Progressive Hedging

Jean-Philippe Chancelier and SOWG
CERMICS, École des Ponts ParisTech

Interface course, CIRM, Luminy
November 2019

Presentation Outline

Outline of the presentation

We consider two energy production units

We consider two energy *production* units

- a "cheap" limited one
  which can produce a quantity q_0, with $0 \leq q_0 \leq q_0^\sharp$,
  at cost $c_0 q_0$
- an "expensive" unlimited one
  which can produce quantity q_1, with $0 \leq q_1$,
  at cost $c_1 q_1$, with $c_1 > c_0$

We handle conomic dispatch as a cost-minimization problem under supply-demand balance

- On the *consumption* side, the demand is $D \geq 0$
- We express the *supply-demand balance objective* as ensuring at least the demand, that is,

$$q_0 + q_1 \geq D$$

- This objective is to be achieved at least cost, so that the *optimization* problem is

$$\min_{q_0, q_1} \underbrace{c_0 q_0 + c_1 q_1}_{\text{total costs}}$$

We express so-called “measurability” constraints

- the quantity q_0 is decided *before* the demand D materializes
  - open-loop control
- the quantity q_1 is decided *after* knowing the demand D (recourse)
  - feedback control $q_1 = \gamma(D)$

We arrive at a stochastic optimization problem

- We introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
- The demand $\boldsymbol{D}$ is a random variable, with known probability distribution
- We consider the stochastic optimization problem

$$\min_{q_0, \boldsymbol{q}_1} \mathbb{E}[c_0 q_0 + c_1 \boldsymbol{q}_1]$$

$$\text{under the constraints}$$

$$\begin{array}{rl} 0 & \leq q_0 \leq q_0^\sharp \\ 0 & \leq \boldsymbol{q}_1 \\ \boldsymbol{D} & \leq q_0 + \boldsymbol{q}_1 \\ \boldsymbol{q}_1 & \text{depends upon } \boldsymbol{D} \end{array}$$

Outline of the presentation

We recall the one day newsvendor problem

- We recall that the minimization problem

$$\min_{u \in \mathbb{R}_+} J(u) = \mathbb{E}_{\boldsymbol{W}}[j(u, \boldsymbol{W})]$$

where

$$j(u, w) = c_M w + (c - c_M)u + (c_M + c_S)(u - w)_+$$

- can be written as a linear program

We consider a finite probability space

- We suppose that the demand $\boldsymbol{W}$
  can take a finite number S of possible values w_s, $s \in \mathbb{S}$
- where s denotes a *scenario* in the finite set $\mathbb{S}$ (S=card($\mathbb{S}$))
- and we denote π_s the probability of scenario s, with

$$\sum_{s \in \mathbb{S}} \pi_s = 1 \text{ and } \pi_s \geq 0 \ , \quad \forall s \in \mathbb{S}$$

We write the one day newsvendor problem as a linear program

$$\min_{u \in \mathbb{U}, (r_s)_{s \in S} \in \mathbb{V}^S} \sum_{s \in S} \pi_s \big(c_M w_s + (c - c_M) u + (c_M + c_S) r_s \big)$$

$$\text{subject to}$$

$$r_s \geq u - w_s \ , \quad \forall s \in S$$

$$r_s \geq 0 \ , \quad \forall s \in S$$

$$u \geq 0$$

- From a nonlinear optimization problem
  - with scalar decision variable $u \in \mathbb{R}_+$
- To a linear program with
  - $1 + |S|$ decision variables : $\big(u, (r_s)_{s \in S}\big) \in \mathbb{R}^{1+|S|}$
  - $2|S| + 1$ constraints

We express the measurability constraint (I)

As the control $u \in \mathbb{U}$ must be the same
for all realizations of the demand w_s,
we introduce a new control $u_s \in \mathbb{U}$ for each scenario
(duplication of variables) and force all the control to be equal,
that is, we add a constraint $u_s = \overline{u}$ for all $s \in S$

$$\min_{\overline{u} \in \mathbb{U}, (u_s)_{s \in S} \in \mathbb{U}^S, (r_s)_{s \in S} \in \mathbb{V}^S} \sum_{s \in S} \pi_s \big(c_M w_s + (c - c_M) u_s + (c_M + c_S) r_s \big)$$

$$\text{subject to}$$

$$r_s \geq u_s - w_s \ , \quad \forall s \in S$$

$$r_s \geq 0 \ , \quad \forall s \in S$$

$$u_s \geq 0$$

$$u_s = \overline{u} \ , \quad \forall s \in S$$

We express the measurability constraint (II)

