

# Programmation stochastique

## Approximation par moyenne d'échantillonnage

Fabian Bastin

IFT-6512 – Hiver 2011

Considérons le problème stochastique de la forme

$$\min_{z \in S} g(z) = E_P [G(z, \xi)],$$

où  $z \in \mathbb{R}^m$  est un vecteur de variables de décision,  $S$  est un sous-ensemble compact de  $\mathbb{R}^m$  représentant les solutions réalisables du problème ci-dessus,  $\xi$  est un vecteur aléatoire réel défini sur l'espace de probabilité  $(\Xi, \mathcal{F}, P)$  et prenant des valeurs dans  $(\mathbb{R}^k, \mathcal{B}^k)$  ( $\mathcal{B}^k$  est la mesure de Borel),  $G : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}$  est une fonction à valeurs réelles, et  $E_P[\cdot]$  est l'espérance par rapport à la mesure  $P$ .

Nous supposons que pour chaque  $z \in S$ , le fonction de valeur espérée  $g(z)$  est bien définie, c'est-à-dire que la fonction  $G(z, \cdot)$  est  $\mathcal{F}$ -mesurable et  $P$ -intégrable.

Pour la simplicité, nous nous restreindrons dans un premier temps au cas où **S est déterministe**.

## Cadre de travail (2)

Si la fonction de répartition de  $\xi$  est continue ou discrète avec un grand nombre de réalisations possibles,  $g(z)$  est généralement très difficile à évaluer.

Résoudre le précédent problème devient dès lors difficile et nous devons nous tourner vers des approximations telles que les méthodes Monte Carlo : le problème originale est remplacé par des approximations successives obtenues en générant des tirs  $\xi_1, \dots, \xi_N$ . L'approximation pour un échantillon de taille  $N$  est

$$\min_{z \in S} \hat{g}_N(z) = \frac{1}{N} \sum_{i=1}^N G(z, \xi_i).$$

Nous désignerons parfois le programme original comme le **programme réel**, ou **programme en valeur espérée**, et le programme approché comme **approximation par moyenne d'échantillonnage** (**SAA** : sample average approximation).

# Convergence au premier ordre

Investigons la convergence des solutions et des valeurs optimales de la séquence de problèmes SAA vers une solution et une valeur optimale du programme en valeur espérée pour  $N \rightarrow \infty$ .

Soit  $z_N^*$  un point critique au premier ordre pour le problème approché, c'est-à-dire que les conditions KKT sont satisfaites en  $z_N^*$ . Afin de souligner la dépendance de  $z_N^*$  vis-à-vis des tirs successifs  $\xi_1, \dots, \xi_N$ , nous utiliserons souvent la notation  $z_N^*(\xi_1, \dots, \xi_N)$ , ou  $z_N^*(\bar{\xi})$ , puisque  $(\xi_1, \dots, \xi_N)$  peut être vu que la troncature finie d'une séquence infinie  $\bar{\xi} := \{\xi_k\}_{k=1}^\infty$ .

Puisque  $S$  est un ensemble compact, la séquence de solutions SAA a un ensemble un ensemble non-vide de points limites (finis). Sous quelle condition un tel point limite est-il un point critique au premier ordre pour le véritable problème ?

Puisque l'ensemble des points limites dépend de la séquence de réalisations  $\bar{\xi}$ , qui n'est pas connue a priori, nous devons introduire un espace de probabilité adapté sur lequel nous pouvons définir des variables aléatoires dont les réalisations sont de telles séquences (infinies).

Considérons le processus stochastique

$$\bar{\xi} = \{\xi_k\}_{k=1}^{\infty},$$

appelé plus tard le processus d'échantillonnage, où les vecteurs aléatoires  $\xi_k$ ,  $k = 1, \dots, \infty$ , sont supposés indépendants et identiquement distribués (i.i.d.).

# Espace de probabilité de dimension infinie

De la propriété i.i.d. et le théorème de consistance de Kolmogorov (voir par exemple Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1967, Chapitre V, Théorème 5.1), nous pouvons construire l'espace de probabilité de dimension infinie

$$(\Xi_{\infty}, \mathcal{F}_{\infty}, P_{\infty}),$$

où la mesure  $P_{\infty}$  a la propriété que pour n'importe quel naturel non nul  $j$ ,

$$P_{\infty}[B] = \prod_{i=1}^j P[B_i],$$

pour n'importe quel ensemble  $B = \prod_{i=1}^j B_i \times \prod_{i=j+1}^{\infty} \Xi$ , avec  $B_i \in \mathcal{F}$ ,  $i = 1, \dots, j$ .

En d'autres termes, les mesures marginales définies sur  $\prod_{i=1}^j (\Xi, \mathcal{F})$ , avec  $j$  fini  $j$  ( $j = 1, \dots$ ), correspondent aux mesures produits  $\prod_{i=1}^j P$ , comme attendu.

Nous pouvons dès lors voir  $\bar{\xi}$  comme une variable aléatoire sur  $(\Xi_{\infty}, \mathcal{F}_{\infty}, P_{\infty})$ , dont les réalisations sont les processus formés par les tirs successifs  $\xi_k$ ,  $k = 1, \dots, \infty$ , i.e.  $\bar{\xi} = \{\xi_k\}_{k=1}^{\infty}$ .

## Notations :

- $\xrightarrow{p.s.}$  : convergence presque sûre ;
- $\xrightarrow{P}$  : convergence en probabilité ;
- $\Rightarrow$  : convergence en distribution.

