Proximal Analytic Center Cutting Plane Method

Olivier du Merle (Air France)
Jean-Philippe Vial (Logilab, University of Geneva)

(F. Babonneau, C. Beltran, O. Epelly, Y. Hachez, O. Péton, C. Tadonki)

IMA, Oct 14-19, 2002

Goals of this work

Primary goal Develop an efficient continuous optimization solver to solve nondifferentiable convex problems.

Secondary goal Adapt the solver for the particular needs of column generation or Lagrangian relaxation

“If Lagrangian relaxation does not work, try column generation!”

- Are the problems to be solved different?
- Are the solution methods different?
- Both?

Problem of interest

$$P^* = \min\{c^T x \mid Bx = b, x \in X\}$$

X is an **arbitrary** set.

Goals:

1. Define an efficient relaxation (small integrality gap)
2. Solve the relaxation

Relaxation 1: Lagrangian relaxation

Lagrangian $\quad L(x, y) = c^T x - u^T(Bx - b)$

Lagrangian dual $\quad \mathcal{L}(u) = \min_{x \in X} L(x, u).$

Dual problem

$$\mathcal{L}^* = \max \mathcal{L}(u) \tag{1}$$

Relaxation

$$\boxed{\mathcal{L}^* \leq P^*}$$

Formulation via extreme points

Since $\mathcal{L}$ is the result of the minimization of a linear function over a set X, we can replace X by

$$\bar{X} = \{x^k\}_{k \in K} = \text{Set of extreme points of the convex hull.}$$

The index set K may be **infinite**.

Equivalent problem

$$\mathcal{L}(u) = \min_{x \in \bar{X}} L(x, u).$$

Alternative formulation of (1)

$$\mathcal{L}^* = \max\{z \mid z \leq c^T x^k - u^T(Bx^k - b), \text{ for all } k \in K\}. \qquad (2)$$

Large or even semi-infinite Linear Programming problem.

The dual takes the form

$$
\begin{aligned}
\text{min} \quad & \{G(\lambda) = \textstyle\sum_{k\in K}(c^T x^k)\lambda_k\}, \\
\text{s. t.} \quad & \textstyle\sum_{k\in K}(Bx^k)\lambda_k = b \\
& \textstyle\sum_{k\in K} \lambda_k = 1, \\
& \lambda \geq 0.
\end{aligned} \tag{3}
$$

Relaxation 2: Column generation

Replace X in (1) by its convex hull:

$$\text{conv}(X) = \text{conv}(\{x^k\}_K) = \begin{cases} x = \sum \lambda_k x^k, \\ \sum \lambda_k = 1,\ \lambda \geq 0, \\ \forall x^k \in \bar{X}. \end{cases} \tag{4}$$

$$\begin{aligned} G^* = \quad \min \quad & \{G(\lambda) = \textstyle\sum_{k \in K} (c^T x^k) \lambda_k\}, \\ \text{s. t.} \quad & \textstyle\sum_{k \in K} (B x^k) \lambda_k = b \\ & \textstyle\sum_{k \in K} \lambda_k = 1, \\ & \lambda \geq 0. \end{aligned} \tag{5}$$

(5) and (1) are dual: $\longrightarrow$ $\boxed{\mathcal{L}^* = G^* \leq P^*}$

Lagrangian relaxation vs. column generation (standard view)

Lagrangian relaxation

Relaxation (partial dualization) + subgradient optimization (solution method)

Column generation

Relaxation (convex hull of X) + Generate columns by solving the restricted master to optimality (solution method)

Column generation is commonly associated with Dantzig-Wolfe decomposition: decomposition principle (relaxation) and decomposition algorithm (solution method).

Lagrangian relaxation = Column generation

"Formulation of the relaxation" and "Solution method" are two different issues. Suggestion: $\rightarrow$ use the terminology to designate the problem transformation.

Step 1. Formulate the relaxation in the dual space (Lagrangian relaxation) or in the primal space (column generation). Both yield the same problem, which is essentially a nondifferentiable convex program.

