Approximations of stochastic optimization problems subject to measurability constraints.

Pierre Carpentier and SOWG[1]

ENSTA — ENPC — EDF

August 30, 2007

XI. International Conference on Stochastic Programming

[1] **Systems and Optimization Working Group**:

K. Barty, J.-P. Chancelier, G. Cohen, A. Dallagi, M. de Lara, B. Seck & C. Strugarek

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\xi,\mathcal{F}) = \min_{\mathbf{u}\in L^2(\Omega,\mathcal{A},\mathbb{P};U)} \mathbb{E}\Big[j(\mathbf{u},\xi)\Big] , \tag{1a}$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} . \tag{1b}$$

- $(\Omega,\mathcal{A},\mathbb{P})$: probability space.
- ξ : random variable on $\Xi = \mathbb{R}^q$ (noise).
- $\mathbf{u}$: random variable on $U = \mathbb{R}^p$ (control).
- $\mathcal{F}$: subfield of $\mathcal{A}$, usually generated by a r.v. $\mathbf{y}$ (observation).

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] , \tag{1a}$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} . \tag{1b}$$

$\rightsquigarrow$ easily extended to the sequential control problem:

$$\min \mathbb{E}\left[\sum_{t=0}^{T-1} L_{t+1}\big(\mathbf{x}_t, \mathbf{u}_t, \xi_{t+1}\big) + K\big(\mathbf{x}_T\big)\right] ,$$

$$\text{subject to} \quad \begin{cases} \mathbf{x}_0 & = f_0\big(\xi_0\big) \\ \mathbf{x}_{t+1} & = f_{t+1}\big(\mathbf{x}_t, \mathbf{u}_t, \xi_{t+1}\big) \end{cases} ,$$

$$\mathbf{u}_t \text{ is } \sigma\big(\xi_0, \ldots, \xi_t\big)-\text{measurable} .$$

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] , \tag{1a}$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} . \tag{1b}$$

In order to obtain a tractable approximation of problem (1),

1. the random variable ξ in (1a) must be discretized,
2. and the σ-field $\mathcal{F}$ in (1b) must be discretized.

These two discretizations are a priori independent.

The first discretization is somewhat traditional (Monte Carlo), whereas the last one is not so well-known. . .

Strong convergence topology of σ-fields (Neveu)

Coarsest topology such that conditional expectation is continuous with respect to the σ-field:

$$\lim_{n\to+\infty} \mathcal{F}_n = \mathcal{F} \iff \lim_{n\to+\infty} \left\| \mathbb{E}\left[f \mid \mathcal{F}_n\right] - \mathbb{E}\left[f \mid \mathcal{F}\right] \right\|_{L^1} = 0 \;\; \forall f \in L^1 .$$

This notion of strong convergence, given using $L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R})$ can be equivalently defined using $L^r(\Omega, \mathcal{A}, \mathbb{P}; U)$, for $r \geq 1$ (Piccinini).

Main properties of the strong topology (Cotter)

1. The strong convergence topology is metrizable.
2. The set of σ-fields generated by a finite partition is dense.
3. If $\mathbf{y}_n \xrightarrow{\mathbb{P}} \mathbf{y}$ and $\sigma(\mathbf{y}_n) \subset \sigma(\mathbf{y})$, then $\sigma(\mathbf{y}_n) \to \sigma(\mathbf{y})$.

Known results

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] \ ,$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} \ .$$

In most discretization schemes (e.g. Pennanen '05 and Barty '04), the approximations of ξ and $\mathcal{F}$ are linked together...

How to devise a discretization scheme independent in ξ and $\mathcal{F}$?

More precisely, can we use the Monte Carlo method in order to discretize ξ, as for open-loop problems (Dupacova-Wets) ?

Known results

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] \ ,$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} \ .$$

In most discretization schemes (e.g. Pennanen '05 and Barty '04), the approximations of ξ and $\mathcal{F}$ are linked together...