# Rappels de probabilité : types de convergence

## Définition (Convergence presque sûre (ou avec probabilité un))

*Une séquence de variables aléatoires  $\{\xi_n\}_{n=1}^{+\infty}$  converge presque sûrement vers une variable aléatoire  $\xi$  si*

$$P_{\xi}[\{\omega \mid \xi_n(\omega) \rightarrow \xi(\omega) \text{ as } n \rightarrow \infty\}] = 1.$$

Un mode plus faible de convergence est défini ci-après.

## Définition (Convergence en probabilité)

*La séquence de variables aléatoires  $\{\xi_n\}_{n=1}^{+\infty}$  converge en probabilité vers la variable aléatoire  $\xi$  si pour tout  $\epsilon > 0$ ,*

$$\lim_{n \rightarrow \infty} P_{\xi}[\{\omega \mid |\xi_n(\omega) - \xi(\omega)| \geq \epsilon\}] = 0.$$

## Rappels de probabilité : types de convergence (2)

La convergence presque sûre implique la convergence en probabilité ; cependant, l'inverse n'est pas vrai.

### Définition (Convergence en distribution)

*Etant donné une séquence de variables aléatoires  $\{\xi_n\}_{n=1}^{\infty}$  et une variable aléatoire  $\xi$ ,  $\xi_n$  converge en distribution vers  $\xi$  si  $\xi_n$  a la fonction de répartition  $F_n$  et  $\xi$  a la fonction de répartition  $F$ , et  $F_n(x) \rightarrow F(x)$  ponctuellement pour chaque  $x \in C$ , où  $C \subseteq \mathbb{R}$  est l'ensemble des points où  $F$  est continue.*

*Nous noterons aussi  $\xi_n \Rightarrow \xi$ .*

Dire qu'une séquence de variables aléatoires converge en distribution signifie seulement que la variable aléatoire limite a la distribution donnée. Si à la fois  $\xi$  et  $\zeta$  suivent la distribution spécifiée par  $F$ , alors  $\xi_n \Rightarrow \xi$  et  $\xi_n \Rightarrow \zeta$  sont des énoncés équivalents.

Dans ce qui suit, et à moins que ce soit explicitement énoncé, nous supposons que les termes *presque partout* et *presque sûrement* réfèrent à l'espace de dimension infinie  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$ , ce qui nous permet la considération explicite des ensembles de réalisations dont les éléments sont de la forme  $\{\xi_N\}_{N=1}^{\infty}$ .

(En d'autres mots, les résultats exprimés en ces termes tiennent pour presque chaque processus échantionnal.)

Une référence à un autre espace de probabilité sera dénoté en préfixant les termes *presque partout* et *presque sûrement* par la mesure définie sur cet espace de probabilité, et l'expression *presque tout processus échantionnal* sera préfixée par la mesure de probabilité associée à l'espace de probabilité de chaque élément du processus.

**A.0** Les tirs aléatoires  $\{\xi_k\}_{k=1}^{\infty}$  sont indépendants et identiquement distribués.

**A.1** Pour  $P$ -presque tout  $\xi$ , la fonction  $G(\cdot, \xi)$  est continûment différentiable sur  $S$ .

**A.2** La famille  $G(z, \xi)$ ,  $z \in S$ , est dominée par une fonction  $P$ -intégrable  $K(\xi)$ , i.e.  $E_P[K]$  est finie et  $|G(z, \xi)| \leq K(\xi)$  pour tout  $z \in S$  et  $P$ -presque chaque  $\xi$ .

**A.1** implique évidemment que  $G(\cdot, \xi)$  est  $P$ -presque sûrement continue. Ce fait, et **A.2**, sont des hypothèses typiques en théorie de programmation stochastique (voir par exemple Rubinstein et Shapiro, *Discrete Event Systems*, John Wiley & Sons, 1993). La forme plus forte de **A.1** est justifiée par notre intérêt dans les conditions d'optimalité au premier ordre, lesquelles sont exprimées en termes du gradient de l'objectif.

# Loi uniforme des grands nombres

**A.0–A.2** ensemble implique qu'il existe une **loi uniforme des grands nombres (LUGN)** sur  $S$ , pour l'approximation  $\hat{g}_N(z)$  de  $g(z)$  :

$$\sup_{z \in S} |\hat{g}_N(z) - g(z)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } N \rightarrow \infty.$$

Elle implique aussi que  $g(z)$  est alors continue sur  $S$ .

Rappel : la loi des grands nombres traite de la convergence d'une moyenne empirique vers l'espérance de la variable aléatoire étudiée. La convergence peut se faire en probabilité (loi faible) ou presque sûrement (loi forte).

Notre motivation d'étudier les conditions au premier ordre nous conduit également à ajouter une hypothèse supplémentaire sur le gradient.

**A.3** Chaque composante du gradient  $\frac{\partial}{\partial [z]_l} G(z, \xi)$  ( $l = 1, \dots, m$ ),  $z \in S$ , est dominée par une fonction  $P$ -intégrable.

# Conséquences

Cette nouvelle hypothèse nous permet de déduire que la fonction en valeur espérée  $g(z)$  est continûment différentiable sur  $S$ , et que les opérateurs de gradient et d'espérance peuvent être interchangés dans l'expression du gradient, donnant

$$\nabla_z g(z) = E_P [\nabla_z G(z, \xi)].$$

Ceci implique aussi que  $\nabla \hat{g}_N(z^*)$  est un estimateur sans biais de  $\nabla g(z^*)$ .

La **convergence au premier ordre** peut être dérivée des inégalités variationnelles stochastiques, comme présentées dans A. Shapiro, *Monte Carlo sampling methods*, dans A. Shapiro et A. Ruszczyński, éditeurs, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 353–425. Elsevier, 2003.

