L'Omniprésence de Lagrange

ou

La Dualité sans Larmes

Claude Lemaréchal, Inria (Grenoble)
Journées franciliennes, VI 2004

I– THÉORIE

II– ALGORITHMES

Cadre

$$\boxed{\max f(x)\,, \quad x \in X\,, \quad c(x) = 0 \in \mathbb{R}^m} \qquad \text{(P)}$$

f, X, c complètement arbitraires mais

Pour une raison quelconque on préfère résoudre

$$\boxed{\max f(x) - u^\top c(x)\,, \quad x \in X} \qquad \text{(L)}$$

à u fixé

Typiquement (L) est "facile"

Exemple naïf: LP 0-1

$$\max b^\top x\,, \quad \underbrace{x \in \{0,1\}^n}_{x \in X}, \quad \underbrace{Ax - a = 0}_{c(x)=0}$$

Alors

$$\min_{x \in \{0,1\}^n} \left(b - A^\top u\right)^\top x + a^\top u \qquad \text{(L)}$$

est trivial

Example: Energy Production

- $p_i = (p_i^1, \dots, p_i^T)$ production level of plant i
- $p_i \in P_i$ operating constraints ”
- $C_i(p_i)$ operating cost ”
- $\sum_i p_i = d = (d^1, \dots, d^T)$

$$x = (p_1, \dots, p_n)\,, \qquad X = \prod_i P_i$$
$$f(x) = -\sum_i C_i(p_i)\,, \qquad c(p) = -\sum_i p_i + d$$

Key observation

$$\begin{aligned}(\mathrm{L}) \quad := \quad & \min_{p \in \prod_i P_i} \sum_i C_i(p_i) + u^\top \sum_i p_i - u^\top d \\ = \quad & \sum_i \left(\min_{p_i \in P_i} C_i(p_i) + u^\top p_i \right) - u^\top d\end{aligned}$$

Souplesse

X et c peuvent être décidés à volonté

Example: Assignment-type problems

Assign jobs i to machines j so that
– total efficiency is maximized
– availability-type constraints are satisfied

$$\left\{\begin{array}{lll} \max \sum_{ij} b_{ij}x_{ij}\,, & x_{ij}\in\{0,1\}\,, & \\ \sum_j x_{ij}=1\ \forall i\,, & & \text{each job on 1 machine} \\ \sum_i c_{ij}x_{ij}\leqslant t_j\ \forall j & & \text{availability of machines} \end{array}\right.$$

Two more possible dualizations:

$$\left\{\begin{array}{l} \max \sum_j (b_{ij}-u_j c_{ij})x_{ij}\,, \\ x_{ij}\in\{0,1\}\,,\ \sum_j x_{ij}=1 \end{array}\right. \qquad \text{each } i$$

trivial

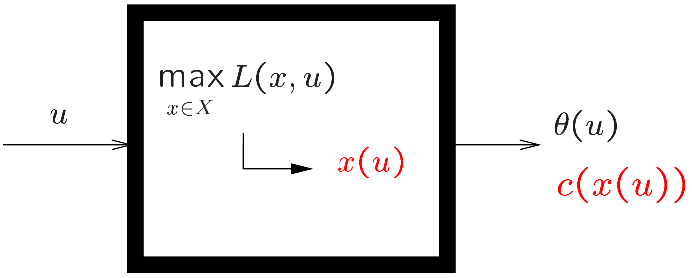
$$\left\{\begin{array}{l} \max \sum_i (b_{ij}-u_i)x_{ij}\,, \\ x_{ij}\in\{0,1\}\,,\ \sum_i c_{ij}x_{ij}\leqslant t_j \end{array}\right. \qquad \text{each } j$$

independent knapsacks

Notation

- $f(x) - u^\top c(x) =: L(x,u)$ Lagrangian
- $\max_{x\in X} L(x,u) =: \theta(u)$ $\boxed{\text{dual function}}$

  Z
- $x(u) \in \operatorname{Argmax} L(\cdot,u)$
- $c(x(u))$constraint-value
- black box or oracle



N.B.- $\theta(u)$ well-defined but
$c(x(u))$ ambiguous:
$x(u)$ exists? unique?

