

Smooth and flexible dual optimal inequalities

N. Haghani, C. Contardo,
J. Yarkony

G-2019-94

December 2019

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

Citation suggérée : N. Haghani, C. Contardo, J. Yarkony (Décembre 2019). Smooth and flexible dual optimal inequalities, Rapport technique, Les Cahiers du GERAD G-2019-94, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2019-94>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

Suggested citation: N. Haghani, C. Contardo, J. Yarkony (December 2019). Smooth and flexible dual optimal inequalities, Technical report, Les Cahiers du GERAD G-2019-94, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (<https://www.gerad.ca/en/papers/G-2019-94>) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2019
– Bibliothèque et Archives Canada, 2019

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2019
– Library and Archives Canada, 2019

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Télec. : 514 340-5665
info@gerad.ca
www.gerad.ca

Smooth and flexible dual optimal inequalities

Naveed Haghani ^{a,b}

Claudio Contardo ^{c,d}

Julian Yarkony ^b

^a University of Maryland, College Park, Maryland
20742, USA

^b Verisk Computational and Human Intelligence
Laboratory, Jersey City, New Jersey 07310

^c ESG UQAM, Montréal (Québec), Canada, H2X
3X2

^d GERAD, Montréal (Québec), Canada, H3T 2A7

claudio.contardo@gerad.ca

December 2019
Les Cahiers du GERAD
G–2019–94

Copyright © 2019 GERAD, Haghani, Contardo, Yarkony

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs. Les auteurs conservent leur droit d'auteur et leurs droits moraux sur leurs publications et les utilisateurs s'engagent à reconnaître et respecter les exigences légales associées à ces droits. Ainsi, les utilisateurs:

- Peuvent télécharger et imprimer une copie de toute publication du portail public aux fins d'étude ou de recherche privée;
- Ne peuvent pas distribuer le matériel ou l'utiliser pour une activité à but lucratif ou pour un gain commercial;
- Peuvent distribuer gratuitement l'URL identifiant la publication.

Si vous pensez que ce document enfreint le droit d'auteur, contactez-nous en fournissant des détails. Nous supprimerons immédiatement l'accès au travail et enquêterons sur votre demande.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*. Copyright and moral rights for the publications are retained by the authors and the users must commit themselves to recognize and abide the legal requirements associated with these rights. Thus, users:

- May download and print one copy of any publication from the public portal for the purpose of private study or research;
- May not further distribute the material or use it for any profit-making activity or commercial gain;
- May freely distribute the URL identifying the publication.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Abstract: We address the problem of accelerating column generation (CG) for set-covering formulations via dual optimal inequalities (DOI). DOI use knowledge of the dual solution space to derive inequalities that might be violated by intermediate solutions to a restricted master problem, and as such are efficient at reducing the number of iterations and the oscillations of the dual variables commonly observed in column generation procedures. We study two novel classes of DOI which are referred to as Flexible DOI (F-DOI) and Smooth-DOI (S-DOI), respectively (and jointly as SF-DOI). F-DOI provide rebates for covering items more than necessary. S-DOI describe the payment of a penalty to permit the under-coverage of items in exchange for the over-inclusion of other items. Unlike other classes of DOI from the literature, the S-DOI and F-DOI rely on very little to no problem-specific knowledge, and as such have the potential to be applied to a vast number of problem domains. In particular, we illustrate the efficiency of the new inequalities by embedding them within a column generation solver for the single source capacitated facility location problem (SSCFLP). A speed-up of a factor of up to $130\times$ can be observed as when compared to a non-stabilized variant of the same CG procedure to achieve the linear relaxation lower bound on problems with dense columns and structured assignments costs.

1 Introduction

Column generation (CG) (Gilmore and Gomory 1961, Lübbecke and Desrosiers 2005, Yarkony et al. 2019) is a core technique to solve linear programs (LP) with a large number of variables. CG mimics the simplex method by adding a pricing step (just like in the simplex method) to work with a reduced basis and enlarge it only as needed. CG has proven particularly efficient when the underlying mathematical formulation is the result of applying Dantzig-Wolfe decomposition to a compact formulation, in such a way that some complicating constraints are moved to a subproblem with some rich structure. Indeed, it is known that when the pricing subproblem is an integer program without the integrality property (Geoffrion 1974), the resulting dual bounds may be strictly better than those associated with the original compact formulation of the problem. That is the case of some remarkable applications in logistics and transportation like vehicle routing (Costa et al. 2019) and facility location (Diaz and Fernández 2002).

CG has been applied historically with great success in a vast number of applications: stock (material) cutting (Gilmore and Gomory 1961), vehicle routing (Costa et al. 2019), crew scheduling (Barnhart et al. 1996), and facility location (Barahona and Jensen 1998), among others. Recently, CG has been applied to computer vision problems including image segmentation (Yarkony and Fowlkes 2015), multi-object tracking (Wang et al. 2017, Leal-Taixe et al. 2012), and multi-person pose estimation (Wang et al. 2018). When embedded within a branch-and-bound framework, CG becomes branch-and-price, an exact method that guarantees the finding of integer solutions.

Dual optimal inequalities (DOI, Yarkony et al. 2019, Ben Amor et al. 2006, Gschwind and Irnich 2016) are a core tool to accelerate CG. DOI reduce the dual feasible solution space and thus correspond to relaxations of the constraints in the primal. The dual space is reduced such that it contains at least one dual optimal solution of the non-reduced space. DOI are constructed to be provably inactive at termination of CG. In the primal problem, DOI permit the violation of constraints often in exchange for a penalty. DOI are associated with a repair procedure that maps any solution violating the original primal constraints to one that does not. The solution produced by the repair procedure has no greater cost than that of the original solution.

Typically, DOI exploit the structure of the underlying problems to derive conditions (cuts) satisfied by the dual variables in an optimal solution. For instance, in bin packing and cutting stock items of equal weight/size are indistinguishable. Thus, DOI exploit this knowledge to derive inequalities that impose this equivalence between indistinguishable items in the dual space (Ben Amor et al. 2006, Gschwind and Irnich 2016). It has not been until recently that Lokhande et al. (2020) introduced a new class of DOI for a general class of weighted set-packing problems arising in computer vision called Flexible dual-optimal inequalities (F-DOI). In the primal problem F-DOI permit the over-covering of items in exchange for a penalty. In the dual problem the F-DOI bind the dual variables to lie in a hypercube. The penalty for over-covering an item v with a column l is an upper bound on the increase in cost induced by removing v from any column consisting of a subset of the items of l including v .

2 Our contributions

Our contributions expand upon previous work on DOI (particularly F-DOI). Most notably, this article contributes to narrowing the gap of applicability of general-purpose DOI by proposing two novel classes of DOI for general set-covering problems. Our contributions can be summarized as follows:

1. We introduce a novel class of dual optimal inequality, namely the Smooth-DOI (S-DOI). In the primal problem, S-DOI permit the under-covering of items in exchange for the over-covering of other items plus a penalty. Each individual S-DOI is associated with a pair of items (u, v) . The penalty associated with a given S-DOI for items (u, v) is an upper bound on the increase in cost induced by replacing u in a column not including v with v . In the dual problem the S-DOI produce a smoothing in the oscillation behavior of the dual vector, thus the motivation for calling them *Smooth DOI*.

2. We adapt the F-DOI of Lokhande et al. (2020) to weighted set-covering formulations. In our case the F-DOI induce a reward for over-including an element v using column l . This reward is a lower bound on the improvement of the objective induced by removing v from any column constructed from a subset of the items of l including v .
3. We provide a unified framework for the simultaneous addition of S-DOI and F-DOI. The resulting inequalities are referred to as *Smooth-Flexible DOI* and denoted SF-DOI.
4. We propose an efficient implementation for the SF-DOI (and therefore for the S-DOI and F-DOI separately as well), which allow us to control the growth of the resulting linear programs, and to perform pricing in a minimally invasive manner. In fact, we show that the same pricing problem used for a non-stabilized CG can also be applied within the new framework.
5. We illustrate the efficiency of our stabilization schemes with an application to the single-source capacitated facility location problem (SSCFLP), and show using an exhaustive computational campaign that the new classes of DOI entail substantial improvements in terms of number of iterations and of CPU time required to achieve the linear relaxation lower bounds, as compared to a non-stabilized implementation of the same algorithm.

We outline this document as follows. In Section 3 we review the most relevant literature in dual stabilization applied to decomposition methods, with an emphasis in CG and DOI. In Section 4 we present a standard CG method using a set-covering formulation. In Sections 5–7 we present the S-DOI, F-DOI and the SF-DOI. In Section 8 we consider the problem of pricing in the presence of the new DOI, and show that it can be performed as usual by ignoring the impact of the new DOI in the pricing without compromising the correctness of the method. In Section 9 we present a case study by applying the SF-DOI to SSCFLP. In Section 10 provide empirical results on SSCFLP. We conclude in Section 11.