As $u_s = \overline{u}$ for all $s \in S$ implies that $\overline{u} = \sum_{s' \in S} \pi_{s'} u_{s'}$, we obtain

$$\begin{aligned}
\min_{(u_s)_{s\in S}\in\mathbb{U}^S,(r_s)_{s\in S}\in\mathbb{V}^S} \; & \sum_{s\in S} \pi_s\big(c_M w_s + (c - c_M)u_s + (c_M + c_S)r_s\big) \\
& \text{subject to} \\
& r_s \geq u_s - w_s \;, \quad \forall s \in S \\
& r_s \geq 0 \;, \quad \forall s \in S \\
& u_s \geq 0 \\
& u_s - \sum_{s'\in S} \pi_{s'} u_{s'} = 0 \quad , \quad \forall s \in S
\end{aligned}$$

We dualize the constraint and use multipliers

- For all $s \in S$, we dualize the constraint $u_s - \sum_{s' \in S} \pi_{s'} u_{s'} = 0$
- Using the property that

$$\sum_{s \in S} \pi_s \langle \lambda_s \, , u_s - \sum_{s' \in S} \pi_{s'} u_{s'} \rangle = \sum_{s \in S} \pi_s \langle \lambda_s - \sum_{s' \in S} \pi_{s'} \lambda_{s'} \, , u_s \rangle$$

- we obtain a dual problem (a lower bound of the original problem)

$$\min_{u_s \in \mathbb{U}^S, (r_s)_{s \in S} \in \mathbb{V}^S} \sum_{s \in S} \pi_s \Big(c_M w_s + (c - c_M) u_s + (c_M + c_S) r_s + \langle \lambda_s - \sum_{s' \in S} \pi_{s'} \lambda_{s'} \, , u_s \rangle \Big)$$

$$r_s \geq u_s - w_s \; , \quad \forall s \in S$$

$$r_s \geq 0 \; , \quad \forall s \in S$$

$$u_s \geq 0$$

There are decomposed subproblems scenario by scenario

For given multipliers, the problem is decomposed scenario by scenario as, for scenario s, we have to solve

$$\min_{u_s \in \mathbb{U}, r_s \in \mathbb{V}} \Big(c_M w_s + (c - c_M) u_s + (c_M + c_S) r_s + \big\langle \lambda_s - \sum_{s' \in S} \pi_{s'} \lambda_{s'} \, , u_s \big\rangle \Big)$$

$$r_s \geq u_s - w_s$$

$$r_s \geq 0$$

$$u_s \geq 0$$

- We obtain $|S|$ linear problems (LP) to solve in parallel
- Each LP has 2 variables and 3 constraints

How to chose multipliers in order to recover a solution of the original problem ?

Outline of the presentation

A generic minimization problem

$$\min_{u \in \mathcal{U}, \Theta(u) \in -C} f(u)$$

- $\Theta : \mathbb{U} \to \mathbb{K}$, vector space $\mathbb{K}$ in duality with $\mathbb{K}^\star$
  - example : $\mathbb{K} = \mathbb{K}^* = \mathbb{R}^n$ with usual $\langle x \, , y \rangle$
- $C \in \mathbb{K}$ a closed convex,
  which is salient, that is, $C \cap -C = \{0\}$
- $C^\star \in \mathbb{K}^\star$, where $C^\star = \{u^\star \in \mathbb{U}^\star \mid \langle u' \, , u \rangle \geq 0 \; \forall u \in C\}$
  - example : $\Theta(u) = Ku$, where $K : \mathbb{R}^p \to \mathbb{R}^n$,
    $Ku = 0 \rightsquigarrow Ku \in -C$,
    with $C = \{0\}$, $C^\star = \mathbb{R}^n$
  - example : $Ku \leq 0 \rightsquigarrow Ku \in -C$, with $C = \mathbb{R}^n_+$, $C^\star = \mathbb{R}^n_+$

Lagrangian

We introduce the Lagrangian

$$
\begin{aligned}
L(u,\lambda) : \mathbb{U} \times \mathbb{K}^\star &\to \overline{\mathbb{R}} \\
(u,\lambda) &\to f(u) + \langle \lambda \, , \Theta(u) \rangle
\end{aligned}
$$

We consider the Lagrangian restricted to $u \in \mathcal{U}$ and $\lambda \in C^\star$

