Duality for Unit-Commitment

Primal Heuristics

Noisy Oracle

Claude Lemaréchal
Inria (Grenoble)

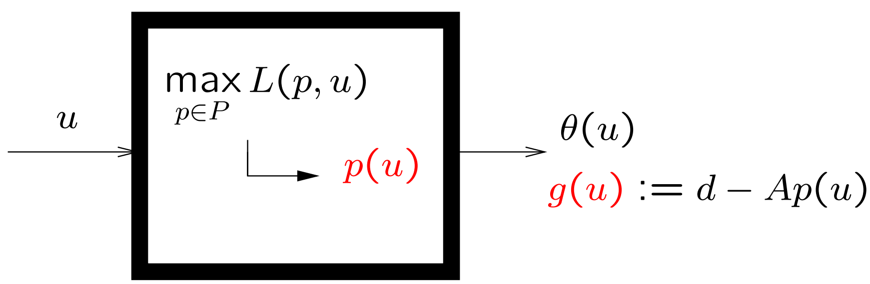
Recalls I: Lagrangian Relaxation

Hard Problem

$$\max_{p \in P} c^\top p\,, \qquad Ap = d \in \mathbb{R}^m$$

Easy pb: Given $u \in \mathbb{R}^m$

$$\left[\max_{p \in P} L(p,u) := c^\top p + u^\top (d - Ap)\right] =: \theta(u)$$



New pb: find appropriate u^*
$+$ recover appropriate p^*

Theory

- θ convex
- $\partial\theta(u) = \text{conv}\,\{d - Ap\}_{p=p(u)}$

Idea find p^* as convex combination of $p(u)$'s

Want $\sum_k \alpha_k(d - Ap_k) = 0$ i.e. $0 \in \partial\theta$

Minimize θ and

Exhibit optimality condition. At u^*:

$$\exists\{\alpha_k\}, \{p_k(u^*)\} : \underbrace{\sum_k \alpha_k \Big(d - Ap_k(u^*)\Big)}_{d - Ap^*} = 0$$

$$p^* = \textstyle\sum_k \alpha_k p_k(u^*) = \boxed{\text{pseudo-schedule}}$$

Theorem p^* solves

$$\max c^\top p\,, \quad p \in \text{conv} P\,, \quad Ap = d$$

Recalls II: Dual Algs.

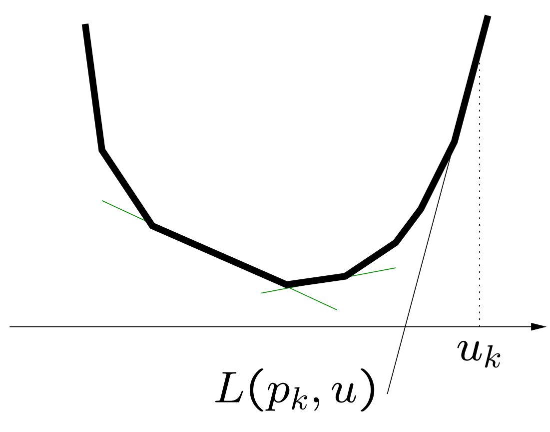
Working Horse Cutting planes

At each u_k, linearize θ: $\forall k$ $p_k := p(u_k)$

$$\theta(u) \geqslant \underbrace{\theta(u_k)}_{L(p_k,u_k)} + \underbrace{g(u_k)}_{d-Ap_k}^\top (u - u_k)$$
$$= c^\top p(u_k) + u^\top(d - Ap_k)$$
$$= L(p_k, u)$$

Model of θ: given $\{u_k\}_{k=1,\dots,K}$

$$\theta^K(u) := \underbrace{\max_k L(p_k, u)}_{\leqslant \max_p L(p,u)}$$



$\theta^K \leqslant \theta$!