3 Literature review

In this section we review the related work on dual stabilization so as to provide context for our contributions. Dual stabilization approaches can be understood as imposing a Bayesian prior on the values of the dual variables and can be classified as trust region based methods, smoothing methods, interior-point methods and dual-optimal inequality based methods. Trust region based methods (Marsten et al. 1975, Du Merle et al. 1999) prevent or discourage the dual solution from departing a window around the current optimal dual solution. They often require an explicit change in the underlying optimization models and rely on periodic updates of the stabilization parameters, without which the methods are not guaranteed to converge to a proven dual optimum. Interior point methods (Mitchell 1994, Rousseau et al. 2007) average across multiple dual optimal solutions to produce well centered solutions. Smoothing methods (Pessoa et al. 2018), on the other hand, do not impose any explicit extra conditions in the models, but rely on the computation of a smoothing dual center to decrease the ℓ_2 -norm of the dual solution used for pricing purposes. As in trust-region methods, a periodic update of the stabilization parameters is often required to ensure convergence. DOI based methods (Ben Amor et al. 2006, Lokhande et al. 2020, Gschwind and Irnich 2017), in contrast, restrict the dual solution from leaving a subspace that provably contains at least one dual optimal solution. In Sections 3.1–3.3 we provide a brief overview of the first three stabilization techniques. In Section 3.4 we provide a detailed review of dual-optimal inequalities for CG and related literature. Finally, we present in Section 3.5 the application of dual stabilization to a method beyond CG, specifically the Benders decomposition (Benders 1962).

3.1 Trust region based methods

One key example of trust region based methods is Marsten et al. (1975), which restricts the dual variables to lie in a box around the previous dual solution at each step of CG. This prevents the dual solution from rapidly traveling outside of the space where the current set of columns provide meaningful information concerning bounds on the dual variables. Marsten et al. (1975) is further developed in

Du Merle et al. (1999) where an ℓ_1 penalty on the distance from a box around the tightest dual solution found thus far is imposed. Here the tightness of the bound is computed by adding negative reduced cost terms to the restricted master problem (RMP) objective so as to provide a lower bound on the optimal solution to the master problem (MP). A key problem with these approaches is the requirement of user defined hyper-parameters for optimization. Related to trust-region methods are Bundle methods (Frangioni 2002, Briant et al. 2008, van Ackooij and Frangioni 2018) that allow dual variables to leave the trust region by paying a penalty. Bundle methods make use of non-linear penalty functions (often quadratic) for the penalties, and have desirable convergence properties at the expense of larger computing times per iteration associated with the non-linear penalties.

3.2 Interior point methods

Interior point methods are based on the assumption that, when multiple dual optima exist for a given problem, the average of those dual vectors tends to be more balanced and thus have lower ℓ_2 norm. Rousseau et al. (2007) exploit this observation by solving each iteration of the restricted master problem (RMP) multiple times, to hopefully find alternative dual optima. The average of those optima is used to execute the pricing. Gondzio et al. (2013) introduce the primal-dual column generation method, which uses an interior point method to avoid solving the RMP to proven optimality at every iteration. Instead, the primal-dual method computes a well-centered primal-dual feasible set of vectors. The associated dual vector is then used as input to the pricing subproblem. Later, Gondzio and González-Brevis (2015) proposed a warm-start strategy to accelerate the re-execution of the primal-dual method. The authors provide further evidence of the usefulness of their method later in Gondzio et al. (2016).

3.3 Smoothing methods

Smoothing in CG can be seen as the use of a dual solution of low ℓ_2 norm as input to the pricing sub-problems. Given a dual variable vector α of a linear program, the center α^0 is a vector of the same dimensions as α but with a lower ℓ_2 norm. Instead of using α directly to search for columns of negative reduced cost, one uses instead $\alpha' = \lambda\alpha + (1 - \lambda)\alpha^0$, where λ is a parameter used to control the smoothing of the vector α' . Pessoa et al. (2018) presented the use of this technique and provide conditions guaranteeing a correct —yet automatic— update of the stabilization parameters to ensure convergence. Unlike trust-region methods, no changes in the structure of the optimization models is necessary along the execution of the algorithm, and therefore is much easier to implement than trust region methods, yet providing competitive and usually better convergence properties.

3.4 Dual optimal inequalities

The work of Ben Amor et al. (2006) applies dual optimal inequalities (DOI) for CG. The authors provide conditions satisfied by the dual variables when they are associated to indistinguishable items in the primal, this is items that if swapped, result in a solution totally equivalent. In cutting stock and bin packing, this equivalence is measured according to the items' sizes or weights. Two items of identical weight/size are indistinguishable and the dual variables associated with their coverage restrictions can be assumed to be identical. This entails a series of identity equations in the dual space. As a result, dual oscillation between indistinguishable items is totally eliminated and the resulting CG algorithms converge substantially faster in the presence of a large amount of such symmetries. The authors also show that in cutting stock/bin packing that for any pair of items u, v , where the size of u is larger than that of v , that the dual variable corresponding to u may be no smaller than that corresponding to v . The work of (Gschwind and Irnich 2016) extends this knowledge to problems with richer structure and less symmetries, such as bin packing with conflicts and vector packing.

Recently, DOI have made a notable contribution in computer vision applications of CG (Yarkony et al. 2019). In the applications considered by the authors, the fact the cost of a column is a sparse quadratic function of its members is exploited as its cost changes smoothly as observations are removed from it. This observation motivates the following class of DOI, namely the *invariant DOIs*, which is

briefly explained as follows. The dual value associated with a given item can be bounded by an easily computed upper bound on the cost increase observed when removing that item from the column. The authors expand on this with *varying DOI* that make use of the current RMP solution to provide bounds that are tighter given the columns in the RMP. Invariant/Varying DOIs can be understood as relaxing the packing constraint and replacing it with a penalty for violating it.

In the same line as the work of Yarkony et al. (2019), the work of Lokhande et al. (2020) introduced the *flexible DOI* (F-DOI) for set packing, which further improves the varying DOIs and is the basis for our work. Varying DOI suffer from the fact that the penalty for using the associated primal variable depends only on the item being removed, and as such do not exploit any specific knowledge of the active columns in the RMP solution. F-DOI circumvent this issue by considering that the penalty may depend not only on the item, but also on the column, to compute bounds on the cost change when removing an item from that specific column.

DOI can also be found in (hierarchical) correlation clustering problems arising in image segmentation (Yarkony and Fowlkes 2015, Yarkony et al. 2012). The authors report that in the absence of dual stabilization, CG suffers from severe convergence issues. In the presence of DOI, on the other hand, convergence is achieved several orders of magnitude faster. In Yarkony et al. (2012) DOIs are able to bound most of the dual variables to be zero for real problem instances. This is not the case in Yarkony and Fowlkes (2015), which is closely related to Yarkony et al. (2012), implying that it is not only the restriction of the dimensionality of the dual space which provides for the effectiveness of DOIs in Yarkony et al. (2012).

3.5 Dual stabilization in Benders decomposition

Dual stabilization is not only a relevant issue in CG. It also occurs in the domain of Benders decomposition (Benders 1962) as a part of Magnanti-Wong cuts (Magnanti and Wong 1981). Magnanti-Wong cuts are Benders cuts that are optimal (or near optimal) with respect to the original objective and yet attempt to optimize a separate auxiliary objective. These cuts have been adapted to computer vision problems for multi-person pose estimation (Wang et al. 2018) and image segmentation (Keuper et al. 2019). In logistics applications, dual stabilization of Benders decomposition has proven effective in hub-location (Contreras et al. 2011) and network design (Zetina et al. 2019). In these settings the auxiliary objective may be an ℓ_1 norm as in Wang et al. (2018), or a random objective (Keuper et al. 2019). Remarkably these Magnanti-Wong cuts are often necessary for convergence of Benders decomposition as reported in Keuper et al. (2019).

4 Standard column generation

In this section we present a standard CG method over a set-covering formulation. We are given a set N of items to be covered. The formulation includes a continuous variable $\theta_l \geq 0$ for every column $l \in \Omega$ where Ω is the set of all feasible columns. It is assumed that every column can cover an item at most once. In the context of CG, the set Ω typically is exponentially large in the size of the input, and therefore it is impractical to explicitly consider the entire set during optimization. For every column $l \in \Omega$ and for every item $u \in N$ we let $a_{ul} \in \{0, 1\}$ be a binary constant equal to 1 if l covers u , and otherwise let $a_{ul} = 0$. We consider problem:

$$\min_{\theta} \sum_{l \in \Omega} c_l \theta_l \tag{1}$$

subject to

$$\sum_{l \in \Omega} a_{ul} \theta_l \geq 1 \quad u \in N \tag{2}$$

$$\theta_l \geq 0 \quad l \in \Omega. \tag{3}$$

Given the unscalability of enumerating the set Ω explicitly, CG consists in considering a subset $\Omega_R \subset \Omega$ at any given time, and thus the RMP consists in the solution of problem (1)-(3) but restricted to the columns in Ω_R . Let us denote that problem $\text{RMP}(\Omega_R)$. Now let us denote $(\alpha_u)_{u \in N}$ as the dual variables associated with the constraints (2) in problem $\text{RMP}(\Omega_R)$. The reduced cost of a column $l \in \Omega \setminus \Omega_R$ can be computed as $\bar{c}_l = c_l - \sum_{u \in N} \alpha_u a_{ul}$. Problem $\text{RMP}(\Omega_R)$ provides a proven optimal solution of (1)-(3) if $\min\{\bar{c}_l : l \in \Omega \setminus \Omega_R\} \geq 0$.

