^2 \big(c_5 \psi^{(k)} + c_6 + c_7 \psi^{(k+1)} \big) \,. \end{aligned}$$

Cette dernière relation s'écrit en terme de variables aléatoires :

$$\begin{aligned} \boldsymbol{\Psi}^{(k+1)} - \boldsymbol{\Psi}^{(k)} \leq & \ \epsilon^{(k)} \big\langle \boldsymbol{G}^{(k)} \,, u^\sharp - \boldsymbol{U}^{(k)} \big\rangle + \epsilon^{(k)} \big\langle p^\sharp \,, \Theta(u^\sharp) - \Theta(\boldsymbol{U}^{(k)}) \big\rangle \\ & + (\epsilon^{(k)})^2 \big(c_5 \boldsymbol{\Psi}^{(k)} + c_6 + c_7 \boldsymbol{\Psi}^{(k+1)} \big) \,, \end{aligned} \tag{4.15}$$

On rappelle que $\boldsymbol{W}^{(k+1)}$ est indépendante des $\boldsymbol{W}^{(l)}$ pour $l \leq k$, et que $\boldsymbol{U}^{(k)}$ et $\boldsymbol{\Psi}^{(k)}$ sont par construction des variables aléatoires mesurables par rapport à la tribu $\mathcal{F}^{(k)}$. On en déduit que :

$$\mathbb{E}\big(\big\langle \boldsymbol{G}^{(k)} \,, u^\sharp - \boldsymbol{U}^{(k)} \big\rangle \ \big|\ \mathcal{F}^{(k)} \big) = \big\langle \nabla J(\boldsymbol{U}^{(k)}) \,, u^\sharp - \boldsymbol{U}^{(k)} \big\rangle \leq J(u^\sharp) - J(\boldsymbol{U}^{(k)}) \,,$$

la dernière inégalité provenant de la convexité de J. Prenant de part et d'autre de l'inégalité (4.15) l'espérance conditionnelle par rapport à la tribu $\mathcal{F}^{(k)}$ engendrée par les k variables aléatoires $(\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(k)})$, on obtient :

$$\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq \alpha^{(k)}\boldsymbol{\Psi}^{(k)} + \beta^{(k)} + \gamma^{(k)}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big) + \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \ , \quad (4.16)$$

où $\alpha^{(k)}$, $\beta^{(k)}$ et $\gamma^{(k)}$ sont les termes de trois séries convergentes. Le terme $L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)$ est toujours négatif ou nul (inégalité de droite caractérisant le point selle $(u^\sharp, p^\sharp)$ de L).

3. **Analyse de convergence.** Une application directe du corollaire 3.6 du théorème de Robbins-Siegmund permet de montrer que la suite de variables aléatoires $\{\boldsymbol{\Psi}^{(k)}\}_{k\in\mathbb{N}}$ converge presque sûrement vers une variable aléatoire $\boldsymbol{\Psi}^\infty$ bornée presque sûrement, et que l'on a :

$$\sum_{k=0}^{+\infty} \epsilon^{(k)}\big(L(\boldsymbol{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp)\big) < +\infty \ , \quad \mathbb{P}\text{-p.s.} \ . \qquad (4.17)$$

4. **Limites des suites.** Du fait que la suite $\{\boldsymbol{\Psi}^{(k)}\}_{k\in\mathbb{N}}$ est bornée presque sûrement, on déduit de (4.12) que les suites $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ et $\{\boldsymbol{P}^{(k)}\}_{k\in\mathbb{N}}$ sont bornées presque sûrement. Par la relation (4.13), il en est de même pour la suite $\{\|\boldsymbol{U}^{(k+1)} - \boldsymbol{U}^{(k)}\|/\epsilon^{(k)}\}_{k\in\mathbb{N}}$. Cette dernière propriété, associée à la relation (4.17) et au fait que les fonctions J et Θ soient Lipschitziennes, permet d'utiliser le lemme 3.7 : on en déduit alors que la suite $\{L(\boldsymbol{U}^{(k)}, p^\sharp)\}_{k\in\mathbb{N}}$ converge presque sûrement vers $L(u^\sharp, p^\sharp)$.
   On note alors Ω_0 le sous-ensemble (de mesure nulle) de Ω sur lequel la suite $\{\boldsymbol{\Psi}^{(k)}\}_{k\in\mathbb{N}}$ n'est pas bornée, et Ω_1 le sous-ensemble (de mesure nulle lui aussi) de Ω sur lequel la relation (4.17) n'est pas vérifiée.
   Soit $\omega \notin \Omega_0 \cup \Omega_1$. La suite des réalisations $\{u^{(k)}\}_{k\in\mathbb{N}}$ associée à cet élément ω est bornée et chaque $u^{(k)}$ appartient à U^{ad}, partie fermée de $\mathbb{U}$. Par un argument de compacité, on conclut que l'on peut extraire de la suite $\{u^{(k)}\}_{k\in\mathbb{N}}$ une sous-suite convergente $\{u^{(\Phi(k))}\}_{k\in\mathbb{N}}$. Soit $\bar{u}$ la limite de la suite $\{u^{(\Phi(k))}\}_{k\in\mathbb{N}}$. La semi-continuité inférieure du Lagrangien L fait que l'on a :

$$L(\bar{u}, p^\sharp) \leq \liminf_{k\to+\infty} L(u^{(\Phi(k))}, p^\sharp) = L(u^\sharp, p^\sharp) \ .$$

   On en déduit que, presque sûrement, $\overline{u}$ est solution du problème de minimisation sur U^{ad} du Lagrangien L pour $p = p^\sharp$ fixé. La stabilité du Lagrangien implique alors que $\overline{u} \in U^\sharp$.

Pour conclure, on notera que dans le cas où la fonction J est strictement convexe, l'ensemble des solutions du problème (4.1) est un singleton :

$$U^\sharp = \{u^\sharp\} \ .$$

Le Lagrangien L est alors stable, et toute réalisation de la suite $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ engendrée par l'algorithme 4.6 a un *unique* point d'accumulation et converge donc toute entière vers $u^\sharp$.

