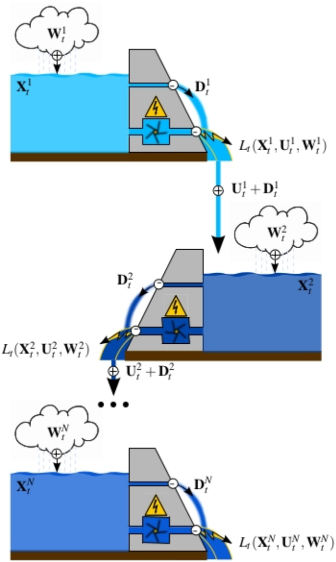
Constraint Qualification Conditions in Stochastic Optimal Problem

V. Leclère (ENPC), P. Carpentier (ENSTA)

January 17, 2013



- We are interested in the problem of managing a group of power stations.
- We want to find strategies (optimal feedback control).
- The problem is of too high dimension to be addressed by Dynamic Programming and thus we aim at decomposing it.
- In order to do that we need to dualize the coupling constraints.

Presentation Outline

Contents

Some Convex Analysis Facts and Notations

- If $f : E \to \overline{\mathbb{R}}$, where E is a Banach space, we note f^* the Fenchel conjugate of f defined on the topological dual of E noted E^* by
$$f^*(x^*) = \sup_x \langle x^*, x\rangle - f(x).$$
- The double conjugate of a function f (noted f^{**}) is the greatest convex l.s.c function lower than f, in particular $f^{**} \leq f$.
- For future use we say that $\Theta : E \to F$ is $C-$convex (with $C \subset F$) if for all $x, x' \in E$ and all $\alpha \in [0,1]$ whe have
$$\Theta\big(\alpha x + (1-\alpha)x'\big) - \big(\alpha\Theta(x) + (1-\alpha)\Theta(x')\big) \in -C.$$
- If Θ is $C-$convex and continuous then $\mathcal{U}^{ad} = \{u \in \mathcal{U} | \Theta(u) \in -C\}$ is a closed convex set.

Dualization by Perturbation

We assume that $\mathcal{U}$,$\mathcal{U}^*$, $\mathcal{Y}, \mathcal{Y}^*$ are paired spaces (for example Banach and their topological dual). We consider the family of perturbed minimization problem, $p \in \mathcal{Y}$ being the perturbation

$$(\mathcal{P}_p) \qquad \varphi(p) = \min_{u \in \mathcal{U}} \; G(u, p) \; .$$

The original problem we are interested in is $(\mathcal{P}_0)$. Then the dual problem is defined as

$$(\mathcal{D}_p) \qquad \varphi^{**}(p) = \max_{p^* \in \mathcal{Y}^*} \langle p^*, p \rangle - G^*(0, p^*).$$

With those definitions we have the classic duality inequality

$$\inf(\mathcal{P}_p) \geq \sup(\mathcal{D}_p)$$

.

Regularity of Value Function and Dual Problem

The regularity of the value function φ at 0 gives information on the dual problem $(\mathcal{D}_0)$. In a first place we have

$$\varphi^{**}(0) = \sup(\mathcal{D}_0) \; ,$$
$$\arg\max(\mathcal{D}_0) = \partial\varphi^{**}(0) \; .$$

Moreover if the value function φ is convex (which is the case if the perturbed cost G is jointly convex in (u, p)) we have

- No duality gap (i.e $\inf(\mathcal{P}_0) = \sup(\mathcal{D}_0)$) iff φ is l.s.c at 0.
- Existence of dual solution iff φ is subdifferentiable at 0, i.e $\inf(\mathcal{P}_0) = \max(\mathcal{D}_0)$ and $\arg\max \mathcal{D}_0 \neq \emptyset$.