Aim of the game find $\{u_k\}_k$

(i) close together:

(ii) such that $\sum_k \alpha_k(d - Ap(u_k))$ small

$$
\begin{aligned}
\theta(u_*) &\simeq \theta(u_k) && [\theta \text{ continuous}] \\
&\simeq c^\top p_k + u_*(d - Ap_k) && [u_* \simeq u_k] \\
&\simeq c^\top \sum \alpha_k p(u_k) && [\text{conv. comb.}] \\
&= c^\top p_*
\end{aligned}
$$

$\Longrightarrow u_*$ approx. optimal [weak duality]

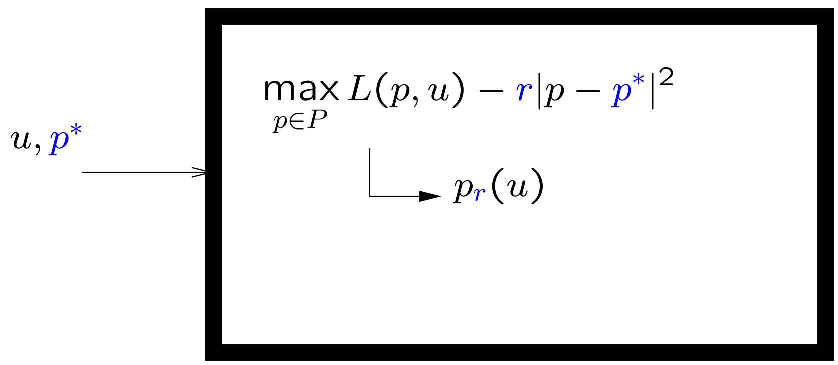
u_+ obtained as

Kelley	bundle
$\min \theta^K(u)$	$\min \theta^K(u) + \mu\lvert u - \hat{u}\rvert^2$
(ii) automatic	**(ii)** asymptotic

Primal Recovery

Pseudo-Schedule $\left\{ \begin{array}{l} \text{good balance} \\ \text{not a schedule} \end{array} \right.$

Idea Force $p(u)$ toward p_*



- Variant, knowing that $Ap = \sum_j A_j p_j$:

$$L_r(p, \lambda) = L(p, u) - r \sum_j |A_j(p_j - p_j^*)|^2$$

closer to augmented Larangian

$$-r\left|\sum_j A_j(p_j - p_j^*)\right|^2$$

= Phase II

- New Lagrangian still decomposable
- Same dual algorithm as Phase I
- Constructs schedules, retain best one

- Numerical results: **?**

Compared with current "augmented Lagrangian"

  - 2002: consistently | faster
more accurate
more reliable

  - 2012: 

- Putting the method in perspective . . .

General Framework

Primal prox (Bertsekas 1979)

Theoretical $p^{k+1} = p_r(p^k)$, with

$$p_r(q) := \operatorname*{Argmax}_{p \in P} \Big\{ \underbrace{c^\top p - r|p - q|^2}_{\text{more concave}} : Ap = d \Big\}$$

Theorem Cluster point p^∞ is

– global max if r small

– any local max if r large

- N.B. Local max probably close to p^1

Implementable $p^{k+1} = \tilde{p}_r(p^k)$

from Lagrangian relaxation

$$\min \theta_r(u) := \max_{p \in P} \underbrace{c^\top p + u^\top (d - Ap) - r|p - q|^2}_{\text{decomposable}}$$

Theorem Cluster point $\tilde{p}^\infty$ is

– "usually" p^* if r small

– any local max if r large   probably p^1

- In practice, r must balance $c^\top p$ and $|p - p^*|^2$

Conclusion

Iterating = probably bad idea

Initialization tricky $p^1 \neq p^*$!

