Large Scale Mixed-Integer Optimization: a Solution Method with Supply Chain Applications

Robin Vujanic, Paul Goulart, Sebastien Mariethoz,
Peyman Mohajerin Esfahani, Manfred Morari

Automatic Control Laboratory
Department of Electrical Engineering
Swiss Federal Institute of Technology (ETHZ)

August 16, 2014

Outline

Large MIPs: Problems' Structure

We investigate large scale problems in the form:

$$(P) : \begin{cases} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \leq b \\ & x_i \in X_i \qquad\qquad \forall i \in I, \end{cases}$$

with X_i compact mixed integer polyhedral sets

$$X_i = \left\{ x \in \mathbb{R}^{r_i} \times \mathbb{Z}^{z_i} \;\middle|\; A_i x \leq d_i \right\}.$$

Interested in $|I| \gg$ coupling constraints.

Large MIPs: Problems' Structure

We investigate large scale problems in the form:

$$(P):\begin{cases} \min_x & \sum_{i\in I} c_i x_i \\ \text{s.t.} & \sum_{i\in I} H_i x_i \leq b \\ & x_i \in X_i \qquad\qquad \forall i \in I, \end{cases}$$

with X_i compact mixed integer polyhedral sets

$$X_i = \left\{ x \in \cancel{\mathbb{R}^{r_i}} \times \mathbb{Z}^{z_i} \;\middle|\; A_i x \leq d_i \right\}.$$

Interested in $|I| \gg$ coupling constraints.

Application Example: Partial Shipments

- distributor has to supply M products to N customers
  - usually, $N \gg M$
- often, avail. inventory $<$ total demand
  - demand uncertainty and high storage costs
  - restrictions on manufacturing capacity
- thus
  - **EITHER** fully supply a smaller set of customers
  - **OR** partially supply a larger set
- but shipments cannot be too small

Partial Shipments Problem [Dawande '06]

Allocate products to customers in the presence of shipping restrictions.

Application Example: Partial Shipments

- distributor has to supply M products to N customers
  - usually, $N \gg M$
- often, avail. inventory $<$ total demand
  - demand uncertainty and high storage costs
  - restrictions on manufacturing capacity
- thus
  - **EITHER** fully supply a smaller set of customers
  - **OR** partially supply a larger set
- but shipments cannot be too small

Partial Shipments Problem [Dawande '06]

Allocate products to customers in the presence of shipping restrictions.

Optimization Problem Model (1/2) [Dawande '06]

$$\begin{array}{ll} i = \text{customer}, & \text{j} = \text{product} \\ w_i = \text{ship to i (yes/no)}, & S_i^j = \text{fraction shipped} \end{array}$$

$$\left\{ \begin{array}{lll} \max & \sum\limits_i \text{reward for } w_i + \sum\limits_{i,j} \text{reward for } S_i^j & \\ \text{s.t.} & \sum\limits_i S_i^j \leq \text{inventory budget of prod. j} & \\ & x_i = \begin{bmatrix} w_i \\ S_i^j \end{bmatrix} \in \left\{ \begin{array}{lll} \text{if } w_i = 1 & \Rightarrow & \sum_j S_i^j \geq \text{min. shipment} \\ & & \text{and } S_i^j \leq \text{demand of } j \\ \text{if } w_i = 0 & \Rightarrow & S_i^j = 0 \\ w_i \in \{0, 1\} & & \end{array} \right\} & \end{array} \right.$$

Compare with generic model P:

$$\left\{ \begin{array}{lll} \min_x & \sum_{i \in I} c_i x_i & \\ \text{s.t.} & \sum_{i \in I} H_i x_i \leq b & \\ & x_i \in X_i & \forall i \in I, \end{array} \right.$$

More examples

- power systems
  - (CDC) electric vehicles charging coordination
  - demand side management of TCLs
  - scheduling a *large* fleet of generators
- many traditional combinatorial problems
  - (multidimendional) knapsack problems
  - generalized assignments
  - ...
- portfolio models for small investors
- automatic reformulations
- maybe your application?