Useful Duality (1)

Theorem weak duality

$f(x) \leqslant \theta(u)$ for all feasible x and u

$$\begin{array}{ccccc} f(x) & = & L(x,u) & \leqslant & \theta(u) \\ x \text{ feas.} & \uparrow & & \uparrow & \text{def. of } \theta \end{array}$$

Corollary (L) provides upper bounds for (P)

fundamental for combinatorial optimization

Motivation for

$$\min_{u \in \mathbb{R}^m} \theta(u) \qquad \text{(D)}$$

$\boxed{\text{Problem no longer in } x \in X \text{ but in } u \in \mathbb{R}^m}$

Duality gap opt.val.(D) − opt.val.(P) $\geqslant 0$

Aim: duality gap as small as possible

Relaxation works $\Longleftrightarrow$ duality gap is 0

Example: LP

$$\max b^\top x\,, \quad \underbrace{x \geqslant 0}_{x \in X}, \quad \underbrace{Ax - a = 0}_{c(x)=0}$$

$$\begin{aligned}
\theta(u) &= \max_{x \geqslant 0} \overbrace{b^\top x - u^\top Ax + u^\top a}^{L(x,u)} \\
&= u^\top a + \max_{x \geqslant 0} (b - A^\top u)^\top x \\
&= \begin{cases} u^\top a & \text{if } b - A^\top u \leqslant 0 \\ +\infty & \text{otherwise} \quad Z \end{cases}
\end{aligned}$$

(D) is clearly $\min_{A^\top u \geqslant b} a^\top u$

Can also take "full" Lagrangian

$$L(x, u, v) = b^\top x - u^\top (Ax - a) - v^\top x$$

exercise (PTO)

Exemple: Inégalités

$$\max f(x)\,, \quad x \in X\,, \quad c(x) \leqslant 0$$

Variables d'écart:

- $x \in X \rightsquigarrow (x, y) \in X \times \mathbb{R}^m_+$
- $c(x) \leqslant 0 \rightsquigarrow c(x) + y = 0$
- $L(x, u) \rightsquigarrow L(x, u) - u^\top y$
- $\max_x L(x, u) \rightsquigarrow \max_x L(x, u) + \max_{y \geqslant 0} (-u)^\top y$

$$\theta(u) \rightsquigarrow \begin{cases} \theta(u) & \text{si } u \geqslant 0 \\ +\infty & \text{sinon} \end{cases}$$

$$\min_{u \geqslant 0} \theta(u) \qquad \text{(D)}$$

Plus généralement

$$c(x) \in G \quad \rightsquigarrow \quad \theta(u) + \underbrace{\max_{y \in G} (-u)^\top y}_{\text{fonction d'appui}}$$

Exemple: Relaxation SDP

$$\max b^\top x\,, \quad \underbrace{x_i(1-x_i)=0}_{c(x)=0}\,, \quad \underbrace{Ax-a=0}_{x\in X}$$

(pourquoi pas ?!)

$$L(x,u) = x^\top \operatorname{Diag}(u)x + (b-u)^\top x$$

$$\theta(u) = \begin{cases} +\infty & \text{si } \operatorname{Diag}(u) \not\succcurlyeq 0 \text{ sur } \operatorname{Ker} A \\ \dots & \text{sinon} \end{cases}$$

Maxcut + facile car pas de terme linéaire

$$\max x^\top E x\,, \quad x \in \mathbb{R}^n = X\,, \quad \underbrace{x_i^2 = 1}_{c(x)=e}$$

(D): $\min e^\top u\,, \quad E - \operatorname{Diag}(u) \preccurlyeq 0$

= eigenvalue bound = Goemans-Williamson

ϑ **Lovász** même jeu

Example: Column Generation

$$\max b^\top x\,, \quad x \in X\,, \quad Ax = a$$

Idea

Reduce X to manageable convex polyhedron:

- Restrict to columns $x_1, \ldots, x_K$ in X
- Convexify: solve restricted master

$$\max b^\top x\,, \;\; x \in \mathrm{conv}\{x_1, \ldots, x_K\}\,, \;\; Ax = a \quad \text{(M)}$$

How generate x_{K+1}?