5 Smooth DOI

Let $s = (u, v) \in N \times N, u \neq v$ be a pair of items satisfying the following property: for every column $l \in \Omega$ such that $a_{ul} = 1, a_{vl} = 0$ the column l' constructed from swapping u for v from l also lies in Ω (the converse may not be true). Let us denote $l' = \text{swap}(l, s)$ as the outcome of the above swap operation, and let us denote by $\Omega(s)$ the set of columns $l \in \Omega$ such that $a_{ul} = 1, a_{vl} = 0$. Let us define, for a given s satisfying this property, a penalty $\rho_s \in \mathbb{R}$ such that $\rho_s \geq \max\{c_{l'} - c_l : l \in \Omega(s), l' = \text{swap}(l, s)\}$. Let us consider a set \mathcal{S} such that for every $s_1 = (u, v), s_2 = (v, w) \in \mathcal{S}$, that $s_3 = (u, w)$ is also in \mathcal{S} and $\rho_{s_1} + \rho_{s_2} \geq \rho_{s_3}$ holds (triangle inequality). For a pair $s = (u, v)$, we denote $s^+ = v$ and $s^- = u$. Moreover, we denote $\mathcal{S}_u^- = \{s \in \mathcal{S} : s^- = u\}$, $\mathcal{S}_v^+ = \{s \in \mathcal{S} : s^+ = v\}$.

Let us consider the following stabilized form of problem (1)-(3). In addition to the notation introduced before, we add one non-negative variable ω_s for every $s \in \mathcal{S}$. We consider the problem as follows:

$$\min_{\theta, \omega} \quad \sum_{l \in \Omega} c_l \theta_l + \sum_{s \in \mathcal{S}} \rho_s \omega_s \quad (4)$$

subject to

$$\sum_{l \in \Omega} a_{ul} \theta_l + \sum_{s \in \mathcal{S}_u^+} \omega_s - \sum_{s \in \mathcal{S}_u^-} \omega_s \geq 17 \quad u \in N \quad (5)$$

$$\theta_l \geq 07 \quad l \in \Omega. \quad (6)$$

$$\omega_s \geq 07 \quad s \in \mathcal{S}. \quad (7)$$

Proposition 1 *Problem (4)-(7) admits an optimal solution (θ^*, ω^*) such that $\omega_s^* = 0$ for every $s \in \mathcal{S}$.*

Proof. Let (θ^*, ω^*) be an optimal solution to problem (4)-(7). If multiple optima exist, let (θ^*, ω^*) be the one attaining the lowest possible value of $\Sigma(\theta^*, \omega^*) = \sum\{\theta_l^* : l \in \Omega\} + \sum\{\omega_s^* : s \in \mathcal{S}\}$. We will prove, by contradiction, that no variable ω_s^* can take a strictly positive value. Let us assume that $\omega_s^* > 0$ for a certain $s \in \mathcal{S}$ and let us consider the following three scenarios:

- **There exists $t \in \mathcal{S}$ such that $\omega_t^* > 0, s^- = t^+, s^+ \neq t^-$.**

Let $r = (t^-, s^+)$. Because of the triangle inequality, r lies in \mathcal{S} and $\rho_r \leq \rho_s + \rho_t$. Let $\Delta = \min\{\omega_s^*, \omega_t^*\}$. We let $\omega'_s \leftarrow \omega_s^* - \Delta, \omega'_t \leftarrow \omega_t^* - \Delta, \omega'_r \leftarrow \omega_r^* + \Delta$ and $\omega'_k \leftarrow \omega_k^*$ for all $k \in \mathcal{S} \setminus \{s, t, r\}$. It is easy to see that (θ^*, ω') is primal feasible. The marginal contribution of replacing ω^* by ω' is $\Delta(\rho_r - \rho_s - \rho_t)$, which is non-positive. If negative (meaning that $\Delta(\rho_r - \rho_s - \rho_t) < 0$) we obtain a contradiction with the optimality of (θ^*, ω^*) . If zero, on the other hand, the operation provides an alternate optimal solution (θ^*, ω') such that $\Sigma(\theta^*, \omega') < \Sigma(\theta^*, \omega^*)$, which is also a contradiction.

- **There exists $t \in \mathcal{S}$ such that $\omega_t^* > 0, s^- = t^+, s^+ = t^-$.**

In this case we let $\Delta = \min\{\omega_s^*, \omega_t^*\}$ and let $\omega'_s \leftarrow \omega_s^* - \Delta, \omega'_t \leftarrow \omega_t^* - \Delta$, and $\omega'_k \leftarrow \omega_k^*$ for all $k \in \mathcal{S} \setminus \{s, t\}$. It is easy to see that the solution (θ^*, ω') is also primal feasible and the marginal contribution of this operation is $-\Delta(\rho_s + \rho_t)$. The term $(\rho_s + \rho_t)$ is non-negative because of the following observation. From the definition of $\Omega(\cdot)$ we have that $l \in \Omega(s), l' = \text{swap}(l, s) \Leftrightarrow l' \in \Omega(t), l = \text{swap}(l', t)$. Then, for any such pair (l, l') we have —from the conditions satisfied by ρ — that $\rho_s + \rho_t \geq [c_{l'} - c_l] + [c_l - c_{l'}] \geq 0$. If $\rho_s + \rho_t > 0$, this operation results in a contradiction

with the optimality of (θ^*, ω^*) . If zero, on the other hand, the solution obtained is an alternate optimum and (θ^*, ω') is such that $\Sigma(\theta^*, \omega') < \Sigma(\theta^*, \omega^*)$, which is also a contradiction.

- **For every $t \in \mathcal{S}$ such that $s^- = t^+$, $\omega_t = 0$.**

Since the primal problem is feasible and no $t \in \mathcal{S}$ exists satisfying $\omega_t^* > 0, s^- = t^+$, there must exist a column $l \in \Omega(s)$ such that $\theta_l^* > 0$. Let $\Delta = \min\{\omega_s^*, \theta_l^*\}$ and let $l' = \text{swap}(l, s)$. We construct new variables θ', ω' with all components equal to those of (θ^*, ω^*) except for $\omega'_s \leftarrow \omega_s^* - \Delta$, $\theta'_l \leftarrow \theta_l^* - \Delta$, $\theta'_{l'} \leftarrow \theta_{l'}^* + \Delta$. This operation entails an increase in the objective of $\Delta(c_{l'} - c_l - \rho_s)$ which by definition of ρ_s is non-positive. If negative, this would contradict the optimality of (θ^*, ω^*) . If zero, on the other hand, it would entail an alternate optimum such that $\Sigma(\theta', \omega') < \Sigma(\theta^*, \omega^*)$ which is also a contradiction. □

6 Flexible DOI

In this section we present the Flexible DOI (F-DOI), which use the potential benefit attained when removing items from a column to bound the solution space of the dual variables $(\alpha_u)_{u \in N}$. In Section 6.1 we introduce F-DOI in the general case. In Section 6.2 we present an efficient implementation of the F-DOI to address some scalability issues.

6.1 The general case

For a column $l \in \Omega$ and an item $u \in N$ we let $\sigma_{ul} \geq 0$ such that $a_{ul} = 0 \Rightarrow \sigma_{ul} = 0$. The set of values $(\sigma_{ul})_{u \in N, l \in \Omega}$ is assumed to satisfy the following property. For every column $l \in \Omega_R$ and for every subset $S \subseteq N(l) = \{u \in N : a_{ul} = 1\}$, $\sum_{u \in S} \sigma_{ul} \leq c_l - c_{l'}$, where $l' \in \Omega$ is the column constructed from removing the items in S from l . We denote $l' = \text{remove}(l, S)$. For any given $u \in N$, we let $\Lambda_u = \{\sigma_{ul} : l \in \Omega_R\}$ be the set of all possible values of σ_{ul} across the restricted set Ω_R .