Outline

Problem's Decomposition

Obtain decomposition using duality:

$$\begin{array}{ll} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in X_i \end{array} \Rightarrow \underbrace{\sum_{i \in I} \left\{ \min_{x_i \in X_i} c_i x_i + \lambda'(H_i x_i) \right\} - \lambda' b}_{\doteq d(\lambda)}$$

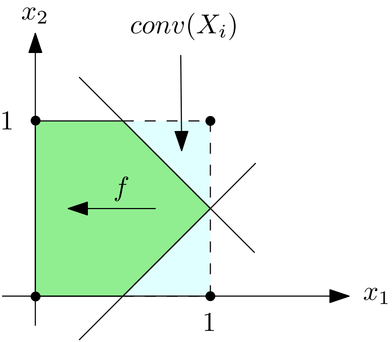
- Lagrangian dual (or outer) problem:

$$(D) : \left\{ \begin{array}{ll} \sup_\lambda & d(\lambda) \\ \text{s.t.} & \lambda \succeq 0 \end{array} \right.$$

Consider solutions to **inner problem** as candidate solutions for P

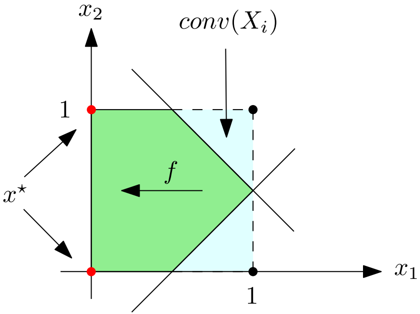
$$x_i(\lambda) \in \arg\min_{x_i \in X_i} \left\{ (c_i + \lambda' H_i) x_i \right\} \quad \lambda \in \mathbb{R}^m_+.$$

Issues affecting inner solutions



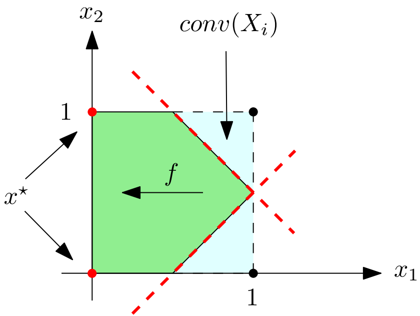
$$
\begin{array}{ll}
\min & -x_1 \\
\text{s.t.} & x_1 - x_2 \leq 0.5 \\
 & x_1 + x_2 \leq 1.5 \\
 & x_1, x_2 \in \{0, 1\}
\end{array}
$$

Issues affecting inner solutions



$$
\begin{array}{ll}
\min & -x_1 \\
\text{s.t.} & x_1 - x_2 \leq 0.5 \\
& x_1 + x_2 \leq 1.5 \\
& x_1, x_2 \in \{0, 1\}
\end{array}
$$

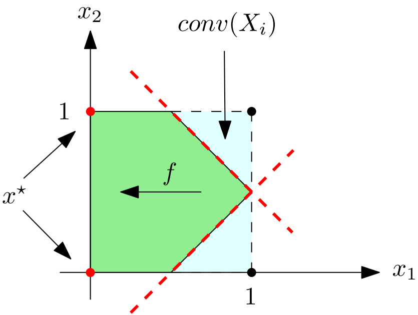
Issues affecting inner solutions



$$\begin{array}{ll} \min & -x_1 \\ \text{s.t.} & x_1 - x_2 \leq 0.5 \\ & x_1 + x_2 \leq 1.5 \\ & x_1, x_2 \in \{0,1\} \end{array}$$

$$\Rightarrow \max_{\lambda \geq 0} \min_{x_1, x_2 \in \{0,1\}} \quad -x_1 + \lambda_1 \underline{(x_1 - x_2 - 0.5)} + \lambda_2 \underline{(x_1 + x_2 - 1.5)}$$