4.5 Cas fortement convexe : utilisation de grands pas

Lorsque la fonction J est *fortement convexe*, on dispose d'une variante intéressante de l'algorithme 4.6 car il est alors possible de remettre à jour les multiplicateurs $p^{(k)}$ avec un pas ρ *constant* (au lieu de pas $\epsilon^{(k)}$ décroissants vers zéro), pourvu que l'on soit capable de localiser les multiplicateurs dans une boule $B(0,R)$ de centre en 0 et de rayon R suffisamment grand pour que les multiplicateurs optimaux $p^\sharp$ se trouvent dans cette boule.

L'intérêt de cette variante est que l'utilisation de « grands pas » ρ permet une convergence plus rapide des variables duales $p^{(k)}$. L'évolution des variables primales $u^{(k)}$ se fait toujours quant à elle avec des « petits pas » $\epsilon^{(k)}$ afin de permettre l'effet de moyenne nécessaire pour obtenir la solution du problème initial. L'algorithme associé à cette variante est le suivant.

Algorithme 4.10. (Algorithme du PPA stochastique à grands pas)

1. Choisir $u^{(0)} \in U^{\mathrm{ad}}$ et $p^{(0)} \in C^\star \cap B(0,R)$, choisir un réel $\rho > 0$ et une suite $\{\epsilon^{(k)}\}_{k\in\mathbb{N}}$ de réels positifs.
2. À l'itération k, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire $\boldsymbol{W}$.
3. Calculer $u^{(k+1)}$ solution du problème auxiliaire (4.9a).
4. Calculer $p^{(k+1)} = \mathrm{proj}_{C^\star\cap B(0,R)}\big(p^{(k)} + \rho\,\Theta(u^{(k+1)})\big)$.
5. Incrémenter l'indice k de 1 et retourner à l'étape 2.

La convergence de cet algorithme est précisée par le théorème suivant.

Théorème 4.11.
En plus des hypothèses du théorème 4.8, on suppose que :
- *la fonction J est fortement convexe*[5] *de module a,*
- *la σ-suite $\{\epsilon^{(k)}\}_{k\in\mathbb{N}}$ est décroissante,*
- *le coefficient ρ est tel que $0 < \rho \leq a/L_\Theta^2$.*

Alors, toute réalisation de la suite $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ engendrée par l'algorithme 4.10 converge presque sûrement vers $u^\sharp$, unique solution du problème (4.1).

Preuve. La démonstration suivant un démarche très similaire à celle du théorème 4.8, on reprend les éléments de cette preuve en indiquant les différences qui apparaissent dans le cadre de cette variante.

5. condition qui implique que J est coercive et que L est stable

1. **Choix de la fonction de Lyapunov**. Soit $(u^\sharp, p^\sharp) \in U^\sharp \times P^\sharp$ un point selle du problème (4.1). On choisit la fonction de Lyapunov de telle sorte que :
$$\psi^{(k)} = K(u^\sharp) - K(u^{(k)}) - \left\langle \nabla K(u^{(k)})\,, u^\sharp - u^{(k)} \right\rangle + \frac{\epsilon^{(k)}}{2\rho} \left\| p^{(k)} - p^\sharp \right\|^2 .$$
Comme dans le théorème 4.8, on a la majoration :
$$\left\| u^{(k)} - u^\sharp \right\|^2 \leq \frac{2}{b}\, \psi^{(k)} . \tag{4.18}$$
Par contre, la majoration $\left\| p^{(k)} - p^\sharp \right\|^2 \leq 2\psi^{(k)}$ n'est plus valide ; c'est pourquoi il faut imposer dans l'algorithme 4.10 le fait que $\left\| p^{(k)} \right\|$ soit majorée par R.
2. **Majorations**.
   a. Afin de majorer la variation $\left\| u^{(k+1)} - u^{(k)} \right\|$, on procède comme dans la preuve du théorème 4.8, et on obtient :
   $$b\left\| u^{(k+1)} - u^{(k)} \right\| \leq \epsilon^{(k)} \left(\left\| g^{(k)} \right\| + L_\Theta \left\| p^{(k)} \right\| \right) .$$
   L'hypothèse GLB, la majoration (4.18) et le fait que la norme du multiplicateur $p^{(k)}$ soit majorée par R impliquent l'existence de constantes positives c_3 et c_4 telles que :
   $$\left\| u^{(k+1)} - u^{(k)} \right\| \leq \epsilon^{(k)} \left(c_3 \sqrt{\psi^{(k)}} + c_4 \right) . \tag{4.19}$$
   b. Pour la majoration de la variation $\left\| p^{(k+1)} - p^\sharp \right\|^2$, on utilise les mêmes arguments que dans le théorème 4.8, d'où :
   $$\begin{aligned} \left\| p^{(k+1)} - p^\sharp \right\|^2 \leq \left\| p^{(k)} - p^\sharp \right\|^2 &+ \left(\rho\, L_\Theta \right)^2 \left\| u^{(k+1)} - u^\sharp \right\|^2 \\ &+ 2\rho \left\langle p^{(k)} - p^\sharp \,, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \right\rangle . \end{aligned}$$
   Multipliant de part et d'autre de cette inégalité par $\epsilon^{(k)}/\rho$, et utilisant le fait que la suite $\left\{ \epsilon^{(k)} \right\}_{k\in\mathbb{N}}$ est décroissante, on obtient :
   $$\begin{aligned} \frac{\epsilon^{(k+1)}}{\rho} \left\| p^{(k+1)} - p^\sharp \right\|^2 \leq \frac{\epsilon^{(k)}}{\rho} \left\| p^{(k)} - p^\sharp \right\|^2 &+ \epsilon^{(k)} \rho L_\Theta^2 \left\| u^{(k+1)} - u^\sharp \right\|^2 \\ &+ 2\epsilon^{(k)} \left\langle p^{(k)} - p^\sharp \,, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \right\rangle . \end{aligned}$$
   c. Enfin, pour la variation $\left(\psi^{(k+1)} - \psi^{(k)} \right)$, comme dans la preuve du théorème 4.8, on a :
   $$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} \leq \epsilon^{(k)} \left\langle g^{(k)} \,, u^\sharp - u^{(k)} \right\rangle &+ \epsilon^{(k)} \left\langle p^\sharp \,, \Theta(u^\sharp) - \Theta(u^{(k)}) \right\rangle \\ &+ \epsilon^{(k)} \left\| u^{(k+1)} - u^{(k)} \right\| \left(\left\| g^{(k)} \right\| + L_\Theta \left\| p^\sharp \right\| \right) \\ &+ \frac{\epsilon^{(k)} \rho L_\Theta^2}{2} \left\| u^{(k+1)} - u^\sharp \right\|^2 . \end{aligned}$$

