Decentralized converging algorithm for binary optimization: face-reweighted Lagrangian method

Diego Cifuentes [1] Santanu S. Dey [1] **Jingye Xu** [1]

[1]Georgia Institute of Technology

Summary

Tree-structured binary optimization is a important class of discrete optimization, which covers multi-stage stochast integer optimization with binary state variables. Lagrangian dual for tree-structured binary optimization based on relaxing nonanticipativity constraints is a well-known decentralized approach to provide a high-quality dual bound. Unfortunately, due to the non-convexity, classics Lagrangian methods may have a non-zero dual gap. In this poster, we provide a new Lagrangian method, which is both tight and decomposible.

Introduction

- Consider the following binary optimization:

$$\begin{aligned} v_p := \min\ & C^1(\mathbf{x}^1) + C^2(\mathbf{x}^2) \\ \text{s.t. } & \mathbf{x}^1, \mathbf{x}^2 \in \{0,1\}^n \\ & \mathbf{x}^1 = \mathbf{x}^2 \end{aligned} \tag{1}$$

  All the results presented below can be extended.
- Let $\mathcal{F}$ denote the set of all faces of hypercube $[0,1]^n$. Given a subsets of faces $\widehat{\mathcal{F}} \subseteq \mathcal{F}$, consider the following extended formulation of (1)

$$\begin{aligned} v_p^{ex} := \min\ & C^1(\mathbf{x}^1) + C^2(\mathbf{x}^2) \\ \text{s.t. } & \mathbf{x}^i \in \{0,1\}^n, w_F^i = \begin{cases} 1 & \text{if } \mathbf{x}^i \in F \\ 0 & \text{otherwise} \end{cases}, \forall i \in \{1,2\} \\ & \mathbf{w}_F^1 = \mathbf{w}_F^2, \forall F \in \widehat{\mathcal{F}} \end{aligned} \tag{2}$$

  and its Lagrangian dual is

$$\begin{aligned} L^{ex}(\lambda) := \min\ & C^1(\mathbf{x}^1) + C^2(\mathbf{x}^2) + \sum_{F \in \widehat{\mathcal{F}}} \lambda_F(\mathbf{w}_F^1 - \mathbf{w}_F^2) \\ \text{s.t. } & \mathbf{x}^i \in \{0,1\}^n, w_F^i = \begin{cases} 1 & \text{if } \mathbf{x}^i \in F \\ 0 & \text{otherwise} \end{cases}, \forall i \in \{1,2\}, \\ v_d^{ex} :=\ & \max_{\lambda} L^{ex}(\lambda) \end{aligned} \tag{3}$$

- Classical Lagrangian methods can be viewed as a special case of (2) where $\widehat{\mathcal{F}}$ is the collection of all facets of $[0,1]^n$.
- (3) can be solved by subgradient or bundle methods and computing subgradient of (3) is a decomposable binary optimization.

Theoretical results

- The classics pseudo-boolean analysis states that every boolean function $C^i(\mathbf{x}) : \{0,1\}^n \to \mathbb{R}$ admits a multilinear representation.

$$C^i(\mathbf{x}) = \sum_{S \subseteq [n]} f_S^i X_S \text{ where } X_S := \Pi_{i \in S} x_i$$

- **Theorem 1 (strong duality)** If $\widehat{\mathcal{F}}$ includes all vertices of $[0,1]^n$, then $v_d^{ex} = v_p$. If each C^i is a polynomial of degree at most k and $\widehat{\mathcal{F}}$ includes all faces of dimension $n-k$ and higher, then $v_d^{ex} = v_p$.
- **Theorem 2 (convergence rate)** If each C^i is a polynomial of degree at most k and $\widehat{\mathcal{F}}$ includes all faces of dimension $n-k$ and higher, there exists a subgradient method to solve (3) to achieve ϵ additive optimally in at most $O\left(\frac{n^k \sum (f_S^i)^2}{\epsilon^2}\right)$ iterations even if f_S^i is unknown in advance.

Proof sketch:

1. One can construct optimal dual variable λ^* such that $v_d^{ex} = v_p$. For example, when $\widehat{\mathcal{F}}$ is collection of all vertices, one can construct optimal dual variable λ^* where $\lambda^*_{\mathbf{v}} = C^2(\mathbf{v}), \forall \mathbf{v} \in \{0,1\}^n$. This proves **Theorem 1**.
2. **Theorem 2** directly comes from the classic convergence analysis of subgradient method and constructive proof of **Theorem 2**.

