Global optimality conditions for discrete and nonconvex optimization, with applications to Lagrangian heuristics, core problems, and column generation

Michael Patriksson

(with Torbjörn Larsson, Linköping University)

Summary

- Illustration: new *radical* set covering heuristic
- Global optimality conditions for general problems, including integer ones
  - $\sim$ convex saddle-point conditions
  - Lagrangian perturbations: near-optimality, near-complementarity
  - Analysis of and guidelines for Lagrangian heuristics
- Applications
  - Core problems; column generation
  - In both cases: additional near-complementarity constraints

A general problem

$$f^* := \text{minimum} \quad f(\boldsymbol{x}), \tag{1a}$$
$$\text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m, \tag{1b}$$
$$\boldsymbol{x} \in X \tag{1c}$$

$f : \mathbb{R}^n \mapsto \mathbb{R}$, $\boldsymbol{g} : \mathbb{R}^n \mapsto \mathbb{R}^m$ cont., $X \subset \mathbb{R}^n$ compact

$$\theta(\boldsymbol{u}) := \underset{\boldsymbol{x}\in X}{\text{minimum}} \left\{ f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \right\}, \quad \boldsymbol{u} \in \mathbb{R}^m \tag{2}$$

$$\theta^* := \underset{\boldsymbol{u}\in\mathbb{R}^m_+}{\text{maximum}} \ \theta(\boldsymbol{u}) \tag{3}$$

Duality gap: $\Gamma := f^* - \theta^*$.

Lagrangian heuristic,

Started at some vector $\overline{\boldsymbol{x}}(\boldsymbol{u}) \in X$, adjust it through a finite number of steps with properties

1. sequence utilize information from the Lagrangian dual problem,
2. sequence remains within X, and
3. terminal vector, if possible, primal feasible, hopefully also near-optimal in (2)

Conservative: initial vector near $\boldsymbol{x}(\boldsymbol{u})$; local moves

Radical: allows the resulting vector to be far from $\boldsymbol{x}(\boldsymbol{u})$; includes starting far away; solving restrictions (e.g., Benders' subproblem)

Lagrangian heuristic, 2

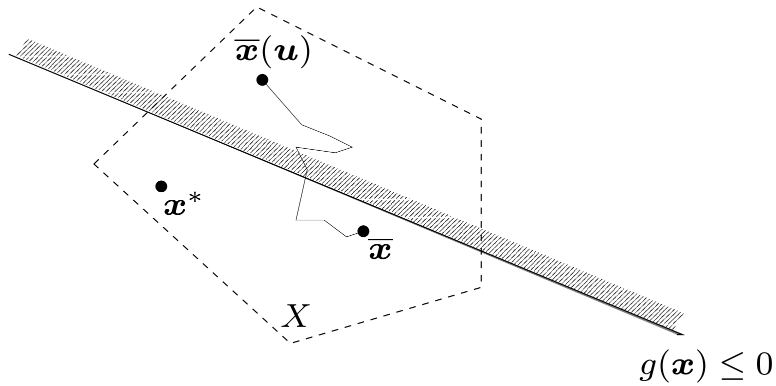


Figure 1: A Lagrangian heuristic

The set covering problem

$$f^* := \text{minimum} \sum_{j=1}^{n} c_j x_j, \tag{4a}$$

$$\text{subject to} \sum_{j=1}^{n} \boldsymbol{a}_j x_j \geq \mathbf{1}^m, \tag{4b}$$

$$\boldsymbol{x} \in \{0,1\}^n, \tag{4c}$$

Lagrangian: $L(\boldsymbol{x}, \boldsymbol{u}) := (\mathbf{1}^m)^{\mathrm{T}} \boldsymbol{u} + \bar{\boldsymbol{c}}^{\mathrm{T}} \boldsymbol{x}$, $\boldsymbol{u} \in \mathbb{R}^m$

Reduced cost vector $\bar{\boldsymbol{c}} := \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}} \boldsymbol{u}$.

Lagrangian dual

$$\theta^* := \text{maximum} \quad \theta(\boldsymbol{u}), \qquad \text{subject to} \quad \boldsymbol{u} \geq \mathbf{0}^m$$

$$\theta(\boldsymbol{u}) := (\mathbf{1}^m)^{\mathrm{T}} \boldsymbol{u} + \sum_{j=1}^{n} \underset{x_j \in \{0,1\}}{\text{minimum}} \; \bar{c}_j x_j, \qquad \boldsymbol{u} \geq \mathbf{0}^m$$

$$x_j(\boldsymbol{u}) \begin{cases} = 1, & \text{if } \bar{c}_j < 0, \\ \in \{0,1\}, & \text{if } \bar{c}_j = 0, \\ = 0, & \text{if } \bar{c}_j > 0 \end{cases}$$

We consider a classic type of polynomial heuristic.

