

4 Gradient stochastique généralisé avec contraintes

Comme dans les paragraphes précédents, on se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire \mathbf{W} à valeurs dans l'espace probabilisé $(\mathbb{W}, \mathcal{W})$. On se donne encore 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 $u \in U^{\text{ad}}$) :

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

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

$$(\mathbf{P}_C) \quad \min_{u \in U^{\text{ad}}} J(u) \quad \text{sous la contrainte} \quad \Theta(u) \in -C . \quad (37)$$

Remarque 7. 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. Il existe des extensions du gradient stochastique au cas où Θ est elle-même l'espérance d'une fonction aléatoire θ définie sur $\mathbb{U} \times \mathbb{W}$ à valeurs dans \mathbb{V} (voir [7]). \square

4.1 Rappels du cas déterministe

On consultera le §10 et le §11 pour plus de détails sur les résultats évoqués dans ce paragraphe.

4.1.1 Dualité

Notant C^* le cône dual²⁴ de C , on définit le lagrangien L du problème (\mathbf{P}_C) , défini sur $U^{\text{ad}} \times C^*$ à valeurs réelles 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 (\mathbf{P}_C) se fait en maxi-minimisant le lagrangien.

4.1.2 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 consiste à alterner sur le lagrangien *phases d'optimisation complète* pour la minimisation en u et *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 :

$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} J(u) + \langle p^{(k)}, \Theta(u) \rangle , \quad (38a)$$

$$p^{(k+1)} = \text{proj}_{C^*} (p^{(k)} + \rho \Theta(u^{(k+1)})) . \quad (38b)$$

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

$$u^{(k+1)} = \text{proj}_{U^{\text{ad}}} \left(u^{(k)} - \epsilon (\nabla J(u^{(k)}) + (\Theta'(u^{(k)}))^{\top} p^{(k)}) \right) , \quad (39a)$$

$$p^{(k+1)} = \text{proj}_{C^*} (p^{(k)} + \rho \Theta(u^{(k+1)})) . \quad (39b)$$

²⁴ $C^* = \{p \in \mathbb{V}, \text{ tel que } \langle p, v \rangle \geq 0 \forall v \in C\}$.

4.1.3 Principe du problème auxiliaire

Une extension du principe du problème auxiliaire au cas de la recherche du point selle d'un lagrangien consiste de nouveau à **choisir** une fonction K définie sur l'espace \mathbb{U} à valeurs dans \mathbb{R} et à remplacer la résolution du problème (\mathbf{P}_C) par la résolution d'une suite de problèmes auxiliaires $(\mathbf{PA}_C^{(k)})$ 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 :

$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon \nabla J(u^{(k)}) - \nabla K(u^{(k)}), u \rangle + \epsilon \langle p^{(k)}, \Theta'(u^{(k)}) \cdot u \rangle, \quad (40a)$$

$$p^{(k+1)} = \text{proj}_{C^*} (p^{(k)} + \rho \Theta(u^{(k+1)})) . \quad (40b)$$

Remarque 8. On peut dans la phase de minimisation en u de l'algorithme ci-dessus remplacer l'approximation du premier ordre $\langle p^{(k)}, (\Theta'(u^{(k)}))u \rangle$ par le terme non linéaire $\langle p^{(k)}, \Theta(u) \rangle$ d'où provient l'approximation précédente, ce qui conduit à une étape de minimisation en u de la forme :

$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon \nabla J(u^{(k)}) - \nabla K(u^{(k)}), u \rangle + \epsilon \langle p^{(k)}, \Theta(u) \rangle . \quad (40c)$$

□

On montre alors que, comme dans le cas sans contrainte explicite, le PPA permet :

- de retrouver les algorithmes d'Uzawa (avec les choix $K(u) = J(u)$, $\epsilon = 1$, et en utilisant la remarque 8 ci-dessus) et d'Arrow-Hurwicz (avec le choix $K(u) = \frac{1}{2}\|u\|^2$),
- de décomposer l'étape de minimisation en u de chaque problème auxiliaire pourvu que le noyau K soit choisi additif par rapport à la décomposition de l'espace \mathbb{U} (si l'on veut utiliser l'étape de minimisation donnée dans la remarque 8, il faut que la fonction Θ soit elle aussi additive par rapport à cette décomposition).

4.2 Extension stochastique de l'algorithme d'Uzawa ?

Une première tentative pour appliquer l'idée du gradient stochastique au problème (\mathbf{P}_C) est, dans le cadre de l'algorithme d'Uzawa, de remplacer dans (38) la valeur $J(u)$ ($= \mathbb{E}(j(u, \mathbf{W}))$) par la valeur $j(u, w^{(k+1)})$, où $w^{(k+1)}$ est un tirage de la variable aléatoire \mathbf{W} .²⁵ Cette manière de procéder conduit à l'algorithme suivant.

Algorithme 2. (*Algorithme d'Uzawa stochastique*)

1. Choisir un point initial $(u^{(0)}, p^{(0)}) \in U^{\text{ad}} \times C^*$ ainsi qu'une suite $\{\rho^{(k)}\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l'itération k de l'algorithme, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire \mathbf{W} suivant sa loi, indépendamment des tirages $(w^{(1)}, \dots, w^{(k)})$ des itérations précédentes.
3. Calculer le point $(u^{(k+1)}, p^{(k+1)})$ par :
$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} j(u, w^{(k+1)}) + \langle p^{(k)}, \Theta(u) \rangle,$$

$$p^{(k+1)} = \text{proj}_{C^*} (p^{(k)} + \rho^{(k)} \Theta(u^{(k+1)})).$$
4. Incrémenter l'indice k de 1 et retourner à l'étape 2.

²⁵On notera que l'on remplace ici 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 ; ce n'est donc pas exactement la même idée qui est employée.

La question est de savoir si cet algorithme peut converger vers la solution du problème (\mathbf{P}_C) .

Plutôt que de mener une étude de convergence exhaustive de cet algorithme, on va se donner un exemple simple sur lequel il sera possible d'effectuer facilement les calculs et on va montrer sur cet exemple que l'algorithme proposé ne permet pas d'obtenir la solution du problème de départ.

Contre-exemple. On considère le problème défini de la manière suivante.

- $\mathbb{U} = \mathbb{R}^2$ et $U^{\text{ad}} = \mathbb{U}$; on note : $u = (u_1, u_2)$.
- $\mathbb{V} = \mathbb{R}$ et $C = \{0\}$ (contrainte égalité).
- $\mathbb{W} = \mathbb{R}^4$ et $\mathbf{W} : \Omega \rightarrow \mathbb{W}$; la variable aléatoire $\mathbf{W} = (\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2)$ est telle que :
 - \mathbf{A}_1 et \mathbf{A}_2 sont intégrables strictement positives,
 - \mathbf{B}_1 et \mathbf{B}_2 sont intégrables.
- Le critère et la contrainte sont définis par :

$$j(u, w) = \frac{1}{2} \left(a_1 u_1^2 + a_2 u_2^2 \right) + b_1 u_1 + b_2 u_2 ,$$

$$\Theta(u) = \theta_1 u_1 + \theta_2 u_2 ,$$

θ_1 et θ_2 étant des constantes réelles non nulles.