How to devise a discretization scheme independent in ξ and $\mathcal{F}$?

More precisely, can we use the Monte Carlo method in order to discretize ξ, as for open-loop problems (Dupacova-Wets) ?

Known results

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\left[j(\mathbf{u}, \xi)\right] ,$$

$$\text{subject to} \quad \mathbf{u} \text{ is } \mathcal{F}-\text{measurable} .$$

In most discretization schemes (e.g. Pennanen '05 and Barty '04), the approximations of ξ and $\mathcal{F}$ are linked together...

How to devise a discretization scheme independent in ξ and $\mathcal{F}$?

More precisely, can we use the Monte Carlo method in order to discretize ξ, as for open-loop problems (Dupacova-Wets) ?

Formulation

- $\mathbf{x}$ and $\mathbf{w}$: independent uniformly distributed random variables on $[-1,1]$ (initial state and noise): $\xi = (\mathbf{x}, \mathbf{w})$.
- $\mathbf{u}$: random variable on $\mathbb{R}$ (control), measurable with respect to the initial state $\mathbf{x}$: $\mathcal{F} = \sigma(\mathbf{x})$.
- $\mathbf{z} = \mathbf{x} + \mathbf{u} + \mathbf{w}$ (final state).
- The problem is formulated on $\left([-1,1]^2, \mathcal{B}_{[-1,1]^2}, \mu\right)$:

$$\min_{\mathbf{u} \text{ is } \mathcal{F}-\text{measurable}} \mathbb{E}\left[\epsilon \mathbf{u}^2 + \mathbf{z}^2\right] \ . \tag{2}$$

Exact resolution using dynamic programming

$$u^\sharp(x) = -\frac{x}{1+\epsilon} \quad \text{and} \quad J^\sharp = V(\xi, \mathcal{F}) = \frac{1}{3}\left(1 + \frac{\epsilon}{1+\epsilon}\right) \ .$$

Formulation

- $\mathbf{x}$ and $\mathbf{w}$: independent uniformly distributed random variables on $[-1,1]$ (initial state and noise): $\xi = (\mathbf{x}, \mathbf{w})$.
- $\mathbf{u}$: random variable on $\mathbb{R}$ (control), measurable with respect to the initial state $\mathbf{x}$: $\mathcal{F} = \sigma(\mathbf{x})$.
- $\mathbf{z} = \mathbf{x} + \mathbf{u} + \mathbf{w}$ (final state).
- The problem is formulated on $\left([-1,1]^2, \mathcal{B}_{[-1,1]^2}, \mu\right)$:

$$\min_{\mathbf{u} \text{ is } \mathcal{F}-\text{measurable}} \mathbb{E}\left[\epsilon \mathbf{u}^2 + \mathbf{z}^2\right] . \tag{2}$$

Exact resolution using dynamic programming

$$u^\sharp(x) = -\frac{x}{1+\epsilon} \quad \text{and} \quad J^\sharp = V(\xi, \mathcal{F}) = \frac{1}{3}\left(1 + \frac{\epsilon}{1+\epsilon}\right) .$$

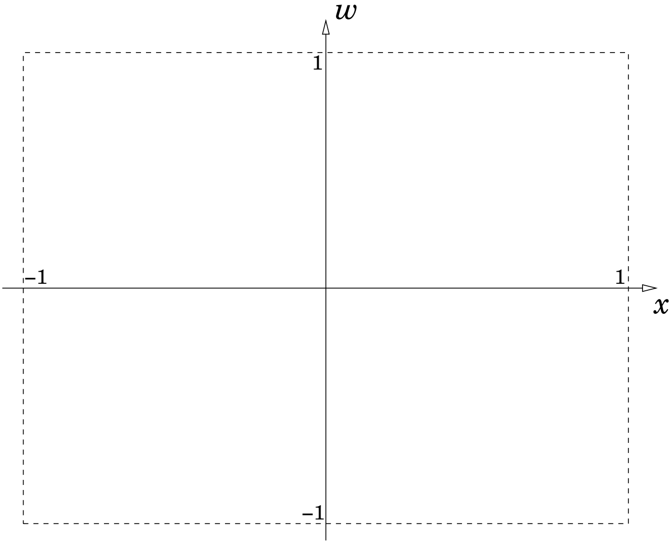


Figure: Discretization scheme.