Noise

$\theta(u) \rightsquigarrow \theta_o(u) = \theta(u) - \eta$

Noisy model unchanged

$\theta_o^K = \max_k L(p_k, \cdot) = \theta^K$

Cutting-plane paradigm still OK

Specific value $\underbrace{\theta(u)}_{\text{sampling point}}$ useless

Except for

- stopping test: $\theta_o(u_+) \leqslant \theta^K(u_+) + \eta$

  nothing to conclude

- bundle: $\hat{u}_+ =$

$$\begin{cases} \text{null-step } \hat{u} & \text{if } \theta_o(u_+) \geqslant \theta_o(\hat{u}) \\ \text{descent-step } u_+ & \text{if } \theta_o(u_+) \ll \theta_o(\hat{u}) \end{cases}$$

possibly misleading

Bundle Features

1) Informative: $\hat{u}$ optimal within

$$\eta_{\hat{u}} := \theta(\hat{u}) - \theta_o(\hat{u})$$

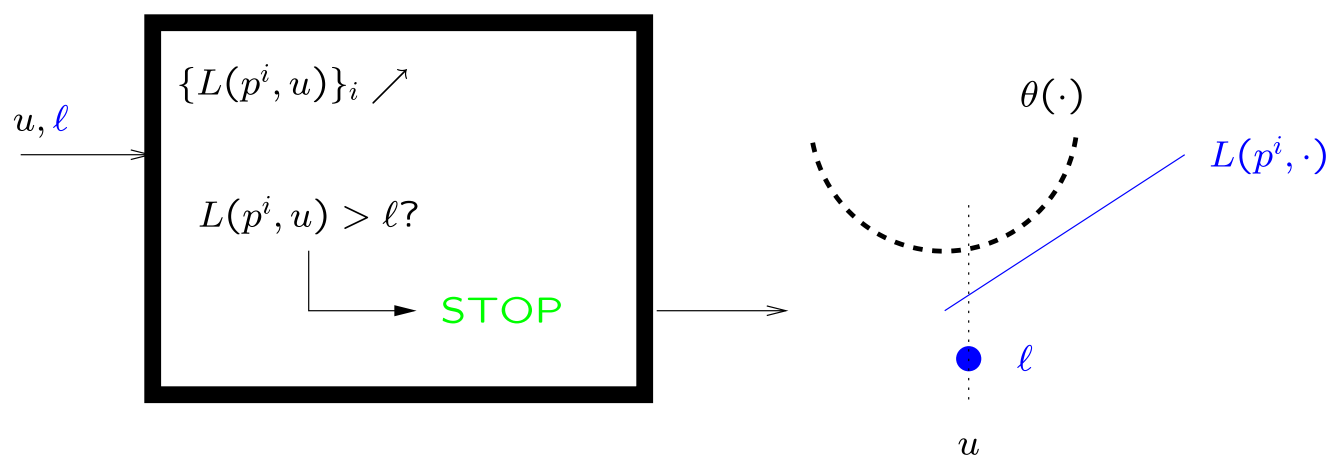
2) Friendly:

Noise possibly harmful only when

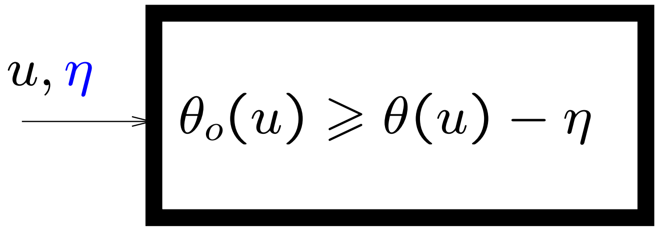
$$\theta_o(u) \leqslant \ell \quad [" = " \theta_o(\hat{u})] \qquad \text{then } \hat{u} \text{ updated}$$

Sends level to oracle

Free if turns out that $\theta_o(u) > \ell$



3) Case of smart oracle: choose η



expensive if η small

cheap if $\eta = +\infty$

Sensible strategy

Call accurate oracle sometimes only

a) At each descent-step Gaudioso etal., Kiwiel

- yields optimality
- expensive oracle too often

b) After optimality if $\widehat{\eta}$ too large

- very cheap
- convergence?

Synthesis

Malick, Oliveira, Zaourar

- Call accurate oracle at each $\hat{u}$-update **BUT**
- View computation of $\hat{u}_+$ as a

cutting-plane subalgorithm

– sequence of null-steps = a)

– full bundle = b)

– in between: anything

coarse bundle, pure Kelley,. . .

Convergence OK

Potentially substantial improvement