In addition to the notation introduced in Section 4, let us introduce non-negative variables $\xi_{u\sigma}$ for every $u \in N$ and for every $\sigma \in \Lambda_u$. A variable $\xi_{u\sigma}$ represents the number of columns l from which u will be removed with a rebate of σ . We also let $\beta_{ul\sigma}$ be a binary constant equal to 1 if $\sigma_{ul} = \sigma$. The following is a stabilized version of (1)–(3):

$$\min_{\theta, \xi} \quad \sum_{l \in \Omega} c_l \theta_l - \sum_{u \in N, \sigma \in \Lambda_u} \sigma \xi_{u\sigma} \quad (8)$$

subject to

$$\sum_{l \in \Omega} a_{ul} \theta_l - \sum_{\sigma \in \Lambda_u} \xi_{u\sigma} \geq 1 \quad u \in N \quad (9)$$

$$\xi_{u\sigma} - \sum_{l \in \Omega} \beta_{ul\sigma} \theta_l \leq 0 \quad u \in N, \sigma \in \Lambda_u \quad (10)$$

$$\theta_l \geq 0 \quad l \in \Omega \quad (11)$$

$$\xi_{u\sigma} \geq 0 \quad u \in N, \sigma \in \Lambda_u. \quad (12)$$

Intuitively, the F-DOI allow the over-coverage of some items for a reward in the objective. If this reward is sufficiently small, it will not be sufficient to compensate the over-coverage and thus will be non-basic in an optimal solution of problem (8)–(12). The following proposition formalizes this result.

Proposition 2 *Problem (8)–(12) admits an optimal solution (θ^*, ξ^*) such that $\xi_{u\sigma}^* = 0$ for every $u \in N, \sigma \in \Lambda_u$.*

Proof. Let (θ^*, ξ^*) be an optimal solution to problem (8)–(12). If multiple such optima exist, let (θ^*, ξ^*) be the one that minimizes $\Sigma(\theta, \xi) = \sum\{\theta_l : l \in \Omega\} + \sum\{\xi_{u\sigma} : u \in N, \sigma \in \Lambda_u\}$. We will prove

by contradiction that if $\xi_{u\sigma}^* > 0$ for some $u \in N, \sigma \in \Lambda_u$ then either (θ^*, ξ^*) cannot be optimum, or that an alternate optimum (θ', ξ') exists such that $\Sigma(\theta', \xi') < \Sigma(\theta^*, \xi^*)$.

Let $\xi_{u\sigma}^* > 0$ for some $u \in N, \sigma \in \Lambda_u$. Constraints (10) assure the existence of a column $l \in \Omega$ such that $\beta_{ul\sigma} = 1$ and $\theta_l^* > 0$. Let $\Delta = \min\{\xi_{u\sigma}^*, \theta_l^*\}$. Let $l' = \text{remove}(l, \{u\})$ be the column resulting from removing u from l . Now, let us consider a solution (θ', ξ') with all entries equal to those of (θ^*, ξ^*) except for the entries $\xi'_{u\sigma} \leftarrow \xi_{u\sigma}^* - \Delta$, $\theta'_l \leftarrow \theta_l^* - \Delta$, $\theta'_{l'} \leftarrow \theta_{l'}^* + \Delta$. This operation entails a feasible solution with a marginal contribution to the objective equal to $\Delta(c_{l'} - c_l + \sigma)$. Either this quantity is negative—which would contradict the optimality of (θ^*, ξ^*) —or $\Sigma(\theta', \xi') < \Sigma(\theta^*, \xi^*)$ which is also not possible. \square

6.2 An efficient implementation

Problem (8)–(12) may contain an exponentially large number of ξ variables as indexed by all the possible values of σ . Let $\Lambda_u^R \subseteq \Lambda_u$ be a (typically much) smaller subset of σ values which includes the largest value in Λ_u . Because any rebate vector σ' such that $\sigma'_{ul} \leq \sigma_{ul}$ for every $u \in N, l \in \Omega$ is a valid rebate vector, one could replace every rebate $\sigma \in \Lambda_u$ by $\sigma' \leftarrow \max\{\sigma' \in \Lambda_u^R : \sigma' \leq \sigma\}$. Therefore, one may consider problem (8)–(12) by replacing Λ_u by Λ_u^R , thus reducing the number of variables ξ and the number of constraints (10), (12) without compromising the conditions required for Proposition 2. Such a set Λ_u^R can be constructed by selecting the M quantities in Λ_u , where M is a user defined parameter ($M=20$ works well in practice). Observe that adding columns to the RMP changes the quintiles and hence would seem to require resolving the RMP from scratch during each iteration of CG. Since resolving the RMP from scratch at each iteration of CG is too time intensive we only change the quintiles periodically. We add the value zero to $\Lambda_u^R \forall u \in \mathcal{N}$ so as to model the case of a value σ that is smaller than any quintile value.

7 Smooth-Flexible DOI

In this section we address the problem of combining both S-DOI and F-DOI within an unified framework. We refer to the resulting inequalities as Smooth-Flexible DOI and denote them as SF-DOI. Using the same notation and decision variables introduced in Sections 4-6, we consider the following optimization problem:

$$\min_{\theta, \omega, \xi} \quad \sum_{l \in \Omega} c_l \theta_l + \sum_{s \in \mathcal{S}} \rho_s \omega_s - \sum_{u \in N, \sigma \in \Lambda_u} \sigma \xi_{u\sigma} \quad (13)$$

subject to

$$\sum_{l \in \Omega} a_{ul} \theta_l + \sum_{s \in \mathcal{S}_u^+} \omega_s - \sum_{s \in \mathcal{S}_u^-} \omega_s - \sum_{\sigma \in \Lambda_u} \xi_{u\sigma} \geq 1 \quad u \in N \quad (14)$$

$$\xi_{u\sigma} - \sum_{l \in \Omega} \beta_{ul\sigma} \theta_l \leq 0 \quad u \in N, \sigma \in \Lambda_u \quad (15)$$

$$\theta_l \geq 0 \quad l \in \Omega \quad (16)$$

$$\omega_s \geq 0 \quad s \in \mathcal{S} \quad (17)$$

$$\xi_{u\sigma} \geq 0 \quad u \in N, \sigma \in \Lambda_u. \quad (18)$$

The following proposition formalizes the correctness of the SF-DOI by proving that, at optimality, variables ω, ξ are non-basic.

Proposition 3 *Problem (13)–(18) admits an optimal solution $(\theta^*, \omega^*, \xi^*)$ such that $\omega_s^* = 0$ for every $s \in \mathcal{S}$ and that $\xi_{u\sigma}^* = 0$ for every $u \in N, \sigma \in \Lambda_u$.*

Proof. Let $(\theta^*, \omega^*, \xi^*)$ be an optimal solution to problem (13)–(18). If multiple optima exist, let it be one that minimizes $\Sigma(\theta, \omega, \xi) = \sum_{l \in \Omega} \theta_l + \sum_{s \in \mathcal{S}} \omega_s + \sum_{u \in N, \sigma \in \Lambda_u} \xi_{u\sigma}$. The proof follows by applying

the same arguments provided for the proofs of correctness for the S-DOI and F-DOI in sequence. First we assume that $\omega_s^* > 0$ for some $s \in \mathcal{S}$ to arrive to a contradiction. Then, assuming that $\omega_s^* = 0$ for every $s \in \mathcal{S}$, we assume that $\xi_{u\sigma}^* > 0$ for some $u \in N, \sigma \in \Lambda_u$ to arrive to another contradiction. \square

Remark 1 *The same arguments used to reduce the problem size associated with the F-DOI can be applied to reduce the problem size of (13)–(18) by replacing Λ by a restricted subset $\Lambda_u^R \subseteq \Lambda_u$.*

8 Efficient pricing

Let us consider the mathematical problem (13)–(18) including the SF-DOI. Let (α, γ) be a vector of dual variables associated with the constraints (14)–(15). For a column $l \in \Omega$, its reduced cost \bar{c}_l can be computed as follows:

$$\bar{c}_l = c_l - \sum_{u \in N} a_{ul} \alpha_u + \sum_{u \in N, \sigma \in \Lambda_u} \beta_{u\sigma} \gamma_{u\sigma}. \quad (19)$$

As shown by Lokhande et al. (2020) in the context of set-packing, the dual variables γ can be ignored by the pricing algorithm, yet without compromising the correctness of the CG method. This result is important since it allows us to use a standard pricing algorithm instead of having to adapt one for the consideration of the new dual variables γ . The following proposition formalizes this result.

Proposition 4 *Let (α^*, γ^*) be an optimal solution to the dual RMP obtained from dualizing problem (13)–(18) and where Ω is replaced by a restricted set Ω_R , and such that $\min\{c_l - \sum_{u \in N} a_{ul} \alpha_u^* : l \in \Omega\} \geq 0$. Then $\sum_{u \in N} \alpha_u^* = z^*$, where z^* is the optimal value associated with problem (1)–(3).*

Proof. Since $\Omega_R \subseteq \Omega$ we have that $z^* \leq \sum_{u \in N} \alpha_u^*$. Also, since $\min\{c_l - \sum_{u \in N} a_{ul} \alpha_u^* : l \in \Omega\} \geq 0$ it follows that α^* is also feasible for the dual of problem (1)–(3), which in turn implies that $\sum_{u \in N} \alpha_u^* \leq z^*$. \square

9 Case study: single-source capacitated facility location

In this section we provide an efficient implementation of S-DOI, F-DOI and SF-DOI to the solution of the single-source capacitated facility location problem (SSCFLP). CG remains one of the most efficient methods to compute high quality dual bounds for the SSCFLP (Diaz and Fernández 2002), although recently cut-and-solve methods have proven particularly efficient for its solution (Yang et al. 2012, Gadegaard et al. 2018). The motivation to consider the SSCFLP in this case study comes from the observation made by other authors in the past that CG applied for the SSCFLP suffers from serious degeneracy issues for problems with dense optimal columns (Gadegaard et al. 2018).