- Get multiplier u of $Ax = a$ in (M)
- Maximize over $x \in X$:

$$\begin{array}{ll} b^\top x - u^\top Ax & \text{or equivalently} \\ b^\top x - u^\top Ax + u^\top a & \\ = L(x, u) & \text{i.e. the oracle!!} \end{array}$$

Example: Energy Production

Formulate

$$\begin{cases} \min C(p) \\ p \in X = \{p^k\}_{k=1}^{\infty} \\ Ap - d = 0 \end{cases} \qquad \begin{array}{c} \min \sum_i C_i(p_i) \\ p_i \in P_i \\ \sum_i p_i = d \end{array}$$

as (Dantzig-Wolfe master program)

$$\begin{cases} \min \sum_{k=1}^{\infty} \alpha^k p^k \\ \alpha^k \in \{0,1\}, \qquad \sum_{k=1}^{\infty} \alpha^k = 1 \\ \sum_{k=1}^{\infty} \alpha^k (Ap^k - d) = 0 \end{cases}$$

Restricted master:

- restrict ∞ to K
- convexify $\{0,1\}$ to $[0,1]$

Lessons

- CG & LR: same problems, same idea
  often hard to see because
  - (M) insists on $c(x) = 0$, softens X
  - (D) insists on $x \in X$, softens c

- (M) has to be a minimization algorithm

  (of θ)

- any other such algorithm possible

  instead of (M)

- Re: optimization has 50 years of history

Useful Duality (1')

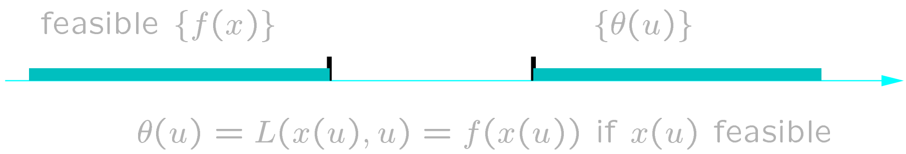
New Problem Find appropriate u, i.e.

- solution from (L) solves (P)
- i.e. $x(u)$ solves (P)
- hence in particular $c(x(u)) = 0$

Ambiguous

- some $x(u)$ feasible?
- any $x(u)$ feasible?

Corollary Anyway u has to solve (D)



Lesson u appropriate

- $\Leftrightarrow$ some $x(u)$ optimal
- $\Leftrightarrow$ some $x(u)$ feasible
- $\Rightarrow$ u solves (D)
- $\Leftarrow$? needs useful duality theory (3)

Useful Duality (2)

Theorem

- θ is a convex function
- $-c(x(u))$ $[= \nabla_u L(x(u), u)]$ $\in \partial\theta(u)$

$$\begin{array}{rcl} \theta(v) & \geqslant & L(x(u), v) \quad \text{for all } v \\ & = & L(x(u), u) - (v-u)^\top c(x(u)) \\ & \uparrow & \text{linearity of } L(x, \cdot) \\ & = & \theta(u) - (v-u)^\top c(x(u)) \end{array}$$

- If θ differentiable then $\nabla\theta(u) = -c(x(u))$

differentiate $\theta(u) = L(x(u), u)$

$$\nabla\theta = \nabla_u L + \underbrace{\nabla_x L}_{=0} \frac{\mathrm{d}x(u)}{\mathrm{d}u}$$

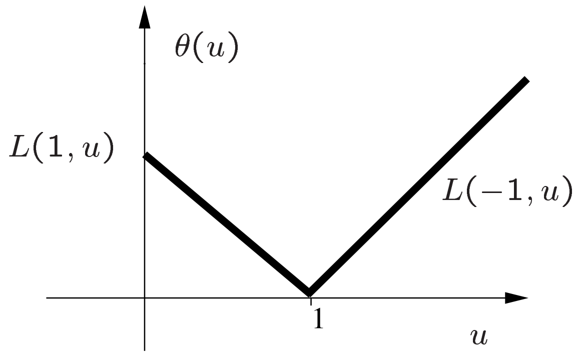
Remarque

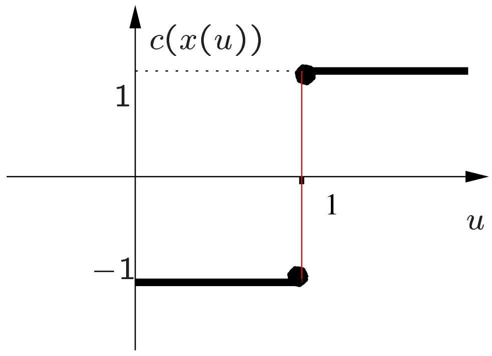
θ differentiable $\Rightarrow$ $x(u)$ unique!