Utilisant l'hypothèse GLB et la majoration (4.18), et écrivant l'inégalité en termes de variables aléatoires, on en déduit l'existence de constantes positives c_5 et c_6 telles que :

$$\begin{aligned}\boldsymbol{\Psi}^{(k+1)} - \boldsymbol{\Psi}^{(k)} \leq \epsilon^{(k)}\big\langle \boldsymbol{G}^{(k)}\,, u^\sharp - \boldsymbol{U}^{(k)}\big\rangle + \epsilon^{(k)}\big\langle p^\sharp\,, \Theta(u^\sharp) - \Theta(\boldsymbol{U}^{(k)})\big\rangle \\ + (\epsilon^{(k)})^2\big(c_5\boldsymbol{\Psi}^{(k)} + c_6\big) + \frac{\epsilon^{(k)}\rho\, L_\Theta^2}{2}\big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 .\end{aligned}$$

Prenant de part et d'autre de cette inégalité l'espérance conditionnelle par rapport à la tribu $\mathcal{F}^{(k)}$ engendrée par les k variables aléatoires $(\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(k)})$, et utilisant la forte convexité de la fonction J, [6] on obtient :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \bigm| \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ + \frac{\epsilon^{(k)}}{2}\Big(\rho\, L_\Theta^2\big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 - a\big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2\Big) \\ + (\epsilon^{(k)})^2\Big(c_5\boldsymbol{\Psi}^{(k)} + c_6\Big) .\end{aligned}$$

Grâce à la condition $\rho\, L_\Theta^2 \leq a$, on obtient :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \bigm| \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ + \frac{\epsilon^{(k)}a}{2}\Big(\big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 - \big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2\Big) \\ + (\epsilon^{(k)})^2\Big(c_5\boldsymbol{\Psi}^{(k)} + c_6\Big) ,\end{aligned}$$

qui s'écrit encore [7] :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \bigm| \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ + \frac{(\epsilon^{(k)})^2}{2}\Big(\big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 + \big\|(\boldsymbol{U}^{(k)} - u^\sharp\big\|^2\Big) \\ + \frac{a^2}{2}\big\|\boldsymbol{U}^{(k+1)} - \boldsymbol{U}^{(k)}\big\|^2 \\ + (\epsilon^{(k)})^2\Big(c_5\boldsymbol{\Psi}^{(k)} + c_6\Big) .\end{aligned}$$

6. à savoir $J(u^\sharp) \geq J(\boldsymbol{U}^{(k)}) + \big\langle \boldsymbol{G}^{(k)}\,, u^\sharp - \boldsymbol{U}^{(k)}\big\rangle + (a/2)\big\|u^\sharp - \boldsymbol{U}^{(k)}\big\|$
7. Notant $X = \boldsymbol{U}^{(k+1)} - u^\sharp$ et $Y = \boldsymbol{U}^{(k)} - u^\sharp$, on utilise les majorations :

$$\begin{aligned}\epsilon^{(k)}\, a\big(\, \|X\|^2 - \|Y\|^2\,\big) &= \epsilon^{(k)}\, a\,\langle X+Y\,, X-Y\rangle \\ &\leq (\epsilon^{(k)})^2\,\|X+Y\|^2 + a^2\,\|X-Y\|^2 \\ &\leq (\epsilon^{(k)})^2\,\|X\|^2 + (\epsilon^{(k)})^2\,\|Y\|^2 + a^2\,\|X-Y\|^2 \ .\end{aligned}$$

Utilisant (4.18) et (4.19), on en déduit l'existence de constantes positives c_7, c_8 et c_9 telles que :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq{}& \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ &+ (\epsilon^{(k)})^2\Big(c_7\boldsymbol{\Psi}^{(k)} + c_8 + c_9\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big)\Big) \ .\end{aligned}$$

3. **Analyse de convergence**. La suite de la démonstration est identique à celle du théorème 4.8 : l'application directe du corollaire 3.6 montre la suite de variables aléatoires $\{\boldsymbol{\Psi}^{(k)}\}_{k\in\mathbb{N}}$ converge presque sûrement vers une variable aléatoire $\boldsymbol{\Psi}^{\infty}$ bornée presque sûrement, et on a :

$$\sum_{k=0}^{+\infty} \epsilon^{(k)}\big(L(\boldsymbol{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp)\big) < +\infty \ , \quad \mathbb{P}\text{-p.s.} \ .$$

4. **Limites des suites**. Comme dans le théorème 4.8, avec de plus la forte convexité de J et donc l'unicité de la solution $u^\sharp$, on conclut que la suite $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ engendrée par l'algorithme 4.10 est bornée presque sûrement et converge vers $u^\sharp$.

Remarque 4.12. Il est clair que la forte convexité de J est nécessaire pour pouvoir prendre un pas ρ constant dans l'étape de mise à jour des multiplicateurs, car sans cette condition, la fonction duale, tout en restant différentiable, ne serait pas forcément à gradient lipschitzien si bien qu'il faudrait utiliser des « petits pas » pour les mises à jour de $p^{(k)}$.