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 $(u_1^\#, u_2^\#)$ la solution primale et $p^\#$ le multiplicateur optimal associé à la contrainte, les conditions d'optimalité du problème s'écrivent :

$$\begin{aligned} \mathbb{E}(\mathbf{A}_1) u_1^\# + \mathbb{E}(\mathbf{B}_1) + \theta_1 p^\# &= 0 , \\ \mathbb{E}(\mathbf{A}_2) u_2^\# + \mathbb{E}(\mathbf{B}_2) + \theta_2 p^\# &= 0 , \\ \theta_1 u_1^\# + \theta_2 u_2^\# &= 0 . \end{aligned}$$

On en déduit que la valeur du multiplicateur optimale est :

$$p^\# = - \frac{\theta_1 \frac{\mathbb{E}(\mathbf{B}_1)}{\mathbb{E}(\mathbf{A}_1)} + \theta_2 \frac{\mathbb{E}(\mathbf{B}_2)}{\mathbb{E}(\mathbf{A}_2)}}{\frac{\theta_1^2}{\mathbb{E}(\mathbf{A}_1)} + \frac{\theta_2^2}{\mathbb{E}(\mathbf{A}_2)}} \quad (41)$$

2. Mise en œuvre d'Uzawa stochastique.

La k -ème itération de l'algorithme 2 revient à résoudre en $(u_1^{(k+1)}, u_2^{(k+1)})$ 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 :

$$p^{(k+1)} = p^{(k)} + \rho^{(k)} \left(\theta_1 u_1^{(k+1)} + \theta_2 u_2^{(k+1)} \right) .$$

On en déduit l'équation de récurrence stochastique décrivant l'évolution du multiplicateur au cours de ces itérations :

$$\mathbf{P}^{(k+1)} = \mathbf{P}^{(k)} - \rho^{(k)} \left(\frac{\theta_1^2}{\mathbf{A}_1^{(k+1)}} + \frac{\theta_2^2}{\mathbf{A}_2^{(k+1)}} \right) \mathbf{P}^{(k)} - \rho^{(k)} \left(\theta_1 \frac{\mathbf{B}_1^{(k+1)}}{\mathbf{A}_1^{(k+1)}} + \theta_2 \frac{\mathbf{B}_2^{(k+1)}}{\mathbf{A}_2^{(k+1)}} \right).$$

Utilisant la méthode de l'équation différentielle moyenne du §15, par le corollaire 2, on sait que cette équation de récurrence stochastique ne peut converger que vers l'unique point p^* d'équilibre stable de l'équation différentielle ordinaire associée, à savoir :

$$p^\# = - \frac{\theta_1 \mathbb{E} \left(\frac{\mathbf{B}_1}{\mathbf{A}_1} \right) + \theta_2 \mathbb{E} \left(\frac{\mathbf{B}_2}{\mathbf{A}_2} \right)}{\frac{\theta_1^2}{\mathbb{E}(\mathbf{A}_1)} + \frac{\theta_2^2}{\mathbb{E}(\mathbf{A}_2)}}. \quad (42)$$

L'examen des relations (41) et (42) montre que, dès que les variables aléatoires \mathbf{A}_1 et \mathbf{B}_1 ne sont pas *indépendantes*, on peut avoir $\mathbb{E} \left(\frac{\mathbf{B}_1}{\mathbf{A}_1} \right) \neq \frac{\mathbb{E}(\mathbf{B}_1)}{\mathbb{E}(\mathbf{A}_1)}$, et donc $p^\# \neq p^*$.

Conclusion. Le contre-exemple précédent montre que, même si l'algorithme d'Uzawa stochastique 2 converge, il ne fournit pas toujours la solution du problème initial. L'exemple choisi (linéaire-quadratique) est suffisamment simple et générique pour pouvoir exclure cet algorithme de la panoplie permettant de résoudre le problème (\mathbf{P}_C) et 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 se rend compte que, 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, ce qui explique son 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 (40) le terme $\nabla J(u^{(k)})$ (correspondant à une espérance sur la variable aléatoire \mathbf{W}) par le terme $\nabla_u j(u^{(k)}, w^{(k+1)})$ (correspondant à un tirage de la variable \mathbf{W}) : on substitue à la résolution du problème (\mathbf{P}_C) la résolution d'une suite de problèmes auxiliaires ($\mathbf{PA}_C^{(k)}$), avec :

$$\begin{aligned} (\mathbf{PA}_C^{(k)}) \quad u^{(k+1)} &= \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}), u \rangle + \epsilon^{(k)} \langle p^{(k)}, \Theta'(u^{(k)})u \rangle, \\ p^{(k+1)} &= \text{proj}_{C^*} (p^{(k)} + \epsilon^{(k)} \Theta(u^{(k+1)})) . \end{aligned}$$

Remarque 9. Comme dans le cas déterministe, il est possible de remplacer dans la phase de minimisation en u l'approximation linéaire $\langle p^{(k)}, \Theta'(u^{(k)})u \rangle$ par le terme non linéaire $\langle p^{(k)}, \Theta(u) \rangle$. L'étape de minimisation en u dans le problème ($\mathbf{PA}_C^{(k)}$) devient alors :

$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}), u \rangle + \epsilon^{(k)} \langle p^{(k)}, \Theta(u) \rangle .$$

On utilisera cette dernière forme plutôt que celle basée sur la différentielle de Θ , ce qui permet de faire des hypothèses moins fortes dans les théorèmes de convergence. \square

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

Algorithme 3. (Algorithme du PPA stochastique)

1. Choisir un point initial $(u^{(0)}, p^{(0)}) \in U^{\text{ad}} \times C^*$ ainsi qu'une suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l'itération k de l'algorithme, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire \mathbf{W} suivant sa loi, indépendamment des tirages $(w^{(1)}, \dots, w^{(k)})$ des itérations précédentes.
3. Calculer le point $(u^{(k+1)}, p^{(k+1)})$ par :
$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}), u \rangle + \epsilon^{(k)} \langle p^{(k)}, \Theta(u) \rangle,$$

$$p^{(k+1)} = \text{proj}_{C^*} (p^{(k)} + \epsilon^{(k)} \Theta(u^{(k+1)})).$$
4. 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 $K(u) = \frac{1}{2} \|u\|^2$, et en utilisant la forme linéarisée en Θ , la minimisation en u dans le problème auxiliaire $(\mathbf{PA}_C^{(k)})$ se met sous la forme :

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

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

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

et correspond exactement à un pas de l'extension "naturelle" au cas stochastique de l'algorithme d'Arrow-Hurwicz déterministe (comparer avec les relations (39)).