Toward Duality Theory (3)

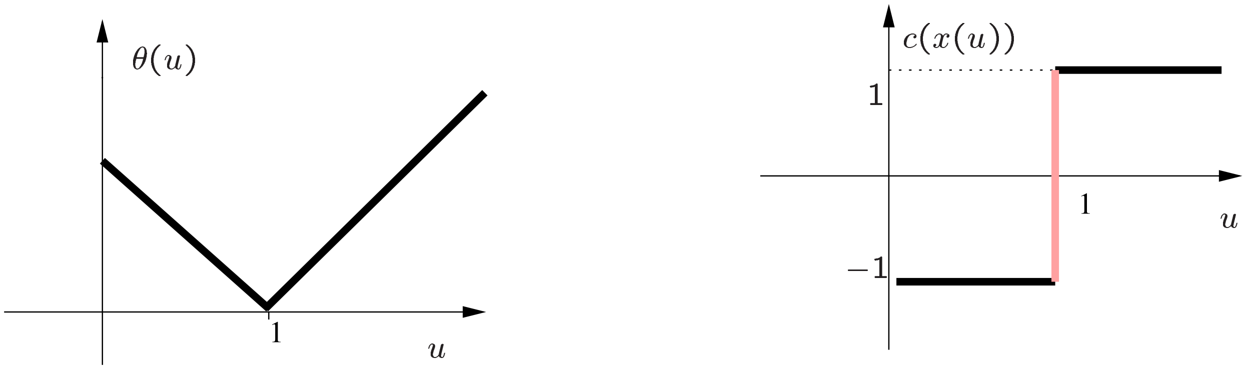
Three schematic examples

	Ex. 1	Ex. 2	Ex. 3
$f(x)$	x	x	$x-\frac{\varepsilon}{2}x^2$
X	$\{-1,+1\}$	$[-1,+1]$	$[-1,+1]$
$c(x)$	$x=0$	$x=0$	$x=0$
x^*	-1	0	0
f^*	-1	0	0
$L(x,u)$	$(1-u)x$	$(1-u)x$	$(1-u)x-\frac{\varepsilon}{2}x^2$

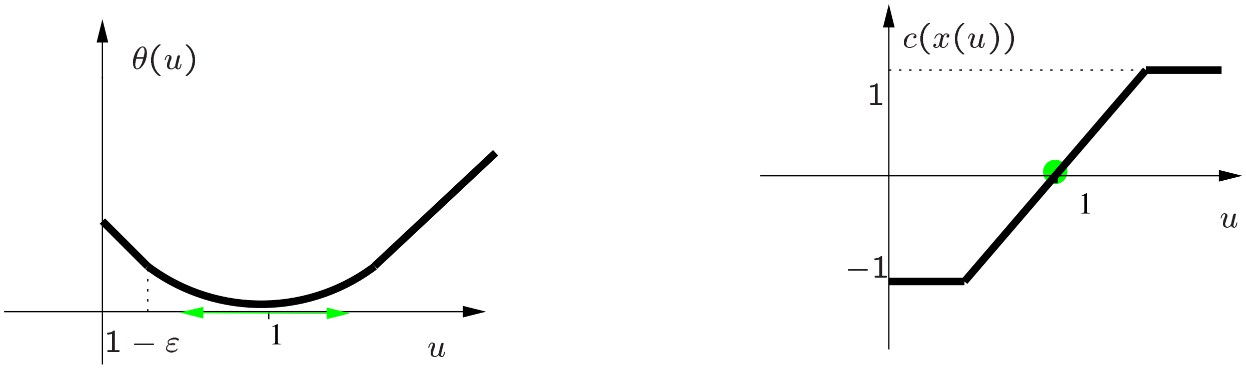




Ex 1. $\exists$ no appropriate u



Ex 2. some $x(u^*)$ is feasible



Ex 3. $c(x(u^*))$ singleton

Duality Theory (3)

At given u, $G(u)$:= set of all possible $[\nabla_u L =] - c(x(u))$, for all possible $x(u)$

- we know that $G(u) \subset \partial\theta(u)$
- hence convex hull $\overline{G}(u) \subset \partial\theta(u)$

Filling Property holds when conversely

$$\boxed{\overline{G}(u) = \partial\theta(u)}$$

"almost" always true

Then

- Let u^* minimize θ $[0 \in \partial\theta(u^*)]$
- assume filling property $[\,0 \in \overline{G}]$

✓ $c(x(u^*))$ is unique $= -\nabla\theta(u^*) = 0$!