Primal greedy heuristic

(**Input**) $\bar{\boldsymbol{x}} \in \{0,1\}^n$, cost vector $\boldsymbol{p} \in \mathbb{R}^n$
(**Output**) $\hat{\boldsymbol{x}} \in \{0,1\}^n$, feasible in (1)
(**Starting phase**) Given $\bar{\boldsymbol{x}}$, delete covered rows, delete variables x_j with $\bar{x}_j = 1$
(**Greedy insertion**) Identify variable x_τ with minimum p_j relative to number of uncovered rows covered. Set $x_\tau := 1$. Delete covered rows, delete x_τ. Unless uncovered rows remain, stop; $\tilde{\boldsymbol{x}} \in \{0,1\}^n$ feasible solution.
(**Greedy deletion**) Identify variable x_τ with $\tilde{x}_\tau = 1$ present only in over-covered rows and maximum p_j relative to k_j. Set $\tilde{x}_\tau := 0$. Repeat.

Instances

Classic heuristics:

(I) Let $\bar{\boldsymbol{x}} := \mathbf{0}^n$ and $\boldsymbol{p} := \boldsymbol{c}$
Chvátal (1979)

(II) Let $\bar{\boldsymbol{x}} := \mathbf{0}^n$ and $\boldsymbol{p} := \bar{\boldsymbol{c}}$, at dual vector $\boldsymbol{u}$
$\sim$ Balas and Ho (1980)

(III) Let $\bar{\boldsymbol{x}} := \boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{p} := \boldsymbol{c}$
Beasley (1987, 1993) and Wolsey (1998)

(IV) Let $\bar{\boldsymbol{x}} := \boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{p} := \bar{\boldsymbol{c}}$
$\sim$ Balas and Carrera (1996)

New primal greedy heuristic

To be motivated later:
Combination of $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$ (or Lagrangian and complementarity) { here, $\lambda \in [1/2, 1]$ }

$$\boldsymbol{p}(\lambda) := \lambda\bar{\boldsymbol{c}} + (1-\lambda)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \lambda[\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}] + (1-\lambda)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}$$

(I) & (III): $\lambda = 1/2$ (original cost)
(II) & (IV): $\lambda = 1$ (Lagrangian cost)

Test both $\bar{\boldsymbol{x}} := \mathbf{0}^n$ ("radical") and $\bar{\boldsymbol{x}} := \boldsymbol{x}(\boldsymbol{u})$ ("conservative")
Test case: `rail507`, with bounds $[172.1456, 174]$ ($n = 63,009$; $m = 507$)

$\boldsymbol{u}$ generated by a subgradient algorithm

Test 1: varying λ

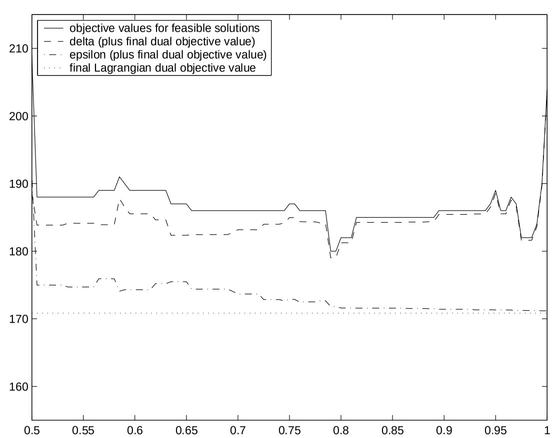


Figure 2: Objective value vs. value of λ

Test 2: Conservative vs. radica

$\lambda = 0.9$
Ran three heuristics from iterations $t = 200$ to $t = 500$ of the subgradient algorithm.

1. (III): $\bar{\boldsymbol{x}} := \boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{p}(1/2) = \boldsymbol{c}$. Conservative.
2. $\bar{\boldsymbol{x}} := \boldsymbol{x}(\boldsymbol{u})$ and $\boldsymbol{p}(0.9)$. Conservative.
3. $\bar{\boldsymbol{x}} := \mathbf{0}^n$ and $\boldsymbol{p}(0.9)$. Radical.