Remarque 4.13. Une légère modification de la preuve permet de choisir le « grand pas » ρ tel que :

$$0 < \rho < \frac{2a}{L_\Theta^2} \ .$$

En effet, de la condition $\rho < 2a/L_\Theta^2$, on déduit l'existence d'un $\delta > 0$, que l'on peut supposer inférieur à $1/2$, tel que :

$$\rho L_\Theta^2 \leq 2a(1-\delta) \ . \tag{4.20}$$

Utilisant le fait la fonction $L(\cdot, p^\sharp)$ est fortement convexe de module a, et que son unique minimum est atteint au point $u^\sharp$, on obtient que :

$$L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp) \leq -\frac{a}{2}\big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2 \ .$$

Pour tout $\delta \in \,]0, 1/2[$, on dispose donc de la majoration :

$$\begin{aligned}L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp) \leq{}& -(1-2\delta)\frac{a}{2}\big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2 \\ &+ 2\delta\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \ .\end{aligned}$$

L'utilisation de cette majoration dans l'inégalité suivante, obtenue à l'étape 2.c de la preuve :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} &\leq \epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ &\quad + \frac{\epsilon^{(k)}}{2}\Big(\rho\, L_\Theta^2 \big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 - a\big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2\Big) \\ &\quad + (\epsilon^{(k)})^2\Big(c_5\boldsymbol{\Psi}^{(k)} + c_6\Big) \ ,\end{aligned}$$

et la relation (4.20) conduisent alors à :

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \mid \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} &\leq 2\delta\epsilon^{(k)}\big(L(u^\sharp, p^\sharp) - L(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ &\quad + \epsilon^{(k)}a(1-\delta)\Big(\big\|\boldsymbol{U}^{(k+1)} - u^\sharp\big\|^2 - \big\|\boldsymbol{U}^{(k)} - u^\sharp\big\|^2\Big) \\ &\quad + (\epsilon^{(k)})^2\Big(c_5\boldsymbol{\Psi}^{(k)} + c_6\Big) \ .\end{aligned}$$

La suite de la preuve est inchangée.

4.6 Cas simplement convexe et Lagrangien augmenté

On se place maintenant dans le cas où la fonction J est *simplement convexe*. On dispose encore d'un algorithme de résolution du problème (4.1) et du théorème de convergence associé pourvu que l'on fasse appel au Lagrangien augmenté plutôt qu'au Lagrangien ordinaire. Le Lagrangien augmenté a été présenté au §A.3. Pour le problème (4.1), il s'écrit :

$$L_c(u,p) = J(u) + \zeta_c\big(p, \Theta(u)\big) \ , \tag{4.21a}$$

l'expression de la fonction ζ_c étant donnée par :

$$\zeta_c(p,\theta) = \frac{1}{2c}\big(\big\|\mathrm{proj}_{C^\star}(p + c\theta)\big\|^2 - \big\|p\big\|^2\big) \ . \tag{4.21b}$$

On rappelle que la fonction ζ_c est concave en p, convexe en θ et qu'elle est différentiable, l'expression de ses gradients partiels étant :

$$\nabla_p\zeta_c(p,\theta) = \frac{1}{c}\big(\mathrm{proj}_{C^\star}(p + c\theta) - p\big) \ , \tag{4.21c}$$

$$\nabla_\theta\zeta_c(p,\theta) = \mathrm{proj}_{C^\star}(p + c\theta) \ . \tag{4.21d}$$

On a vu au §A.3 dans le cas *déterministe* un algorithme utilisant à la fois le principe du problème auxiliaire et le Lagrangien augmenté. La k-ème itération de cet algorithme s'écrit :

$$\begin{aligned}u^{(k+1)} &= \arg\min_{u\in U^{\mathrm{ad}}} K(u) + \big\langle \epsilon\nabla J(u^{(k)}) - \nabla K(u^{(k)}) \,, u\big\rangle \\ &\qquad + \epsilon\big\langle \nabla_\theta\zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) \,, \Theta(u)\big\rangle \ , \\ p^{(k+1)} &= p^{(k)} + \rho\nabla_p\zeta_c\big(p^{(k)}, \Theta(u^{(k+1)})\big) \ ,\end{aligned}$$

et son principal avantage est que l'on peut prouver sa convergence sans qu'il soit nécessaire de faire une hypothèse de forte convexité sur la fonction J.

L'extension au cas stochastique de cet algorithme consiste à remplacer le gradient $\nabla J(u^{(k)})$ par $\nabla_u j(u^{(k)}, w^{(k+1)})$; la résolution du problème (4.1) est alors remplacée par la résolution de la suite de problèmes auxiliaires :

$$\begin{aligned} u^{(k+1)} \in \arg\min_{u \in U^{\mathrm{ad}}} K(u) &+ \left\langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}) \, , u \right\rangle \\ &+ \epsilon^{(k)} \left\langle \nabla_\theta \zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) \, , \Theta(u) \right\rangle . \qquad (4.22a) \end{aligned}$$

$$p^{(k+1)} = p^{(k)} + \epsilon^{(k)} \nabla_p \zeta_c\big(p^{(k)}, \Theta(u^{(k+1)})\big) . \qquad (4.22b)$$

Algorithme 4.14. (Algorithme du PPA stochastique régularisé)

1. Choisir $(u^{(0)}, p^{(0)}) \in U^{\mathrm{ad}} \times C^\star$, et une suite $\big\{\epsilon^{(k)}\big\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l'itération k, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire $\boldsymbol{W}$.
3. Calculer $u^{(k+1)}$ solution du problème auxiliaire (4.22a).
4. Calculer $p^{(k+1)}$ par la formule de mise à jour (4.22b).
5. Incrémenter l'indice k de 1 et retourner à l'étape 2.

Remarque 4.15. Du point de vue algorithmique, on notera les différences suivantes entre l'utilisation du Lagrangien augmenté et celle d'un Lagrangien ordinaire.