- Cet algorithme ne peut *jamaïs* 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 $(\mathbf{PA}_C^{(k)})$, supposant que l'espace \mathbb{U} et l'ensemble admissible U^{ad} se mettent sous forme de produits cartésiens $\mathbb{U}_1 \times \dots \times \mathbb{U}_N$ et $U_1^{\text{ad}} \times \dots \times U_N^{\text{ad}}$, avec pour tout i l'inclusion $U_i^{\text{ad}} \subset \mathbb{U}_i$, et en choisissant le noyau K sous forme additive par rapport à cette décomposition :

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

l'étape de minimisation en u s'écrit :

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

qui se décompose en N sous-problèmes d'optimisation indépendants, le i -ème de ces sous-problèmes s'écrivant :

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

Remarque 10. Dans le cas où la fonction Θ est elle aussi additive par rapport à la décomposition de l'espace :

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

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

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.²⁶ On s'intéressera ensuite au cas où la fonction J est simplement convexe et on présentera un nouvel algorithme, ainsi que le résultat de convergence associé, basé sur l'utilisation du lagrangien augmenté.

4.4 Théorème de convergence

On se place dans le cas où le lagrangien du problème déterministe est stable. La question de la convergence de l'algorithme 3 est réglée par le résultat suivant.

Théorème 2.

On suppose que les hypothèses suivantes sont vérifiées.

H0 U^{ad} est une partie convexe fermée non vide de \mathbb{U} , espace de Hilbert de dimension finie ;
 \mathbf{W} est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace (\mathbb{W}, \mathbb{W}) , de loi μ ;
 C est un cône convexe fermé saillant²⁷ de \mathbb{V} , espace de Hilbert de dimension finie.

H1 La variable aléatoire $j(u, \mathbf{W}) : \Omega \rightarrow \mathbb{R}$ est mesurable et son espérance existe pour tout $u \in U^{\text{ad}}$.

H2 La fonction $j(\cdot, w) : \mathbb{U} \rightarrow \mathbb{R}$ est propre, convexe, semi-continue inférieurement, différentiable pour tout $w \in \mathbb{W}$, et son gradient partiel en u est noté $\nabla_u j(u, w)$.

H3 La fonction $j(\cdot, w)$ est à gradient linéairement borné (GLB) uniformément en w :

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

H4 La fonction Θ est C -convexe, lipschitzienne de rapport L_{Θ} .

H5 La fonction J est coercive sur U^{ad} , les contraintes sont qualifiées et le lagrangien L est stable.

H6 La fonction K est propre, fortement convexe de module b , différentiable.

H7 La suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ est une σ -suite.

On a alors les conclusions suivantes.

C1 Le problème $(\mathbf{P}_{\mathbf{C}})$ admet un ensemble de points selle $U^{\sharp} \times P^{\sharp}$ non vide.

C2 La minimisation en u dans le problème $(\mathbf{P}^{\mathbf{A}^{(k)}})$ admet une solution $\mathbf{U}^{(k+1)}$ unique.

C3 les suites $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ et $\{\mathbf{P}^{(k)}\}_{k \in \mathbb{N}}$ engendrées par l'algorithme 3 sont bornées presque sûrement, tout point d'accumulation d'une réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ appartient presque sûrement à U^{\sharp} , ensemble des solutions du problème $(\mathbf{P}_{\mathbf{C}})$.

²⁶Voir le §10.2.2 pour toutes les notions touchant à l'optimisation sous contraintes et à la dualité.

²⁷ $C \cap -C = \{0\}$

C4 $\forall p^\sharp \in P^\sharp$, $\lim_{k \rightarrow +\infty} L(\mathbf{U}^{(k)}, p^\sharp) = L(u^\sharp, p^\sharp)$ presque sûrement.

Remarque 11. L'hypothèse de stabilité du lagrangien dans **H5** 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 $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$. \square

Preuve. La démonstration des points **C1** et **C2** découle des théorèmes généraux relatifs à l'optimisation avec contraintes explicites. La démonstration des points **C3** et **C4** se fait en quatre étapes, à savoir :

- choix d'une fonction de Lyapunov opérant sur $(u^{(k)}, p^{(k)})$,
- majoration de la variation de cette fonction d'une itération sur l'autre,
- convergence par un argument de (quasi-)martingale,
- analyse des limites des suites.

Pour alléger les écritures, on utilise la notation :

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

On fait la preuve en utilisant la forme non linéarisée en Θ dans la minimisation en u des problèmes $(\mathbf{PA}_C^{(k)})$. Le fait que $u^{(k+1)}$ soit solution de cette minimisation en u est caractérisé par la condition d'optimalité suivante :

$$\forall u \in U^{\text{ad}}, \left\langle \nabla K(u^{(k+1)}) - \nabla K(u^{(k)}) + \epsilon^{(k)} r^{(k)}, u - u^{(k+1)} \right\rangle + \epsilon^{(k)} \left\langle p^{(k)}, \Theta(u) - \Theta(u^{(k+1)}) \right\rangle \geq 0 . \quad (43)$$

1. Fonction de Lyapunov.

Soit $(u^\sharp, p^\sharp) \in U^\sharp \times P^\sharp$ un point selle du problème (\mathbf{P}_C) . On choisit la fonction de Lyapunov Λ de la forme :

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

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

$$\|u - u^\sharp\|^2 \leq \frac{2}{b} \Lambda(u, p) \quad , \quad \|p - p^\sharp\|^2 \leq 2\Lambda(u, p) . \quad (44)$$

On utilisera dans la suite la notation :

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

2. Majorations.

- On majore d'abord la variation $\|u^{(k+1)} - u^{(k)}\|$.

Partant des conditions d'optimalité (43) écrites au point $u = u^{(k)}$, puis utilisant la forte convexité de K ainsi que le caractère lipschitzien de Θ , on obtient :

$$b \|u^{(k+1)} - u^{(k)}\|^2 \leq \epsilon^{(k)} \|r^{(k)}\| \|u^{(k)} - u^{(k+1)}\| + \epsilon^{(k)} \|p^{(k)}\| L_\Theta \|u^{(k)} - u^{(k+1)}\| ,$$

et donc :

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

L'hypothèse GLB et les majorations (44) impliquent alors l'existence de constantes c_3 et c_4 telles que :

$$\|u^{(k+1)} - u^{(k)}\| \leq \epsilon^{(k)} \left(c_3 \sqrt{\psi^{(k)}} + c_4 \right) . \quad (45)$$

- On majore ensuite la variation $\|p^{(k)} - p^\sharp\|^2$.