# Inégalités variationnelles stochastiques

Considérons un mapping  $\Phi : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^m$  et une multi-fonction  $\Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ . Supposons que l'espérance  $\phi(z) := E_P[\Phi(z, \xi)]$  est bien définie. Nous désignons à présent

$$\phi(z) \in \Gamma(z)$$

comme l'équation généralisée réelle, ou en valeur espérée, et dirons qu'un point  $z^* \in \mathbb{R}^m$  est une solution de l'équation généralisée si  $\phi(z^*) \in \Gamma(z^*)$ .

Si  $\{\xi_1, \dots, \xi_N\}$  est un échantillon aléatoire, nous désignons

$$\hat{\phi}_N(z) \in \Gamma(z)$$

comme l'équation SAA généralisée, où  $\hat{\phi}_N(z) = N^{-1} \sum_{i=1}^N \Phi(z, \xi_i)$ . Nous dénotons par  $S^*$  et  $S_N^*$  l'ensemble de (toutes) les solutions des équations généralisées en valeur espérée et SAA, respectivement.

# Convergence au premier ordre

Dénotons par  $d(x, A) := \inf_{x' \in A} \|x - x'\|$ , la distance de  $x \in \mathbb{R}^m$  à  $A$ , et  $D(A, B) := \sup_{x \in A} D(x, B)$ , la déviation de l'ensemble  $A$  à partir de l'ensemble  $B$ . Nous avons le résultat suivant (Shapiro, 2003).

## Théorème (Equations généralisées)

*Soit  $S$  un sous-ensemble compact de  $\mathbb{R}^m$  tel que  $S^* \subseteq S$ .*

*Supposons que*

- (a) la multi-fonction  $\Gamma(z)$  est fermée, c'est-à-dire que si  $z_k \rightarrow z$ ,  $y_k \in \Gamma(z_k)$  et  $y_k \rightarrow y$ , alors  $y \in \Gamma(z)$ ,*
- (b) le mapping  $\phi(z)$  est continu sur  $S$ ,*
- (c) presque sûrement,  $\emptyset \neq S_N^* \subseteq S$  pour  $N$  suffisamment grand, et*
- (d)  $\hat{\phi}_N(z)$  converge vers  $\phi(z)$  presque sûrement uniformément sur  $S$  comme  $N \rightarrow \infty$ .*

*Alors  $D(S_N^*, S^*) \rightarrow 0$  presque sûrement comme  $N \rightarrow \infty$ .*

# Conditions d'optimalité au premier ordre.

Notons l'ensemble réalisable d'un programme mathématique donné par  $\mathcal{A}$ . On supposera simplement ici que  $\mathcal{A}$  est fermé.

**Condition nécessaire** : si  $x^*$  est un minimum local,

$$-\nabla_x f(x^*) \in \mathcal{N}_{\mathcal{A}}(x^*),$$

où  $\mathcal{N}_{\mathcal{A}}(x^*)$  est le cône normal à  $\mathcal{A}$  en  $x^*$ .

**Question** : définition d'un cône ? (Encore des) définitions. . .

Un sous-ensemble  $C$  d'un espace vectoriel  $V$  est un **cône** (linéaire) si et seulement si  $\alpha x$  appartient à  $C$  pour n'importe quel  $x \in C$  et n'importe quel scalaire strictement positif  $\alpha$  de  $V$ . Il est pointé si  $\alpha$  peut être nul.

Un **cône convexe** est un cône qui est fermé sous les combinaisons convexe, i.e. si et seulement  $\alpha x + \beta y \in C, \forall \alpha, \beta$  non-négatifs, avec  $\alpha + \beta = 1$ .

# Cône tangent, cône normal

Cadre général. Nous dirons tout d'abord qu'un vecteur  $w \in \mathbb{R}^n$  est tangent à  $\mathcal{A}$  en  $x \in \mathcal{A}$  si pour toutes les séquences de vecteur  $\{x_i\}$  avec  $x_i \rightarrow x$ , et  $x_i \in \mathcal{A}$ , et toutes les séquences de scalaires positifs  $t_i \downarrow 0$ , il existe une séquence  $w_i \rightarrow w$  tel que  $x_i + t_i w_i \in \mathcal{A}$  pour tout  $i$ .

Le **cône tangent**  $T_{\mathcal{A}}(x)$  est la collection de tous les vecteurs tangents à  $\mathcal{A}$  en  $x$ .

Le **cône normal**  $N_{\mathcal{A}}(x)$  est le complément orthogonal au cône tangent, c'est-à-dire

$$N_{\mathcal{A}}(x) = \{v \mid v^T w \leq 0, \forall w \in T_{\mathcal{A}}(x)\}.$$

Oui, mais, en pratique ???

## Cône tangent, cône normal (2)

La définition du cône normal se simplifie fortement si  $\mathcal{A}$  est convexe. En effet, on a alors

$$N_{\mathcal{A}}(x) = \{v \mid v^T(x - x_0) \geq 0, \forall x_0 \in \mathcal{A}\}.$$

La condition nécessaire d'optimalité devient alors  $\nabla_x f(\hat{x})(x - \hat{x}) \geq 0, \forall x \in \mathcal{A}$ .

Si  $f$  est convexe (i.e. on est dans le cadre de la programmation convexe), cette condition est de plus suffisante.

# Contraintes déterministes et convexes

Quand  $S$  est convexe, nous pouvons donc réécrire les conditions de criticalité au premier ordre pour un certain point  $z^*$  comme l'exigence que  $-\nabla_z g(z^*)$  appartienne au cône normal à  $S$  en  $z^*$ , dénoté par  $\mathcal{N}_S(z^*)$ .

Si  $S$  est de plus déterministe, les ensembles réalisables sont les mêmes pour les problèmes réel et SAA.