- Dans l'étape (4.22a) de minimisation en u, on utilise dans le dernier produit scalaire le terme $\nabla_\theta \zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) = \mathrm{proj}_{C^\star}\big(p^{(k)} + c\Theta(u^{(k)})\big)$ plutôt que le terme $p^{(k)}$: ceci correspond à une sorte d'anticipation dans l'utilisation du multiplicateur pour le calcul de $u^{(k+1)}$.
- L'expression de $\nabla_p \zeta_c\big(p^{(k)}, \Theta(u^{(k+1)})\big)$ permet d'interpréter la remise à jour de $p^{(k+1)}$ comme la succession des deux opérations suivantes :

$$\text{projection :}\ \ p^{(k+\frac{1}{2})} = \mathrm{proj}_{C^\star}\big(p^{(k)} + c\Theta(u^{(k+1)})\big) , \qquad (4.23a)$$

$$\text{relaxation :}\ \ p^{(k+1)} = \big(1 - (\epsilon^{(k)}/c)\big) p^{(k)} + \big(\epsilon^{(k)}/c\big) p^{(k+\frac{1}{2})} . \qquad (4.23b)$$

  On en déduit qu'initialiser l'algorithme avec un $p^{(0)} \in C^\star$ conduit à des multiplicateurs $p^{(k)} \in C^\star$ pourvu que le cône C soit convexe (et que les pas $\epsilon^{(k)}$ restent plus petits que le coefficient c).

Pour étudier cet algorithme, on le formule en termes de variables aléatoires. On considère donc un échantillon $\{\boldsymbol{W}^{(k)}\}_{k \in \mathbb{N}}$ de taille infinie de la variable aléatoire $\boldsymbol{W}$, et les étapes 3 et 4 de l'algorithme 4.14 prennent la forme :

$$\begin{aligned} \min_{u \in U^{\mathrm{ad}}} K(u) &+ \left\langle \epsilon^{(k)} \nabla_u j(\boldsymbol{U}^{(k)}, \boldsymbol{W}^{(k+1)}) - \nabla K(\boldsymbol{U}^{(k)}) \, , u \right\rangle \\ &+ \epsilon^{(k)} \left\langle \mathrm{proj}_{C^\star}\big(\boldsymbol{P}^{(k)} + c\Theta(\boldsymbol{U}^{(k)})\big) \, , \Theta(u) \right\rangle , \qquad (4.24a) \end{aligned}$$

$$\boldsymbol{P}^{(k+1)} = \boldsymbol{P}^{(k)} + \epsilon^{(k)} \nabla_p \zeta_c\big(\boldsymbol{P}^{(k)}, \Theta(\boldsymbol{U}^{(k+1)})\big) . \qquad (4.24b)$$

La minimisation dans (4.24a) et les projections dans (4.24a) et (4.24b) sont effectuées ω par ω. Le résultat de la minimisation (4.24a) dépend de ω et est noté $\boldsymbol{U}^{(k+1)}$. Le théorème suivant précise les conditions de convergence de l'algorithme 4.14.

Théorème 4.16.
On suppose que les hypothèses suivantes sont vérifiées.

1. U^{ad} *est une partie convexe fermée non vide d'un l'espace de Hilbert* $\mathbb{U}$*, et* C *est un cône convexe fermé saillant d'un autre espace de Hilbert* $\mathbb{V}$*.*
2. *La fonction* $j : \mathbb{U} \times \mathbb{W} \to \mathbb{R}$ *est une intégrande normale, et l'espérance de* $j(u, \boldsymbol{W})$ *existe pour tout* $u \in U^{\mathrm{ad}}$*.*
3. *Pour tout* $w \in \mathbb{W}$*, la fonction* $j(\cdot, w) : \mathbb{U} \to \mathbb{R}$ *est propre, convexe, semi-continue inférieurement, et est différentiable sur un sous-ensemble ouvert contenant* U^{ad}*, et son gradient partiel par rapport à* u *est noté* $\nabla_u j(u, w)$*.*
4. *La fonction* $j(\cdot, w)$ *est à gradient linéairement borné (GLB), uniformément en* w *:*
$$\exists c_1 > 0,\ \exists c_2 > 0,\ \forall w \in \mathbb{W},\ \forall u \in U^{\mathrm{ad}}\ ,\quad \|\nabla_u j(u, w)\| \le c_1 \|u\| + c_2\ .$$
5. *La fonction* J *est Lipschitzienne, coercive sur l'ensemble* U^{ad}*.*
6. *La fonction* Θ *est* C*-convexe, Lipschitzienne de rapport* L_Θ*.*
7. *Les contraintes sont qualifiées.*
8. *La fonction* K *est propre, fortement convexe de module* b*, semi-continue inférieurement, et est différentiable sur un sous-ensemble ouvert contenant* U^{ad}*.*
9. *La suite* $\{\epsilon^{(k)}\}_{k\in\mathbb{N}}$ *est une* σ*-suite.*

On a alors les conclusions suivantes.

1. *Le problème* (4.1) *admet un ensemble de points selle* $U^\sharp \times P^\sharp$ *non vide.*
2. *Le problème* (4.24a) *admet une solution* $\boldsymbol{U}^{(k+1)}$ *unique.*
3. *Les suites* $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ *et* $\{\boldsymbol{P}^{(k)}\}_{k\in\mathbb{N}}$ *engendrées par l'algorithme* 4.14 *sont bornées presque sûrement, et tout point d'accumulation d'une réalisation de la suite* $\{\boldsymbol{U}^{(k)}\}_{k\in\mathbb{N}}$ *appartient à* $U^\sharp$*, ensemble des solutions du problème.*

Preuve. Pour alléger les écritures, on utilise les notations :
$$\begin{aligned} g^{(k)} &= \nabla_u j\big(u^{(k)}, w^{(k+1)}\big)\ ,\\ q^{(k)} &= \operatorname{proj}_{C^\star}\big(p^{(k)} + c\Theta(u^{(k)})\big)\ ,\\ p^{(k+\frac{1}{2})} &= \operatorname{proj}_{C^\star}\big(p^{(k)} + c\Theta(u^{(k+1)})\big)\ . \end{aligned}$$