Contents

Application to Constrained Optimization

We consider the following problem, with J convex proper and l.s.c, $\mathcal{U}^{ad}$ closed convex, C a closed convex cone and Θ is $C-$convex and continuous.

$$\min_{\substack{u \in \mathcal{U}^{ad} \\ \Theta(u) \in -C}} \quad J(u)$$

We choose to embedd it in the following family of perturbed problem

$$(\mathcal{P}_p) \qquad \varphi(p) = \min_{u \in \mathcal{U}^{ad}} \quad \underbrace{J(u) + \chi_{\{\Theta(u) - p \, \in -C\}}}_{G(u,p)},$$

$$\text{or} \qquad \min_{u \in \mathcal{U}^{ad}} \max_{p^* \in C^*} \quad J(u) + \langle p^*, \Theta(u) - p \rangle.$$

The dual problem $(\mathcal{D}_p)$ reads

$$(\mathcal{D}_p) \qquad \max_{p^* \in C^*} \min_{u \in \mathcal{U}^{ad}} J(u) + \langle p^*, \Theta(u) - p \rangle.$$

Sufficient Condition of Qualification

In this framework we have that φ is convex and l.s.c, and consequently there is no duality gap. However we need some condition to ensure subdifferentiability, and thus existence of solution of the dual problem $\mathcal{D}_0$. Those solutions are called optimal multipliers.

The following sufficient condition (CQC) of qualification is equivalent to $0 \in \operatorname{int} \operatorname{dom} \varphi$.

$$(CQC) \qquad 0 \in \operatorname{int}\Big(\Theta\big(\mathcal{U}^{ad} \cap \operatorname{dom}(J)\big) + C\Big)$$

Note that we could easily choose $\mathcal{U}^{ad} \subset \operatorname{dom}(J)$.

Contents

Some Notations

- We note in capital bold letter (e.g $\mathbf{U}$) any random variables.
- $\mathbf{U} \preceq \mathcal{F}$ means that $\mathbf{U}$ is a $\mathcal{F}-$measurable random variable :

$$\sigma(\mathbf{U}) \subset \mathcal{F} \ .$$

- As far as a dynamic system is concerned we will use
  - $\mathbf{X} = (\mathbf{X}_t) = (\mathbf{X}_0, \cdots, \mathbf{X}_T)$ for the state of a dynamic system,
  - $\mathbf{W} = (\mathbf{W}_t)$ is an exogeneous random noise,
  - and $\mathbf{U} = (\mathbf{U}_t)$ for the control.
- $\mathcal{F}_t = \sigma(\mathbf{W}_0, \cdots, \mathbf{W}_t)$ is the σ-algebra generated by the noises up to time t, and consequently $(\mathcal{F}_t)_{t \in [\![0,T]\!]}$ is a filtration.

An Optimization Problem

Let us consider the following dynamic optimization problem

$$
\begin{array}{rl}
\min\limits_{\mathbf{X},\mathbf{U}} & \mathbb{E}\Big(\sum_{t=0}^{T} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t)\Big) \\
\text{(dynamic equation)} & \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t) \\
\text{(measurability constraints)} & \mathbf{U}_t \preceq \mathcal{F}_t \\
\text{(other constraints)} & \mathbf{U}_t \in \mathcal{U}_t^{ad} \\
\text{(``coupling constraint'')} & \Theta_t(\mathbf{U}_t) = 0
\end{array}
$$

where $(\mathbf{W}_t)$ is a noise, $(\mathbf{X}_t)$ is the state process, $(\mathbf{U}_t)$ is the control process, and $\mathcal{F}_t = \sigma(\mathbf{W}_0, \cdots, \mathbf{W}_t)$. $\mathcal{U}^{ad}$ is typically used to denote bound constraints on the control.
Note that there is no state constraint.

The Large Scale System Version

This problem is inspired from the following problem that we want to decompose.

$$\min_{\mathbf{X},\mathbf{U}} \quad \mathbb{E}\Big(\sum_{t=0}^{T} \sum_{i=1}^{N} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i) \Big)$$

$$\forall\, i \in [\![1, n]\!], \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i)$$

$$\forall\, i \in [\![1, n]\!], \quad \mathbf{X}_0^i = x_0^i$$

$$\forall\, i \in [\![1, n]\!], \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}$$

$$\forall\, i \in [\![1, n]\!], \quad \mathbf{U}_t^i \preceq \mathcal{F}_t^i$$

$$\text{(coupling constraint)} \quad \sum_{i=1}^{N} \Theta_t^i(\mathbf{U}_t^i) = 0$$

The Large Scale System Version: dualized constraint

Under some assumptions we can dualize the coupling constraint.