Issues affecting inner solutions

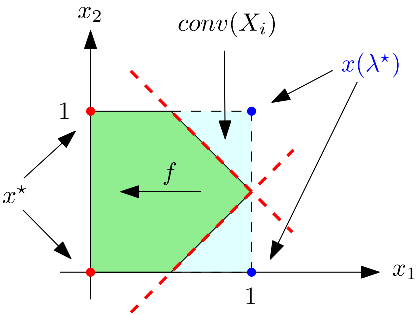


$$\begin{array}{ll} \min & -x_1 \\ \text{s.t.} & x_1 - x_2 \leq 0.5 \\ & x_1 + x_2 \leq 1.5 \\ & x_1, x_2 \in \{0,1\} \end{array}$$

$$\Rightarrow \max_{\lambda \geq 0} \min_{x_1, x_2 \in \{0,1\}} \quad -x_1 + \lambda_1 (x_1 - x_2 - 0.5) + \lambda_2 (x_1 + x_2 - 1.5)$$

$$\Rightarrow d^\star = -1 \qquad \lambda^\star = (0.25, 0.25)$$

Issues affecting inner solutions

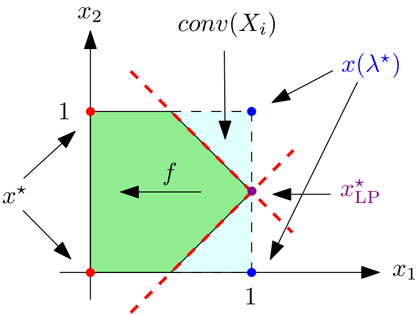


$$\begin{aligned} \min \quad & -x_1 \\ \text{s.t.} \quad & x_1 - x_2 \leq 0.5 \\ & x_1 + x_2 \leq 1.5 \\ & x_1, x_2 \in \{0, 1\} \end{aligned}$$

$$\Rightarrow \max_{\lambda \geq 0} \min_{x_1, x_2 \in \{0,1\}} \quad -x_1 + \lambda_1(x_1 - x_2 - 0.5) + \lambda_2(x_1 + x_2 - 1.5)$$

$$\Rightarrow d^\star = -1 \qquad \lambda^\star = (0.25, 0.25)$$

Issues affecting inner solutions



$$\begin{array}{ll} \min & -x_1 \\ \text{s.t.} & x_1 - x_2 \leq 0.5 \\ & x_1 + x_2 \leq 1.5 \\ & x_1, x_2 \in \{0, 1\} \end{array}$$

$$\Rightarrow \max_{\lambda \geq 0} \min_{x_1, x_2 \in \{0,1\}} \quad -x_1 + \lambda_1 (x_1 - x_2 - 0.5) + \lambda_2 (x_1 + x_2 - 1.5)$$

$$\Rightarrow d^\star = -1 \qquad \lambda^\star = (0.25, 0.25)$$

Main Idea

The following convexification of P is strongly related to D

$$(P_{\mathrm{LP}}) : \begin{cases} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in \mathrm{conv}(X_i) \quad \forall i \in I \end{cases}$$

Known that:

- $J^\star_{P_{LP}} = J^\star_D$ (**surprising!**)
- $J^\star_D \to J^\star_P$ as $|I| \to \infty$, vanishing duality gap [Aubin '76]

We show

1. **an optimal solution $x^\star_{\mathrm{LP}}$ to P_{LP} is useful**
   - feasible w.r.t. the complicating constraints
   - $x^\star_{\mathrm{LP}}$ is non-integer in at most m subsystems
   - Shapley–Folkman–Starr Theorem
2. **we can compute it**, borrowing methods from conv. opt.

$x^\star_{\mathrm{LP}}$ can be computed

1. solve (D) using a subgradient method

$$\begin{aligned} \lambda^{[1]} &= 0 \\ \lambda^{[k+1]} &= \mathrm{P}_+\left(\lambda^{[k]} + s^{[k]} \cdot \gamma^{[k]}\right) \end{aligned} \qquad (1)$$

   - $s^{[k]}$ steplength, pick $s^{[k]} = \alpha/k$.
   - $\gamma^{[k]}$ a subgradient, pick $\gamma^{[k]} = \sum_{i\in I} H_i x_i(\lambda^{[k]}) - b$.