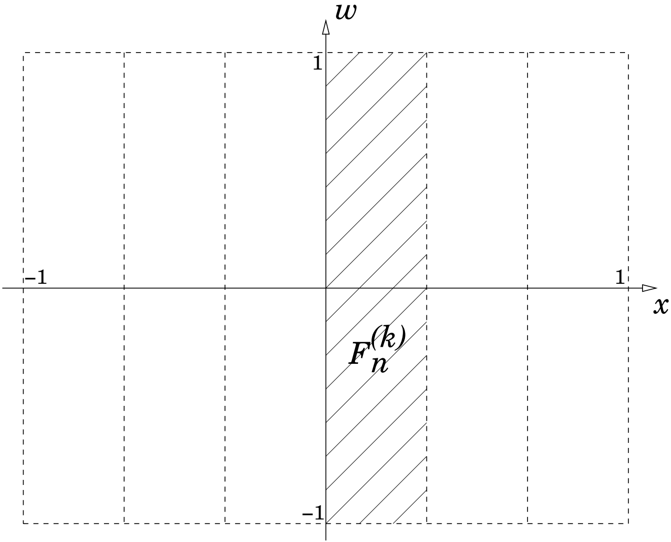


Figure: Discretization scheme.

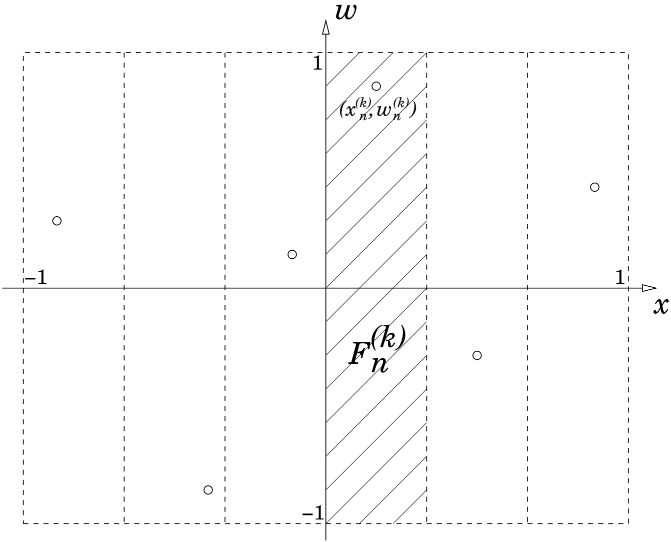


Figure: Discretization scheme.

Information

Let $n \in \mathbb{N}^\star$. Let $\left(F_n^{(1)}, \dots, F_n^{(n)}\right)$ be a partition of $[-1,1]^2$, with

$$F_n^{(k)} = \left(\frac{2(k-1)}{n} - 1, \frac{2k}{n} - 1 \right] \times [-1,1] \, .$$

Let $\mathcal{F}_n$ be the sub σ-field generated by $\left(F_n^{(1)}, \dots, F_n^{(n)}\right)$.

- $\left(\mathcal{F}_n\right)_{n\in\mathbb{N}}$ *strongly converges* to $\mathcal{F}$,
- $\mathbf{u}$ is $\mathcal{F}_n-$ measurable $\iff$ $\mathbf{u}$ is constant over each $F_n^{(k)}$

$$\iff \mathbf{u}(x,w) = \sum_{k=1}^{n} u_n^{(k)} \mathbb{I}_{F_n^{(k)}}(x,w).$$

Random variable

Let $(\zeta_n)_{n\in\mathbb{N}}$ be a deterministic sequence of elements in $[-1,1]^2$ such that the associated sequence of empirical probability laws *weakly converges* to μ. For $n \in \mathbb{N}^\star$ and $k \in \{1,\dots,n\}$, let

