# Convex Optimization and Column Generation

## Stabilized Column Generation

Chungmok Lee

SOLAB, KAIST

June 12, 2008

# Danzig-Wolfe Decomposition



- Consider the optimization problems of the form

$$\min cx, \; Ax \geq b \in R^m, \; x \in X := \{x^i | i \in I\} \subset R^n \quad (1)$$

  , where the index set $I$ is assumed finite

- An equivalent formulation is Danzig-Wolfe's *master problem*

$$\min \sum_{i \in I} (cx^i)\lambda_i, \; \sum_{i \in I} (Ax^i)\lambda_i \geq b \in R^m, \; \sum_{i \in I} \lambda_i = 1, \; \lambda_i \in \{0,1\}, \; i \in I \quad (2)$$

# Danzig-Wolfe Decomposition. *Relaxed*

- Actually, we try to solve

$$\min cx, \;\; Ax \geq b \in R^m, \;\; x \in conv(X) \tag{3}$$

  , where the index set $I$ is assumed finite

- An equivalent formulation is Danzig-Wolfe's *master problem*

$$\min \sum_{i \in I} (cx^i)\lambda_i, \;\; \sum_{i \in I} (Ax^i)\lambda_i \geq b \in R^m, \;\; \sum_{i \in I} \lambda_i = 1, \;\; \lambda_i \geq 0, \;\; i \in I \tag{4}$$

# Standard Column Generation

- *Restricted Master Problem*(RMP) : restricting $I$ to some $I^k \in I$ in (2) or (3). $X^k := \{x^i | i \in I^k\}$

$$\min cx, \; Ax \geq b \in R^m, \; x \in conv(X^k) \tag{5}$$

In linear form,

$$\min \sum_{i \in I} (cx^i)\lambda_i, \; \sum_{i \in I} (Ax^i)\lambda_i \geq b \in R^m, \; \sum_{i \in I} \lambda_i = 1, \; \lambda_i \geq 0, \; i = 1, \ldots, k \tag{6}$$

- RMP gives
  - a primal solution $\hat{x} = \sum \hat{\lambda}_i x^i$, which is a candidate to solving (2) or (3)
  - a dual solution $(u^k, r^k) \in R^m_+ \times R$ associated with constraints $Ax \geq b$ and $\sum_i \lambda_i = 1$, respectively
- The dual solution is used to price out $I \backslash I^k$ : *Oracle*
  - $\min_{x \in conv(X)} (c - uA)x$ at $u = u^k$

## Standard Column Generation-Remarks



- Some tolerance can be inserted, to stop the algorithm the smallest reduced cost is *not too negative*
- We want to find $x^*$ from (2), but the key role is actually played by $u^k$
- Good $u^k$'s are those that are close to optimal multipliers for (2) or (3)
- LP dual of RMP being *dual restricted master*

$$\max ub - r,\ uAx^i - r \leq cx^i,\ i = 1, \ldots, k,\ (u, r) \in R^m_+ \times R \tag{7}$$

# Column Generation seen in the Dual Space

- From now, free oneself from the LP duality and use general duality
- To (2) we associate the *Lagrange function*

$$conv(X) \times R^m_+ \ni (x,u) \mapsto L(x,u) := cx + u(b - Ax) \\ = (c - uA)x + ub \quad (8)$$

- The so-called *dual function*

$$R^m_+ \ni u \mapsto \theta(u) := \min_{x \in X} L(x,u) = \min_{x \in conv(X)} L(x,u) \\ = (c - uA)x(u) + ub \quad (9)$$

, where $x(u)$ denotes the answer of the oracle

- The dual problem associated with (2) is then

$$\max\{\theta(u) | u \in R^m_+\} = \max_{(u,s) \in R^m_+ \times R} \{s | s \leq ub + (c - uA)x^i, i \in I\} \quad (10)$$

# Column Generation seen in the Dual Space

- We introduce the *restricted dual function*

$$R^m_+ \ni u \mapsto \theta^k(u) := \min_{x \in X^k} L(x,u) = ub + \min_{i=1,\ldots,k} (c - uA)x^i \tag{11}$$

- The dual of (5) is then

$$\begin{aligned} &\max\{\theta^k(u) | u \in R^m_+\} \\ &= \max_{(u,s) \in R^m_+ \times R} \{s | s \le ub + (c-uA)x^i, i = 1, \ldots, k\} \end{aligned} \tag{12}$$

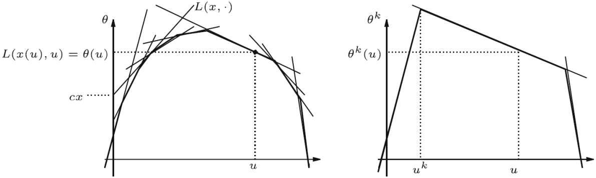


Figure: The dual function $\theta$ (left) and its restricted version $\theta^k$ for $k = 3$ (right)