Step 2. Use your favorite solution method for convex nondifferentiable optimization (THE REAL ISSUE!)

Methods

- Subgradient method
- Ellipsoid
- DW or Kelley-Goldstein-Cheney
- Regularized DW
- Bundle method
- Level set method
- Volumetric center
- ACCPM
- ...

General framework for nondifferentiable optimization.

Canonical problem

$$\max_u f(u).$$

$f : R^n \to R$ concave. Information on f is given by an oracle

blackbox returning $f(u)$ and $\xi \in -\partial(-f(u))$

Cutting plane

$$f(u') \leq f(u) + \xi^T(u' - u), \text{ for all } u'.$$

The localization set.

Iteration record:

- Query points: u^1, $u^2, \dots u^k$
- Oracle responses: $(f^1 = f(u^1), \xi^1), \dots (f^k = f(u^k), \xi^k)$
- Best query point $u^{*k} = u^j$; best response $f^{*k} = f(u^j) = \max_{i \le k} f^i$.

Set of localization

$$\boxed{\mathcal{S}_k = \{(u, z) \mid z \le f^j + (u - u^j)^T \xi^j,\ j = 1, \dots k\}}$$

Generic cutting plane

1. Select query point: $u^k \in \mathcal{S}_{k-1}$ (Dual feasible to the restricted master in the Column generation context. Not necessarily optimal!)
2. Call oracle at u^k and get $(f^k = f(u^k), \xi^k)$
3. Update $\mathcal{S}_k = \mathcal{S}_{k-1} \cap \{z \leq f^k + (u - u^k)^T \xi^k\}$
4. Test termination

Method: generates query points u^k, $k = 1, \ldots$ in the localization set (or a subset of it)	$\longrightarrow$ $\longleftarrow$	**Oracle**: generates (anti-)subgradient $\longrightarrow$ (cutting planes, columns) in response to query point u^k

Lagrangian relaxation framework

Define $\boxed{x(u) = \arg\min_x L(x,u)}$

with

- $L(x,u) = b^T u + f(u)$
- $f(u) = -(Bx(u))^T u + c^T x(u)$
- $\xi = -Bx(u) \in -\partial(-f(u))$

In response to $(u^1, \ldots u^k)$, we get $x^j = x(u^j)$, $j \le k$

- $A = (\xi^1, \ldots \xi^k) = (-Bx^1, \ldots, -Bx^k)$
- $c = (c_1, \ldots, c_j)$ with $c_j = f^j - (\xi^j)^T u^j$

Localization set intersected with objective cut

$$\boxed{\begin{array}{l} \mathcal{S}_k = \{(u,z) \mid ez - A^T u \le c\} \\ z + b^T u \end{array}}$$

Kelley-Goldstein-Cheney

Query point selection

$$u^{k+1} = \arg\max\{z + b^T u \mid (z, u) \in \mathcal{S}_k\} \tag{6}$$

To make it work, $\mathcal{S}_k$ must be made compact, for instance by adding bounding box on u (penalized artificial variables in a column generation view of the world).

Remark The optimal value z^{k+1} is a valid upper bound. With the lower bound $\max_{j \leq k} f^j$ one can compute a relative duality gap at each iteration.

Dantzig-Wolfe for column generation

Take the dual of (6)

$$\begin{aligned} \min \quad & c^T x \\ & -Ax = b \\ & e^T x = 1. \end{aligned}$$

(Recall: $A = -(Bx^1, \ldots, Bx^k)$)

Take the best combination of columns. Use dual optimal variables to query the oracle.

Proximal ACCPM

Bounding objective cut $z + b^T u \geq f^{*k}$ with $f^{*k} = \max_{j \leq k} f(u^j)$

Modified localization set $\widetilde{\mathcal{S}}_k = \mathcal{S}_k \cap \{z + b^T u \geq f^{*k}\}$

Proximal analytic center of the set $\widetilde{\mathcal{S}}_k$

$$u = \arg\min_{u,z}\{G(u,z) = \frac{1}{2}(u - \bar{u})^T Q(u - \bar{u}) + F_{\widetilde{\mathcal{S}}_k}(u,z)\}$$

$F_{\widetilde{\mathcal{S}}_k}(u,z)$: (weighted) logarithmic barrier on the localization set

$\bar{u}$: the proximal center ($\bar{u} = u^j$ with $f^{*k} = f^j = f(u^j)$).