In the SSCFLP, we are given a set of customers N and a set of potential facilities I , and such that $I \cap N = \emptyset$. With each facility $i \in I$ is associated a fixed opening cost f_i and a capacity K_i . With each customer $u \in N$ is associated a demand $d_u \geq 0$. With each pair $(i, u) \in I \times N$ is associated an assignment cost $c_{iu} \geq 0$. Without loss of generality, we may assume that all parameters are integer-valued. It is assumed that $\sum\{K_i : i \in I\} \geq \sum\{d_u : u \in N\}$ as otherwise the problem is trivially infeasible. The problem consists in selecting a subset $I' \subseteq I$ of facilities to open, and to assign every customer to exactly one open facility in such a way that the capacities of the selected facilities are respected, at minimum total cost (fixed costs + assignment costs).

The SSCFLP can be modeled as a set-covering problem, as follows. For every facility $i \in I$, let Ω_i be the set of pairs (i, S) where $S \subseteq N$ is a customer subset that can be assigned fully to i without it exceeding its capacity (thus such that $\sum\{d_u : u \in S\} \leq K_i$). The set of all such pairs (i, S) is denoted Ω . The cost of an assignment pattern $l = (i, S)$ is given by $c_l = f_i + \sum\{c_{iu} : u \in S\}$. For an assignment pattern $l = (i, S)$ let $a_{ul} \geq 0$ be a binary constant taking the value 1 if $u \in S$. Let $\theta_l \geq 0$

be a continuous variable representing the number of times that the assignment pattern l appears in the solution. The set-covering formulation for the SSCFLP is as follows:

$$\min_{\theta} \sum_{l \in \Omega} c_l \theta_l \quad (20)$$

subject to

$$\sum_{l \in \Omega} a_{ul} \theta_l \geq 1 \quad u \in N \quad (21)$$

$$\sum_{l \in \Omega_i} \theta_l \leq 1 \quad i \in I \quad (22)$$

$$\theta_l \geq 0 \quad l \in \Omega. \quad (23)$$

To embed this problem within the proposed framework, we need to provide efficient mechanisms to compute the rebates σ (for the F-DOI) and the penalties ρ (for the S-DOI). We propose the following formulae for the computation of those quantities:

$$\rho_{uv} \leftarrow \max\{c_{iv} - c_{iu} : i \in I\} \quad u, v \in N, u \neq v, d_u \geq d_v \quad (24)$$

$$\sigma_{ul} \leftarrow c_{iu} \quad u \in N, l = (i, S) \in \Omega. \quad (25)$$

It is easy to see that these quantities satisfy the conditions necessary for the validity of the SF-DOI using formulation (13)–(18). Moreover, they are computationally cheap to compute. The pricing subproblem, as usual for the SSCFLP, remains a 0-1 knapsack problem.

10 Computational results

In this section we present computational evidence of the impact of the DOI introduced in this article for the particular case of the SSCFLP. We consider two classical benchmark datasets —namely the Holmberg et al. and the Yang et al. datasets— plus two additional synthetic datasets generated by us. The performance of the DOI is assessed according to the speedup provided with respect to a classical (non-stabilized) CG algorithm for the same problem. Each algorithm is executed until proving optimality of the CG, thus the linear relaxation is solved fully. The algorithms have been coded in Matlab and we use CPLEX as general-purpose MIP solver for solving the RMP. Our machine is equipped with a 8-core AMD Ryzen 1700 CPU @3.0 GHz and 32 GB of memory running Windows 10.

When employing F-DOI we assign the values in Λ_u^R as 20 evenly spaced quantiles over the distribution of values in Λ_u . As more columns enter the RMP this distribution changes. We update Λ_u^R periodically to reflect this change and more accurately represent the distribution. We update on iterations 1, 5, 25, 100, 200, 500 and every 500 iterations onward. When employing S-DOI we save computational time in solving the RMP by only including a subset of the DOI. Specifically, we include only the DOI variables associated to the smallest 25% of ρ_s values.

For pricing, each facility induces a 0-1 knapsack problem capable of producing a negative reduced cost column. We choose to return the first 20 columns with negative reduced cost found while iterating over the facilities. If all facilities are cycled through before 20 negative reduced cost columns are found, the algorithm terminates pricing and returns all the negative reduced cost columns found by that point. If through a single round pricing no negative reduced cost columns are found after all facilities have been cycled through, then pricing is terminated and CG is then complete. The 0-1 knapsack problems are solved using dynamic programming.

10.1 Results on the Holmberg et al. dataset

In our first set of experiments we test our DOI on problems from the SSCFLP benchmark dataset defined in (Holmberg et al. 1999). We focus on the 16 largest problems (numerically indexed 56–71 in

the original paper) where $|N| = 200$ and $|I| = 30$. In those instances, the capacities and demands are assigned randomly such that the ratio of total capacity over total demand (K_{total}/d_{total}) ranges from 1.97 to 3.95. The facility fixed costs are distributed over a range from 500 to 1500 and the assignment costs are proportional to the Euclidean distance between a customer and a facility.

We execute each algorithm variant on all 16 problem instances in this dataset. In Table 1 we report the CPU times required to achieve the linear relaxation lower bound of problem (20)–(22) for each algorithm variant, and the associated speedup as a factor of the CPU taken for the standard CG. In Figure 1 we report the following data. In the two top figures, we plot average gap (across all 16 problem instances) as a function of the CPU time (left-most figure) and the number of iterations (right-most figure). The two bottom figures report the speedup (the time ratio of employing a given DOI as a factor of employing no DOI) as a function of run time (left-most figure) and number of iterations (right-most figure). All three DOI on average provide a speedup over standard CG with S-DOI performing the best with an average speedup of 3.3. The S-DOI provides a positive speedup for all 16 instances while the F-DOI and the SF-DOI show more mixed results in a case by case analysis. The F-DOI provide a positive speedup in 11 out of the 16 instances while the SF-DOI provide a positive speedup in 12 out of the 16 instances. All three DOI categorically required fewer CG iterations to converge than the standard CG, however employing these DOI comes at greater computational cost in solving the RMP. This computational liability is enough to negate the benefit of the DOI in those instances where the DOI provided no benefit. The SF-DOI notably required the fewest iterations to converge on average but still, for the most part, is outperformed by the S-DOI when judging for time. We note that employing the F-DOI largely comes at a significantly greater computational cost than employing the S-DOI, which accounts for its lower performance benefit.

Table 1: Holmberg et al. dataset ($|N| = 200$) runtime results

Instance	Time (sec)				Speedup		
	Standard	S-DOI	F-DOI	SF-DOI	S-DOI	F-DOI	SF-DOI
56	9.3	5.9	11.7	9.5	1.6	0.8	1.0
57	9.4	6.9	21.8	21.8	1.4	0.4	0.4
58	9.8	8.6	48.5	41.3	1.1	0.2	0.2
59	15.4	10.2	42.6	34.6	1.5	0.4	0.4
60	32.8	8.8	10.9	8.5	3.7	3.0	3.8
61	43.7	9.4	21.8	10.8	4.6	2.0	4.0
62	33.7	11.4	45.1	43.8	3.0	0.7	0.8
63	69.0	12.9	21.0	11.8	5.3	3.3	5.8
64	55.5	10.6	14.9	6.0	5.2	3.7	9.2
65	87.1	13.1	16.5	11.0	6.6	5.3	7.9
66	55.1	16.6	37.8	41.6	3.3	1.5	1.3
67	72.5	66.3	25.1	18.0	1.1	2.9	4.0
68	29.4	8.5	13.9	10.4	3.5	2.1	2.8
69	44.9	11.2	19.2	10.3	4.0	2.3	4.4
70	81.2	18.5	68.3	30.3	4.4	1.2	2.7
71	45.1	15.0	27.8	21.3	3.0	1.6	2.1
mean	43.4	14.6	27.9	20.7	3.3	2.0	3.2
median	44.3	10.9	21.8	14.9	3.4	1.8	2.8

10.2 Results on the Yang et al. dataset

In our second set of experiments we use the benchmark dataset presented in Yang et al. (2012). This dataset contains a number of large instances. We focus here on a subset of 10 instances where $|N| = 200$. Instances 1-5 have $|I| = 30$ while instance 6-10 have $|I| = 60$. Customer and facility locations are randomly assigned on the unit square. Costs are set as the euclidean distance between a particular customer and facility multiplied by 10 and then rounded. The ratio of total capacity to total demand ranges from 1.8 to 3.5.