De la relation $p^\sharp = \text{proj}_{C^\star} \left(p^\sharp + \epsilon^{(k)} \Theta(u^\sharp) \right)$,²⁸ de la relation définissant $p^{(k+1)}$ à partir de $p^{(k)}$ dans l'algorithme 3 et du fait que l'opérateur de projection $\text{proj}_{C^\star}(\cdot)$ est contractant, on obtient :

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

Élevant cette inégalité au carré, développant le terme de droite et comme Θ est lipschitzienne, il vient :

$$\|p^{(k+1)} - p^\sharp\|^2 - \|p^{(k)} - p^\sharp\|^2 \leq \left(\epsilon^{(k)} L_\Theta \right)^2 \|u^{(k+1)} - u^\sharp\|^2 + 2\epsilon^{(k)} \left\langle p^{(k)} - p^\sharp, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \right\rangle . \quad (46)$$

²⁸L'égalité $p^\sharp = \text{proj}_{C^\star} (p^\sharp + \rho \Theta(u^\sharp))$ est vraie pour tout $\rho > 0$: on écrit le problème d'optimisation associé à cette projection, et la première inégalité du point selle prouve que p^\sharp en est solution.

- On majore enfin la variation $(\psi^{(k+1)} - \psi^{(k)})$.

Formant la différence $\psi^{(k+1)} - \psi^{(k)}$, la forte convexité de K et la relation (46) impliquent :

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

Utilisant alors les conditions d'optimalité (43) du problème $(\mathbf{PA}^{(k)})$ au point $u = u^\#$, on obtient :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} &\leq -\frac{b}{2} \left\| u^{(k+1)} - u^{(k)} \right\|^2 + \epsilon^{(k)} \left\langle r^{(k)}, u^\# - u^{(k+1)} \right\rangle + \\ &\quad \epsilon^{(k)} \left\langle p^\#, \Theta(u^\#) - \Theta(u^{(k+1)}) \right\rangle + \frac{1}{2} \left(\epsilon^{(k)} L_\Theta \right)^2 \left\| u^{(k+1)} - u^\# \right\|^2. \end{aligned}$$

Négligeant le terme $-\frac{b}{2} \left\| u^{(k+1)} - u^{(k)} \right\|^2$ (négatif), cette inégalité se met sous la forme :

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

Par (44), (45) et avec l'hypothèse GLB, on en déduit l'existence de constantes c_5 , c_6 et c_7 telles que :

$$\psi^{(k+1)} - \psi^{(k)} \leq \epsilon^{(k)} \left\langle r^{(k)}, u^\# - u^{(k)} \right\rangle + \epsilon^{(k)} \left\langle p^\#, \Theta(u^\#) - \Theta(u^{(k)}) \right\rangle + (\epsilon^{(k)})^2 (c_5 \psi^{(k)} + c_6 + c_7 \psi^{(k+1)}),$$

relation que l'on peut aussi écrire en terme de variables aléatoires :

$$\psi^{(k+1)} - \psi^{(k)} \leq \epsilon^{(k)} \left\langle \mathbf{R}^{(k)}, u^\# - \mathbf{U}^{(k)} \right\rangle + \epsilon^{(k)} \left\langle p^\#, \Theta(u^\#) - \Theta(\mathbf{U}^{(k)}) \right\rangle + (\epsilon^{(k)})^2 (c_5 \psi^{(k)} + c_6 + c_7 \psi^{(k+1)}), \quad (47)$$

3. Convergence.

- La suite $\{\mathbb{E}(\boldsymbol{\psi}^{(k)})\}_{k \in \mathbb{N}}$ est bornée par une constante $M > 0$.

Prenant de part et d'autre de l'inégalité (47) l'espérance conditionnelle par rapport à la tribu $\mathcal{F}^{(k)}$ engendrée par les k premières variables aléatoires $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(k)})$, remarquant que $\mathbb{E}(\mathbf{R}^{(k)} | \mathcal{F}^{(k)}) = \nabla J(\mathbf{U}^{(k)})$ et utilisant la convexité de la fonction J , on obtient :

$$\mathbb{E}(\boldsymbol{\psi}^{(k+1)} | \mathcal{F}^{(k)}) - \boldsymbol{\psi}^{(k)} \leq \epsilon^{(k)} (L(u^\#, p^\#) - L(\mathbf{U}^{(k)}, p^\#)) + (\epsilon^{(k)})^2 (c_5 \boldsymbol{\psi}^{(k)} + c_6 + c_7 \mathbb{E}(\boldsymbol{\psi}^{(k+1)} | \mathcal{F}^{(k)})). \quad (48)$$

De l'inégalité de droite du point selle du lagrangien L et du fait que $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ est une σ -suite, on déduit :

$$\mathbb{E}(\boldsymbol{\psi}^{(k+1)} | \mathcal{F}^{(k)}) - \boldsymbol{\psi}^{(k)} \leq \alpha^{(k)} \boldsymbol{\psi}^{(k)} + \beta^{(k)} + \gamma^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k+1)} | \mathcal{F}^{(k)}), \quad (49)$$

où $\alpha^{(k)}$, $\beta^{(k)}$ et $\gamma^{(k)}$ sont les termes de séries convergentes. Prenant alors l'espérance de part et d'autre de cette dernière inégalité, on obtient :

$$\mathbb{E}(\boldsymbol{\psi}^{(k+1)}) - \mathbb{E}(\boldsymbol{\psi}^{(k)}) \leq \alpha^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k)}) + \beta^{(k)} + \gamma^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k+1)}).$$

Par le lemme 1 (voir §3.4), on conclut que la suite $\{\mathbb{E}(\boldsymbol{\psi}^{(k)})\}_{k \in \mathbb{N}}$ est positive et bornée par un réel $M > 0$.

- La suite $\{\boldsymbol{\psi}^{(k)}\}_{k \in \mathbb{N}}$ converge presque sûrement.

Le processus $\{\boldsymbol{\psi}^{(k)}\}_{k \in \mathbb{N}}$ est clairement adapté à la filtration $\{\mathcal{F}^{(k)}\}_{k \in \mathbb{N}}$. De plus :

- $\inf_{k \in \mathbb{N}} \mathbb{E}(\boldsymbol{\psi}^{(k)}) > -\infty$ car $\boldsymbol{\psi}^{(k)} \geq 0$ d'après la relation (44) ;
- notant $G^{(k)}$ l'ensemble $\{\omega \in \Omega, \mathbb{E}(\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)} | \mathcal{F}^{(k)}) > 0\}$ et utilisant la relation (49), il vient :

$$\begin{aligned} \sum_{k \in \mathbb{N}} \mathbb{E}(\mathbf{1}_{G^{(k)}} \cdot (\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)})) &= \sum_{k \in \mathbb{N}} \mathbb{E}(\mathbf{1}_{G^{(k)}} \cdot \mathbb{E}(\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)} | \mathcal{F}^{(k)})) \\ &\leq \sum_{k \in \mathbb{N}} \mathbb{E}(\mathbf{1}_{G^{(k)}} \cdot (\alpha^{(k)} \boldsymbol{\psi}^{(k)} + \beta^{(k)} + \gamma^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k+1)} | \mathcal{F}^{(k)}))) \\ &\leq \sum_{k \in \mathbb{N}} \alpha^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k)}) + \beta^{(k)} + \gamma^{(k)} \mathbb{E}(\boldsymbol{\psi}^{(k+1)}) \\ &\leq \sum_{k \in \mathbb{N}} (\alpha^{(k)} + \gamma^{(k)}) M + \beta^{(k)}, \end{aligned}$$

et cette dernière quantité est bornée car $\alpha^{(k)}$, $\beta^{(k)}$ et $\gamma^{(k)}$ sont les termes de séries convergentes.