# Column Generation seen in the Dual Space - Remarks



- The column generation algorithm seen in the dual is a cutting-plane procedure to maximize $\theta$ : at each iteration, we maximize $\theta^k$ instead.

$$\theta(u) \leq \theta^k(u) \leq \theta^k(u^k) \text{ for all } u \in R_+^m \qquad (13)$$

- At each iteration, we try to find the most violated inequality such that

$$\theta^k(u^k) = s^k > u^k b + (c - u^k A)x^{k+1} = \theta(u^k)$$

- The best dual

$$\hat{u} := \arg\max\{\theta(u^i) | i = 1, \ldots, k\}$$

# Primal-dual Based on Convex Analysis Language

- The set of optimal solutions in (11) (for given $u$)

$$\hat{X}^k(u) := \{x \in conv(X^k) | L(x,u) = \theta^k(u)\} \qquad (14)$$

### Theorem

*An optimal solution $u^k$ of the restricted master (12) is characterized by the existence of some $\hat{x} \in \hat{X}^k(u^k)$ satisfying followings*

$$A\hat{x} - b \geq 0, \; u^k \geq 0, \; u^k(A\hat{x} - b) = 0$$

### Proof.

Elementary convex alalysis says that $\theta^k$ is concave and its subdifferential (the set of $g \in R^m$ such that $\theta^k(v) \leq \theta^k(u) + (v-u)g$ for all $v \in R^m_+$) is

$$\partial\theta^k(u) = \bigcup_{x \in \hat{X}^k(u)} \{g = b - Ax\}$$

# Conclusion on Standard Column Generation

- It consists in merely restricting X to $X^k$ in (2), leaving everything else unchanged. In the dual space, $\theta$ is correspondingly *impoverished* to $\theta^k$. In nonlinear optimization, this is known as the cutting-plane method of Kelley or Cheney-Goldstein
- It needs a number of columns to start
- It can be desperately slow: an example constructed by A.S. Nemirovskii
- Note that the method of subgradient is also a valid candidate to maximize $\theta$, as well as the ellipsoid method. Both methods behave poorly in practice,though.

# Motivation

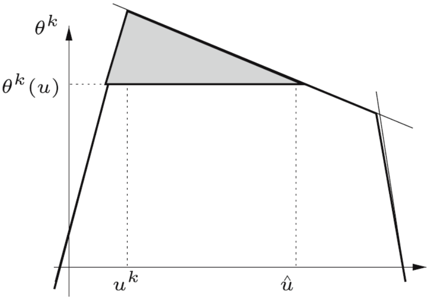


Figure: The safeguard polyhedron $P$ for $k = 3$. $\theta^k(u) = \theta(\hat{u})$

- Because $\theta^k \geq \theta$, the optimal $(u^*, \theta(u^*))$ clearly lies somewhere in the safeguard polyhedron $P$
- Standard column generation chooses $u^k$ as the highest point in $P$
- Stabilized column generation : $u^k$ is replaced by some less high but more central point in $P$

# Schematic Stabilized Column Generation

The dual restricted master (12) being replaced by

$$\max_{u \in R^m_+} \tilde{\theta}^k(u), \text{ where } \tilde{\theta}^k(u) := \theta^k(u) - S(u - \hat{u}) \tag{15}$$

1: **procedure** SCHEMATIC STABILIZED COLUMN GENERATION
2: Choose an initial stability center $\hat{u}$. Select an initial set of $k$ columns
3: Compute an optimal solution $u^k$ of the *stabilized dual restricted master problem* (15).
4: Call the oracle (4) at $u^k$ to obtain the new column $x^{k+1}$ and the dual value $\theta(u^k) = L(x^{k+1}, u^k)$. Perform the stopping test
5: If $\theta(u^k) > \theta(\hat{u})$ set $\hat{u} = u^k$
6: Increase $k$ by 1 and go to 2
7: **end procedure**

- $u^k$ no longer maximizes $\theta^k$
- $\tilde{x}$ should be constructed from (15)

# Conjugate Calculus

- Introduce *conjugate function*

$$S^*(g) := \max_{v \in R^m_+} \{vg - S(v)\} \tag{16}$$

- $g$ shall be interpreted as a slack for constraints $Ax \geq b$
- $v$ stands for the deviation between $u$ and the stability center $\hat{u}$
- *Fenchel dual* to (15)

$$\min cx + \hat{u}g + S^*(g), \; Ax \geq b - g, \; (x, g) \in conv(X^k) \times R^m \tag{17}$$

- Standard column generation uses $S \equiv 0$, whose conjugate is clearly

$$S^*(g) = \begin{cases} 0, & \text{if } g = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

# Primal-dual Based on Convex Analysis Language Revisited

Assume that S is a convex function over $R^m$. Consider an optimal solution $u^k$ of (15) such that $S(v) < +\infty$ for all $v$ close to $u^k - u$. Then