Le précédent théorème nous permet de prouver facilement la convergence au premier ordre. Considérons le choix  $\Gamma(\cdot) = \mathcal{N}_S(\cdot)$ ;  $\phi(z^*)$  appartient à  $\Gamma(z^*)$  si et seulement si

$$\langle \phi(z^*), u - z^* \rangle \leq 0, \quad \forall u \in S.$$

Nous désignerons de telles inégalités variationnelles comme inégalités variationnelles stochastiques.

L'hypothèse (a) du théorème tient toujours dans ce cas. Prenons  $\Phi(z, \xi) = -\nabla_z G(z, \xi)$  et représentons par  $S^*$  et  $S_N^*$  les ensembles de points critiques au premier ordre des équations généralisées réelle et SAA, respectivement.

- L'hypothèse (a) tient pas la définition de cône normal.
- Alors sous **A.0–A.3**, nous avons que  $\phi(z) = -\nabla_z g(z)$ , et que  $\phi(z)$  est un vecteur aléatoire continu sur  $S$ , portant l'hypothèse (b).
- L'hypothèse (d) résulte de la LUGN, tandis que **A.1** et la compacité de  $S$  assure l'hypothèse (c) en posant  $\mathcal{S} = S$ .

Le précédent théorème assure dès lors la criticalité au premier ordre à la limite comme  $N \rightarrow \infty$ , presque sûrement.

Sous de plus fortes hypothèses, il est aussi possible de prouver la convergence au premier ordre presque sûrement quand  $S$  est non-convexe ou non-déterministe.

Nous supposons à présent que l'ensemble réalisable peut être décrit par des contraintes d'égalité et d'inégalité. Le problème original devient alors

$$\begin{aligned} \min_{z \in V} g(z) &= E_P[G(z, \xi)], \\ \text{tel que } c_j(z) &\geq 0, \quad j = 1, \dots, k, \\ c_j(z) &= 0, \quad j = k + 1, \dots, M, \end{aligned}$$

où  $V$  est un sous-ensemble compact de  $\mathbb{R}^m$ .

## Contraintes stochastiques (2)

Le problème SAA correspondant se définit comme

$$\begin{aligned} \min_{z \in V} \hat{g}_N(z), \\ \text{tel que } \hat{c}_{jN}(z) \geq 0, \quad j = 1, \dots, k, \\ \hat{c}_{jN}(z) = 0, \quad j = k + 1, \dots, M. \end{aligned}$$

Ici, pour chaque  $j = 1, \dots, M$ ,  $\{\hat{c}_{jN}(\cdot)\}$  est une séquence de fonctions (aléatoires) réelles convergeant asymptotiquement vers la fonction correspondante  $c_j(\cdot)$  comme  $N \rightarrow \infty$ .

Nous supposons que les fonctions  $c_j(\cdot)$  peuvent être représentées sous forme d'espérances :

$$c_j(z) = E_P[H_j(z, \xi)], \quad j = 1, \dots, M.$$

## Contraintes stochastiques (3)

Ces fonctions peuvent alors être estimées par les fonctions de moyenne échantillonnale correspondantes :

$$\hat{c}_{jN}(z) = \frac{1}{N} \sum_{i=1}^N H_j(z, \xi_i).$$

Pour la simplicité, nous considérerons le problème mathématique paramétrique plus général

$$\begin{aligned} & \min_{z \in V} \hat{g}(z, \epsilon), \\ & \text{tel que } \hat{c}_j(z, \epsilon) \geq 0, \quad j = 1, \dots, k, \\ & \quad \quad \hat{c}_j(z, \epsilon) = 0, \quad j = k + 1, \dots, M, \end{aligned}$$

où  $\epsilon$  est un vecteur aléatoire de paramètres donnant les perturbations du programme approché, et  $g(\cdot)$ ,  $\hat{g}(\cdot, \epsilon)$ ,  $c_j(\cdot)$ ,  $\hat{c}_j(\cdot, \epsilon)$  sont supposant de classe  $C^2$  par rapport à  $z$ .

# Perturbations aléatoires

Nous supposons que la perturbation est de la forme

$$\epsilon = \epsilon(z, \bar{\xi}) = \left( \epsilon_g, \epsilon_{c_1}, \dots, \epsilon_{c_M}, \epsilon_{\nabla g}^T, \epsilon_{\nabla c_1}^T, \dots, \epsilon_{\nabla c_M}^T \right)^T,$$

où chaque composante est une fonction de  $\mathbb{R}^m \times \prod_{i=1}^{\infty} \mathbb{R}^k$  vers  $\mathbb{R}$  ou  $\mathbb{R}^m$ , et

$$\begin{aligned} \hat{g}(z, \epsilon) &= g(z) + \epsilon_g, \quad \nabla_z \hat{g}(z, \epsilon) = \nabla_z g(z) + \epsilon_{\nabla g}, \\ \hat{c}_j(z, \epsilon) &= c_j(z) + \epsilon_{c_j}, \quad \nabla_z \hat{c}_j(z, \epsilon) = \nabla_z c_j(z) + \epsilon_{\nabla c_j}, \quad j = 1, \dots, M. \end{aligned}$$

Nous définirons aussi  $\epsilon_N(z, \bar{\xi})$  comme

$$\epsilon_N(z, \bar{\xi}) = \begin{pmatrix} \hat{g}_N(z) - g(z) \\ \hat{c}_{jN}(z) - c_j(z), \quad j = 1, \dots, M \\ \nabla_z \hat{g}_N(z) - \nabla_z g(z) \\ \nabla_z \hat{c}_{jN}(z) - \nabla_z c_j(z), \quad j = 1, \dots, M \end{pmatrix},$$

## Perturbations aléatoires (2)

Nous dénoterons le vecteur aléatoire correspondant par  $\epsilon_N(z, \bar{\xi})$ , et nous supposerons que la LUGN tient pour l'objectif et les contraintes, ainsi que pour leurs dérivées correspondantes.