On en déduit que le processus $\{\psi^{(k)}\}_{k \in \mathbb{N}}$ est une quasi-martingale (voir propriété 12, §13.2).

Enfin, utilisant de nouveau $\psi^{(k)} \geq 0$, on a : $\sup_{k \in \mathbb{N}} \mathbb{E} \left((\psi^{(k)})^- \right) < +\infty$: la quasi-martingale $\{\psi^{(k)}\}_{k \in \mathbb{N}}$ converge donc presque sûrement vers une variable aléatoire intégrable (voir théorème 16, §13.2).

4. Limites.

- Les suites $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ et $\{\mathbf{P}^{(k)}\}_{k \in \mathbb{N}}$ sont bornées presque sûrement.

Ceci est une conséquence directe de la convergence presque sûre de la suite $\{\psi^{(k)}\}_{k \in \mathbb{N}}$ et des relations (44).

- La suite $\{L(\mathbf{U}^{(k)}, p^\sharp)\}_{k \in \mathbb{N}}$ converge presque sûrement vers $L(u^\sharp, p^\sharp)$.

En effet, des relations (48) et (49), on déduit :

$$\epsilon^{(k)} \left(L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp) \right) \leq (1 + \alpha^{(k)})\psi^{(k)} + \beta^{(k)} - (1 - \gamma^{(k)})\mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) .$$

Prenant l'espérance de part et d'autre de l'inégalité, sommant pour k compris entre 1 et N , ré-arrangeant les termes de la somme et utilisant la valeur M bornant la suite $\{\mathbb{E}(\psi^{(k)})\}_{k \in \mathbb{N}}$, on obtient :

$$\sum_{k=1}^N \epsilon^{(k)} \mathbb{E} \left(L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp) \right) \leq \sum_{k=1}^N \left((\alpha^{(k)} + \gamma^{(k)})M + \beta^{(k)} \right) + \mathbb{E} \left(\psi^{(1)} \right) - \mathbb{E} \left(\psi^{(N+1)} \right) .$$

La série de terme général $\epsilon^{(k)} \mathbb{E} \left(L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp) \right)$ est donc convergente, et comme, d'après les inégalités du point selle, chacun des termes $L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp)$ est positif, la série de terme général $\epsilon^{(k)} \left(L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp) \right)$ est presque sûrement convergente :

$$\sum_{k \in \mathbb{N}} \epsilon^{(k)} \left(L(\mathbf{U}^{(k)}, p^\sharp) - L(u^\sharp, p^\sharp) \right) < +\infty .$$

Appliquant le lemme 2 (voir §3.4), on déduit que $L(\mathbf{U}^{(k)}, p^\sharp)$ converge vers $L(u^\sharp, p^\sharp)$ presque sûrement.

- Tout point d'accumulation d'une réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ est presque sûrement solution de (\mathbf{P}_C) . La suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ est bornée presque sûrement dans l'ensemble fermé U^{ad} . Par un argument de compacité, on en déduit que, sauf sur un ensemble de mesure nulle, toute réalisation $\{u^{(k)}\}_{k \in \mathbb{N}}$ de cette suite admet au moins un point d'accumulation dans U^{ad} . Soit $\{u^{(\Phi(k))}\}_{k \in \mathbb{N}}$ une sous-suite de $\{u^{(k)}\}_{k \in \mathbb{N}}$ convergeant vers \bar{u} . Le caractère s.c.i. du lagrangien implique que l'on a :

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

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

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

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

Le lagrangien L est alors toujours stable, et toute réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ engendrée par l'algorithme 3 a un *unique* point d'accumulation et converge donc toute entière vers u^\sharp . \square

4.5 Variantes de l'algorithme

4.5.1 Variante dans le cas où J est fortement convexe

Lorsque la fonction J est *fortement convexe*, on dispose d'une variante intéressante de l'algorithme 3 car on montre qu'il est alors possible de remettre à jour les multiplicateurs $p^{(k)}$ avec un pas ρ *constant* (au lieu de pas $\epsilon^{(k)}$ formant une σ -suite), 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. (Algorithme du PPA stochastique “à grands pas”)

1. Choisir $u^{(0)} \in U^{\text{ad}}$ et $p^{(0)} \in C^* \cap B(0, R)$,
choisir un réel $\rho > 0$ ainsi qu’une suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l’itération k de l’algorithme, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire \mathbf{W} suivant sa loi, indépendamment des tirages $(w^{(1)}, \dots, w^{(k)})$ des itérations précédentes.
3. Calculer le point $(u^{(k+1)}, p^{(k+1)})$ par :
$$u^{(k+1)} = \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon^{(k)} \nabla_{uj}(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}), u \rangle + \epsilon^{(k)} \langle p^{(k)}, \Theta(u) \rangle,$$

$$p^{(k+1)} = \text{proj}_{C^* \cap B(0, R)} (p^{(k)} + \rho \Theta(u^{(k+1)})).$$
4. Incrémenter l’indice k de 1 et retourner à l’étape 2.

Les conditions de convergence de cet algorithme sont précisées par le théorème suivant.

Théorème 3.

En plus des hypothèses faites dans le théorème 2, on suppose les conditions suivantes vérifiées :

- la fonction J est fortement convexe de module a ,²⁹
- la σ -suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ est décroissante,
- le coefficient ρ vérifie : $0 < \rho < \frac{2a}{L_{\Theta}^2}$.

Alors, toute réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ engendrée par l’algorithme 4 converge presque sûrement vers $u^{\#}$, unique solution du problème $(\mathbf{P}_{\mathbf{C}})$.

Preuve. La démonstration suivant exactement le même schéma que celle du théorème 2, on repart de cette preuve et on indique les différences qui apparaissent dans le cadre de cette variante.