Three equivalent minimization problems

- The three following problems have the same solutions

$$\min_{u \in \mathcal{U}, \Theta(u) \in -C} f(u)$$

$$\min_{u \in \mathcal{U}} \Big(f(u) + \delta_{-C}\big(\Theta(u)\big)\Big)$$

$$\min_{u \in \mathcal{U}} \sup_{\lambda \in C^\star} L(u, \lambda)$$

- where

$$\delta_A(u) = \begin{cases} 0 & \text{if } u \in A \\ +\infty & \text{if } u \notin A \end{cases}$$

Sketch of proof ignore on first read

$$\delta_{-C}\big(\Theta(u)\big) = \sup_{\lambda \in C^\star} \langle \lambda , \Theta(u) \rangle$$

- If $\Theta(u) \in -C$, then $\langle \lambda , \Theta(u) \rangle \leq 0$ for all $\lambda \in C^\star$, and $\langle \lambda , \Theta(u) \rangle = 0$ when $\lambda = 0 \in C^\star$: hence

$$\sup_{\lambda \in C^\star} \langle \lambda , \Theta(u) \rangle = 0$$

- (As C is a closed convex cone, we have $C^{\star\star} = C$
Thus, if $\langle \lambda , \Theta(u) \rangle \leq 0$ for all $\lambda \in C^\star$,
we have that $\Theta(u) \in C^{\star\star}$, and thus $\Theta(u) \in C$)
- If $\Theta(u) \notin C$, then there exists $\lambda_0 \in C^\star$ such that
$\langle \lambda_0 , \Theta(u) \rangle < 0$
Using the fact that $C^\star$ is a cone, we get

$$\sup_{\lambda \in C^\star} \langle \lambda , \Theta(u) \rangle \geq \sup_{\mu \in \mathbb{R}_+} \langle \lambda_0 , \Theta(u) \rangle = +\infty$$

We introduce the so-called dual function

- We always have that

$$\sup_{\lambda \in C^\star} \inf_{u \in \mathcal{U}} L(u, \lambda) \leq \inf_{u \in \mathcal{U}} \sup_{\lambda \in C^\star} L(u, \lambda) = \min_{u \in \mathcal{U}, \Theta(u) \in -C} f(u)$$

- We can obtain a a lower bound
  by maximizing the dual function

$$\sup_{\lambda \in C^\star} \phi(\lambda) \quad \text{where} \quad \phi(\lambda) = \inf_{u \in \mathcal{U}} L(u, \lambda)$$

- A possible algorithm is to
  maximize $\phi(\lambda)$ by the projected gradient algorithm

$$\lambda^{(k+1)} = P_{C^\star}\big(\lambda^{(k)} + \rho \Theta(u^{(k+1)})\big)$$

Saddle point

- Let $f : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}}$ and $X \times Y \subset \mathbb{X} \times \mathbb{Y}$
  $(x^\sharp, y^\sharp) \in \mathbb{X} \times \mathbb{Y}$ is a saddle point of f on $\mathbb{X} \times \mathbb{Y}$ if

$$\forall (x, y) \in X \times Y \, , f(x^\sharp, y) \leq f(x^\sharp, y^\sharp) \leq f(x, y^\sharp)$$

- Result : $(x^\sharp, y^\sharp) \in \mathbb{X} \times \mathbb{Y}$ is a saddle point of f if and only if

$$\begin{aligned} f(x^\sharp, y^\sharp) = \sup_{y \in \mathbb{Y}} f(x^\sharp, y) &= \min_{x \in \mathbb{X}} \sup_{y \in \mathbb{Y}} f(x, y) \\ &= \max_{y \in \mathbb{Y}} \inf_{x \in \mathbb{X}} f(x, y) = \inf_{x \in \mathbb{X}} f(x, y^\sharp) \end{aligned}$$

  - $\sup \inf$ and $\inf \sup$ commute
  - we have $\sup \inf = \max \inf$ and $\inf \sup = \min \sup$

Saddle points of the Lagrangian

- If $(u^\star, \lambda^\star)$ is a saddle point of $L(u, \lambda)$ on $\mathcal{U} \times C^\star$, then $u^\star$ is solution of the primal problem $\min_{u \in \mathcal{U}, \Theta(u) \in -C} f(u)$
- $(u^\star, \lambda^\star)$ is a saddle point if and only if

$$\max_{\lambda \in C^\star} \inf_{u \in \mathcal{U}} L(u, \lambda) = \min_{u \in \mathcal{U}} \sup_{\lambda \in C^\star} L(u, \lambda)$$