Optimality conditions

$$\begin{aligned} G(u,z) &= \frac{1}{2}(u-\bar{u})^T Q(u-\bar{u}) - \sum_{j=0}^{k} w_j \log s_j \\ s_0 &= b^T u + z - f^{*k} \\ s_j &= c_j + (A^T u)_j - z, \quad j = 1, \ldots k. \end{aligned}$$

Thus

$$\begin{aligned} Q(u-\bar{u}) - Ax - bx_0 &= 0 \\ e^T x - x_0 &= 0 \\ b^T u + z - s_0 &= f^{*k} \\ -(A^T u)_j + z - s_j &= c_j, \quad j = 1, \ldots k \\ x_j s_j &= w_j, \quad j = 0, \ldots k. \end{aligned}$$

Dual problem

Define the augmented barrier

$$H(x, x_0) = \frac{1}{2}(Ax + bx_0)^T Q^{-1}(Ax + bx_0) + (c + A^T\bar{u})^T x + (b^T\bar{u} - f^{*k})x_0 - \sum_{j=0}^{k} w_j \log x_j.$$

Dual problem

$$\min\{H(x, x_0) \mid e^T x - x_0 = 0\}.$$

Primal Newton method

First order optimality conditions for $\min\{H(x,x_0) \mid e^T x - x_0 = 0\}$

$$
\begin{aligned}
H'_x(x,x_0) - e\zeta &= 0, \\
H'_{x_0}(x,x_0) + \zeta &= 0, \\
e^T x - x_0 &= 0,
\end{aligned}
$$

where $\zeta \in R$ is some appropriate scalar. The Newton direction associated with a non-optimal point ($H'(x,x_0) - e\zeta = -r \neq 0$ and $H'_{x_0}(x,x_0) + \zeta = \sigma \neq 0$, but w.l.o.g. $e^T x - x_0 = 0$) solves

$$
\begin{aligned}
H''_{xx}(x,x_0)dx + H''_{xx_0}(x,x_0)dx_0 - ed\zeta &= r, \\
H''_{x_0x}(x,x_0)dx + H''_{x_0x_0}(x,x_0)dx_0 + d\zeta &= -\sigma, \\
e + dx - dx_0 &= 0.
\end{aligned}
$$

Computation of Newton direction

$$H'_x(x) = A^T Q^{-1}(Ax + bx_0) + (c + A^T\bar{u}) - wx^{-1}$$
$$H'_{x_0}(x) = b^T Q^{-1}(Ax + bx_0) + (-f^{*k} + b^T\bar{u}) - w_0 x_0^{-1}$$
$$H''(x) = \begin{pmatrix} A^T Q^{-1} A & A^T Q^{-1} b \\ b^T Q^{-1} A & b^T Q^{-1} b \end{pmatrix} + \begin{pmatrix} WX^{-2} & 0 \\ 0 & w_0 x_0^{-1} \end{pmatrix}.$$

Two cases are in order

- $m < n$ (less cuts than the dimension of the space)

$$(A, B)^T Q^{-1}(A, B) = \begin{pmatrix} A^T Q^{-1} A & A^T Q^{-1} B \\ B^T Q^{-1} A & B^T Q^{-1} B \end{pmatrix}.$$

- $m \geq n$ (more cuts than the dimension of the space)

$$[A^T Q^{-1} A + WX^{-2}]^{-1} = W^{-1}X^2 \\ -W^{-1}X^2 A^T \left(Q + AW^{-1}X^2A^T\right)^{-1} AW^{-1}X^2.$$

Retrieve analytic center in the u-space

We compute a center x, but we want to query the oracle at the AC u of the localization set.