2. while iterating (1), calculate average

$$\bar{x}^{[k]} = \frac{1}{k}\sum_{j=1}^{k} x(\lambda^{[j]})$$

Theorem (primal + dual convergence)

- $\lambda^{[k]} \to \lambda^\star \in \Lambda^\star$,
- *if $x^\star_{\mathrm{LP}}$ is the unique solution to $\mathcal{P}_{\mathrm{LP}}$, $\bar{x}^{[k]} \to x^\star_{\mathrm{LP}}$.*

Computing approximate solutions to (P)

Assume $H_i x_i \geq 0$, and $0 \in X_i$ (relaxed in [Vujanic '13]).

- choices of x_i always "consume" budget

Procedure to compute approx. $\hat{x}^\star$ to P (distributed)

1. compute $x^\star_{\mathrm{LP}}$ using averaging
2. find $I_1 \subseteq I$ such that $(x^\star_{\mathrm{LP}})_i \in \mathrm{vert}(X_i)$
   - $|I_1| \geq |I| - m - 1$
3. for $i \in I_1$, set $\hat{x}^\star_i = (x^\star_{\mathrm{LP}})_i$
4. for the remaining $i \in I \setminus I_1$,
   - EITHER $\hat{x}^\star_i = 0$
   - OR set $b \to b - \sum_{i \in I_1} H_i \hat{x}^\star_i$, and solve smaller dimensional problem

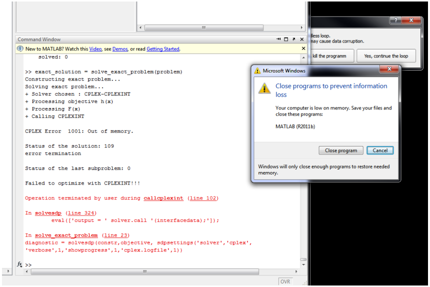
Then, $\hat{x}^\star$ is feasible for P, and satisfies

$$J_{\mathcal{P}}(\hat{x}^\star) - J^\star_{\mathcal{P}} \leq (m+1) \max_{i \in I} \max_{x_i \in X_i} c_i^\top x_i .$$

Outline

Partial Shipment is NP-Hard

- partial shipment problem is NP-Hard (reduction to knapsack)



CPLEX with "just" $N = 300$, $M = 75$

- greedy strategies perform poorly
- tailored MINTO setup also not good (see [Dawande '06])
  - at $N = 300$, $M = 75$
    avg opt. gap $= 6.2\%$ and computation time 6h

Results using Proposed Method

- generate instances of industrial size, data from [Dawande]

		Proposed Method		CPLEX
N	M	Gap (%)	Time (s)	Time (s)
100	25	0.11–0.38–0.60	16.7–18.5–21.5	10.3–63.5–202.5
300	75	0.01–0.03–0.07	61.0–62.5–63.4	–
500	50	0.04–0.06–0.11	95.0–101.0–110.6	–
600	50	0.01–0.06–0.10	113.7–116.6–119.5	–
1000	100	0.00–0.03–0.05	225.3–245.2–288.8	–

- works well because the integer part converges very quickly

Thank you for your attention!

Questions?

References:

- R. Vujanic, P. Mohajerin Esfahani, P. Goulart, S. Mariethoz and M. Morari, *Vanishing Duality Gap in Large Scale Mixed-Integer Optimization: a Solution Method with Power System Applications*, submitted to Mathematical Programming (2013).
- J. P. Aubin and I. Ekeland, *Estimates of the duality gap in nonconvex optimization*, Mathematics of Operations Research **1** (1976), no. 3, 225-–245.
- Dimitri P. Bertsekas, G. Lauer, N. Sandell, and T. Posbergh, *Optimal short-term scheduling of large-scale power systems*, IEEE Transactions on Automatic Control **28** (1983), no. 1, 1– 11.
- Milind Dawande, Srinagesh Gavirneni, and Sridhar Tayur, *Effective heuristics for multiproduct partial shipment models*, Operations Research **54** (2006), no. 2, 337–352 (en).