- In the convex case ($+$ technical conditions), if $u^\star$ is solution of the primal problem, there exists $\lambda^\star$ such that $(u^\star, \lambda^\star)$ is a saddle point of the Lagrangian

Outline of the presentation

Back to the newsvendor problem

$$\min_{(u_s,r_s)_{s\in S}\in(\mathbb{U}\times\mathbb{V})^S} \sum_{s\in S} \pi_s \overbrace{\big(c_M w_s + (c - c_M)u_s + (c_M + c_S)r_s\big)}^{f_s\big((u_s,r_s)\big)}$$

$$\text{subject to}$$

$$(u_s, r_s) \in \overline{\mathcal{U}}_s \subset \mathbb{U} \times \mathbb{V}$$

$$u_s - \sum_{s'\in S} \pi_{s'} u_{s'} = 0 \ , \ \ \forall s \in S$$

where $\overline{\mathcal{U}}_s \subset \mathbb{U} \times \mathbb{V}$ is defined by

$$r_s \geq u_s - w_s \ , \ \ \forall s \in S$$

$$r_s \geq 0 \ , \ \ \forall s \in S$$

$$u_s \geq 0$$

Data for a minimization problem

- $\mathbb{U} = \prod_{s=1}^{n} \mathbb{U}_s$ with generic element $u = \{u_s\}_{s=1,\dots,n}$
- equipped with a scalar product
  $\langle u \, , u' \rangle = \sum_{s=1}^{n} \pi_s \, \langle u_s \, , u'_s \rangle_s$ ($\pi_s > 0$ for all $i \in \{1, \dots, n\}$)
- $\Pi : \mathbb{U} \to \mathbb{V} \subset \mathbb{U}$ an orthognal projection on $\mathbb{V}$, a subspace of $\mathbb{U}$
  $\mathbb{V} = \{u \in \mathbb{U} \mid Ku = 0\}$ where $K = Id - \Pi$
- $f : \mathbb{U} \to \mathbb{R} \cup +\infty$ such that $f(u) = \sum_{s=1}^{n} \pi_s f_s(u_s)$
- $\mathcal{U} \subset \mathbb{U}$ such that $\mathcal{U} = \prod_{s=1}^{n} \mathcal{U}_s$ with $\mathcal{U}_s \subset \mathbb{U}_s$

Minimization problem

- The minimization problem is

$$\min_{u \in \mathcal{U} \cap \mathbb{V}} f(u) = \min_{\{u_s\}_{s=1,\dots,n} \in \mathcal{U} \cap \mathbb{V}} \sum_{s=1}^{n} \pi_s f_s(u_s)$$

- Without the coupling constraint $u \in \mathbb{V}$, we would have

$$\min_{u \in \mathcal{U}} f(u) = \sum_{s=1}^{n} \pi_s \min_{u_s \in \mathcal{U}_s} f_s(u_s)$$

- The coupling constraint $u \in \mathbb{V}$ can be written $Ku = 0$

Abstract Version of P.H.

- Measurability constraint is $Ku = 0$, where $K = Id - \Pi$
- $\Pi : \mathbb{U} \to \mathbb{U}$ is a projection

$$\Pi\big((u_1, \ldots, u_n)\big) = \big((\sum_{i=1}^{n} \pi_i u_i), \ldots, (\sum_{i=1}^{n} \pi_i u_i)\big)$$

- The subspace $\mathbb{V}$ is

$$\mathbb{V} = \big\{(u_1, \ldots, u_n) \in \mathbb{U} \,\big|\, u_1 = \ldots = u_n\big\}$$

- The orthogonal subspace $\mathbb{V}^\perp$ is

$$\mathbb{V}^\perp = \big\{(\lambda_1, \ldots, \lambda_n) \in \mathbb{U} \,\big|\, \sum_{i=1}^{n} \pi_i \lambda_i = 0\big\}$$

Abstract Version of P.H (II)