1. Fonction de Lyapunov.

Soit $(u^{\#}, p^{\#}) \in U^{\#} \times P^{\#}$ un point selle du problème $(\mathbf{P}_{\mathbf{C}})$. On choisit la fonction de Lyapunov Λ de la forme suivante :

$$\Lambda(u^{(k)}, p^{(k)}) = K(u^{\#}) - K(u^{(k)}) - \langle \nabla K(u^{(k)}), u^{\#} - u^{(k)} \rangle + \frac{\epsilon^{(k)}}{2\rho} \|p^{(k)} - p^{\#}\|^2,$$

et l’on note encore :

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

On déduit de la forte convexité de K et de la définition de Λ la relation :

$$\|u^{(k)} - u^{\#}\|^2 \leq \frac{2}{b} \psi^{(k)},$$

Par rapport au théorème 2, on a perdu la propriété $\|p^{(k)} - p^{\#}\|^2 \leq 2\psi^{(k)}$, mais on a fait l’hypothèse que $\|p^{(k)}\|$ est bornée par R .

2. Majorations.

- Majoration de la variation $\|u^{(k+1)} - u^{(k)}\|$.

Par un raisonnement identique à celui fait dans la preuve du théorème 2, on a :

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

L’hypothèse GLB et le fait que le multiplicateur $p^{(k)}$ soit majoré en norme par R impliquent l’existence de constantes positives c_3 et c_4 telles que :

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

²⁹condition qui implique que J est coercive et que L est stable

- Majoration de la variation $\|p^{(k)} - p^\sharp\|^2$.

Avec les mêmes arguments que dans le théorème 2, on a :

$$\|p^{(k+1)} - p^\sharp\|^2 - \|p^{(k)} - p^\sharp\|^2 \leq (\rho L_\Theta)^2 \|u^{(k+1)} - u^\sharp\|^2 + 2\rho \langle p^{(k)} - p^\sharp, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \rangle.$$

Divisant de part et d'autre de cette inégalité par $\frac{\epsilon^{(k)}}{\rho}$, et la suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ étant décroissante, on obtient :

$$\frac{\epsilon^{(k+1)}}{\rho} \|p^{(k+1)} - p^\sharp\|^2 - \frac{\epsilon^{(k)}}{\rho} \|p^{(k)} - p^\sharp\|^2 \leq \epsilon^{(k)} \rho L_\Theta^2 \|u^{(k+1)} - u^\sharp\|^2 + 2\epsilon^{(k)} \langle p^{(k)} - p^\sharp, \Theta(u^{(k+1)}) - \Theta(u^\sharp) \rangle.$$

- Majoration de la variation $(\psi^{(k+1)} - \psi^{(k)})$.

Comme dans la preuve du théorème 2, on a :

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

et donc, en négligeant le terme $-\frac{b}{2} \|u^{(k+1)} - u^{(k)}\|^2$:

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

Alors, avec l'hypothèse GLB, il existe des constantes positives c_5 et c_6 telles que :

$$\begin{aligned} \psi^{(k+1)} - \psi^{(k)} &\leq \epsilon^{(k)} \langle r^{(k)}, u^\sharp - u^{(k)} \rangle + \epsilon^{(k)} \langle p^\sharp, \Theta(u^\sharp) - \Theta(u^{(k)}) \rangle + \\ &\quad \frac{1}{2} \epsilon^{(k)} \rho L_\Theta^2 \|u^{(k+1)} - u^\sharp\|^2 + (\epsilon^{(k)})^2 (c_5 \psi^{(k)} + c_6). \end{aligned}$$

3. Convergence.

Prenant de part et d'autre de l'inégalité précédente l'espérance conditionnelle par rapport à la tribu $\mathcal{F}^{(k)}$ et utilisant la forte convexité de J , on obtient :

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

On a : $L(u^\sharp, p^\sharp) - L(\mathbf{U}^{(k)}, p^\sharp) \leq -\frac{a}{2} \|\mathbf{U}^{(k)} - u^\sharp\|^2$.³⁰ Avec la condition supplémentaire :

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

on peut trouver un coefficient $\delta > 0$ tel que l'on ait $\rho \frac{L_\Theta^2}{2} \leq a(1 - \delta)$ et on obtient :

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

Notant $a' = a(1 - \delta)$ et utilisant la majoration :

$$\begin{aligned} \epsilon^{(k)} a' \left(\|\mathbf{U}^{(k+1)} - u^\sharp\|^2 - \|\mathbf{U}^{(k)} - u^\sharp\|^2 \right) &= a' \epsilon^{(k)} \langle \mathbf{U}^{(k+1)} - \mathbf{U}^{(k)}, (\mathbf{U}^{(k+1)} - u^\sharp) + (\mathbf{U}^{(k)} - u^\sharp) \rangle \\ &\leq \frac{(a')^2}{2} \|\mathbf{U}^{(k+1)} - \mathbf{U}^{(k)}\|^2 + \frac{(\epsilon^{(k)})^2}{2} \|(\mathbf{U}^{(k+1)} - u^\sharp) + (\mathbf{U}^{(k)} - u^\sharp)\|^2 \\ &\leq \frac{(a')^2}{2} \|\mathbf{U}^{(k+1)} - \mathbf{U}^{(k)}\|^2 + (\epsilon^{(k)})^2 \|\mathbf{U}^{(k+1)} - u^\sharp\|^2 + (\epsilon^{(k)})^2 \|(\mathbf{U}^{(k)} - u^\sharp)\|^2 \end{aligned}$$

on conclut à l'existence de constantes positives c_7 , c_8 et c_9 telles que :

$$\mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) - \psi^{(k)} \leq \epsilon^{(k)} 2\delta \left(L(u^\sharp, p^\sharp) - L(\mathbf{U}^{(k)}, p^\sharp) \right) + (\epsilon^{(k)})^2 (c_7 \psi^{(k)} + c_8 + c_9 \mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right)).$$

La suite de la démonstration est identique à celle du théorème 2, et on en déduit que la suite $\{\mathbb{E}(\psi^{(k)})\}_{k \in \mathbb{N}}$ est positive et bornée par un réel $M > 0$ et que le processus $\{\psi^{(k)}\}_{k \in \mathbb{N}}$ est une quasi-martingale convergente.

³⁰car l'application $u \rightarrow L(u, p^\sharp)$ est fortement convexe de même module que J et atteint son minimum en u^\sharp

4. Limites.

Comme dans le théorème 2, avec de plus la forte convexité de J et donc l'unicité de la solution u^\sharp , on conclut que la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ engendrée par l'algorithme 4 est bornée presque sûrement et qu'elle converge vers u^\sharp (unicité du point d'accumulation). □

Remarque 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 $p^{(k)}$. □

4.5.2 Cas du lagrangien augmenté

On se place maintenant dans le cas où la fonction J est *simplement convexe*. On dispose encore d'un algorithme de calcul des solutions du problème (\mathbf{P}_C) et du théorème de convergence associé pourvu que l'on fasse appel au lagrangien augmenté plutôt qu'au lagrangien simple.³¹ On rappelle que le lagrangien augmenté du problème (\mathbf{P}_C) s'écrit :

$$L_c(u, p) = J(u) + \zeta_c(\Theta(u), p) ,$$

l'expression de la fonction ζ_c étant donnée par :

$$\zeta_c(\theta, p) = \frac{1}{2c} (\|\text{proj}_{C^*}(p + c\theta)\|^2 - \|p\|^2) ,$$

et l'expression de ses gradients partiels étant :

$$\begin{aligned} \nabla_p \zeta_c(\theta, p) &= \frac{1}{c} (\text{proj}_{C^*}(p + c\theta) - p) , \\ \nabla_\theta \zeta_c(\theta, p) &= \text{proj}_{C^*}(p + c\theta) . \end{aligned}$$

On sait écrire dans le cas *déterministe* un algorithme utilisant à la fois le principe du problème auxiliaire et le lagrangien augmenté (voir [4, Ch. 4] pour une présentation complète). La k -ème itération de cet algorithme s'écrit :

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

et sa principale caractéristique est que l'on peut prouver sa convergence sans qu'une hypothèse de forte convexité soit nécessaire sur la fonction J .