✓ G is convex $\exists x(u^*)$ s.t. $-c(x(u^*)) = 0$
some $x(u^*)$ solves (P)

✓ $0 \notin G$ crash duality gap > 0

Duality Theory (3')

Theorem Linear case:

$$\max b^\top x\,,\quad x \in X\,,\quad Ax = a$$

+ "tolerable" assumptions, e.g.
X polyhedral
or X bounded

Then dualization equivalent to

$$\max b^\top x\,,\quad x \in \mathrm{conv}(X)\,,\quad Ax = a$$

cf Ex. 1-2: no difference seen from dual space

[Falk, '69],
later [Geoffrion, '74], [Magnanti el al., '76]

Summary I

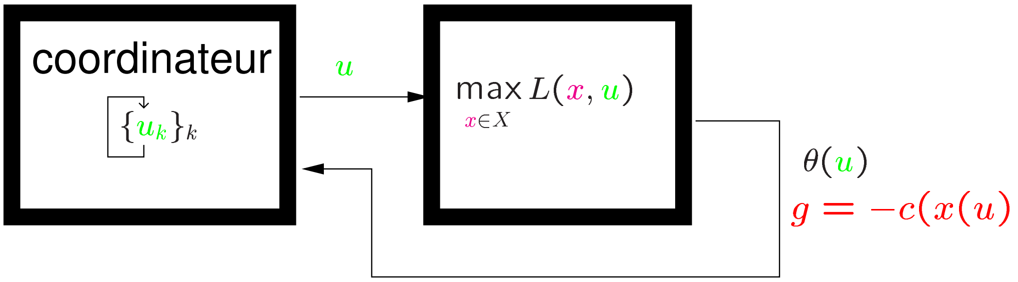
- Dual function $\theta = \max L$
  - bounds opt.val.(P)
  - is convex (and must be minimized)
  - has known subgradients $\nabla_u L(x(u), u)$
- $\min \theta$ convexifies (P)
- no duality gap when (P) satisfies
  - convexity assumptions
  - $+$ technical assumptions ("always true")
- $\boxed{\text{Lagrangian relaxation} \equiv \text{column generation}}$

Méthodes Numériques

Minimiser la fonction convexe θ:

$$\min \theta(u) \quad \Leftrightarrow \quad \min_{u \in \mathbb{R}^m} \max_{x \in X} L(x, u)$$

à l'aide d'un oracle délivrant $\theta(u)$ et $g \in \partial\theta(u)$



Coordinateur = algorithme d'optimisation

Dorénavant Cas linéaire

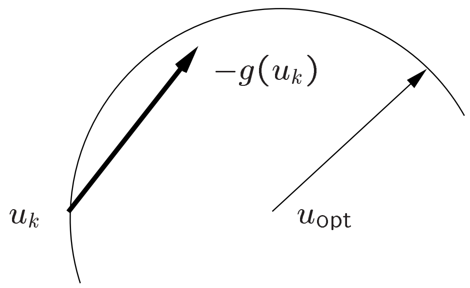
$$\max \underbrace{b^\top x}_{f(x)}, \quad x \in X, \quad \underbrace{Ax - a}_{c(x)} = 0$$

$$\begin{aligned} L(x, u) &= \left(b - A^\top u\right)^\top x + a^\top u \\ g(u) &= a - Ax(u) \end{aligned}$$

Subgradient and Ellipsoid

$$0 > \theta(u_{\text{opt}}) - \theta(u_k) \geqslant g_k^\top (u_{\text{opt}} - u_k)$$

For small $t > 0$, $u_k - tg_k$ better than u_k
i.e. closer to optimum point
(n.b. not in terms of θ)