$$
\begin{aligned}
\max_{\boldsymbol{\lambda}} \min_{\mathbf{X},\mathbf{U}} \quad & \sum_{i=1}^{N} \sum_{t=0}^{T} \mathbb{E}\Big(L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i) \Big) + \langle \boldsymbol{\lambda}_t, \Theta_t^i(\mathbf{U}_t^i) \rangle \\
& \forall\, i \in [\![1, n]\!], \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i) \\
& \forall\, i \in [\![1, n]\!], \quad \mathbf{X}_0^i = x_0^i \\
& \forall\, i \in [\![1, n]\!], \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad} \\
& \forall\, i \in [\![1, n]\!], \quad \mathbf{U}_t^i \preceq \mathcal{F}_t^i
\end{aligned}
$$

The Large Scale System Version: decomposed problem

Once the constraint dualized the problem is spatially decomposed.

$$\max_{\boldsymbol{\lambda}} \quad \sum_{i=1}^{N} \quad \min_{\mathbf{X}^i, \mathbf{U}^i} \quad \sum_{t=0}^{T} \mathbb{E}\Big(L_t^i\big(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i\big)\Big) + \langle \boldsymbol{\lambda}_t, \Theta_t^i(\mathbf{U}_t^i)\rangle$$

$$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t^i)$$

$$\mathbf{X}_0^i = x_0^i$$

$$\mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}$$

$$\mathbf{U}_t^i \preceq \mathcal{F}_t^i$$

Back to the General Optimization Problem

In the remaining of the talk we will focus on this version of the problem

$$\begin{aligned}
& \min_{\mathbf{X},\mathbf{U}} && \mathbb{E}\left(\sum_{t=0}^{T} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t)\right) \\
& \text{(dynamic equation)} && \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t) \\
& \text{(measurability constraints)} && \mathbf{U}_t \preceq \mathcal{F}_t \\
& \text{(other constraints)} && \mathbf{U}_t \in \mathcal{U}_t^{ad} \\
& \text{(``coupling constraint'')} && \Theta_t(\mathbf{U}_t) = 0
\end{aligned}$$

Dualization

Formally we dualize the coupling constraint $\Theta_t(\mathbf{U}_t) = 0$ on the control $\mathbf{U}_t$, with multiplier $\boldsymbol{\lambda}_t$ we obtain

$$\min_{\substack{\mathbf{X},\mathbf{U} \in \mathcal{U}^{ad} \\ \mathbf{U}_t \preceq \mathcal{F}_t}} \quad \max_{\boldsymbol{\lambda}} \quad \mathbb{E}\left(\sum_{t=0}^{T} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t) \right) + \sum_{t=0}^{T} \langle \boldsymbol{\lambda}_t, \Theta_t(\mathbf{U}_t) \rangle \ ,$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t)$$

where $\langle \boldsymbol{\lambda}_t, \Theta_t(\mathbf{U}_t) \rangle$ is a duality pairing to be precised.

Contents

Stochastic Uzawa Algorithm

We assume that the duality pairing $\langle \boldsymbol{\lambda}_t, \Theta_t(\mathbf{U}_t) \rangle$ can be written

$$\langle \boldsymbol{\lambda}_t, \Theta_t(\mathbf{U}_t) \rangle = \mathbb{E}\big(\boldsymbol{\lambda}_t \Theta_t(\mathbf{U}_t)\big) .$$

Then under constraint qualification condition we can exchange min and max and use the following algorithm. At step k we have a process $\boldsymbol{\lambda}^{(k)}$ and solve

$$\min_{\substack{\mathbf{X},\mathbf{U} \in \mathcal{U}^{ad} \\ \mathbf{U}_t \preceq \mathcal{F}_t}} \quad \mathbb{E}\Big(\sum_{t=0}^{T} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t) + \boldsymbol{\lambda}_t^{(k)} \Theta_t(\mathbf{U}_t) \Big) .$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_t)$$

Finally we determine $\boldsymbol{\lambda}^{(k+1)}$ by a gradient step

$$\boldsymbol{\lambda}_t^{(k+1)} = \boldsymbol{\lambda}_t^{(k)} + \rho \Theta_t(\mathbf{U}_t^{(k)}) .$$

Which spaces ?

- Until now we have not defined in which spaces lies our random variables.
- Depending on the duality choosen there might exist an optimal multiplier or not.
- Uzawa is given in Hilbert spaces framework, which would lead to choose $\mathbf{U}$ as well as the constraint function $\Theta(\mathbf{U})$ to be in L^2 (or Sobolev) spaces. However there might not be an optimal multiplier in those spaces.
- In order to assure the existence of multiplier, we could assume that the state and control variables are essentially bounded (in L^∞).
- Note that we restrict ourselves to multiplier process $\boldsymbol{\lambda}$ adapted to $(\mathcal{F}_t)_{t\in[\![0,T]\!]}$.