Nous supposons finalement que les ensembles réalisables pour le problème original et approché sont non vides. Les lagrangiens du programme réel et du programme approché sont respectivement

$$\mathcal{L}(z, \lambda) = g(z) - \sum_{j=1}^M [\lambda]_j c_j(z) \quad \text{et} \quad L(z, \lambda, \epsilon) = \hat{g}(z, \epsilon) - \sum_{j=1}^M [\lambda]_j \hat{c}_j(z, \epsilon).$$

Soit  $z^*(\epsilon)$  un point critique au premier ordre du programme approché, et supposons que  $z^*(\epsilon)$  appartient à l'intérieur de  $V$ , que nous dénoterons par  $\overset{o}{V}$ .

Il existe alors des multiplicateurs de Lagrange  $\lambda^*(\epsilon)$  tels que  $(z^*(\epsilon), \lambda^*(\epsilon))$  satisfait les conditions KKT ; en d'autres termes  $(z^*(\epsilon), \lambda^*(\epsilon))$  est solution du système

$$\nabla_z L(z, \lambda, \epsilon) = 0,$$

$$[\lambda]_j \hat{c}_j(z, \epsilon) = 0, \quad j = 1, \dots, M,$$

$$\hat{c}_j(z, \epsilon) = 0, \quad j = k + 1, \dots, M,$$

$$\hat{c}_j(z, \epsilon) \geq 0, \quad j = 1, \dots, k,$$

$$[\lambda]_j(\epsilon) \geq 0, \quad j = 1, \dots, k.$$

Considérons maintenant un processus d'échantillonnage particulier  $\bar{\xi}$ . Pour clarifier la dépendance des points critiques au premier ordre au processus d'échantillonnage, nous écrirons  $z_N^*(\bar{\xi})$  pour  $z^*(\epsilon_N)$  et  $\lambda_N^*(\bar{\xi})$  pour  $\lambda^*(\epsilon_N)$ .

Soit  $\mathcal{Z}(\{z_N^*(\bar{\xi})\})$  l'ensemble des points d'accumulation de la séquence  $\{z_N^*(\bar{\xi})\}_{N=1}^{\infty}$ . La compacité de  $V$  implique que pour chaque  $\bar{\xi}$ ,  $\mathcal{Z}(\{z_N^*(\bar{\xi})\})$  est non vide.

Nous pouvons à présent prouver la convergence au premier ordre pour le cas général.

# Convergence au premier ordre

## Théorème

Supposons que pour presque tout  $\bar{\zeta}$  dans  $(\Xi_N, \mathcal{F}_N, P_N)$ ,  $\epsilon_N(z, \bar{\zeta}) \rightarrow 0$  uniformément sur l'ensemble compact  $V$ , comme  $N \rightarrow \infty$ , et soit  $\bar{\xi}$  dans  $(\Xi_N, \mathcal{F}_N, P_N)$  satisfaisant l'hypothèse de convergence uniforme.

Si  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\})$  appartient à  $\overset{o}{V}$ , et une sous-séquence  $\{z_\ell^*(\bar{\xi})\}_{\ell=1}^\infty \subseteq \{z_N^*(\bar{\xi})\}_{N=1}^\infty$  convergeant vers  $z^*$  est associée avec une séquence de multiplicateurs de Lagrange  $\{\lambda_\ell^*(\bar{\xi})\}_{\ell=1}^\infty$  ayant au moins un point limite, alors  $z^*$  est un point critique au premier ordre pour le problème réel.

## Démonstration.

A partir de nos hypothèses, la séquence  $\{(z_\ell^*(\bar{\xi}), \lambda_\ell^*(\bar{\xi}))\}$ ,  $\ell = 1, \dots, \infty$ , a un point limite  $(z^*, \lambda^*)$ . La propriété de convergence uniforme implique que  $(z^*, \lambda^*)$  satisfait les conditions KKT du problème réel, de sorte que  $z^* \in \overset{o}{V}$  est un point critique au premier ordre pour le problème original. □

# Sur les multiplicateurs de Lagrange

L'hypothèse de convergence des multiplicateurs de Lagrange tient toujours si les multiplicateurs restent bornés. Cette hypothèse plus forte permet d'utiliser les résultats de convergence sur les inégalités variationnelles pour prouver la criticalité au premier ordre des points limites (Shapiro, 2003).

Soit  $\mu := (z, \lambda) \in \mathbb{R}^{m+M}$  et  $\mathcal{K} := \mathbb{R}^m \times \mathbb{R}_+^k \times \mathbb{R}^{M-k} \subset \mathbb{R}^{m+M}$ .

Définissons

$$\begin{aligned}\phi(\mu) &= (\nabla_z \mathcal{L}(z, \lambda), c_{k+1}(z), \dots, c_M(z)), \\ \hat{\phi}_N(\mu) &= (\nabla_z L(z, \lambda, \epsilon_N), \hat{c}_{k+1}(z, \epsilon_N), \dots, \hat{c}_M(z, \epsilon_N)).\end{aligned}$$

L'équation généralisée  $\phi(\mu) \in \mathcal{N}_{\mathcal{K}}(\mu)$  représente alors les conditions d'optimalité KKT pour le problème d'optimisation original,  $S^* \subseteq \overset{o}{V}$ , et le théorème sur les équations généralisées implique alors la criticalité au premier ordre presque sûrement, avec  $\Gamma(\mu) := \mathcal{N}_{\mathcal{K}}(\mu)$ .