Algorithm $u_{k+1} = u_k - t_k g_k$

- in theory: bad method
- in practice:
  - t_k difficult to adjust
  - stopping criterion?
  - low accuracy < 3 digits
  - very simple

For the Record

Preconditioner (usual in NLP)

$$u_{k+1} = u_k - t_k W_k g_k$$

W_k: symmetric positive definite matrix

Two specific choices (N.Z. Shor '70's)

- r-algorithm efficient in practice
  hard to tune
  not convergent in theory

- ellipsoid optimal in theory
  very inefficient in practice

Many Variants Camerini et al. '75, Nesterov '90's, Barahona, . . .

Kelley's Method

& Cheney-Goldstein, also called cutting plane

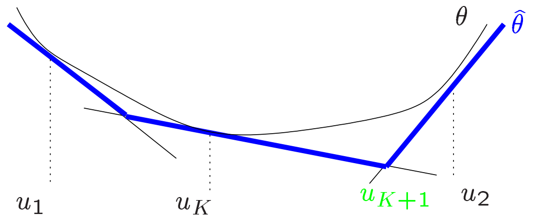
Model of dual function at Kth iteration:

$$\hat{\theta}(u) = \max_{k=1}^{K} \theta(u_k) + g_k^\top (u - u_k) \leqslant \theta(u)$$

$\{\theta(u_k), g_k\}_k$ = "bundle"

Algorithm u_{K+1} minimizes $\hat{\theta}(u)$:

$$\begin{cases} \min z, & z \in \mathbb{R}, u \in \mathbb{R}^m \\ z \geqslant \theta(u_k) + g_k^\top (u - u_k) \end{cases} \qquad \text{(LP)}$$



Remark $\theta(u_k) - u_k^\top g_k = f(x_k)$ hence

$$\hat{\theta}(u) = \max_k L(x_k, u)$$

Primalisation

(LP) s'écrit aussi $\min\limits_{z \geqslant L(x_k,u)} z$

dont le (bi-)dual est

$$\left\{\begin{array}{l} \max \sum_k \alpha_k f(x_k) \\ \alpha_k \geqslant 0,\ \sum_k \alpha_k = 1 \\ \sum_k \alpha_k c(x_k) = 0 \end{array}\right. \qquad \text{(MP)}$$

i.e. (f et c affines)

$$\left\{\begin{array}{l} \max b^\top x \\ x \in \operatorname{conv}\{x_1, \ldots, x_K\} \\ Ax = a \end{array}\right. \qquad \text{(MP)}$$

Leçon Dantzig-Wolfe = Kelley

Difficultés avec Kelley

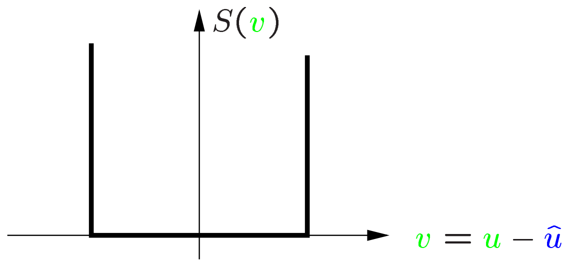
- nécessite un démarreur ($\hat{\theta}$ non borné)
- convergence peut être désastreuse $\left(\frac{1}{\sqrt[m/2]{k}}\right)$
- LP tend à être mal conditionné ($g_{k+1} \simeq g_k$)

Remède: stabiliser

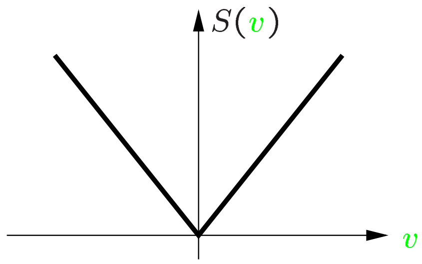
Stabilization

- Choose a penalty function S
- Penalize deviation from best dual point $\hat{u}$
- Replace (LP) by $\min_{z \geqslant L(x_k, u)} z + S(u - \hat{u})$