$$\left(x_n^{(k)}\ ,\ w_n^{(k)}\right) = \left(\frac{2k-1}{n} - 1 + \frac{\zeta_{k,1}}{n}\ ,\ \zeta_{k,2}\right)\ ,$$

and define the approximation $(\mathbf{x}_n, \mathbf{w}_n)$ of $(\mathbf{x}, \mathbf{w})$ by

$$(\mathbf{x}_n, \mathbf{w}_n) = \sum_{k=1}^{n} \left(x_n^{(k)}, w_n^{(k)}\right) \mathbb{I}_{F_n^{(k)}}(\mathbf{x}, \mathbf{w})\ .$$

- $(\mathbf{x}_n, \mathbf{w}_n)$ is constant over each subset $F_n^{(k)}$.
- $(\mathbf{x}_n, \mathbf{w}_n)_{n\in\mathbb{N}}$ *converges in distribution* to $(\mathbf{x}, \mathbf{w})$.

Approximated problem and solution

$$\min_{\left(u_n^{(1)},\dots,u_n^{(n)}\right)\in\mathbb{R}^n} \sum_{k=1}^{n} \int_{F_n^{(k)}} \left(\epsilon\left(u_n^{(k)}\right)^2 + \left(x_n^{(k)} + u_n^{(k)} + w_n^{(k)}\right)^2\right) \mu(\mathrm{d}x\mathrm{d}w) \; .$$

$$\widehat{u}_n^{(k)} = -\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \; .$$

Approximated feedback and associated cost

$$\widehat{\mathbf{u}}_n(x,w) = -\sum_{k=1}^{n} \frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \mathbb{I}_{F_n^{(k)}}(x,w) \;\rightsquigarrow\; \mathbb{E}\left[\epsilon \widehat{\mathbf{u}}_n^2 + \mathbf{z}^2\right] \longrightarrow \frac{2}{3} \; .$$

Discretization fails to asymptotically give the optimal solution.

Approximated problem and solution

$$\min_{\left(u_n^{(1)},\ldots,u_n^{(n)}\right)\in\mathbb{R}^n} \sum_{k=1}^{n} \int_{F_n^{(k)}} \left(\epsilon\left(u_n^{(k)}\right)^2 + \left(x_n^{(k)} + u_n^{(k)} + w_n^{(k)}\right)^2\right) \mu(\mathrm{d}x\mathrm{d}w) \ .$$

$$\widehat{u}_n^{(k)} = -\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \ .$$

Approximated feedback and associated cost

$$\widehat{\mathbf{u}}_n(x,w) = -\sum_{k=1}^{n} \frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \mathbb{I}_{F_n^{(k)}}(x,w) \ \rightsquigarrow \ \mathbb{E}\left[\epsilon\widehat{\mathbf{u}}_n^2 + \mathbf{z}^2\right] \longrightarrow \frac{2}{3} \ .$$

Discretization fails to asymptotically give the optimal solution.

Standard notions of convergence, but

- $\mathcal{F}$ and ξ are *independently* approximated:

  this makes possible to solve each open-loop subproblem using a *unique* sample of the random variable (a very poor way to compute conditional expectations).
- The convergence notion used for ξ is *weak*:

  $\{(\mathbf{x}_n, \mathbf{w}_n)\}_{n\in\mathbb{N}}$ does not converge *in probability* to $(\mathbf{x}, \mathbf{w})$.

Question: can we expect a convergence result when using a stronger convergence notion for the random variable?

Standard notions of convergence, but

- $\mathcal{F}$ and ξ are *independently* approximated:

  this makes possible to solve each open-loop subproblem using a *unique* sample of the random variable (a very poor way to compute conditional expectations).
- The convergence notion used for ξ is *weak*:

  $\{(\mathbf{x}_n, \mathbf{w}_n)\}_{n\in\mathbb{N}}$ does not converge *in probability* to $(\mathbf{x}, \mathbf{w})$.