## Sur les multiplicateurs de Lagrange (2)

Les hypothèses (a) et (d) du théorème sur les équations généralisée sont satisfaites puisque  $\epsilon \rightarrow 0$  presque sûrement, et l'hypothèse (c) tient car les ensembles réalisables pour les problèmes original et approximatifs sont non vides.

Passons à la partie difficile : le second ordre.

# Convergence au second ordre

Pour autant que nous renforçons nos hypothèses, nous pouvons aussi montrer que, presque sûrement, il existe des points limites dans  $\mathcal{Z}(\{z_N^*(\bar{\xi})\})$  qui sont des minimiseurs locaux.

Nous considérons d'abord le cas où  $S$  est déterministe et supposons que, pour un processus d'échantillonnage particulier  $\bar{\xi}$ ,  $z_N^*(\bar{\xi})$  est un minimiseur local de  $\hat{g}_N(z)$  sur  $S$ . Cela revient à dire que

$$\exists \delta_N(\bar{\xi}) \text{ t.q. } \forall z \in B(z_N^*(\bar{\xi}), \delta_N(\bar{\xi})) \cap S, \hat{g}_N(z_N^*(\bar{\xi})) \leq \hat{g}_N(z),$$

où  $B(x, d)$  est la boule ouverte centrée en  $x$  et de rayon  $d$ .

# Hypothèse de rigidité

(Bastin, Cirillo et Toint, 2006, Pasupathy, 2010)

Afin de montrer que  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\})$  est un minimiseur local de  $g(\cdot)$  sur  $S$ , nous devons dès lors montrer que pour une certaine sous-séquence  $\{z_\ell^*(\bar{\xi})\}_{\ell=1}^\infty \subseteq \{z_N^*(\bar{\xi})\}_{N=1}^\infty$  convergeant vers  $z^*$ , le voisinage dans lequel  $z_\ell^*(\bar{\xi})$  est un minimum local ne se réduit pas à un singleton comme  $\ell \rightarrow \infty$ .

Nous exprimons cette exigence avec l'hypothèse technique suivante.

**A.4** Il existe une certaine sous-séquence  $\{z_\ell^*(\bar{\xi})\}_{\ell=1}^\infty$  convergeant vers un certain  $z^*$ , avec  $\{z_\ell^*(\bar{\xi})\}_{\ell=1}^\infty \subseteq \{z_N^*(\bar{\xi})\}_{N=1}^\infty$ , et certaines constantes  $\delta_{z^*\bar{\xi}} > 0$  et  $\ell_{z^*\bar{\xi}} > 0$  telles que pour tout  $\ell \geq \ell_{z^*\bar{\xi}}$ ,

$$\forall z \in B\left(z_\ell^*(\bar{\xi}), \delta_{z^*\bar{\xi}}\right) \cap S, \hat{g}_\ell(z_\ell^*(\bar{\xi})) \leq \hat{g}_\ell(z).$$

# Un théorème basique

## Théorème

*Supposons que **A.0–A.3** tiennent. Alors pour presque tout processus d'échantillonnage  $\bar{\xi}$ ,  $\{z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\})$  satisfaisant **A.4** est un ensemble de minima locaux de  $g(\cdot)$  sur  $S$ .*

## Démonstration.

*Voir Bastin, Cirillo et Toint, *Convergence theory for nonconvex stochastic programming with an application to mixed logit*, Mathematical Programming, 108(2–3) :207–234 (2006). □*

L'hypothèse **A.4** est quelque peu artificielle et il est dès lors intéressant de chercher des conditions plus élégantes.

A la différence des analyses de sensibilité classiques, nous nous concentrons ici sur les conditions sous lesquelles un point limite d'une séquence de solutions approximatives est lui-même une solution du problème original.

Comme précédemment, nous considérons le cas où l'ensemble réalisable est décrit par un ensemble de contraintes d'égalité et d'inégalité, et nous supposons que presque sûrement  $\epsilon_N$  converge vers zéro uniformément sur  $V$ .

# Convergence des multiplicateurs de Lagrange

Considérons un processus d'échantillonnage particulier  $\bar{\xi}$  in  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$ , et  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\})$ . Sous certaines conditions, si la sous-séquence  $\{z_{\ell}^*\}_{\ell=1}^{\infty}$  converge vers  $z^*$ , la séquence de vecteurs de Lagrange associés  $\{\lambda_{\ell}^*\}$  converge aussi vers un certain vecteur de multiplicateur de Lagrange  $\lambda^*$  associé à  $z^*$  pour le problème réel.

## Lemme

*Considérons un processus d'échantillonnage particulier  $\bar{\xi}$  in  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$  tel que  $\epsilon_N(z, \bar{\xi}) \rightarrow 0$  uniformément sur  $V$  comme  $N \rightarrow \infty$ . Si  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\}) \cap \overset{\circ}{V}$ ,  $\{z_{\ell}^*\}_{\ell=1}^{\infty} \subseteq \{z_N^*\}_{N=1}^{\infty} \cap \overset{\circ}{V}$  converge vers  $z^*$ , et qu'il existe un vecteur de multiplicateurs de Lagrange unique associé à  $z^*$  satisfaisant les conditions KKT, alors  $\lambda_{\ell}^*(\bar{\xi})$  converge vers  $\lambda^*$  comme  $\ell \rightarrow \infty$ .*

# Convergence des multiplicateurs de Lagrange (2)

## Démonstration.

De l'unicité de  $\lambda^*$ , la contrainte de qualification de Mangasarian-Fromowitz (MFCQ) tient en  $z^*$ , et dès lors dans un voisinage de  $z^*$ . Il en découle que les multiplicateurs de Lagrange sont uniformément bornés pour  $\epsilon$  proche de zéro. Il est dès lors suffisant de montrer que chaque point limite de la séquence  $\{\lambda_\ell^*(\bar{\xi})\}$ ,  $\ell = 1, \dots, \infty$ , est égal à  $\lambda^*$ .