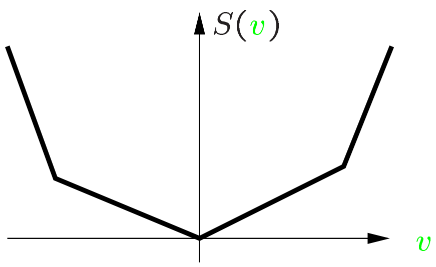
Polyhedral penalties



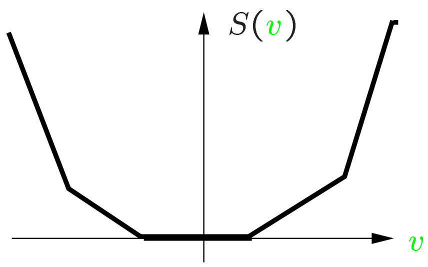
Boxstep - Marsten et al., 1975



Kim et al., 1995



du Merle et al., 1999



Ben-Amor et al., 2003

Quadratic Penalty

Bundle 1974

(LP) replaced by

$$\min_{z \geqslant L(x_k, u)} z + \frac{1}{2t}|u - \hat{u}|^2 \qquad \text{(QP)}$$

Advantages

- NO STARTER needed
- convergence of order $\frac{1}{\sqrt{k}}$ $\forall t > 0$
- allows to clean bundle delete selected k's
- QP easy

Necessary for efficiency

- tune penalty: $t = t_k$ Z $(u - \hat{u})^\top M_k (u - \hat{u})$?
- move $\hat{u}$ only if really worthwhile
  required by theory

Primalization

(bi)-Lagrangian for (QP) to be minimized

$$z - z\sum_k \alpha_k + \frac{1}{2t}|u - \widehat{u}|^2 + \sum_k \alpha_k f(x_k) + u^\top \sum_k \alpha_k g_k$$

$$= \Big(1 - \sum_k \alpha_k\Big) z + \frac{1}{2t}|u - \widehat{u}|^2 + \sum_k \alpha_k L(x_k, u)$$

.

.

.

$$\alpha_k \geqslant 0\,, \ \sum_k \alpha_k = 1\,; \qquad u_+ = \hat{u} - t\sum_k \alpha_k g_k$$

.

.

.

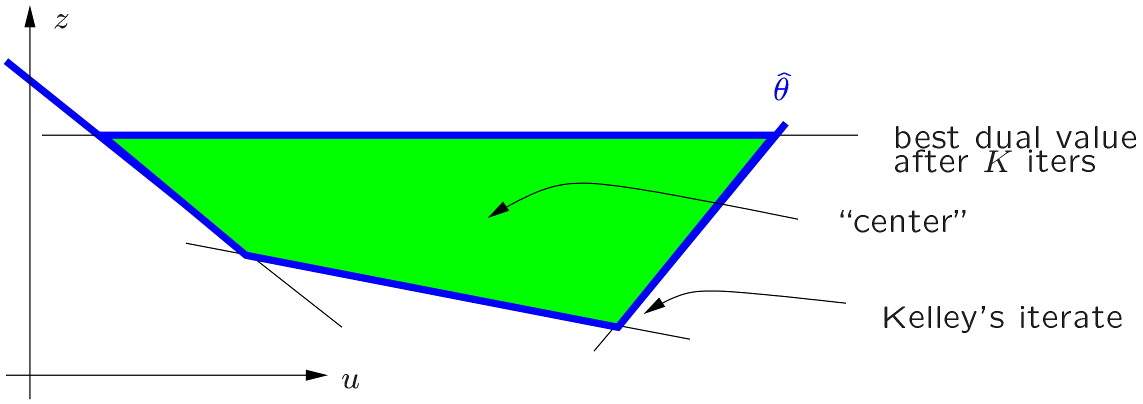
(bi)-dual problem for (QP):

$$\left\{\begin{array}{l} \max \quad \overbrace{L(x, \widehat{u}) - \frac{1}{2}|Ax - a|^2}^{\text{augmented Lagrangian}} \\ x \in \operatorname{conv}\{x_1, \ldots, x_K\} \end{array}\right. \qquad \text{(QMP)}$$

N.B. Complexity $= K$, not dual dimension m

Another Stabilization: ACCPM

compute “center” in bracketting polyhedron



$$
\begin{aligned}
(u_{k+1}, z_{k+1}) &= \text{analytic center} \\
&= \text{min product of slacks} \\
& \text{ or sum of their logs}
\end{aligned}
$$

benefits from quadratic penalty

Illustration

Held-Karp + TSP [oracle: $O(m^2)$]