Bkacup slides

BACKUP SLIDES

Shapley Folkman Theorem

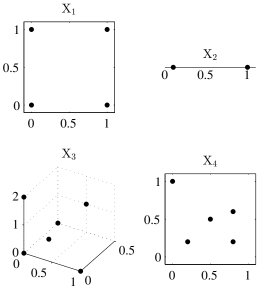
Theorem (, '69)

Let S_i, $i = 1, \ldots, |I|$ be nonempty subsets of $\mathbb{R}^m$, with $|I| > m$, and let $S = S_1 + \cdots + S_{|I|}$. Then every vector $s \in conv(S)$ can be represented as $s = s_1 + \cdots + s_{|I|}$, where $s_i \in conv(S_i)$ for all $i = 1, \ldots, |I|$, and $s_i \notin S_i$ for at most m indices i.

1) $x^{\star}_{\mathrm{LP}}$ is useful

Visualization of the Theorem

$$\begin{cases} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in X_i \qquad\qquad \forall i \in I \end{cases}$$

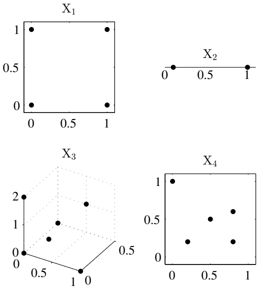


(a) subsystems X_i

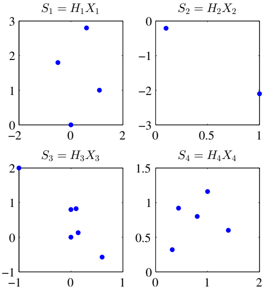
1) $x^\star_{\mathrm{LP}}$ is useful

Visualization of the Theorem

$$\begin{cases} \min_x & \sum_{i\in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in X_i \qquad\qquad \forall i \in I \end{cases}$$



(d) subsystems X_i

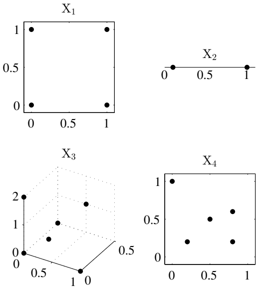


(e) budget consumption $H_i X_i$

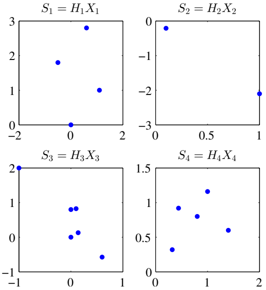
1) $x^\star_{\mathrm{LP}}$ is useful

Visualization of the Theorem

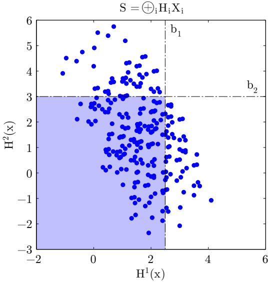
$$\begin{cases} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in X_i \qquad\qquad \forall i \in I \end{cases}$$



(g) subsystems X_i



(h) budget consumption $H_i X_i$

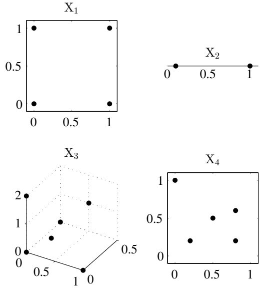


(i) aggregated budget consumption

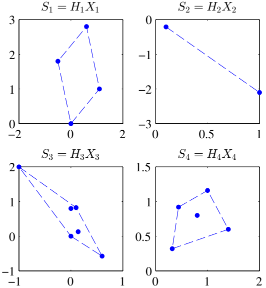
1) $x^\star_{\mathrm{LP}}$ is useful

Visualization of the Theorem

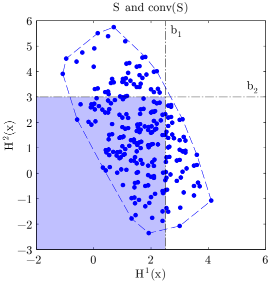
$$\begin{cases} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & x_i \in \color{red}{conv}(X_i) \quad \forall i \in I \end{cases}$$