At x (approximate analytic center) of the dual problem, the first optimality conditions are close to be satisfied

$$\begin{aligned} A^T Q^{-1}(Ax + bx_0) + c + A^T \bar{u} - wx^{-1} - e\zeta &\approx 0 \\ b^T Q^{-1}(Ax + bx_0) - f^{*k} + b^T \bar{u} - w_0 x_0^{-1} + \zeta &\approx 0, \end{aligned}$$

Define

$$\begin{aligned} u &= \bar{u} + Q^{-1}(Ax + bx_0), \\ s &= wx^{-1}. \end{aligned}$$

Let $z = \zeta$. Since $s > 0$, $-A^T u + ez \leq c$ and $b^T u + z \geq f^{*k}$. (u, z) belongs to the localization set.

Upper bound

Recall that the DW problem

$$\min\{c^T x \mid -Ax = b,\ e^T x = 1,\ x \geq 0\}$$

is almost surely infeasible at the begining ($k < n$). The 1st OC at the analytic center $Ax + bx_0 = Q(u - \bar{u})$ provides a solution

$$\tilde{x} = x/x_0 \quad \text{and} \quad r = x_0^{-1} Q(u - \bar{u})$$

to the perturbed problem

$$\min\{c^T x \mid -Ax = b - r,\ e^T x = 1,\ x \geq 0\}.$$

By duality,

$$b^T u + z - r^T u \leq c^T \tilde{x},$$

for all feasible pairs (z, u). We deduce the upper bound

$$c^T \tilde{x} + r^T \bar{u} + ||r||\, ||\bar{u} - u^*||,$$

where u^* is an unknown solution for the Lagrangian dual.

The p-median problem

Data: n points and a distance matrix c, with $c_{ij} \geq 0$ and $c_{ii} = 0$. Assign each of the n points to exactly one out of p medians. The medians are selected among the n points.

$$
\begin{aligned}
z^* = \min \quad & \sum_{i,j} c_{ij} x_{ij} \\
& \sum_{i} x_{ij} = 1,\ j = 1, \ldots n, && (7) \\
& \sum_{i} y_i = p, && (8) \\
& 0 \leq x_{ij} \leq y_i,\ j = 1, \ldots n,\ i = 1, \ldots n && (9) \\
& y \in \{0, 1\}^n. && (10)
\end{aligned}
$$

(7) can be relaxed to '$\geq$' and (8) to '$\leq$'.

Relaxations

- Linear relaxation: $0 \leq x_{ij} \leq y_i \leq 1$.
- Lagrangian relaxation 1 (weak?): dualize the cover constraints (7) and the p-median constraint (8).
- Lagrangian relaxation 2 (strong?): dualize the cover constraints (7) only.

Bad news

The three relaxations have the same optimal value.

Lagrangian relaxation 1; formulation

$$L(x, y; u, v) = \sum_{ij} c_{ij} x_{ij} - \sum_j u_j (\sum_i x_{ij} - 1) + v(\sum_i y_i - p).$$

Lagrangian dual

$$\mathcal{L}(u, v) = \mathcal{M}(u, v) + \sum_j u_j - pv$$

with

$$\mathcal{M}(u, v) = \min_{0 \leq x_{ij} \leq y_i \leq 1} M(x, y; u, v) = \sum_i \left(\min_{0 \leq x_{ij} \leq y_i \leq 1} M_i(x, y; u, v) \right), \tag{11}$$

and

$$M_i(x, y; u, v) = \sum_j (c_{ij} - u_j) x_{ij} + v y_i,$$

Lagrangian relaxation 1; oracle

If $y_i = 0$, then set $x_{ij} = 0$ for all $j = 1, \ldots n$. Otherwise choose x_{ij} as follows:

$$x_{ij} = \begin{cases} 1 & \text{if } c_{ij} - u_j < 0 \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

Thus, the minimum in (11) is achieved by setting

$$y_i = \begin{cases} 1 & \text{if } -\sum_j (c_{ij} - u_j)^- + v < 0 \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