L'extension au cas stochastique de cet algorithme du PPA avec lagrangien augmenté est la suivante.

Algorithme 5. (Algorithme du PPA stochastique "régularisé")

1. Choisir un point initial $(u^{(0)}, p^{(0)}) \in U^{\text{ad}} \times C^*$, ainsi qu'une suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ de réels positifs.
2. À l'itération k de l'algorithme, effectuer un tirage $w^{(k+1)}$ de la variable aléatoire \mathbf{W} suivant sa loi, indépendamment des tirages $(w^{(1)}, \dots, w^{(k)})$ des itérations précédentes.
3. Calculer le point $(u^{(k+1)}, p^{(k+1)})$ par :
$$\begin{aligned} u^{(k+1)} &= \arg \min_{u \in U^{\text{ad}}} K(u) + \langle \epsilon^{(k)} \nabla_u j(u^{(k)}, w^{(k+1)}) - \nabla K(u^{(k)}), u \rangle + \\ &\quad \epsilon^{(k)} \langle \nabla_\theta \zeta_c(\Theta(u^{(k)}), p^{(k)}), \Theta(u) \rangle , \\ p^{(k+1)} &= p^{(k)} + \epsilon^{(k)} \nabla_p \zeta_c(\Theta(u^{(k+1)}), p^{(k)}) . \end{aligned}$$
4. Incrémenter l'indice k de 1 et retourner à l'étape 2.

³¹Voir le §10.3 pour une présentation du lagrangien augmenté et de ces principales propriétés.

Remarque 13. Par rapport à l'utilisation d'un lagrangien ordinaire, on note que :

- dans la minimisation en u , on utilise la valeur $\nabla_{\theta}\zeta_c(\Theta(u^{(k)}), p^{(k)}) = \text{proj}_{C^*}(p^{(k)} + c\Theta(u^{(k)}))$ du multiplicateur (au lieu de la valeur $p^{(k)}$ dans le cas du lagrangien ordinaire) ; ceci correspond à une sorte d'anticipation de la remise à jour du multiplicateur dans le calcul de $u^{(k+1)}$.
- le multiplicateur $p^{(k+1)}$ n'est pas projeté sur le cône dual C^* (alors que c'est le cas lorsque l'on utilise un lagrangien ordinaire) ;
- à l'aide de l'expression de $\nabla_p\zeta_c(\Theta(u^{(k+1)}), p^{(k)})$, on peut interpréter l'étape de remise à jour de $p^{(k+1)}$ comme la succession des deux étapes suivantes :
 - projection : $p^{(k+\frac{1}{2})} = \text{proj}_{C^*}(p^{(k)} + c\Theta(u^{(k+1)}))$,
 - relaxation : $p^{(k+1)} = \left(1 - \frac{\epsilon^{(k)}}{c}\right)p^{(k)} + \frac{\epsilon^{(k)}}{c}p^{(k+\frac{1}{2})}$;

On note que, si l'on choisit $p^{(0)} \notin C^*$, la suite $\{p^{(k)}\}_{k \in \mathbb{N}}$ n'appartient pas nécessairement à C^* (alors que la suite $\{p^{(k+\frac{1}{2})}\}_{k \in \mathbb{N}}$ y appartient toujours). \square

On donne alors les conditions de convergence de l'algorithme 5.

Théorème 4.

On suppose que les hypothèses suivantes sont vérifiées.

- H0** U^{ad} est une partie convexe fermée non vide de \mathbb{U} , espace de Hilbert de dimension finie ; \mathbf{W} est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace (\mathbb{W}, \mathbb{W}) , de loi μ ; C est un cône convexe fermé saillant de \mathbb{V} , espace de Hilbert de dimension finie.
- H1** La variable aléatoire $j(u, \mathbf{W}) : \Omega \rightarrow \mathbb{R}$ est mesurable et son espérance existe pour tout $u \in U^{\text{ad}}$.
- H2** La fonction $j(\cdot, w) : \mathbb{U} \rightarrow \mathbb{R}$ est propre, convexe, semi-continue inférieurement, différentiable pour tout $w \in \mathbb{W}$, et son gradient partiel en u est noté $\nabla_u j(u, w)$.
- H3** La fonction $j(\cdot, w)$ est à gradient linéairement borné (GLB) uniformément en w :
$$\exists c_1 > 0, \exists c_2 > 0, \forall w \in \mathbb{W}, \forall u \in U^{\text{ad}}, \|\nabla_u j(u, w)\| \leq c_1 \|u\| + c_2 .$$
- H4** La fonction Θ est C -convexe, lipschitzienne de rapport L_{Θ} .
- H5** La fonction J est coercive sur U^{ad} et les contraintes sont qualifiées.
- H6** La fonction K est propre, fortement convexe de module b , différentiable.
- H7** La suite $\{\epsilon^{(k)}\}_{k \in \mathbb{N}}$ est une σ -suite.

On a alors les conclusions suivantes.

- C1** Le problème $(\mathbf{P}_{\mathbf{C}})$ admet un ensemble de points selle $U^{\sharp} \times P^{\sharp}$ non vide.
- C2** Le problème de minimisation en u dans l'algorithme 5 admet une solution $u^{(k+1)}$ unique.
- C3** Les deux suites $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ et $\{\mathbf{P}^{(k)}\}_{k \in \mathbb{N}}$ engendrées par l'algorithme 5 sont bornées presque sûrement, tout point d'accumulation d'une réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ appartient presque sûrement à U^{\sharp} , ensemble des solutions du problème $(\mathbf{P}_{\mathbf{C}})$.