(j) subsystems X_i



(k) budget consumption $H_i X_i$

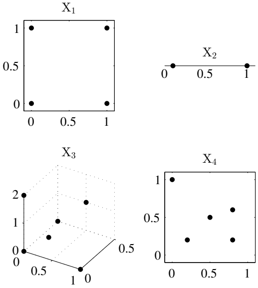


(l) aggregated budget consumption

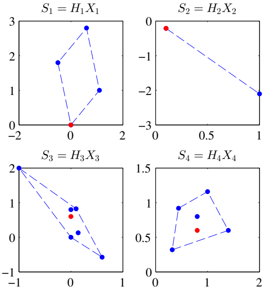
1) $x^{\star}_{\mathrm{LP}}$ is useful

Visualization of the Theorem

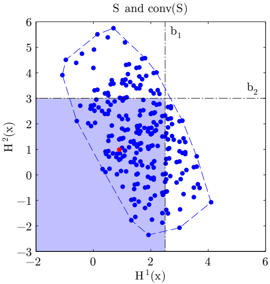
$$
\begin{cases}
\min_x & \sum_{i\in I} c_i x_i \\
\text{s.t.} & \sum_{i\in I} H_i x_i \preceq b \\
 & x_i \in \textcolor{red}{conv(X_i)} \quad \forall i \in I
\end{cases}
$$



(m) subsystems X_i



(n) budget consumption $H_i X_i$



(o) aggregated budget consumption

Result #1: Vanishing Duality Gap

- Duality for programs structured as (3) is **almost** strong:

Theorem (bound on duality gap (Ekeland '76, Bertsekas '83))

Assume that the sets X_i are non-empty and compact, and that for any given $x_i \in \mathrm{conv}(X_i)$, there exists an $\tilde{x}_i \in X_i$ such that $H_i \tilde{x}_i \preceq H_i x_i$. Then

$$J^\star_{\mathcal{P}} - J^\star_{\mathcal{D}} \leq m \cdot \max_{i \in I} \left(\max_{x_i \in X_i} c_i x_i - \min_{x_i \in X_i} c_i x_i \right)$$

- non-convexities do not compound each other indefinitely in P, i.e. the distance between P and P_{LP} does not grow
- Thus, if $J^\star_{\mathcal{P}}$ increases linearly with $|I|$,

$$\frac{J^\star_{\mathcal{P}} - J^\star_{\mathcal{D}}}{J^\star_{\mathcal{P}}} \to 0 \quad \text{as} \quad |I| \to \infty.$$

Result #2: Primal Recovery Scheme

- immunize 3 against disturbances of m subsystems:

$$(\overline{P}):\begin{cases}\min_x & \sum_{i\in I} c_i x_i \\ \text{s.t.} & \sum_{i\in I} H_i x_i \preceq \bar{b} \\ & x_i \in X_i \qquad\qquad \forall i \in I,\end{cases}$$

with $\bar{b} = b - \rho$ and

$$\rho^k = m \cdot \max_{i\in I}\left(\max_{x_i\in X_i} H_i^k x_i - \min_{x_i\in X_i} H_i^k x_i\right)$$

Theorem (feasible solutions)

If $\overline{P}_{\mathrm{LP}}$ and $\overline{D}$ have unique solutions $\bar{x}^\star_{\mathrm{LP}}$ and $\bar{\lambda}^\star$, then any selection $x(\bar{\lambda}^\star) \in \mathcal{X}(\bar{\lambda}^\star)$ is feasible for 3.