La démonstration suivant le même schéma que celle des théorèmes 4.8 et 4.11 et contenant des points très semblables à ceux des démonstrations de ces théorèmes, on détaillera surtout les différences qui apparaissent dans le cadre de cette variante. Le fait que $u^{(k+1)}$ soit solution du problème (4.22a) est caractérisé par le fait que, pour tout $u \in U^{\mathrm{ad}}$,

$$\begin{aligned}\big\langle \nabla K(u^{(k+1)}) - \nabla K(u^{(k)}) + \epsilon^{(k)} g^{(k)} \,, u - u^{(k+1)} \big\rangle \\ + \epsilon^{(k)} \big\langle q^{(k)} \,, \Theta(u) - \Theta(u^{(k+1)}) \big\rangle \geq 0 \,. \quad (4.25)\end{aligned}$$

1. **Choix de la fonction de Lyapunov**. Soit $(u^\sharp, p^\sharp) \in U^\sharp \times P^\sharp$ un point selle du problème (4.1). On choisit la fonction de Lyapunov Λ de la forme :

$$\Lambda(u,p) = K(u^\sharp) - K(u) - \big\langle \nabla K(u) \,, u^\sharp - u \big\rangle + \frac{1}{2} \big\| p - p^\sharp \big\|^2 \,,$$

   et l'on note :
$$\psi^{(k)} = \Lambda(u^{(k)}, p^{(k)}) \,.$$

   On déduit de la forte convexité de K et de la définition de Λ les deux relations :

$$\big\| u^{(k)} - u^\sharp \big\|^2 \leq \frac{2}{b} \, \psi^{(k)} \,, \quad (4.26a)$$

$$\big\| p^{(k)} - p^\sharp \big\|^2 \leq 2 \, \psi^{(k)} \,, \quad (4.26b)$$

   ce qui prouve que Λ est bornée inférieurement et coercive.

2. **Majorations**.
   Le fait que la projection soit contractante, le caractère Lipschitz de Θ et les inégalités (4.26) permettent d'écrire :

$$\begin{aligned}\big\| q^{(k)} \big\| &= \big\| \mathrm{proj}_{C^\star} \big(p^{(k)} + c\Theta(u^{(k)}) \big) \big\| \\ &\leq \big\| p^{(k)} + c\Theta(u^{(k)}) - \big(p^\sharp + c\Theta(u^\sharp) \big) \big\| + \big\| p^\sharp + c\Theta(u^\sharp) \big\| \\ &\leq \big\| p^{(k)} - p^\sharp \big\| + cL_\Theta \big\| u^{(k)} - u^\sharp \big\| + \big\| p^\sharp + c\Theta(u^\sharp) \big\| \,,\end{aligned}$$

   d'où l'existence de coefficients a_1 et a_2 tels que

$$\big\| q^{(k)} \big\| \leq a_1 + a_2 \sqrt{\psi^{(k)}} \,. \quad (4.27)$$

   a. Pour la variation $\big\| u^{(k+1)} - u^{(k)} \big\|$, un raisonnement identique à celui fait dans la démonstration du théorème 4.8 associé à la majoration (4.27) montre que :

$$\big\| u^{(k+1)} - u^{(k)} \big\| \leq \epsilon^{(k)} \big(a_3 \sqrt{\psi^{(k)}} + a_4 \big) \,. \quad (4.28)$$

   b. Dans la mesure où $p^{(0)} \in C^\star$ et donc que $p^{(k)} \in C^\star$ pour tout indice k, comme la projection est contractante et Θ lipschitzienne, on a :

$$\begin{aligned}\big\| p^{(k+\frac{1}{2})} - p^{(k)} \big\| &= \big\| \mathrm{proj}_{C^\star} \big(p^{(k)} + c\Theta(u^{(k+1)}) \big) - p^{(k)} \big\| \\ &\leq cL_\Theta \big\| u^{(k+1)} - u^\sharp \big\| + c \big\| \Theta(u^\sharp) \big\| \,.\end{aligned}$$

De (4.26a), on conclut à l'existence de coefficients b_1 et b_2 tels que [8]

$$\left\|p^{(k+\frac{1}{2})} - p^{(k)}\right\| \leq b_1 + b_2\sqrt{\psi^{(k+1)}} \ . \tag{4.29}$$

La relation (4.23) définissant $p^{(k+1)}$ s'écrit sous la forme :

$$p^{(k+1)} - p^{(k)} = \frac{\epsilon^{(k)}}{c}\big(p^{(k+\frac{1}{2})} - p^{(k)}\big) \ . \tag{4.30}$$

De la majoration (4.29), on déduit l'existence de coefficients b_3 et b_4 tels que :

$$\left\|p^{(k+1)} - p^{(k)}\right\| \leq \epsilon^{(k)}\Big(b_3 + b_4\sqrt{\psi^{(k+1)}}\Big) \ . \tag{4.31}$$

Enfin, utilisant que la projection est contractante et que l'application Θ est lipschitzienne, on a :

$$\left\|p^{(k+\frac{1}{2})} - q^{(k)}\right\| \leq cL_\Theta\left\|u^{(k+1)} - u^{(k)}\right\| \ ,$$

et donc, par (4.28), l'existence de coefficients b_5 et b_6 tels que

$$\left\|p^{(k+\frac{1}{2})} - q^{(k)}\right\| \leq \epsilon^{(k)}\big(b_5 + b_6\sqrt{\psi^{(k)}}\big) \ . \tag{4.32}$$

c. La variation $\psi^{(k+1)} - \psi^{(k)}$ s'écrit :

$$\begin{aligned}\psi^{(k+1)} - \psi^{(k)} = &\underbrace{K(u^{(k)}) - K(u^{(k+1)}) - \big\langle \nabla K(u^{(k)})\,, u^{(k)} - u^{(k+1)}\big\rangle}_{T_1} \\ &+ \underbrace{\big\langle \nabla K(u^{(k)}) - \nabla K(u^{(k+1)})\,, u^\sharp - u^{(k+1)}\big\rangle}_{T_2} \\ &+ \underbrace{\frac{1}{2}\left\|p^{(k+1)} - p^\sharp\right\|^2 - \frac{1}{2}\left\|p^{(k)} - p^\sharp\right\|^2}_{T_3} \ .\end{aligned}$$