Some implementation details

- We consider $\widehat{\mathcal{F}}$ is collection of all vertices and facets of $[0,1]^n$. The Lagrangian takes the following form:

$$\begin{aligned} L^{ex}(\lambda, \mu) := \min\ & C^1(\mathbf{x}^1) + C^2(\mathbf{x}^2) + \sum_{j \in [n]} \mu_i(\mathbf{x}_i^1 - \mathbf{x}_i^2) + \sum_{v \in \{0,1\}^n} \lambda_v(\mathbf{w}_v^1 - \mathbf{w}_v^2) \\ \text{s.t. } & \mathbf{x}^i \in \{0,1\}^n, w_v^i = \begin{cases} 1 & \text{if } \mathbf{x}^i \in v \\ 0 & \text{otherwise} \end{cases}, \forall i \in \{1,2\}, v \in \{0,1\}^n \end{aligned} \tag{4}$$

- For any given λ, computing the subgradient is decomposed to many binary optimization taking form of $\min\ C^i(\mathbf{x}^i) \pm \mu^\top \mathbf{x}^i$ s.t. $\mathbf{x}^i \in \{0,1\}^n$ with additional $\|\lambda\|_0$ many no good cuts.
- Even if the dimension of λ is exponential, if we start with $\lambda = 0$, $\|\lambda\|_0$ increases by at most one for each iteration of subgradient methods or bundle methods.

Numerical experimental

We apply (4) to two-stage stochastic integer optimization. We consider Stochastic multi-knapsack problem in [1]. The problem (1) takes form of

$$\begin{aligned} \min\ & C^0(\mathbf{x}) + \frac{1}{s} \sum_{i \in [s]} C^i(\mathbf{x}) \\ \text{s.t. } & \mathbf{x} \in \{0,1\}^n \end{aligned}$$

where

$$\begin{aligned} C^i(\mathbf{x}) := \min\ & \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{z}^i \\ \text{s.t. } & A^i \mathbf{x} + B^i \mathbf{z}^i \geq \mathbf{g}^i \\ & \mathbf{x} \in \{0,1\}^n, \mathbf{z}^i \in \{0,1\}^n \end{aligned}$$

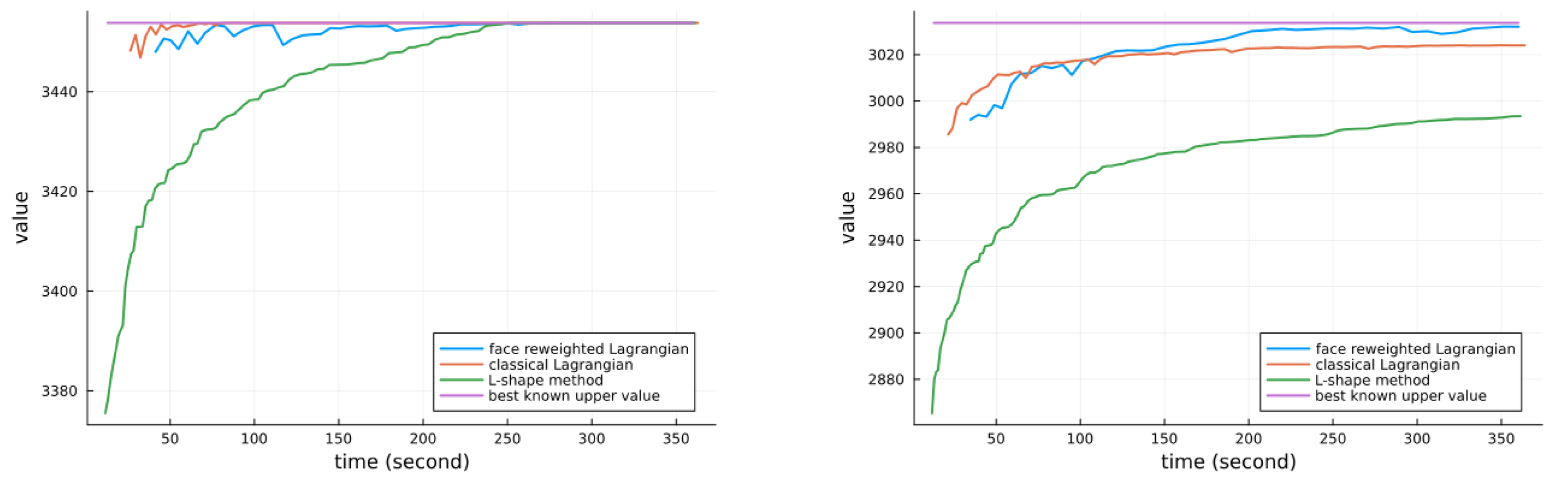

Figure 1. Gap vs time of different methods. $n = 50, s = 20, m = 10$

the number of instances solved in time and final gap				
Method	≤ 120	≤ 240	≤ 360	avg final gap (unsolved)
CL Lagrangian	25/60	36/60	40/60	0.16%
FW Lagrangian	20/60	37/60	53/60	0.12%
L-shape	4/60	7/60	15/60	1.0 %

References

[1] Gustavo Angulo, Shabbir Ahmed, and Santanu S Dey. Improving the integer l-shaped method. *INFORMS Journal on Computing*, 28(3):483–499, 2016.