Soit  $\lambda'$  un tel point limite. Par continuité,  $(z^*, \lambda')$  satisfait les conditions KKT, aussi  $\lambda'$  est égal à  $\lambda^*$ . □

L'unicité de  $\lambda^*$  peut être assurée avec une condition de qualification de contraintes adaptée, comme la condition de qualification de contraintes d'indépendance linéaire (LICQ : linear independence constraint qualification).

Cette qualification de contraintes sera particulièrement pratique pour notre discussion.

Considérons le programme original. Rappelons que l'ensemble actif  $\mathcal{A}(z)$  en n'importe quel point réalisable  $z$  est l'union de l'ensemble des indices des contraintes d'égalité et des indices des contraintes d'inégalité actives :

$$\mathcal{A}(z) = \{i \in \{1, \dots, k\} \mid c_i(z) = 0\} \cup \{k + 1, \dots, M\}.$$

## Définition

*Etant donné le point  $z^*$  et l'ensemble actif  $\mathcal{A}(z^*)$ , nous disons que la LICQ tient en  $z^*$  si l'ensemble des gradients des contraintes actives  $\{\nabla c_j(z^*), j \in \mathcal{A}(z^*)\}$  est linéairement indépendant.*

# Complémentarité stricte

Un autre concept utile pour nos besoins est la condition de complémentarité stricte.

## Définition

*Etant donné  $z^*$  et un vecteur  $\lambda^*$  satisfaisant les conditions KKT, nous disons que la condition de complémentarité stricte tient si exactement un des  $[\lambda]_j^*$  et  $c_j(z^*)$  est nul pour chaque index  $j = 1, \dots, k$ , i.e. nous avons que  $[\lambda]_j^* > 0$  pour chaque  $j \in \{1, \dots, k\} \cap \mathcal{A}(z^*)$ .*

Considérons à nouveau un processus d'échantillonnage particulier  $\bar{\xi}$  dans  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$ , tel que  $\epsilon_N(z, \bar{\xi}) \rightarrow 0$  comme  $N \rightarrow \infty$ . Si les hypothèses du lemme ci-dessus tiennent pour une certaine sous-séquence  $\{z_{\ell}^*\}_{\ell=1}^{\infty} \rightarrow z^*$ , le gradient du Lagrangien  $\nabla L(z_{\ell}^*(\bar{\xi}), \lambda_{\ell}^*(\bar{\xi}))$  converge vers  $\nabla \mathcal{L}(z^*, \lambda^*)$ , quand  $\ell$  tend vers l'infini, avec  $\lambda_{\ell}^*(\bar{\xi}) \rightarrow \lambda^*$ .

## Complémentarité stricte (2)

Supposons que la condition de complémentarité stricte tienne en  $z^*$  pour le problème approché. Nous obtenons que pour  $\ell$  suffisamment grand,  $[\lambda_\ell^*]_j(\bar{\xi})$ ,  $j \in \{1, \dots, k\} \cap \mathcal{A}(z^*)$ , sont strictement positifs et par conséquent, les contraintes correspondantes sont actives en  $z_\ell^*(\bar{\xi})$ .

De plus, vu que  $\epsilon_N(z, \bar{\xi}) \rightarrow 0$ ,  $\hat{c}_j(z_\ell^*(\bar{\xi}), \epsilon_\ell(z_\ell^*(\bar{\xi}), \bar{\xi})) \rightarrow c_j(z^*)$  et, pour  $\ell$  assez grand,  $\mathcal{A}(z_\ell^*(\bar{\xi})) = \mathcal{A}(z^*)$ , aussi la condition de complémentarité stricte tient en  $(z_\ell^*(\bar{\xi}), \lambda_\ell^*(\bar{\xi}))$  pour le problème approché.

Ceci nous permet d'énoncer un résultat de convergence au second ordre.

## Théorème (Second-order convergence)

Supposons que, pour presque tout processus d'échantillonnage  $\bar{\xi}$  dans  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$ , il existe un certain  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\}) \cap \overset{\circ}{V}$  associé à un vecteur de multiplicateurs de Lagrange unique  $\lambda^*$  et une certaine sous-séquence  $\{z_{\ell}^*\}_{\ell=1}^{\infty} \subseteq \{z_N^*\}_{N=1}^{\infty}$ , tel que  $z_{\ell}^*(\bar{\xi}) \rightarrow z^*$ , et

(a)  $\epsilon_N(z_N^*(\bar{\xi}), \bar{\xi}) \rightarrow 0$  uniformément sur  $V$ , comme  $N \rightarrow \infty$ ,

(b)  $z_N^*(\bar{\xi}) \in \overset{\circ}{V}$ ,  $N = 1, \dots$ ,

(c)  $\nabla_{zz}^2 \hat{g}(z_{\ell}^*(\bar{\xi}), \epsilon_{\ell}(z_{\ell}^*(\bar{\xi}), \bar{\xi})) \rightarrow \nabla_{zz}^2 g(z^*)$  comme  $\ell \rightarrow \infty$ ,

(d)  $\nabla_{zz}^2 \hat{c}_j(z_{\ell}^*(\bar{\xi}), \epsilon_{\ell}(z_{\ell}^*(\bar{\xi}), \bar{\xi})) \rightarrow \nabla_{zz}^2 c_j(z^*)$  ( $j = 1, \dots, M$ )  
comme  $\ell \rightarrow \infty$ .