- Le noyau K étant convexe, le terme T_1 est négatif ou nul.
- Par la condition d'optimalité (4.25) évalué en $u = u^\sharp$, on obtient :

  $$T_2 \leq \epsilon^{(k)}\big\langle g^{(k)}\,, u^\sharp - u^{(k+1)}\big\rangle + \epsilon^{(k)}\big\langle q^{(k)}\,, \Theta(u^\sharp) - \Theta(u^{(k+1)})\big\rangle \ .$$

  Comme ζ_c est convexe en θ, et que $q^{(k)} = \nabla_\theta\zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big)$, on a :

  $$\big\langle q^{(k)}\,, \Theta(u^\sharp) - \Theta(u^{(k)})\big\rangle \leq \zeta_c\big(p^{(k)}, \Theta(u^\sharp)\big) - \zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) \ .$$

  Par des calculs élémentaires, on montre alors que :

8. Un calcul analogue permettrait, dans le cas où $p^{(k)} \notin C^\star$, d'obtenir l'existence de coefficients b_1, b_2 et b_3 tels que $\left\|p^{(k+\frac{1}{2})} - p^{(k)}\right\| \leq b_1 + b_2\sqrt{\psi^{(k)}} + b_3\sqrt{\psi^{(k+1)}}$.

$$
\begin{aligned}
T_2 \leq \epsilon^{(k)} \big\langle g^{(k)}\,, u^\sharp - u^{(k)} \big\rangle \\
&+ \epsilon^{(k)} \big(\zeta_c\big(p^{(k)}, \Theta(u^\sharp)\big) - \zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) \big) \\
&\quad + \underbrace{\epsilon^{(k)} \big\| u^{(k+1)} - u^{(k)} \big\| \big(\big\| g^{(k)} \big\| + L_\Theta \big\| q^{(k)} \big\| \big)}_{T_{2,1}} \ .
\end{aligned}
$$

Utilisant (4.28), l'hypothèse GLB et les relations (4.26a) et (4.27), on obtient l'existence de constantes c_1 et c_2 telles que :

$$
T_{2,1} \leq \big(\epsilon^{(k)}\big)^2 \big(c_1 + c_2 \psi^{(k)} \big) \ .
$$

- Le terme T_3 s'écrit :

$$
\begin{aligned}
T_3 &= \frac{1}{2} \big\langle p^{(k+1)} - p^{(k)}\,, p^{(k+1)} + p^{(k)} - 2p^\sharp \big\rangle \\
&= \frac{1}{2} \big\| p^{(k+1)} - p^{(k)} \big\|^2 + \big\langle p^{(k+1)} - p^{(k)}\,, p^{(k)} - p^\sharp \big\rangle \ .
\end{aligned}
$$

Utilisant la relation (4.30) et par des calculs élémentaires, on déduit :

$$
\begin{aligned}
T_3 = \frac{1}{2c^2} \big(\epsilon^{(k)}\big)^2 \big\| p^{(k+\frac{1}{2})} - p^{(k)} \big\|^2 \\
&+ \frac{\epsilon^{(k)}}{c} \big\langle q^{(k)} - p^{(k)}\,, p^{(k)} - p^\sharp \big\rangle \\
&\quad + \frac{\epsilon^{(k)}}{c} \big\langle p^{(k+\frac{1}{2})} - q^{(k)}\,, p^{(k)} - p^\sharp \big\rangle \ .
\end{aligned}
$$

Le gradient partiel $\nabla_p \zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big)$ étant égal à $\big(q^{(k)} - p^{(k)}\big)/c$, on déduit de la concavité en p de ζ_c que :

$$
\frac{1}{c} \big\langle q^{(k)} - p^{(k)}\,, p^{(k)} - p^\sharp \big\rangle \leq \zeta_c\big(p^{(k)}, \Theta(u^{(k)}\big) - \zeta_c\big(p^\sharp, \Theta(u^{(k)})\big) \ .
$$

On obtient finalement :

$$
\begin{aligned}
T_3 \leq \epsilon^{(k)} \big(\zeta_c\big(p^{(k)}, \Theta(u^{(k)})\big) - \zeta_c\big(p^\sharp, \Theta(u^{(k)})\big) \big) \\
&+ \underbrace{\frac{1}{2c^2} \big(\epsilon^{(k)}\big)^2 \big\| p^{(k+\frac{1}{2})} - p^{(k)} \big\|^2}_{T_{3,1}} \\
&\quad + \underbrace{\frac{\epsilon^{(k)}}{c} \big\| p^{(k+\frac{1}{2})} - q^{(k)} \big\| \big\| p^{(k)} - p^\sharp \big\|}_{T_{3,2}} \ .
\end{aligned}
$$

Utilisant les relations (4.29), (4.32) et (4.26b), on en déduit l'existence de coefficients c_3, c_4 et c_5 tels que :

$$
T_{3,1} + T_{3,2} \leq \big(\epsilon^{(k)}\big)^2 \big(c_3 + c_4 \psi^{(k)} + c_5 \psi^{(k+1)} \big) \ .
$$

Regroupant les majorations des termes T_1, T_2 et T_3 et des sous-termes $T_{2,1}$ et $T_{3,1} + T_{3,2}$, on obtient :

$$\begin{aligned}\psi^{(k+1)} - \psi^{(k)} \leq \epsilon^{(k)} \Big(\big\langle g^{(k)} \,, u^\sharp - u^{(k)} \big\rangle \\ + \zeta_c\big(\Theta(u^\sharp), p^{(k)}\big) - \zeta_c\big(\Theta(u^{(k)}), p^\sharp\big)\Big) \\ + \big(\epsilon^{(k)}\big)^2 \big(c_6 + c_7\psi^{(k)} + c_8\psi^{(k+1)}\big) \;,\end{aligned}$$

d'où, par convexité de la fonction J, par définition du Lagrangien augmenté et prenant l'espérance conditionnelle par rapport à la tribu engendrée par $(\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(k)})$:

$$\begin{aligned}\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \bigm| \mathcal{F}^{(k)}\big) - \boldsymbol{\Psi}^{(k)} \leq \epsilon^{(k)}\big(L_c(u^\sharp, \boldsymbol{P}^{(k)}) - L_c(\boldsymbol{U}^{(k)}, p^\sharp)\big) \\ + (\epsilon^{(k)})^2\big(c_6 + c_7\boldsymbol{\Psi}^{(k)} + c_8\mathbb{E}\big(\boldsymbol{\Psi}^{(k+1)} \bigm| \mathcal{F}^{(k)}\big)\big) \;. \quad (4.33)\end{aligned}$$