We report the same data as for the Holmberg et al. dataset. In Table 2 we report the computing times required to achieve the linear relaxation lower bounds. In Figure 2 we report the DOI average

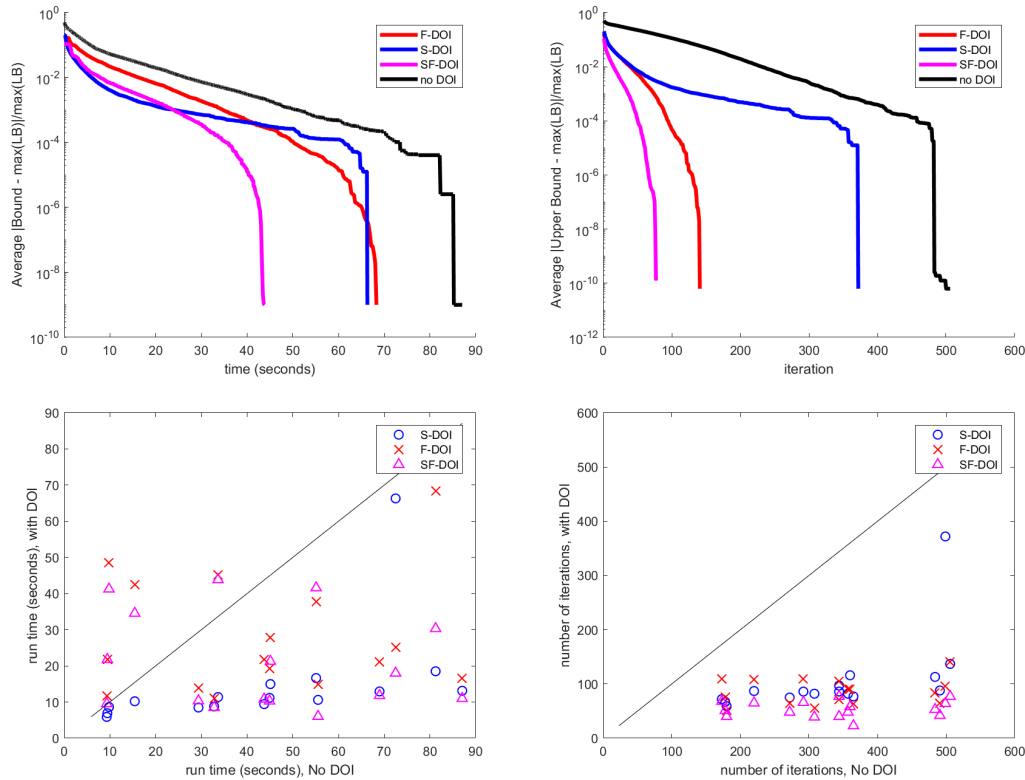


Figure 1: Holmberg et al. dataset ($|N| = 200$), Aggregated plots. Relative gaps are displayed as relative difference between upper and the maximum lower bounds. (Top Left): Average relative gap over 16 problem instances as a function of time. (Top Right): Average relative gap over 16 problem instances as a function of iterations. (Bottom Left): Comparative run times between using DOI and using no DOI for all 16 problem instances. (Bottom Right): Comparative iterations required between using DOI and using no DOI for all 16 problem instances.

performance as a function of time and iterations, just as in Figure 1. Here, although each DOI provides an improvement in the iterations, the F-DOI and the SF-DOI fail to improve upon the convergence time and in fact manage to slow down convergence overall. The computational cost here for the F-DOI is too high when compared to the iterative benefit, leading to a speedup (or a slowdown rather) of 0.2. The S-DOI still manage to provide a positive speedup in all instances, with an average speedup of 1.7. We note here that the F-DOI and the SF-DOI can often have their benefits negated due to their heavier computational cost. They tend to provide better benefits on problems whose solutions contain denser columns relative to the size of the problem. If the problem does not contain dense columns, the benefit cannot manage to outweigh the computational cost, as is seen here.

10.3 DOI and their effect on problems with dense columns

We note that problems with denser columns in the final solution can lead to more degeneracy, allowing the DOI to provide more benefit. Facility capacities provide a hard ceiling on the size of any potential column. We aim here to relax this ceiling and see the effect on the performance of each of the DOI. Specifically, we take the same problem instances from Holmberg et al. (1999) and Yang et al. (2012) in the previous sections and boost each facility capacity by a factor of 2, 3, and 4. For each facility we use $K'_i = LK_i$ as the facility capacity where K_i was the original facility capacity and L is the factor increase.

The speedup results for both datasets are show in Table 3. Here we see a general trend of improvement in speedup as L increases. The F-DOI and SF-DOI show the most marked improvement. The F-DOI goes from an average speedup of 2.0 and 0.2 at $L = 1$ on Holmberg et al. and Yang et al. datasets respectively to 9.2 and 4.4 at $L = 4$. The S-DOI also shows overall improvement, going from

Table 2: Yang et al. dataset ($|N| = 200$) runtime results

Instance	Time (sec)				Speedup		
	Standard	S-DOI	F-DOI	SF-DOI	S-DOI	F-DOI	SF-DOI
1	47.5	29.1	217.5	151.4	1.6	0.2	0.3
2	40.2	26.0	222.1	137.8	1.5	0.2	0.3
3	53.0	28.3	213.2	143.0	1.9	0.2	0.4
4	74.6	36.6	276.2	171.1	2.0	0.3	0.4
5	131.5	89.2	572.6	269.9	1.5	0.2	0.5
6	58.0	30.3	249.4	156.9	1.9	0.2	0.4
7	49.5	27.1	233.8	142.5	1.8	0.2	0.3
8	68.9	40.5	312.3	159.3	1.7	0.2	0.4
9	53.9	27.7	236.5	127.9	1.9	0.2	0.4
10	56.7	33.6	249.6	148.5	1.7	0.2	0.4
mean	63.4	36.8	278.3	160.8	1.8	0.2	0.4
median	55.3	29.7	243.0	149.9	1.8	0.2	0.4

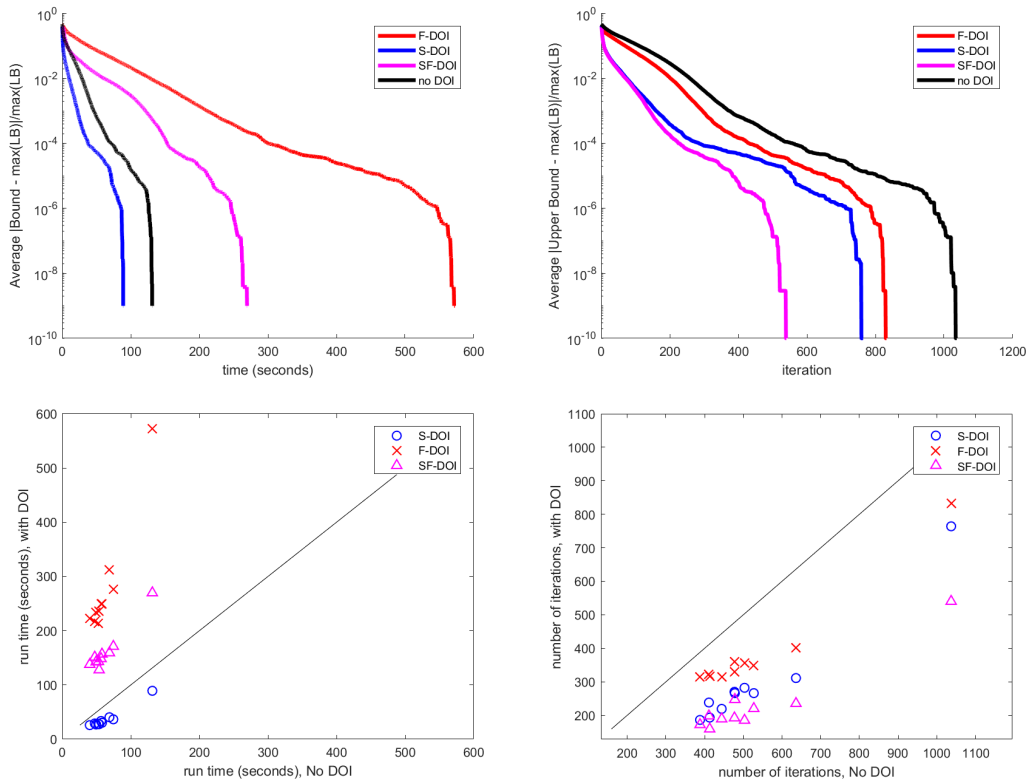


Figure 2: Yang et al. dataset ($|N| = 200$), Aggregated plots. Relative gaps are displayed as the relative difference between upper and maximum lower bound. (Top Left): Average relative gap over 10 problem instances as a function of time. (Top Right): Average relative gap over 10 problem instances as a function of iterations. (Bottom Left): Comparative run times between using DOI and using no DOI for all 10 problem instances. (Bottom Right): Comparative iterations required between using DOI and using no DOI for all 10 problem instances.

an average speedup of 3.3 to 8.6 on Holmberg et al. and from 1.8 to 4.0 on Yang et al.. The SF-DOI shows the greatest average performance at high capacity levels with a speedup of 16.1 and 11.5 at $L = 4$ on Holmberg et al. and Yang et al. datasets, respectively. We note that on most problem instances where S-DOI and F-DOI both perform well, the SF-DOI can often experience significant diminishing returns in employing both DOI. On certain problems, however, this stark pattern of diminishing returns is not as prevalent and the SF-DOI can do better than both other DOI separately. We find this occurs frequently on heavily degenerate problems where standard CG takes particularly long to converge.