- Full optimality

m	K	ℓ	(QMP)	oracle
120	132	42	0.5s	0.5s
442	556	116	98s	45s
1173	502	81	136s	280s
3038	4212	367	3h	5h40
4461	6965	361	3h?	26h?

final gap 10^{-10}; $Ax - a$: 10^{-8} (relative)

ℓ: # active pieces

- Poorman bundle: keep only 3 pieces in $\hat{\theta}$

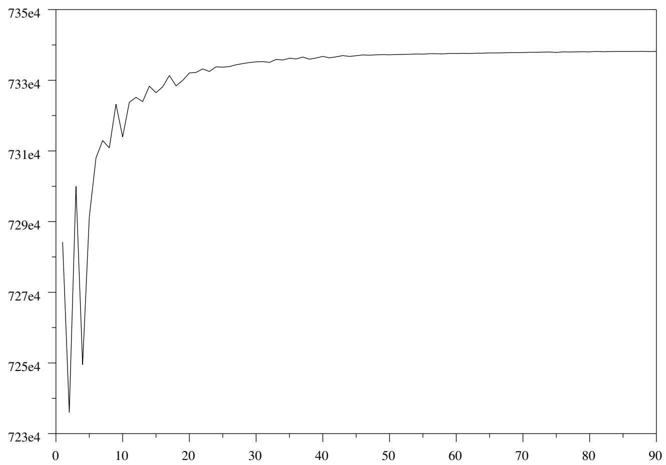
m	K	(QMP)	oracle
120	173	ε	1s
442	233	ε	19s
1173	236	ε	155s
3038	6920	ε	5h45
4461	411	ε	0h45

stopped manually when 4 digits in $\theta(u_K)$

Illustration: EdF

- Friday+Saturday
- 114 thermal + 102 hydro production units
- 296 relaxed constraints
- Run on Sun Blade 900MHz

	Initial	90%	final	
θ (MF)	72.842	73.344	73.385	
CPU	0	99 s	660 s	(all in min L)
# min L	0	30	200	



Evolution of θ

Warning fake data

Summary II

- Lagrangian relaxation (= column generation)
  is nothing but convex minimization
  $\Rightarrow$ many possible algorithms

- Standard Dantzig-Wolfe algorithm
  = Kelley among the worst in the field

- Standard subgradient is simple but
  - low speed ellipsoid at best
  - accuracy not guaranteed
  - hard to tune

- Introduction of quadratic term (bundle)
  opens the way to
  - automatic stability
  - robustness works in theory $\forall t$ Z
  - accuracy $\leftrightarrow$ subgradient
  - flexibility cf. poorman

  and has firm theoretical grounds

Quelques références bibliographiques

C.L.: 4or 1,1(2003) et: In Jünger, Naddef, LNCS 2241(2001)
Chaps. XII (théorie) et XV (algorithmes) de Hiriart-Urruty & C.L.:
"Convex Analysis & Minimization Algorithms" (Springer Verlag)

Deux références simples:
R. Grinold: Lagrangian subgradients, Man.Sc.17,3(1970)
L.S. Lasdon: "Optimization Theory for Large Systems" (MacMillan, 1970) (Appendix)

Quelques articles de base de Geoffrion:
MathProg Study 2(1974) mais aussi
SIAM Review 13, 1-37 (1971),
Manage. Sci., Theory 16, 652-675, 676-691 (1970)

Récupération primale dans la méthode de sous-gradients
K. Anstreicher, L.A. Wolsey: On dual solutions in subgradient optimisation. Unpublished manuscript, CORE, Louvain-la-Neuve (1993)
Larsson, Patriksson, Strömberg, MathProg86,2(1999)

Un article purement technologique:
Chap. 6 de C.r. Reeves, "Modern Heuristic Techniques..." (Blackwell, 1993)

Sur les relaxations sdp:
Poljak, Rendl, Wolkowicz: J. of Global Optimization(1995)
C.L. & Oustry: In Hadjisavvas & Pardalos,
"Advances in Convex Analysis and Global Optimization" (Kluwer,2001)