Question: can we expect a convergence result when using a stronger convergence notion for the random variable?

Framework for the study of problem (1):

- $\boldsymbol{\xi} \in L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$ with $q \in [1, +\infty)$,
- $\mathbf{u} \in L^r(\Omega, \mathcal{A}, \mathbb{P}; U)$ with $r \in [1, +\infty)$,
- $\Delta(\mathcal{F})$ subset of $\mathcal{F}$-measurable control random variables:
$$\Delta(\mathcal{F}) = L^r(\Omega, \mathcal{F}, \mathbb{P}; U) \ ,$$
- j a normal integrand on $U \times \Xi$ and J the associated integral functional:
$$J(\mathbf{u}, \boldsymbol{\xi}) = \mathbb{E}\big[j(\mathbf{u}, \boldsymbol{\xi})\big] \ .$$

$$V(\boldsymbol{\xi}, \mathcal{F}) = \min_{\mathbf{u} \in \Delta(\mathcal{F})} J(\mathbf{u}, \boldsymbol{\xi}) \ .$$

Theorem

Under the following assumptions:

H1 $\{\xi_n\}_{n\in\mathbb{N}}$ converges to ξ in $L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$,

H2 $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ strongly converges to $\mathcal{F}$ and $\mathcal{F}_n \subset \mathcal{F}$,

H3 j is such that:

$$\forall (u,v),\ \forall (\xi,\zeta),\ |j(u,\xi) - j(v,\zeta)| \leq \alpha \|u - v\|_U^r + \beta \|\xi - \zeta\|_\Xi^q \ ,$$

the convergence of the approximated optimal costs holds true:

$$\lim_{n\to+\infty} V\big(\xi_n, \mathcal{F}_n\big) = V\big(\xi, \mathcal{F}\big) \ . \tag{3}$$

1. Same result with:

$$\Delta(\mathcal{F}) = \left\{ \mathbf{u} \in \mathcal{U},\ \mathbf{u}\ \ \mathcal{F} - \text{measurable},\ \mathbf{u}(\omega) \in U^{\mathrm{ad}}\ \ \mathbb{P} - as \right\}\ ,$$

U^{ad} being a closed convex subset of U.

2. Assumption **H3** in our theorem is far from being minimal, and can be alleviated using the tools of epi-convergence (see Chancelier for further details).

3. Numerical point of view:
   - $\left(\Omega_n^{(1)}, \ldots, \Omega_n^{(n)}\right)$ partition generating the σ-field $\mathcal{F}_n$,
   - $\left(\mho_n^{(1)}, \ldots, \mho_n^{(n)}\right)$ partition generated by ξ_n,

$$\min_{(u_n^{(1)}, \ldots, u_n^{(n)}) \in U^n} \sum_{i=1}^{n} \sum_{l=1}^{n} \mathbb{P}\left(\Omega_n^{(i)} \cap \mho_n^{(l)}\right) j\left(u_n^{(i)}, \xi_n^{(l)}\right) .$$

1. Same result with:

$$\Delta(\mathcal{F}) = \{\mathbf{u} \in \mathcal{U},\ \mathbf{u}\ \ \mathcal{F} - \text{measurable},\ \mathbf{u}(\omega) \in U^{\mathrm{ad}}\ \ \mathbb{P} - as\}\ ,$$

U^{ad} being a closed convex subset of U.

2. Assumption **H3** in our theorem is far from being minimal, and can be alleviated using the tools of epi-convergence (see Chancelier for further details).