- $\bar{\lambda}^\star \succeq \lambda^\star \rightarrow$ it is like imposing a "price uplift"
- conservatism reduction possible in some cases

Performance of the Recovered Solutions

- under some technical assumption, ...

Theorem (solutions performance)

The recovered solution $x(\bar{\lambda}^\star)$ is feasible and satisfies

$$J_{\mathcal{P}}(x(\bar{\lambda}^\star)) - J_{\mathcal{P}}^\star \leq (m + ||\rho||_\infty / \zeta) \cdot \left(\max_{x_i \in X_i} c_i x_i - \min_{x_i \in X_i} c_i x_i \right)$$

- if $J_{\mathcal{P}}^\star$ grows linearly with $|I|$, and X_i uniformly bounded

$$\frac{J(x(\bar{\lambda}^\star)) - J_{\mathcal{P}}^\star}{J_{\mathcal{P}}^\star} \to 0 \quad \text{as} \quad |I| \to \infty,$$

- in fact, optimality gap decreases at a "$1/|I|$" rate

Intuitive explanation of proof that $x^\star_{\mathrm{LP}}$ and $x_i(\lambda)$ are connected

- Problem $\mathcal{P}_{\mathrm{LP}}$ is equivalent to

$$
\begin{array}{lll}
\min_p & \sum_{i\in I}\sum_{j\in J_i} p_{ij}(c_i x_i^j) & \\
\text{s.t.} & \sum_{i\in I}\sum_{j\in J_i} p_{ij}(H_i x_i^j) \leq b & \\
 & \sum_{j\in J_i} p_i^j = 1 & \forall i \in I \\
 & p_i^j \geq 0 & \forall i \in I, j \in J_i,
\end{array}
\qquad (\overline{P})
$$

- Inner problems can return solution x_i^j only if its "probability" is $(p^\star)_i^j > 0$.
- For $|I| - m$ subproblems only one vertex $\hat{\jmath}_i$ has probability > 0 (in fact, $(p^\star)_i^{\hat{\jmath}_i} = 1$), so they must be picked.
- Proof uses strict complementarity, which **requires uniqueness**.

Uniqueness assumption

"Nice" assumption

$x^\star_{\mathrm{LP}}$ and $\lambda^\star$ are the unique optimal solutions to $\mathcal{P}_{\mathrm{LP}}$ and (D)

- sufficient but *not necessary*
- it is **stable** (for given P, it holds with probability 1)
- and **sometimes easy to check** with standard LP tools
  - e.g. problems with integrality property, such as KPs and CLPs
- adding δ_i perturbations to cost function to ensure it **substantially enhances dual convergence**
  - $||\delta_i||_2$ 2-3 orders of magnitude smaller than $||c_i||_2$ is sufficient

Detour: Strength of Relaxations (1/2)

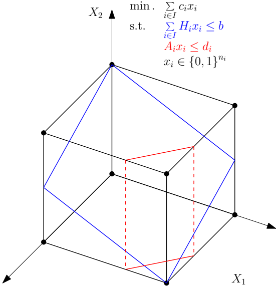
- Assume pure binary P
- If you had to solve it, which convexification would you pick?

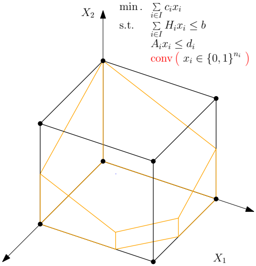
$$P_1: \quad \begin{array}{ll} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & A_i x_i \preceq d_i \\ & \text{conv}(x_i \in \{0,1\}^{n_i}) \end{array} \qquad P_2: \quad \begin{array}{ll} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \sum_{i \in I} H_i x_i \preceq b \\ & \text{conv}\left(\begin{array}{l} A_i x_i \preceq d_i \\ x_i \in \{0,1\}^{n_i} \end{array} \right) \end{array}$$