Preuve. Pour alléger les écritures, on utilise les notations :

$$r^{(k)} = \nabla_u j(u^{(k)}, \omega^{(k+1)}) \quad , \quad q^{(k)} = \text{proj}_{C^*} \left(p^{(k)} + c\Theta(u^{(k)}) \right) \quad , \quad p^{(k+\frac{1}{2})} = \text{proj}_{C^*} \left(p^{(k)} + c\Theta(u^{(k+1)}) \right) .$$

La démonstration des points **C1** et **C2** découle des théorèmes généraux relatifs à l'optimisation avec contraintes explicites. Le fait que $u^{(k+1)}$ soit solution du problème de minimisation en u est caractérisé par la condition d'optimalité :

$$\forall u \in U^{\text{ad}} \quad , \quad \left\langle \nabla K(u^{(k+1)}) - \nabla K(u^{(k)}) + \epsilon^{(k)} r^{(k)} \quad , \quad u - u^{(k+1)} \right\rangle + \epsilon^{(k)} \left\langle q^{(k)} \quad , \quad \Theta(u) - \Theta(u^{(k+1)}) \right\rangle \geq 0 .$$

1. Fonction de Lyapunov.

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

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

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 :

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

ce qui prouve que Λ est bornée inférieurement et coercive. Ces deux inégalités, associées au fait que Θ est lipschitzienne, impliquent l'existence de coefficients a_1, \dots, a_5 tels que :

$$\|q^{(k)}\| \leq a_1 + a_2 \sqrt{\psi^{(k)}} \quad , \quad \|p^{(k+\frac{1}{2})} - p^{(k)}\| \leq a_3 + a_4 \sqrt{\psi^{(k)}} + a_5 \sqrt{\psi^{(k+1)}} \quad ,$$

2. Majorations.

- Variation de $u^{(k)}$.

Par un raisonnement en tout point identique à celui de la démonstration du théorème 2, on montre que :

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

- Variation de $p^{(k)}$.

De la définition de $p^{(k+1)}$ et de la majoration de $\|p^{(k+\frac{1}{2})} - p^{(k)}\|$, on déduit l'existence de coefficients c_5, c_6 et c_7 tels que :

$$\|p^{(k+1)} - p^{(k)}\| \leq \epsilon^{(k)} \left(c_5 \sqrt{\psi^{(k)}} + c_6 \sqrt{\psi^{(k+1)}} + c_7 \right) .$$

- Variation de $\psi^{(k)}$.

On montre enfin la relation :

$$\psi^{(k+1)} - \psi^{(k)} \leq \epsilon^{(k)} \left(\left\langle r^{(k)} \quad , \quad u^\sharp - u^{(k)} \right\rangle + \zeta(\Theta(u^\sharp), p^{(k)}) - \zeta(\Theta(u^{(k)}), p^\sharp) \right) + (\epsilon^{(k)})^2 \left(c_8 \psi^{(k)} + c_9 + c_{10} \psi^{(k+1)} \right) .$$

3. Convergence.

À l'aide des majorations précédentes, on montre que :

$$\mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) - \psi^{(k)} \leq \epsilon^{(k)} \left(L_c(u^\sharp, \mathbf{P}^{(k)}) - L_c(\mathbf{U}^{(k)}, p^\sharp) \right) + (\epsilon^{(k)})^2 \left(c_8 \psi^{(k)} + c_9 + c_{10} \mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) \right) . \quad (50)$$

En combinant les deux inégalités du point selle du lagrangien augmenté L_c , on obtient que :

$$L_c(u^\sharp, \mathbf{P}^{(k)}) - L_c(\mathbf{U}^{(k)}, p^\sharp) \leq 0 \quad ,$$

d'où l'on déduit :

$$\mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) - \psi^{(k)} \leq \alpha^{(k)} \psi^{(k)} + \beta^{(k)} + \gamma^{(k)} \mathbb{E} \left(\psi^{(k+1)} \mid \mathcal{F}^{(k)} \right) \quad ,$$

où $\alpha^{(k)}$, $\beta^{(k)}$ et $\gamma^{(k)}$ sont les termes de séries convergentes.

Par un raisonnement analogue à celui fait dans la démonstration du théorème 2, on montre que la suite $\{\mathbb{E}(\psi^{(k)})\}_{k \in \mathbb{N}}$ est positive et bornée, et que la suite $\{\psi^{(k)}\}_{k \in \mathbb{N}}$ est une quasi-martingale qui converge presque sûrement vers une variable aléatoire intégrable.

4. Limites.

De manière similaire à la démonstration du théorème 2, on montre que les suites $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ et $\{\mathbf{P}^{(k)}\}_{k \in \mathbb{N}}$ sont bornées, et, à l'aide de la majoration (50) et du lemme 2, que la suite $\{L_c(u^\sharp, \mathbf{P}^{(k)}) - L_c(\mathbf{U}^{(k)}, p^\sharp)\}_{k \in \mathbb{N}}$ converge presque sûrement vers zéro. On en déduit que tout point d'accumulation d'une réalisation de la suite $\{\mathbf{U}^{(k)}\}_{k \in \mathbb{N}}$ est solution du problème (\mathbf{P}_C) . □

Remarque 14. On arrive même à montrer dans ce cas que les points d'accumulation de la suite des variables $\{\mathbf{P}^{(k)}\}_{k \in \mathbb{N}}$ ont des propriétés d'optimalité. On consultera [5] pour plus de détails. □

4.6 Conclusions

Le fait de pouvoir traiter des problèmes d'optimisation stochastique sous contraintes déterministes ne couvrent pas l'ensemble des cas que l'on peut rencontrer en pratique dans le cadre de la boucle ouverte. Une première extension conceptuellement simple est de prendre en compte des contraintes en *espérance* :

$$\Theta(u) = \mathbb{E}(\theta(u, \mathbf{W})) .$$

Dans l'esprit du gradient stochastique, ce cas peut être traité en profitant des itérations de gradient pour reconstituer l'espérance du gradient *et* des contraintes (voir [7]).

Cependant, le cas des contraintes en espérance n'est pas fréquent en pratique,³² et on préférerait pouvoir traiter des contraintes presque sûres :

$$\theta(u, \mathbf{W}(\omega)) \in -C \quad \text{pour presque tout } \omega ,$$

ou mieux encore des contraintes en probabilité

$$\mathbb{P}(\theta(u, \mathbf{W}) \in -C) \geq \pi .$$

Les contraintes presque sûres sont en général trop restrictives pour pouvoir être prises en compte dans le cadre de la boucle ouverte. Quant aux contraintes en probabilité, leur traitement direct entraîne de grandes difficultés mathématiques (perte de convexité, voir de connexité de l'ensemble admissible). Cependant, le fait de toujours pouvoir écrire une contrainte en probabilité comme une contrainte en espérance fait que l'on peut dans certains cas espérer, moyennant l'utilisation d'un lagrangien augmenté pour surmonter la non convexité, s'attaquer à ce type de problèmes. Mais il faut alors savoir optimiser par le gradient stochastique des fonctions non linéaires de l'espérance.

Ces différents points seront un peu plus détaillés dans le §6.

³²quel sens donner au respect d'une contrainte de satisfaction d'un besoin en espérance ?