Histograms of objective values

Results of Test 2

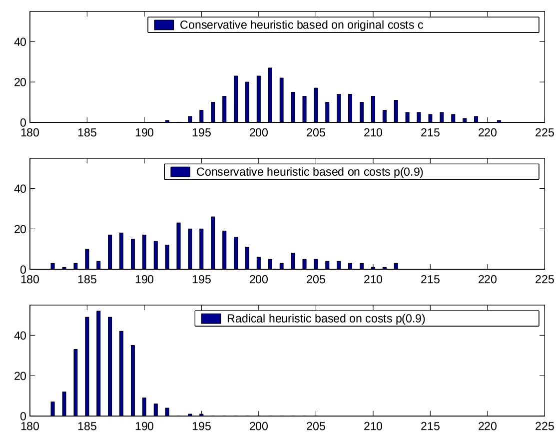


Figure 3: Quality obtained by three greedy heuristics

Conclusion

- Remarkable difference between the heuristics
- Simple modification of (III) improves it
- Radical one consistently provides good solutions

	[(III)]	[p(0.9)/cons.]	[p(0.9)/rad.]
maximum :	221	212	195
mean :	203.99	194.45	186.55
minimum :	192	182	182

Why is it good to (i) use radical Lagrangian heuristics with (ii) an objective function which is neither the original nor the Lagrangian, but a combination?

Global optimality conditions, 1

$(\boldsymbol{x}, \boldsymbol{u}) \in X \times \mathbb{R}^m_+$

$$f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}) \leq \theta(\boldsymbol{u}), \quad \text{(5a)}$$

$$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m, \quad \text{(5b)}$$

$$\boldsymbol{u}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}) = 0 \quad \text{(5c)}$$

Equivalent statements for pair $(\boldsymbol{x}^*, \boldsymbol{u}^*) \in X \times \mathbb{R}^m_+$:

- satisfies (5)
- saddle point of $L(\boldsymbol{x}, \boldsymbol{u}) := f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x})$:

  $$L(\boldsymbol{x}^*, \boldsymbol{v}) \leq L(\boldsymbol{x}^*, \boldsymbol{u}^*) \leq L(\boldsymbol{y}, \boldsymbol{u}^*), \ (\boldsymbol{y}, \boldsymbol{v}) \in X \times \mathbb{R}^m_+$$

- primal–dual optimal and $f^* = \theta^*$

Global optimality conditions,

Further, given any $\boldsymbol{u} \in \mathbb{R}^m_+$,

$$\{\, \boldsymbol{x} \in X \mid (5) \text{ is satisfied} \,\} = \begin{cases} X^*, & \text{if } \theta(\boldsymbol{u}) = f^*, \\ \emptyset, & \text{if } \theta(\boldsymbol{u}) < f^* \end{cases}$$

- Inconsistency if either $\boldsymbol{u}$ is non-optimal *or* there is a positive duality gap!
- Then (5) is inconsistent; no optimal solution is found by applying it from an optimal dual sol.
- Equality constraints: not even a feasible solution is found!
- Why (and when) then are Lagrangian heuristics successful for integer programs?

New global optimality conditions, 1

$(\boldsymbol{x}, \boldsymbol{u}) \in X \times \mathbb{R}^m_+$

$$f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \le \theta(\boldsymbol{u}) + \varepsilon, \quad \text{(6a)}$$
$$\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}^m, \quad \text{(6b)}$$
$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \ge -\delta, \quad \text{(6c)}$$
$$\varepsilon + \delta \le \Gamma, \ \text{(duality gap)} \quad \text{(6d)}$$
$$\varepsilon, \delta \ge 0 \quad \text{(6e)}$$

- (6a): ε-optimality
- (6c): δ-complementarity
- System equivalent to previous one when duality gap is zero

New global optimality conditions, 2

Equivalent statements for pair $(\boldsymbol{x}^*, \boldsymbol{u}^*) \in X \times \mathbb{R}^m_+$:

- satisfies (6)
- $\varepsilon + \delta = \Gamma$; further,

$$L(\boldsymbol{x}^*, \boldsymbol{v}) - \delta \le L(\boldsymbol{x}^*, \boldsymbol{u}^*) \le L(\boldsymbol{y}, \boldsymbol{u}^*) + \varepsilon, \ (\boldsymbol{y}, \boldsymbol{v}) \in X \times \mathbb{R}^m_+$$