Why can we take an adapted multiplier ?

Assume that $\boldsymbol{\lambda}$ is an optimal multiplier process of our problem.
We see that each $\boldsymbol{\lambda}_t$ appears in the problem only in $\mathbb{E}\big(\boldsymbol{\lambda}_t \Theta_t(\mathbf{U}_t)\big)$.
And noting that $\Theta_t(\mathbf{U}_t) \preceq \mathcal{F}_t$ we have

$$\mathbb{E}\big(\boldsymbol{\lambda}_t \Theta_t(\mathbf{U}_t)\big) = \mathbb{E}\Big(\mathbb{E}\big(\boldsymbol{\lambda}_t \Theta_t(\mathbf{U}_t) \mid \mathcal{F}_t\big)\Big)$$

$$= \mathbb{E}\Big(\underbrace{\mathbb{E}\big(\boldsymbol{\lambda}_t \mid \mathcal{F}_t\big)}_{\tilde{\boldsymbol{\lambda}}_t} \Theta_t(\mathbf{U}_t)\Big)$$

Consequently the $(\mathcal{F}_t)$-adapted process $\tilde{\boldsymbol{\lambda}}$ defined, for all time-step t, by $\tilde{\boldsymbol{\lambda}}_t = \mathbb{E}\big(\boldsymbol{\lambda}_t \mid \mathcal{F}_t\big)$ is also an optimal multiplier, and he is adapted to $(\mathcal{F}_t)$.

Contents

Why these Examples

- In a first example (inspired by R. Wets) we show that even on a simple, strongly convex with almost sure inequality problem there might not exist a saddle point in L^2.
- In a second example we show that it might exist a saddle point even if the sufficient qualification condition (CQC) is not satisfied.

Recall that we have

$$(CQC) \qquad 0 \in int\Big(\Theta\big(\mathcal{U}^{ad} \cap dom(J)\big) + C\Big)$$

A first example derived from the Rockafellar-Wets example

Let ξ be a random variable uniform on $[1,2]$, x is a deterministic variable and $\mathbf{Y}$ a random variable measurable with respect to ξ.

$$\begin{aligned} \min_{x,\mathbf{Y}} \quad & \frac{x^2}{2} + \mathbb{E}\frac{(\mathbf{Y}-\alpha)^2}{2} \\ x &\geq a \\ (x-\mathbf{Y}) &\geq \xi \\ \mathbf{Y} &\geq 0 \end{aligned}$$

We consider the associated perturbed problem (in L^2)

$$\begin{aligned} \varphi(\mathbf{P}) = \min_{x,\mathbf{Y}} \quad & \frac{x^2}{2} + \mathbb{E}\frac{(\mathbf{Y}-\alpha)^2}{2} \\ x &\geq a + P_1 \end{aligned}$$

Solutions of the problem

We can compute the optimal solution

$$\varphi(\mathbf{P}) = [\max\{a+P_1, essup(\xi+\mathbf{P}_2-max(\mathbf{P}_3,\alpha))\}]^2+\mathbb{E}\frac{((\mathbf{P}_3-\alpha)^+)^2}{2}$$

Assume that $a < 2$.

- We can show that φ is convex and l.s.c at 0.
- However φ is not subdifferentiable at 0: $\partial\varphi(0) = \emptyset$.
- Consequently there is no duality gap between the primal and dual problem, but the dual problem has no solution in L^2. More precisely a maximizing sequence of the dual problem is of total mass 2 concentrating on the event $\{\xi = 2\}$.
- In fact a multiplier exist in the strong topological dual of L^∞. Moreover L^2 is a dense subset of this dual.

In a nutshell

	L^2	L^∞
Optimal multiplier	No	Yes
Uzawa	Yes	No?