## Theorem

*There exist $\tilde{x} \in \hat{X}^k(u^k)$ and a subgradient $\tilde{g} \in \partial S(u^k - \hat{u})$ such that*

$$A\hat{x} - b + \tilde{g} \geq 0,\ \ u^k \geq 0,\ \ u^k(A\hat{x} - b + \tilde{g}) = 0$$

## Proof.

$$\partial\tilde{\theta}^k(u) = \partial\theta^k(u^k) - \partial S(u^k - \hat{u})$$

$\square$

# Boxstep

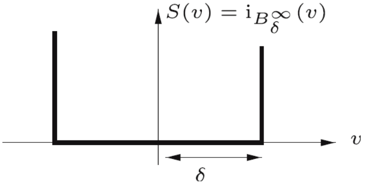


Figure: $l_\infty$ box around $\hat{u}$

- The stabilizing problem is

$$\max\{\theta^k(u) | u \geq 0, \|u - \hat{u}\|_\infty \leq \delta\} = \max_{u \in R^m_+} \theta^k(u) - i_{B^\infty_\delta}(u - \hat{u})$$

- $i^*_{B^\infty_\delta}(g) = \delta \|g\|_1$
- Fenchel dual

$$\min cx + \hat{u}g + \delta\|g\|_1, \;\; Ax \geq b - g, \;\; (x, g) \in conv(X^k) \times R^m$$

# Polyhedral Penalties

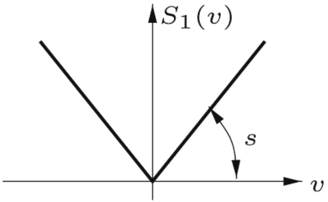


Figure: Stabilization by $l_1$ penalty

- The stabilizing problem is

$$\max_{u \in R_+^m} \theta^k(u) - s \sum_{j=1}^{m} |u_j - \hat{u}_j| = \max_{u \in R_+^m} \theta^k(u) - s\|u_j - \hat{u}_j\|_1$$

- $(s\|\cdot\|_1)^*(g) = i_{B_s^\infty}$
- Fenchel dual

$$\min cx + \hat{u}g, \; Ax \geq b - g, \; \|g\| \leq s, \; (x, g) \in conv(X^k) \times R^m$$

# Polyhedral Penalties - More

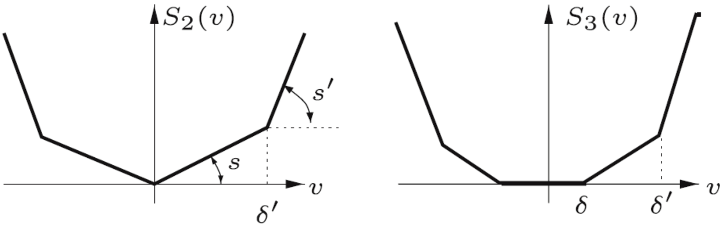


Figure: Stabilization by $l_1$ penalty

- Necessity to manage carefully their parameters $s, \delta$, etc

# Euclidean Penalties



- Stabilizers suggests that they tend to mimic a quadratic function
- In fact, stabilization is a rather fundamental paradigm in nonlinear programming (not limited to Kelley's method) and there are good reasons to use quadratic stabilizers
- The stabilizing problem is

$$\max_{u \in R^m_+} \theta^k(u) - \frac{1}{2t}\|u - \hat{u}\|^2$$

- $(\frac{1}{2t}\|\cdot\|^2)^*(g) = \frac{t}{2}\|g\|^2$
- Fenchel dual

$$\min cx + \hat{u}g + \frac{t}{2}\|g\|^2,\ Ax \geq b - g,\ (x, g) \in conv(X^k) \times R^m$$

- No more LP formulations

# Ok, Show me Results!



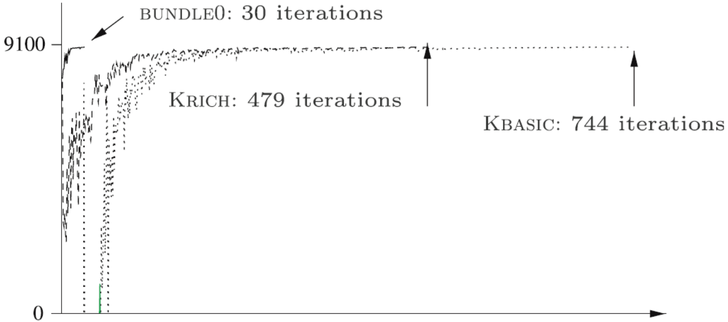


Figure: TSP: Kelley versus bundle on bays29

- Number of iterations is dramatically reduced

# References


- O. Briant, et. al., Comparison of bundle and classical column generation, *Mathematical Programming Ser.A* 299-344, 113, 2008