- primal–dual optimal

Given any $\boldsymbol{u} \in \mathbb{R}^m_+$,

$$\{\, \boldsymbol{x} \in X \mid (6) \text{ is satisfied} \,\} = \begin{cases} X^*, & \text{if } \theta(\boldsymbol{u}) = f^* - \Gamma, \\ \emptyset, & \text{if } \theta(\boldsymbol{u}) < f^* - \Gamma \end{cases}$$

Next up: characterize near-optimal solutions

Relaxed optimality conditions, 1

$$f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \le \theta(\boldsymbol{u}) + \varepsilon, \quad \text{(7a)}$$
$$\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}^m, \quad \text{(7b)}$$
$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \ge -\delta, \quad \text{(7c)}$$
$$\varepsilon + \delta \le \Gamma + \kappa, \quad \text{(7d)}$$
$$\varepsilon, \delta, \kappa \ge 0 \quad \text{(7e)}$$

$\kappa \sim$ sum of non-optimality in primal and dual
If consistent, $\Gamma \le \varepsilon + \delta \le \Gamma + \kappa$

- (Near-optimality) $f(\boldsymbol{x}) \le \theta(\boldsymbol{u}) + \Gamma + \kappa$
  [$\boldsymbol{u}$ optimal: $f(\boldsymbol{x}) \le f^* + \kappa$]
- (Lagrangian near-optimality) $(\boldsymbol{x}, \boldsymbol{u})$ optimal:
  $\theta^* \le f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \le f^*$

Relaxed optimality conditions, 2

$\boldsymbol{u} \in \mathbb{R}^m_+$ α-optimal

$$\{\, \boldsymbol{x} \in X \mid (7) \text{ is satisfied} \,\} = \begin{cases} X^{\kappa-\alpha}, & \text{if } \kappa \ge \alpha, \\ \emptyset, & \text{if } \kappa < \alpha \end{cases} \quad (8)$$

- Characterize optimal solutions when $\kappa = \alpha$!
- Valid for all duality gaps, also convex problems
- Goal: construct Lagrangian heuristics so that (7) is satisfied for small values of κ
- Previous Lagrangian heuristics ignore near-complementarity

Numerical example, 1

$$f^* := \text{minimum} \quad f(\boldsymbol{x}) := -x_2, \tag{9a}$$
$$\text{subject to} \quad g(\boldsymbol{x}) := x_1 + 4x_2 - 6 \le 0, \tag{9b}$$
$$\boldsymbol{x} \in X := \{\, \boldsymbol{x} \in \mathcal{Z}^2 \mid 0 \le x_1 \le 4;\; 0 \le x_2 \le 2 \,\} \tag{9c}$$

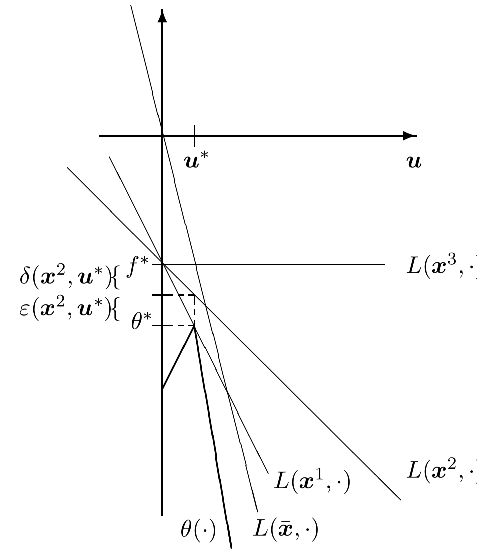
$$L(\boldsymbol{x}, u) = ux_1 + (4u-1)x_2 - 6u$$

$$\theta(u) := \begin{cases} 2u - 2, & 0 \le u \le 1/4, \\ -6u, & 1/4 \le u, \end{cases}$$

$u^* = 1/4$, $\theta^* = -3/2$

Three optimal solutions, $\boldsymbol{x}^1 = (0,1)^{\mathrm{T}}$, $\boldsymbol{x}^2 = (1,1)^{\mathrm{T}}$, and $\boldsymbol{x}^3 = (2,1)^{\mathrm{T}}$; $f^* = -1$; $\Gamma = f^* - \theta^* = 1/2$