Supposons aussi que la condition de complémentarité stricte et la LICQ tiennent en  $(z^*, \lambda^*)$  pour le problème réel.

# Un résultat de convergence au second ordre (2)

## Théorème (Convergence au second ordre (suite))

Alors, pour presque tout processus d'échantillonnage  $\bar{\xi}$ ,

- (i) la LICQ tient en  $z_\ell^*(\bar{\xi})$ , pour  $\ell$  assez grand,
- (ii)  $(z^*, \lambda^*)$  satisfait la condition nécessaire de second-ordre pour le programme réel :

$$w^T \nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*) w \geq 0, \text{ for all } w \in \text{Null} \left[ \nabla_z c_j(z^*)^T \right]_{j \in \mathcal{A}(z^*)}.$$

Si de plus,  $\exists$  une certaine constante  $\alpha_{\bar{\xi}} > 0$  t.q.,  $\forall \ell$  assez grand,

$$w^T \nabla_{zz}^2 L_\ell(z_\ell^*(\bar{\xi}), \lambda_\ell^*(\bar{\xi})) w > \alpha_{\bar{\xi}},$$

pour tout  $w \in \text{Null} \left[ \nabla_z \hat{c}_j(z_\ell^*(\bar{\xi}))^T \right]_{j \in \mathcal{A}(z_\ell^*(\bar{\xi}))}$ ,  $\|w\| = 1$ ,

alors  $(z^*, \lambda^*)$  satisfait presque sûrement les conditions suffisantes pour le problème réel.



# Un résultat de convergence au second ordre (2)

## Théorème (Convergence au second ordre (suite))

*En d'autres termes*

$$(iii) \mathbf{w}^T \nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*) \mathbf{w} > 0,$$

$$\text{for all } \mathbf{w} \in \text{Null} \left[ \nabla_z \mathbf{c}_j(z^*)^T \right]_{j \in \mathcal{A}(z^*)}, \|\mathbf{w}\| = 1.$$

*Cela revient à dire que  $z^*$  est un minimiseur isolé pour le problème original.*

## Démonstration.

Voir Bastin, Cirillo et Toint, 2006. □

Remarquons que la LICQ et la condition de complémentarité stricte impliquent que le minimum est isolé alors que la condition suffisante de second-ordre est généralement utilisée pour caractériser des minima locaux stricts.

Rappel : un minimum local isolé est aussi strict, mais l'inverse n'est pas toujours vérifié.

Si  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\}) \cap \overset{\circ}{V}$  est un minimiseur local strict mais non isolé, chaque voisinage de  $z^*$  contient d'autres minimiseurs locaux qui sont candidats pour être des points limites des séquences de solutions des problèmes SAA, comme  $N$  tend vers l'infini, et  $z^*$  peut dès lors être difficile à identifier.

L'hypothèse de non-dégénérescence

$$w^T \nabla_{zz}^2 L_\ell(z_\ell^*(\bar{\xi}), \lambda_\ell^*(\bar{\xi})) w > \alpha_{\bar{\xi}},$$

pour tout  $w \in \text{Null} \left[ \nabla_z \hat{c}_j(z_\ell^*(\bar{\xi}))^T \right]_{j \in \mathcal{A}(z_\ell^*(\bar{\xi}))}, \|w\| = 1,$

peut aussi être remplacée en exigeant que le Jacobien des équations d'égalité impliquées dans les conditions KKT associé au programme original

$$\begin{aligned} \nabla_z \mathcal{L}(z, \lambda) &= 0, \\ [\lambda]_j c_j(z) &= 0, \quad j = 1, \dots, M, \\ c_j(z) &= 0, \quad j = k + 1, \dots, M \end{aligned}$$

est non-singulier en  $(z^*, \lambda^*)$ .

# Convergence au second-ordre : corollaire

## Corollaire

*Supposons que, pour presque tout processus d'échantillonnage  $\bar{\xi}$  dans  $(\Xi_{\Pi}, \mathcal{F}_{\Pi}, P_{\Pi})$ , il existe un certain  $z^* \in \mathcal{Z}(\{z_N^*(\bar{\xi})\}) \cap \overset{\circ}{V}$  associé avec un unique vecteur de multiplicateurs de Lagrange  $\lambda^*$  et une certaine sous-séquence  $\{z_{\ell}^*\}_{\ell=1}^{\infty} \subseteq \{z_N^*\}_{N=1}^{\infty}$ , tels que  $z_{\ell}^*(\bar{\xi}) \rightarrow z^*$ , que les hypothèses (a)–(d) du théorème de convergence au second-ordre tiennent et que la condition de complémentarité stricte tiennent en  $(z^*, \lambda^*)$  pour le programme original.*

*Supposons de plus que le Jacobien des relations d'égalité des conditions KKT est non-singulier en  $(z^*, \lambda^*)$ . Il existe alors presque sûrement un certain  $z^*$  dans  $\mathcal{Z}(\{z_N^*(\bar{\xi})\}) \cap \overset{\circ}{V}$ , associé un certain vecteur  $\lambda^*$ , qui satisfait les conditions suffisantes au second-ordre pour le programme original.*

## Démonstration.

Voir Bastin, Cirillo et Toint, 2006.

L'implication inverse du théorème de convergence au second-ordre peut être obtenue à partir des résultats classiques en analyse de perturbations (A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, USA (1983), Théorème 3.2.2). D'avantage de développements dans le contexte de la programmation stochastique peut être trouvés dans Rubinstein et Shapiro (2003) et A. Shapiro, *Probabilistic constrained optimization : Methodology and applications*. Dans S. Uryasev, éditeur, *Statistical inference of stochastic optimization problems*, pages 282–304. Kluwer Academic Publishers, 2000.