3. Numerical point of view:
   - $(\Omega_n^{(1)}, \ldots, \Omega_n^{(n)})$ partition generating the σ-field $\mathcal{F}_n$,
   - $(\mho_n^{(1)}, \ldots, \mho_n^{(n)})$ partition generated by ξ_n,

$$\min_{(u_n^{(1)}, \ldots, u_n^{(n)}) \in U^n} \sum_{i=1}^{n} \sum_{l=1}^{n} \mathbb{P}(\Omega_n^{(i)} \cap \mho_n^{(l)})\, j(u_n^{(i)}, \xi_n^{(l)})\ .$$

1. Same result with:

$$\Delta(\mathcal{F}) = \{\mathbf{u} \in \mathcal{U},\ \mathbf{u}\ \ \mathcal{F} - \text{measurable},\ \mathbf{u}(\omega) \in U^{\mathrm{ad}}\ \ \mathbb{P} - as\}\ ,$$

U^{ad} being a closed convex subset of U.

2. Assumption **H3** in our theorem is far from being minimal, and can be alleviated using the tools of epi-convergence (see Chancelier for further details).

3. Numerical point of view:
   - $(\Omega_n^{(1)}, \ldots, \Omega_n^{(n)})$ partition generating the σ-field $\mathcal{F}_n$,
   - $(\mho_n^{(1)}, \ldots, \mho_n^{(n)})$ partition generated by $\boldsymbol{\xi}_n$,

$$\min_{(u_n^{(1)},\ldots,u_n^{(n)})\in U^n} \sum_{i=1}^{n} \sum_{l=1}^{n} \mathbb{P}\big(\Omega_n^{(i)} \cap \mho_n^{(l)}\big)\, j\big(u_n^{(i)}, \xi_n^{(l)}\big)\ .$$

Conclusions

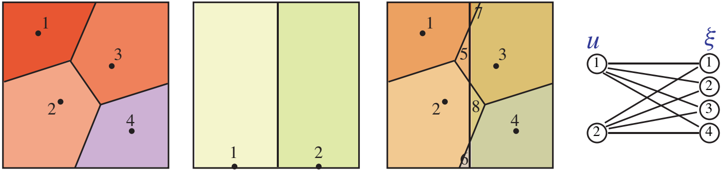
1. Another point of view on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.

Conclusions

1. Another point of view on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.

Conclusions

1. Another point of view on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.



K. Barty
Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique.
Thèse de doctorat de l'ENPC, june 2004.

J.-P. Chancelier and SOWG
Epi-convergence of stochastic optimization problems involving both random variables and measurability constraints approximations.
Preprint CERMICS, october 2006.

K.D. Cotter
Similarity of information and behavior with a pointwise convergence topology.
Journal of Mathematical Economics, 15, 1986.

T. Pennanen
Epi-Convergent Discretizations of Multistage Stochastic Programs.
Mathematics of Operations Research, Vol. 30, No. 1, 2005.

C. Strugarek and SOWG
On the Fortet-Mourier metric for the stability of Stochastic Optimization Problems, an example.
Stochastic Programming E-Print Series, 25, 2004.

Pennanen's discretization scheme for problem (1)

Suppose that $\mathcal{F} = \sigma(\mathbf{y})$, with $\mathbf{y} = h(\xi)$.

1. Approximate ξ by a finitely valued r.v. $\xi_n = q_n(\xi)$,

   and approximate $\mathcal{F}$ by $\mathcal{F}_n$ generated by $\mathbf{y}_n = h(\xi_n)$:

$$V(\xi_n, \mathcal{F}_n) = \min_{\mathbf{u} \text{ is } \mathcal{F}_n-\text{measurable}} \mathbb{E}\big[j(\mathbf{u}, \xi_n)\big] .$$

But... $\mathbf{u}$ is $\mathcal{F}_n-$ measurable $\not\Rightarrow$ $\mathbf{u}$ is $\mathcal{F}-$ measurable

Convergence theorem (Epi-convergence)

Main assumptions:

1. $\xi_n \longrightarrow \xi$ in probability.
2. $\sigma(h \circ q_n) \subset \sigma(h)$.