Numerical example, 2



Numerical example, 3

- For $\boldsymbol{x}^2$, $\varepsilon(\boldsymbol{x}^2, \boldsymbol{u}^*)$ is the vertical distance between the two functions θ and $L(\boldsymbol{x}^2, \cdot)$ at $\boldsymbol{u}^*$
- Remaining vertical distance to f^* is minus the slope of $L(\boldsymbol{x}^2, \cdot)$ at $\boldsymbol{u}^*$ [which is $\boldsymbol{g}(\boldsymbol{x}^2) = -1$] times $\boldsymbol{u}^*$, that is, $\delta(\boldsymbol{x}^2, \boldsymbol{u}^*) = 1/4$
- $\boldsymbol{x}^1$: $\varepsilon = 0$, $\delta = 1/2$; $\boldsymbol{x}^2$: $\varepsilon = 1/4$, $\delta = 1/4$; $\boldsymbol{x}^3$: $\varepsilon = 1/4$, $\delta = 0$. Unpredictable, except that $\varepsilon + \delta = \Gamma$ must hold at an optimal solution
- Candidate vector $\bar{\boldsymbol{x}} := (2,0)^{\mathrm{T}}$: $\varepsilon = 1/2$, $\delta = 1$ [the slope of $L(\bar{\boldsymbol{x}}, \cdot)$ at $\boldsymbol{u}^*$ is -4]; here, $\theta^* + \varepsilon + \delta = f(\bar{\boldsymbol{x}}) = 0 > f^*$, so $\bar{\boldsymbol{x}}$ cannot be optimal

Numerical example, 4 (a

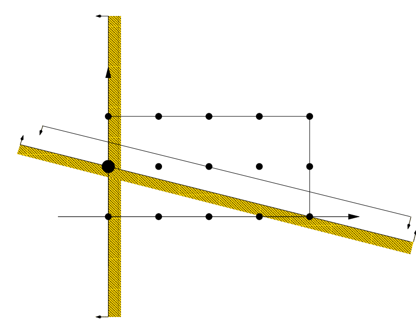


Figure 4: The optimal solution $\boldsymbol{x}^1$ (marked with large circle) is specified by the global optimality conditions (6) for $(\varepsilon, \delta) := (0, 1/2)$. The shaded regions and arrows illustrate the conditions (6a) and (6c) corresponding to $u = u^*$.

Numerical example, 4 (b)

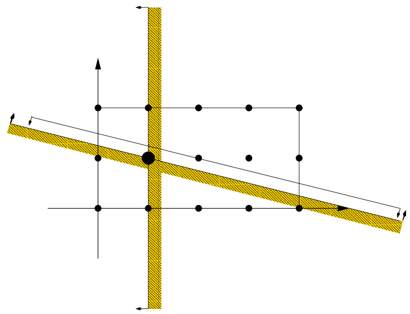


Figure 5: The optimal solution $\boldsymbol{x}^2$ (marked with large circle) is specified by the global optimality conditions (6) for $(\varepsilon, \delta) := (1/4, 1/4)$. The shaded regions and arrows illustrate the conditions (6a) and (6c) corresponding to $u = u^*$.

Numerical example, 4 (c

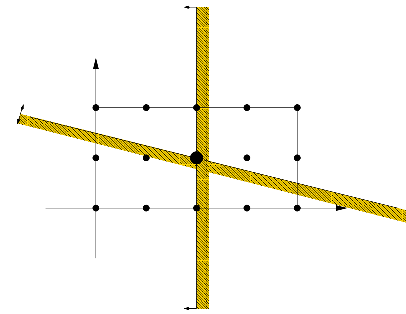


Figure 6: The optimal solution $\boldsymbol{x}^3$ (marked with large circle) is specified by the global optimality conditions (6) for $(\varepsilon, \delta) := (1/2, 0)$. The shaded regions and arrows illustrate the conditions (6a) and (6c) corresponding to $u = u^*$.

A dissection of heuristics

- (**Small duality gap**) $\bar{\boldsymbol{x}}(\boldsymbol{u})$ Lagrangian near-optimal, small complementarity violations $\Rightarrow$ conservative Lagrangian heuristics ***sufficient*** (if they can reduce large complementarity violations)
- (**Large duality gap**) Dual solution far from optimal/large duality gap $\Rightarrow$ radical Lagrangian heuristics ***necessary***

The first experiment agai

- The cost used was $h(\boldsymbol{x}) := \lambda[f(\boldsymbol{x}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})] + (1-\lambda)[-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})], \quad \lambda \in [1/2, 1]$
- Rail problems often have over-covered optimal solutions, hence complementarity is violated substantially; δ large, ε rather small, hence $\lambda \lesssim 1$ a good choice (cf. Figure 1)
- ε still not very close to zero, so radical heuristics better than conservative