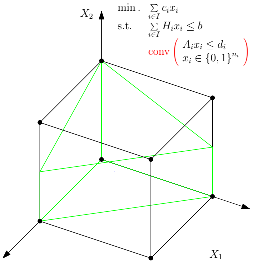
or

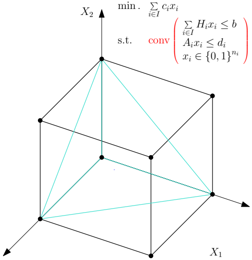
$$P_3: \quad \begin{array}{ll} \min_x & \sum_{i \in I} c_i x_i \\ \text{s.t.} & \text{conv}\left(\begin{array}{l} \sum_{i \in I} H_i x_i \preceq b \\ A_i x_i \preceq d_i \\ x_i \in \{0,1\}^{n_i} \end{array} \right) \end{array}$$

Detour: Strength of Relaxations (2/2)









Performance Guarantees

- Existence and uniqueness of $\bar{x}^\star_{\mathrm{LP}}$ and $\bar{\lambda}^\star$
- There exist $\zeta > 0$ and $\hat{x}_i \in \mathrm{conv}(X_i)$ for all $i \in I$ such that

$$\sum_{i \in I} H_i \hat{x}_i \preceq b - \zeta |I| \mathbb{1}.$$

Then

$$J_{\mathcal{P}}(x(\bar{\lambda}^\star)) - J^\star_{\mathcal{P}} \leq (m + ||\rho||_\infty / \zeta) \cdot \left(\max_{x_i \in X_i} c_i x_i - \min_{x_i \in X_i} c_i x_i \right). \tag{2}$$

This implies that, if $J^\star_{\mathcal{P}}$ grows linearly with $|I|$,

$$\frac{J(x(\bar{\lambda}^\star)) - J^\star_{\mathcal{P}}}{J^\star_{\mathcal{P}}} \to 0 \quad \text{as} \quad |I| \to \infty, \tag{3}$$

i.e. optimality gap decreases at a "$1/|I|$" rate.

Dual convergence

- dual convergence ($d^{(k)} \to d^\star$) is *fast*
- in all instances we tried, well within 30 dual iterations

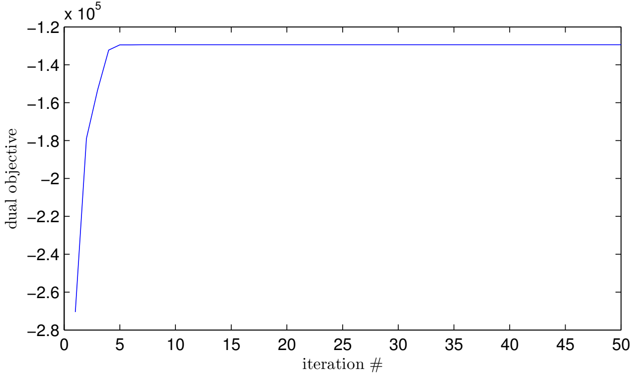


Figure: dual convergence

Ergodic Sequence: Convergence

- ergodic sequence does converge, but rather slowly
- but we don't care: we use it only to detect fractional parts

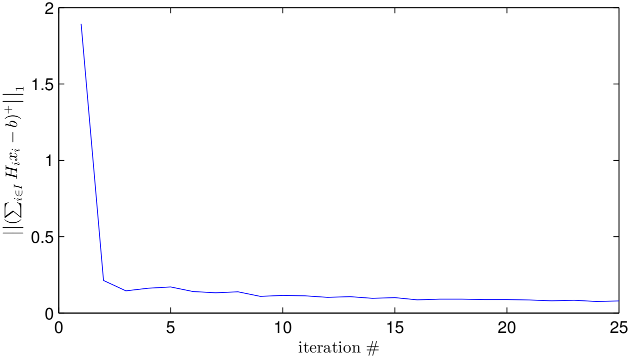


Figure: feasibility violation of ergodic sequence