Barty's discretization scheme for problem (1)

1. Approximate $\mathcal{F}$ by $\mathcal{F}_k$ generated by a finite partition of Ω:

$$V(\xi, \mathcal{F}_k) = \min_{\mathbf{u} \text{ is } \mathcal{F}_k-\text{measurable}} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] .$$

2. Approximate ξ by a finitely valued random variable ξ_n:

$$V(\xi_n, \mathcal{F}_k) = \min_{\mathbf{u} \text{ is } \mathcal{F}_k-\text{measurable}} \mathbb{E}\big[j(\mathbf{u}, \xi_n)\big] .$$

Convergence theorem

1. **Information structure discretization error:**
$\big|V(\xi, \mathcal{F}) - V(\xi, \mathcal{F}_k)\big| \longrightarrow 0$ as $\mathcal{F}_k \longrightarrow \mathcal{F}$ strongly.
2. **Mean computation discretization error:**
$\big|V(\xi, \mathcal{F}_k) - V(\xi_n, \mathcal{F}_k)\big| \longrightarrow 0$ as $\xi_n \longrightarrow \xi$ in distribution.

$\limsup_{n\to+\infty} V\big(\xi_n, \mathcal{F}_n\big) \leq V\big(\xi, \mathcal{F}\big)$

- $\forall \mathbf{u} \in \Delta(\mathcal{F})$, define $\mathbf{u}_n = \mathbb{E}\left[\mathbf{u} \mid \mathcal{F}_n\right]$. Then, $\mathcal{F}_n \to \mathcal{F} \implies \mathbf{u}_n \to \mathbf{u}$. The set-valued mapping Δ is thus lsc.
- J being continuous, we conclude that the marginal function V is u.s.c.

$\liminf_{n\to+\infty} V\big(\xi_n, \mathcal{F}_n\big) \geq V\big(\xi, \mathcal{F}\big)$

- From $J(\mathbf{u}, \xi_n) = J(\mathbf{u}, \xi) + (J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi))$, we obtain:

$$\min_{\mathbf{u}\in\Delta(\mathcal{F}_n)} J(\mathbf{u}, \xi_n) \geq \min_{\mathbf{u}\in\Delta(\mathcal{F}_n)} J(\mathbf{u}, \xi) + \min_{\mathbf{u}\in\Delta(\mathcal{F}_n)} \Big(J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi)\Big) .$$

- Using $\mathcal{F}_n \subset \mathcal{F} \implies \Delta(\mathcal{F}_n) \subset \Delta(\mathcal{F})$, we deduce:

$$V(\xi_n, \mathcal{F}_n) \geq V(\xi, \mathcal{F}) + \min_{\mathbf{u}\in\Delta(\mathcal{F})} \Big(J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi)\Big) .$$

- The conclusion is again a consequence of **H3**.

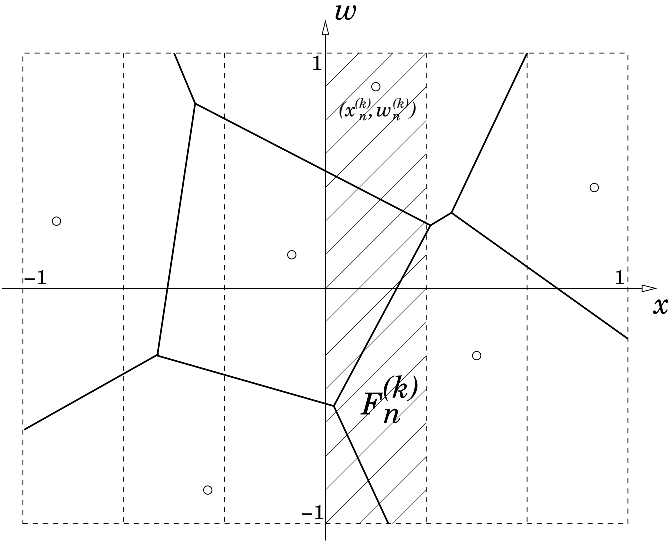


Figure: Voronoi cells.