Anti-subgradient of $\mathcal{M}(u, v)$ at (u, v):

$$\begin{pmatrix} -\sum_i x_{i\cdot}(u, v) \\ \sum_i y_i(u, v) \end{pmatrix}$$

- **Starting point** $u_i = \min_{j \neq i} c_{ij}$, $v = -1$.
- **Warm start** Heuristic.
- **Centering condition** Proximity measure less than 1.
- **Serious step** When lower bound improves. Else, 'null' step.
- **Proximal term** $Q = \text{diag}(1, \ldots, 1, 10^{-4})$. At serious step $Q \to Q/2$; at null step $Q \to 2Q$. Prox center = best point.
- **Weight** $w_i = 1$, $i = 1, \ldots m$. $w_0 = 1$ or $w_0 = m$.
- **Convergence** Relative duality gap $= 10^{-6}$.
- **Aggregation of cutting planes** 'Far away' cuts to make a single cut. Weight = sum of the weights of aggregated cuts.
- **Programming language** Matlab 6.5 (C-file for the oracle)
- **Hardware** Laptop Pentium III 800Mhz, 256Mb

Problem instances

- Set 1. Points on a square grid. $n = q \times q$. 15 problems with dimension ranging from $n = 100$ to $n = 2500$.
- Set 2. 2-D examples with data from the TSP library (distance matrix). 17 problems with dimension ranging from $n = 120$ to $n = 3038$.
- Medians. $p = (2, 3, 5, 10, 20, 30, 50, 70, 100)$

Methods

- Kelley with Matlab interior point LP solver.
- Kelley with Mosek (simplex with warm start).
- Proximal ACCPM.

Tests

- Grid and TSP collection problems with fixed number of medians (10), high (1e-6) and low (1e-3) precision.

- Grid problem 1024 and TSP 1060 solved with a variable number of medians (2,3,5,10,15,20,30,50,70,100)

outer iterations (oracle solves), # inner iterations (Newton steps), # active cuts in the end, CPU time, time spent in the oracle

GRID problems, with 10 medians

n =# variables 100 to 2500, (15 problems)

Accuracy	10^{-6}
CPU (total)	monotonic from 1s. to 428s.
CPU (grand total)	1624s
CPU (oracle)	monotonic from 8.7% to 34%
Outer	57 to 440
Outer/n	monotonic from 93% to 18%
Active cuts	44% (stable)
Inner/outer	2.47

Accuracy	10^{-3}
CPU (total)	monotonic from .9s. to 40s.
CPU (grand total)	172s.
CPU (oracle)	monotonic from 5.7% to 64.7%
Outer	38 to 77 (peak at $n = 1521$)
Outer/n	monotonic from 38% to 3%
Active cuts	between 50% and 65%
Inner/outer	2.7

Problems with TSP data, with 10 medians

n =# variables   120 to 3038, (14 problems) (without outliers)

Accuracy	10^{-6}
CPU (total)	monotonic from 3.7s. to 250s.
CPU (oracle)	monotonic from 8.7% to 34%
Outer	102 to 369
Active cuts	$\approx$ 50%
Inner/outer	2.23

Outliers: 1173, 1748, 1817

Accuracy	10^{-6}
CPU (total)	3005s., 531s., 1861s.
CPU (oracle)	3.5%, 26% , 10%
Outer	1375, 817, 1070
Active cuts	959, 226, 802
Inner	3906, 1763, 2303
CPU (grand total)	6115s.

Problems with TSP data, with 10 medians

n =# variables 120 to 3038, (14 problems) (without outliers)

Accuracy	10^{-3}
CPU (total)	monotonic from 31.35s. to 94s.
CPU (oracle)	monotonic from 10% to 70%
Outer/n	85% to 4%
Active cuts	$\approx$ 45%
Inner/outer	2.4

Outliers: 1173, 1748, 1817

Accuracy	10^{-3}
CPU (total)	91s., 117s., 105s.
CPU (oracle)	37%, 53%, 53%
Outer	282, 258, 220
Active cuts	160, 102, 116
Inner	623, 553, 481
CPU (grand total)	463s.