Table 3: Runtime results for increased capacity. New capacity $K'_i = LK_i$ for each facility

Instance		Time (sec)				Speedup			
		Standard	S-DOI	F-DOI	SF-DOI	S-DOI	F-DOI	SF-DOI	
Holmberg et al.	mean	$L = 1$	43.4	14.6	27.9	20.7	3.3	2.0	3.2
		$L = 2$	136.8	19.7	26.3	14.4	25.3	5.4	11.0
		$L = 3$	240.4	33.8	28.5	16.2	8.3	8.2	14.4
		$L = 4$	300.0	40.5	32.4	19.1	8.6	9.2	16.1
	median	$L = 1$	44.3	10.9	21.8	14.9	3.4	1.8	2.8
		$L = 2$	117.6	12.8	25.5	13.1	16.7	5.8	9.8
		$L = 3$	211.2	24.7	27.4	15.9	8.4	8.4	13.6
		$L = 4$	255.1	29.8	29.0	18.9	9.0	8.6	14.8
Yang et al.	mean	$L = 1$	63.4	36.8	278.3	160.8	1.8	0.2	0.4
		$L = 2$	198.5	80.8	382.2	197.0	2.4	0.5	1.1
		$L = 3$	581.8	191.3	326.6	173.9	3.0	1.9	4.2
		$L = 4$	1269.0	218.6	231.9	123.5	4.0	4.6	11.5
	median	$L = 1$	55.3	29.7	243.0	149.9	1.8	0.2	0.4
		$L = 2$	163.6	72.4	326.0	181.3	2.3	0.5	0.8
		$L = 3$	360.3	134.9	311.0	154.1	2.7	1.1	2.2
		$L = 4$	405.3	159.2	210.2	114.9	2.8	2.3	4.5

10.4 Results on newly generated random instances

To assess our DOI further, we have decided to generate two new datasets. The first dataset has a specific construction where assignment costs are untruncated Euclidean distances between facilities and customers. We refer to these problems as the structured problems. We set $|N| = 250$ and $|I| = 50$ and generate 50 independent and identically distributed random instances. For each instance, customer and facility locations are randomly generated uniformly on the 2-D unit plane. The associated costs between facilities and customers is set as the euclidean distance between their respective locations. Each facility is given an instantiation cost $f_i = 5$ and capacity $K_i = 150$. Each customer is assigned a random demand drawn uniformly over $\{1, 2, 3, 4, 5\}$. Table 4 provides aggregate runtimes and Figure 3 aggregate plots of the DOI performance.

In the second class of random problem instances we abandon the structure defined previously and instead assign costs between facilities and customers devoid of any underlying structure. We refer to these problems as the unstructured problems. Assignment costs are randomly generated over a uniform distribution on the interval $(0,1)$. We set $|N| = 250$ and $|I| = 50$ and generate 50 independent and identically distributed random instances. Each facility is given an instantiation cost $f_i = 5$ and capacity $K_i = 150$. Each customer is assigned a random demand drawn uniformly over $\{1, 2, 3, 4, 5\}$. Runtime results are shown in Table 5 and aggregate plots are shown in Figure 4. Both datasets can be downloaded from <http://claudio.contardo.org>.

Table 4: Average runtime over 50 structured problem instances.

	Time (sec)				Speedup		
	Standard	S-DOI	F-DOI	SF-DOI	S-DOI	F-DOI	SF-DOI
mean	3252.0	37.3	226.6	52.7	113.6	14.4	130.2
median	766.1	35.0	206.5	51.8	22.7	4.4	17.4

Table 5: Average runtime over 50 unstructured problem instances.

	Time (sec)				Speedup		
	Standard	S-DOI	F-DOI	SF-DOI	S-DOI	F-DOI	SF-DOI
mean	89.0	99.5	52.9	80.5	0.9	1.7	1.1
median	88.7	98.7	53.7	82.4	0.9	1.7	1.1

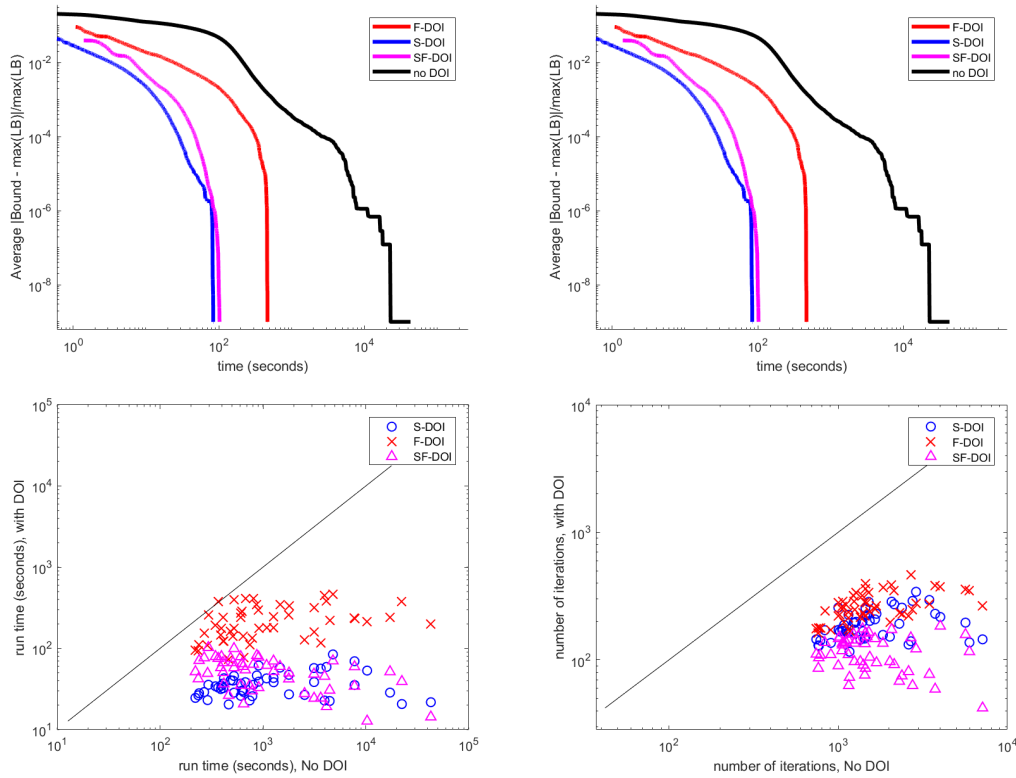


Figure 3: Structured SSCFLP aggregate plots. Relative gaps are displayed as the relative difference between upper and maximum lower bound. (Top Left): Average relative gap over 50 problem instances as a function of time. (Top Right): Average relative gap over 50 problem instances as a function of iterations. (Bottom Left): Comparative run times between using DOI and using no DOI for all 50 problem instances. (Bottom Right): Comparative iterations required between using DOI and using no DOI for all 50 problem instances.

We immediately see from the structured problem dataset the vast improvement achievable with the S-DOI on highly degenerate problems with an underlying structure behind the assignment costs. The S-DOI and the SF-DOI manage median speedups of 22.7 and 17.4 respectively. The F-DOI also do well with a median speedup of 4.4. It can be observed from the relative runtime plots that the S-DOI actually outperform the SF-DOI for most problem instances. However, for particularly difficult problems for standard CG to solve, both DOI appear very competitive with one another. The SF-DOI see a massive improvement in the number of iterations required on these types of challenging problems.

From the unstructured problem dataset we observe a limitation of the S-DOI. The S-DOI fail to provide a significant benefit. This is due to the poor correlation between relative customer costs across facilities. This prevents the S-DOI from finding low ρ_s values to effectively restrict the dual space. We note that the F-DOI do not share this limitation as they perform relatively well on these problems, achieving an average speedup of 1.7. Just as the S-DOI were more robust to problems with lower facility capacities, we see here that the F-DOI are more robust to problems with weaker underlying structure.