3. **Analyse de convergence**. Par le corollaire 3.6 du théorème de Robbins-Siegmund, on en déduit que la suite $\big\{\boldsymbol{\Psi}^{(k)}\big\}_{k\in\mathbb{N}}$ converge presque sûrement vers une variable aléatoire bornée, et que la série de terme général $\epsilon^{(k)}\big(L_c(u^\sharp, \boldsymbol{P}^{(k)}) - L_c(\boldsymbol{U}^{(k)}, p^\sharp)\big)$ est convergente. De la double inégalité du point selle :
$$L_c(u^\sharp, \boldsymbol{P}^{(k)}) \leq L_c(u^\sharp, p^\sharp) \leq L_c(\boldsymbol{U}^{(k)}, p^\sharp) \;,$$
on en déduit qu'il en est de même des deux séries de terme général $\epsilon^{(k)}\big(L_c(u^\sharp, p^\sharp) - L_c(u^\sharp, \boldsymbol{P}^{(k)})\big)$ et $\epsilon^{(k)}\big(L_c(\boldsymbol{U}^{(k)}, p^\sharp) - L_c(u^\sharp, p^\sharp)\big)$.
4. **Limites des suites**. Les majorations (4.28) et (4.31) permettent l'utilisation du lemme 3.7 sur les deux séries dont on a montré la convergence au point 3 ci-dessus. Les suites $\big\{L_c(\boldsymbol{U}^{(k)}, p^\sharp)\big\}_{k\in\mathbb{N}}$ et $\big\{L_c(u^\sharp, \boldsymbol{P}^{(k)})\big\}_{k\in\mathbb{N}}$ convergent donc toutes les deux vers $L_c(u^\sharp, p^\sharp)$. Des majorations (4.26), on déduit que les suites $\big\{\boldsymbol{U}^{(k)}\big\}_{k\in\mathbb{N}}$ et $\big\{\boldsymbol{P}^{(k)}\big\}_{k\in\mathbb{N}}$ sont bornées presque sûrement. Grâce aux propriétés de continuité du Lagrangien L_c, on obtient que tout point d'accumulation d'une réalisation de la suite $\big\{\boldsymbol{U}^{(k)}\big\}_{k\in\mathbb{N}}$ est solution du problème (4.1), et que tout point d'accumulation d'une réalisation de la suite $\big\{\boldsymbol{P}^{(k)}\big\}_{k\in\mathbb{N}}$ est un multiplicateur optimal du Lagrangien.

4.7 Conclusions

On a montré dans ce chapitre comment utiliser la méthode du gradient stochastique généralisé au cas des problèmes d'optimisation soumis à des contraintes déterministes, et on a illustré le fait que cette utilisation était une généralisation de l'algorithme de Arrow-Hurwicz et non de celui d'Uzawa. Cependant, le fait de pouvoir traiter des problèmes d'optimisation stochastique sous contraintes déterministes ne couvre pas l'ensemble des cas que l'on peut rencontrer en pratique.

En fait, on peut distinguer au moins les trois types de contraintes suivantes :

1. les contraintes presque sûres :
$$\theta(u, \boldsymbol{W}) \in -C \quad \mathbb{P}\text{-p.s.} \;,$$
2. les contraintes en probabilité :
$$\mathbb{P}\big(\theta(u, \boldsymbol{W}) \in -C\big) \geq \pi \;.$$
3. les contraintes en espérance :
$$\mathbb{E}\big(\theta(u, \boldsymbol{W})\big) \in -C \;.$$

Les contraintes presque sûres sont en général trop restrictives pour pouvoir être prises en compte dans le cadre de la boucle ouverte, car elles reviennent à écrire une contrainte pour chaque $\omega \in \Omega$. Quant aux contraintes en probabilité, leur traitement direct entraîne de grande difficultés mathématiques (perte de convexité, voir de connexité de l'ensemble admissible). Enfin, le cas des contraintes en espérance n'est pas fréquent en pratique, car il est souvent difficile de donner un sens à la satisfaction d'un besoin en espérance.

Cependant, l'extension des algorithmes vus dans ce chapitre au cas des contraintes en espérance paraît relativement abordable : dans l'esprit du gradient stochastique, il semblerait raisonnable de profiter des itérations de gradient pour reconstituer l'espérance du gradient du critère *et* l'espérance de la contrainte. De plus, le fait de pouvoir écrire une contrainte en probabilité comme une contrainte en espérance :
$$\mathbb{P}\big(\theta(u, \boldsymbol{W}) \in -C\big) = \mathbb{E}\big(\mathbf{1}_{\{\theta(u, \boldsymbol{W}) \in -C\}}\big) \;, \tag{4.34}$$
où $\mathbf{1}$ dénote la fonction indicatrice d'ensemble[9], fait que l'on peut espérer s'attaquer numériquement aux problèmes sous contraintes en probabilité par le biais des contraintes en espérance. Il faudra alors, pour surmonter les difficultés liées à la non convexité engendrée par la fonction indicatrice présente dans (4.34), utiliser un Lagrangien augmenté sur une contrainte en espérance, et donc savoir appliquer le gradient stochastique à des fonctions non linéaires de l'espérance. Ce point sera abordé dans le §5.

9. La fonction indicatrice d'un sous-ensemble mesurable A de Ω est définie par :
$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \;,$$
et donc $\mathbb{P}(A) = \mathbb{E}(\mathbf{1}_A)$.