- The Lagrangian $L : \mathbb{U} \times \mathbb{U}^\star \to \overline{\mathbb{R}}$, associated with $Ku = 0$ is

$$L(u, v) = f(u) + \langle Ku\,, v\rangle$$

- We can in fact consider

$$\begin{aligned} L : \mathbb{U} \times \mathbb{U}^\star &\to \overline{\mathbb{R}} \\ (u, v) &\mapsto L(u, v) = f(u) + \langle u\,, v\rangle \end{aligned}$$

for $u \in \mathcal{U}$ and $\lambda \in K(\mathbb{U})$ (equivalent to $\sum_{s=1}^{n} \pi_s \lambda_s = 0$)

ignore on first read

- $v \in \mathbb{U}$, $v = (Id - \Pi)v + \Pi v$, with $(Id - \Pi)v \in \mathbb{V}$ and $\Pi v \in \mathbb{V}^\perp$
- $L(x, v) = f(u) + \langle Ku\,, Kv + \Pi v\rangle = f(u) + \langle Ku\,, Kv\rangle = L(x, Kv)$
- We can restrict the dual space to $K(\mathbb{U})$ by considering $L : \mathbb{U} \times K(\mathbb{U}) \to \overline{\mathbb{R}}$, that is,
  using dual variables u' satisfying $\Pi u' = 0$
- Assuming that $v \in K(\mathbb{U})$, we have
  $L(x, v) = f(u) + \langle Ku\,, v\rangle = f(u) + \langle Ku\,, Ku'\rangle = f(u) + \langle u\,, K \circ Ku'\rangle = f(u) + \langle u\,, Ku'\rangle = f(u) + \langle u\,, v\rangle$
- We thus consider $L(u, v) = f(u) + \langle u\,, v\rangle$

Augmented Lagrangian

"Augmented Lagrangian methods were developed in part to bring robustness to the dual ascent method, and in particular, to yield convergence without assumptions like strict convexity or finiteness of f"

$$\min_{\Theta(u)=0} f(u) \rightsquigarrow L_r(u, v) = f(u) + \langle v \,, \Theta(u) \rangle + \frac{r}{2} \|\Theta(u)\|_2^2$$

The augmented Lagrangian can be viewed as
the (unaugmented) Lagrangian associated with the minimization problem

$$\min_{\Theta(u)=0} f(u) + \frac{r}{2} \|\Theta(u)\|_2^2$$

Augmented Lagrangian

- The augmented Lagrangian associated with $Ku = 0$ is

$$L_r(u, v) = L(u, v) + r/2\|Ku\|^2$$

.

- As $Ku = u - \Pi u$, we get

$$L_r(u, v) = f(u) + \langle u \,, v \rangle + r/2\|u - \Pi u\|^2$$

- Since $\Pi u = \sum_{i=1}^{n} \pi_i u_i$, we obtain product terms $u_i u_j$ after developing the square

Therefore, at first look, we lose the decomposition property !

The Progressive Hedging Algorithm

1. Given $u^k \in \mathcal{U}$, λ^k such that $\Pi\lambda^k = 0$
2. Compute $\overline{u}^{k+1} = \Pi u^k$
3. Compute u^{k+1} solution of

$$u^{k+1} \in \operatorname*{Argmin}_{u \in \mathcal{U}} f(u) + \left\langle u \, , \lambda^k \right\rangle + r/2\|u - \overline{u}^{k+1}\|^2$$

   - From Linear Programming to Quadratic Programming
   - But we can linearize a quadratic term

4. Update multiplier with $\lambda^{k+} = \lambda^k + rKu^{k+1}$.
   (Note that $\Pi\lambda^{k+1} = \Pi\lambda^k + r\Pi K u^{k+1} = 0$)

Abstract Version of P.H (III)

- Compute u^{k+1} solution of

$$u^{k+1} \in \operatorname*{Argmin}_{u \in \mathcal{U}} f(u) + \left\langle u \, , \lambda^k \right\rangle + r/2 \|u - \overline{u}^{k+1}\|^2$$

- leads to scenario decomposition as

$$u_s^{k+1} \in \operatorname*{Argmin}_{u_s \in \mathcal{U}_s} f(u_s) + \left\langle u_s \, , \lambda_s^k \right\rangle + r/2 \|u_s - \overline{u}^{k+1}\|^2$$

As $u^{k+1} \in \mathcal{U}$ and $\overline{u}^{k+1} \in \mathbb{V}$, thus $\|u^{k+1} - \overline{u}^{k+1}\|^2$ is used to measure how far u^{k+1} is from $\mathcal{U} \cap \mathbb{V}$

Convergence of Progressive Hedging

Rockafellar, R.T., Wets R. J-B.
Scenario and policy aggregation in optimization under uncertainty,
Mathematics of Operations Research, 16, pp. 119-147, 1991

- Extends easily to N-stage problems
- With integer variables, the P.H. is used as heuristic
  (many extensions to improve the integer variable case)