Contents

A toy Example

Let us consider in $L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$ the following problem

$$\min_{\substack{\mathbf{U} \leq 1 \\ \mathbf{U} = 0}} \frac{1}{2} \mathbb{E}(\mathbf{U}^2)$$

the unique admissible solution is $\mathbf{U} = 0$, and the optimal value is accordingly 0. We dualize the constraint $\mathbf{U} = 0$ to obtain the dual problem

$$\sup_{\boldsymbol{\lambda} \in L^2} \min_{\mathbf{U} \leq 1} \mathbb{E}\Big(\frac{\mathbf{U}^2}{2} + \boldsymbol{\lambda} \mathbf{U} \Big) = \sup_{\boldsymbol{\lambda} \in L^2} -\mathbb{E}\Big(\frac{\boldsymbol{\lambda}^2}{2} \Big) = 0.$$

Thus there is no duality gap, and an optimal multiplier exist, namely $\boldsymbol{\lambda}^* = 0$.

With perturbation theory

We choose $\mathcal{U}^{ad} = \{\mathbf{U} \in L^2 | \mathbf{U} \leq 1\}$, $\Theta = \mathrm{Id}$, and $C = \{0\}$. Thus we embed our former problem in the following family of problem

$$\varphi(\mathbf{P}) = \min_{\substack{\mathbf{U} \leq 1 \\ \mathbf{U}=\mathbf{P}}} \frac{1}{2} \mathbb{E}\mathbf{U}^2 ,$$

and easily see that

$$\varphi(\mathbf{P}) = \frac{\|\mathbf{P}\|_2^2}{2} + \chi_{\{\mathbf{P} \leq 1\}}(\mathbf{P})$$

And as $\forall \mathbf{P} \in L^2, \quad \varphi(\mathbf{P}) \geq \varphi(0)$ we have by definition that φ is l.s.c at 0 and that $0 \in \partial\varphi(0)$ thus φ is subdifferentiable at 0. However as $\Theta(\mathcal{U}^{ad}) + C = \mathcal{U}^{ad}$ is of empty (relative) interior this example does not verify the sufficient qualification condition (CQC).

Uzawa with no saddle point ?

As far as we know Uzawa's convergence proof assume the existence of a saddle-point, however :

- As $L^\infty(\Omega,\mathcal{F},\mathbb{P}) \subset \Big(L^\infty(\Omega,\mathcal{F},\mathbb{P})\Big)^*$ the update step of the Uzawa algorithm, i.e $\lambda^{(k+1)} = \lambda^{(k)} + \rho\Theta(u^{(k)})$, have a sense.
- Moreover $L^\infty(\Omega,\mathcal{F},\mathbb{P})$ is dense in $\Big(L^\infty(\Omega,\mathcal{F},\mathbb{P})\Big)^*$.
- Finally the proof of convergence of Uzawa's algorithm does not show the convergence of the multiplier.

Consequently we might hope that we can prove the convergence of the algorithm in the duality $\Big(L^\infty, \big(L^\infty\big)^*\Big)$ with only a non-duality gap assumption ? Relation with $\epsilon-$resolution ?

Multistage duality $\left(L^\infty, L^1\right)$

T.Rockafellar and R.Wets have worked in a series of four papers on the duality $\left(L^\infty, L^1\right)$. The main idea was that measure multiplier came from dynamic induced constraints. In our framework dynamic induced constraints are produced by state constraints. Consequently without state constraints (for example if we choose to penalize those constraints and consider the penalized problem) we are in the (relatively) complete recourse case, and the multiplier should be in L^1.

However the papers were written for a two-stage problem, has it been extended to the multistage case ? Is it worth it ? Is there a way to guarantee that the multiplier will be in L^2 instead of L^1 ?

Comparison with the Optimal Control

In deterministic continuous time optimal control the same problem of existence of multiplier appears in the case of state constraint. They are used to ask for more regularity (Sobolev Spaces), maybe it is worth a look in this direction ?

ROCKAFELLAR, R.T., WETS,R. (1976).
Stochastic Convex Programming : Basic Duality.
Pacific Journal of Mathematics

ROCKAFELLAR, R.T., WETS,R. (1976).
Stochastic Convex Programming : Singular Multipliers and Extended Duality.
Pacific Journal of Mathematics

ROCKAFELLAR, R.T., WETS,R. (1976).
Stochastic Convex Programming : Relatively Complete Recourse and Induced Feasability.
SIAM J. Control and Optimization

EKELAND, I., TÉMAM,R. (1987).
Convex Analysis and Variational Problems.
SIAM

BONNANS, J.F., SHAPIRO,A. (2000).
Perturbation Analysis of Optimization Problems.
Springer

The end

Thank you for your attention !