Equality constraints

$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}^\ell$

$$f(\boldsymbol{x}) + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}) \le \theta(\boldsymbol{v}) + \varepsilon, \qquad (10a)$$
$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}^\ell, \qquad (10b)$$
$$0 \le \varepsilon \le \Gamma \qquad (10c)$$

- Global optimum $\Longleftrightarrow \varepsilon = \Gamma$
- Saddle-type condition for $L(\boldsymbol{x},\boldsymbol{v}) := f(\boldsymbol{x}) + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})$ over $X \times \mathbb{R}^\ell$:

$$L(\boldsymbol{x},\boldsymbol{w}) \le L(\boldsymbol{x},\boldsymbol{v}) \le L(\boldsymbol{y},\boldsymbol{v}) + \varepsilon, \quad (\boldsymbol{y},\boldsymbol{w}) \in X \times \mathbb{R}^\ell$$

Application to core problems

- Core problems used to solve large-scale set-covering and binary knapsack problems.
- Guess which $x_j^* = 1$ or $x_j^* = 0$.
- Often based on the LP reduced costs: $\bar{c}_j \ll 0 \Longrightarrow x_j^* = 1$; $\bar{c}_j \gg 0 \Longrightarrow x_j^* = 0$. Fix according to a threshold value for $\bar{c}_j$.
- The remaining part of $\boldsymbol{x}$ is the "difficult" part of the problem.
- Standard method ignores complementarity.

Application to column generation

$$f^* := \text{minimum} \sum_{j=1}^{n} \boldsymbol{c}_j^{\mathrm{T}}\boldsymbol{x}_j, \qquad (11a)$$
$$\text{subject to} \sum_{j=1}^{n} \boldsymbol{A}_j\boldsymbol{x}_j \ge \boldsymbol{b}, \qquad (11b)$$
$$\boldsymbol{x}_j \in X_j, \qquad j = 1,\dots,n \qquad (11c)$$

$X_j \subset \mathbb{R}^{n_j}$, $j = 1,\dots,n$, are finite
$\boldsymbol{c}_j \in \mathbb{R}^{n_j}$, $\boldsymbol{A}_j \in \mathbb{R}^{m\times n_j}$, $j = 1,\dots,n$, and $\boldsymbol{b} \in \mathbb{R}^m$
$\boldsymbol{u} \in \mathbb{R}^m_+$ multipliers for the side constraints (10b)

Disaggregated master problem

$$f^* = \text{minimum} \sum_{j=1}^{n}\sum_{i=1}^{P_j} (\boldsymbol{c}_j^{\mathrm{T}} x_j^i)\, \lambda_j^i, \qquad (12a)$$
$$\text{subject to} \sum_{j=1}^{n}\sum_{i=1}^{P_j} (\boldsymbol{A}_j x_j^i)\, \lambda_j^i \ge \boldsymbol{b}, \qquad (12b)$$
$$\sum_{i=1}^{P_j} \lambda_j^i = 1, \quad j = 1,\dots,n, \qquad (12c)$$
$$\lambda_j^i \in \{0,1\}, \quad i = 1,\dots,P_j, \quad j = 1,\dots,n \qquad (12d)$$

P_j: number of points in the set X_j, denoted by x_j^i
Let $p_j < P_j$, $\bar{\boldsymbol{u}}$ near-optimal to Lagrangian dual

Restricted master problem

$$
\begin{aligned}
f_r^* := \text{minimum} \; & \sum_{j=1}^{n} \sum_{i=1}^{p_j} (\boldsymbol{c}_j^{\mathrm{T}} x_j^i)\, \lambda_j^i, \\
\text{subject to} \; & \sum_{j=1}^{n} \sum_{i=1}^{p_j} (\boldsymbol{A}_j x_j^i)\, \lambda_j^i \geq \boldsymbol{b}, \\
& \sum_{j=1}^{n} \sum_{i=1}^{p_j} (\bar{\boldsymbol{u}}^{\mathrm{T}} \boldsymbol{A}_j x_j^i)\, \lambda_j^i \leq \bar{\boldsymbol{u}}^{\mathrm{T}} \boldsymbol{b} + \delta, \\
& \sum_{i=1}^{p_j} \lambda_j^i = 1, \quad j = 1, \ldots, n, \\
& \lambda_j^i \in \{0, 1\}, \quad i = 1, \ldots, p_j, \quad j = 1, \ldots, n
\end{aligned}
$$

Complementarity near-fulfillment side constraint