11 Concluding remarks

In this document we introduced a new approach to accelerate the convergence of column generation when applied to set-covering-based formulations. The proposed mechanism uses little-to-no specific problem knowledge to derive two classes of dual-optimal inequalities. The Smooth-DOI (S-DOI) estimates the gain that would be obtained by replacing items from a column by items not originally covered, and applies a penalty for such swapping operation. The Flexible-DOI (F-DOI) describe rebates

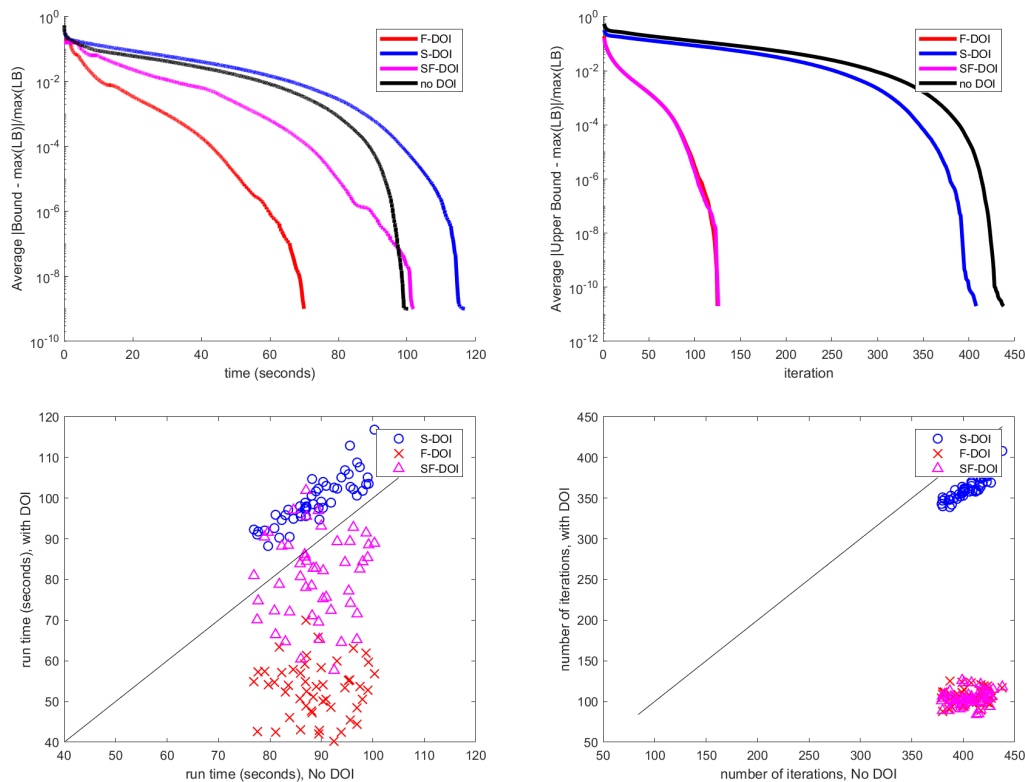


Figure 4: Unstructured SSCFLP aggregate plots. Relative gaps are displayed as the relative difference between upper and maximum lower bound. (Top Left): Average relative gap over 50 problem instances as a function of time. (Top Right): Average relative gap over 50 problem instances as a function of iterations. (Bottom Left): Comparative run times between using DOI and using no DOI for all 50 problem instances. (Bottom Right): Comparative iterations required between using DOI and using no DOI for all 50 problem instances.

for removing items from columns. These rebates are not sufficiently large to weaken the relaxation of the master problem. We show that our DOI can be combined without any additional implementation effort, and refer to such combination as the Smooth-Flexible-DOI (SF-DOI). We have assessed the efficiency of the new DOI to accelerate the convergence of some degenerate problem instances of the single-source capacitated facility location problem. The computational evidence shows that the new dual inequalities are efficient at reducing by a factor of up to $130\times$ the CPU times required to compute the linear relaxation dual bounds on structured problems with high a correlation between node positioning and assignment costs and loose facility capacities. In future work we seek to apply our approach to capacitated vehicle routing and extend our approach to operate in the context of branch and price (Barnhart et al. 1996). The extension of our approach to consider valid inequalities such as subset-row inequalities (Jepsen et al. 2008) presents another promising avenue for exploration.

References

- F. Barahona and D. Jensen. Plant location with minimum inventory. *Mathematical Programming*, 83(1): 101–112, 1998.
- C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46:316–329, 1996.
- H. Ben Amor, J. Desrosiers, and J. M. Valério de Carvalho. Dual-optimal inequalities for stabilized column generation. *Operations Research*, 54(3):454–463, 2006.
- J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

- O. Briant, C. Lemaréchal, P. Meurdesoif, S. Michel, N. Perrot, and F. Vanderbeck. Comparison of bundle and classical column generation. *Math. Program.*, 113(2):299–344, 2008. doi: 10.1007/s10107-006-0079-z.
- I. A. Contreras, J. Cordeau, and G. Laporte. Benders decomposition for large-scale uncapacitated hub location. *Operations Research*, 59(6):1477–1490, 2011. doi: 10.1287/opre.1110.0965.
- L. Costa, C. Contardo, and G. Desaulniers. Exact branch-price-and-cut algorithms for vehicle routing. *Transportation Science*, Forthcoming, 2019.
- J. Diaz and E. Fernández. A branch-and-price algorithm for the single source capacitated plant location problem. *Journal of the Operational Research Society*, 53(7):728–740, 2002.
- O. Du Merle, D. Villeneuve, J. Desrosiers, and P. Hansen. Stabilized column generation. *Discrete Mathematics*, 194(1-3):229–237, 1999.
- A. Frangioni. Generalized bundle methods. *SIAM Journal on Optimization*, 13(1):117–156, 2002. doi: 10.1137/S1052623498342186.
- S. L. Gadegaard, A. Klöse, and L. R. Nielsen. An improved cut-and-solve algorithm for the single-source capacitated facility location problem. *EURO J. Computational Optimization*, 6(1):1–27, 2018. doi: 10.1007/s13675-017-0084-4.
- A. M. Geoffrion. Lagrangean relaxation for integer programming. In *Approaches to integer programming*, pages 82–114. Springer, 1974.
- P. Gilmore and R. Gomory. A linear programming approach to the cutting-stock problem. *Operations Research*, 9(6):849–859, 1961.
- J. Gondzio and P. González-Brevis. A new warmstarting strategy for the primal-dual column generation method. *Math. Program.*, 152(1-2):113–146, 2015. doi: 10.1007/s10107-014-0779-8.
- J. Gondzio, P. González-Brevis, and P. A. Munari. New developments in the primal-dual column generation technique. *European Journal of Operational Research*, 224(1):41–51, 2013. doi: 10.1016/j.ejor.2012.07.024.
- J. Gondzio, P. González-Brevis, and P. A. Munari. Large-scale optimization with the primal-dual column generation method. *Math. Program. Comput.*, 8(1):47–82, 2016. doi: 10.1007/s12532-015-0090-6.
- T. Gschwind and S. Irnich. Dual inequalities for stabilized column generation revisited. *INFORMS Journal on Computing*, 28(1):175–194, 2016. doi: 10.1287/ijoc.2015.0670.
- T. Gschwind and S. Irnich. Stabilized column generation for the temporal knapsack problem using dual-optimal inequalities. *OR Spectrum*, 39(2):541–556, 2017. doi: 10.1007/s00291-016-0463-x.
- K. Holmberg, M. Rönnqvist, and D. Yuan. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*, 113(3):544–559, 1999.
- M. Jepsen, B. Petersen, S. Spoorendonk, and D. Pisinger. Subset-row inequalities applied to the vehicle-routing problem with time windows. *Operations Research*, 56(2):497–511, 2008.
- M. Keuper, M. Singh, and J. Yarkony. Massively parallel benders decomposition for correlation clustering. arXiv preprint arXiv:1902.05659, 2019.
- L. Leal-Taixe, G. Pons-Moll, and B. Rosenhahn. Branch-and-price global optimization for multi-view multi-target tracking. In *Proc. 25th Conference on Computer Vision and Pattern Recognition*, pages 1987–1994, Providence, Rhode Island, 2012.
- V. S. Lokhande, S. Wang, M. Singh, and J. Yarkony. Accelerating column generation via flexible dual optimal inequalities with application to entity resolution, 2020.
- M. E. Lübbecke and J. Desrosiers. Selected topics in column generation. *Operations Research*, 53(6):1007–1023, 2005.
- T. L. Magnanti and R. T. Wong. Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484, 1981.
- R. E. Marsten, W. Hogan, and J. W. Blankenship. The boxstep method for large-scale optimization. *Operations Research*, 23(3):389–405, 1975.
- J. E. Mitchell. An interior point column generation method for linear programming using shifted barriers. *SIAM Journal on Optimization*, 4(2):423–440, 1994. doi: 10.1137/0804023.
- A. A. Pessoa, R. Sadykov, E. Uchoa, and F. Vanderbeck. Automation and combination of linear-programming based stabilization techniques in column generation. *INFORMS Journal on Computing*, 30(2):339–360, 2018. doi: 10.1287/ijoc.2017.0784.
- L.-M. Rousseau, M. Gendreau, and D. Feillet. Interior point stabilization for column generation. *Operations Research Letters*, 35(5):660–668, 2007.

- W. van Ackooij and A. Frangioni. Incremental bundle methods using upper models. *SIAM Journal on Optimization*, 28(1):379–410, 2018. doi: 10.1137/16M1089897.
- S. Wang, S. Wolf, C. Fowlkes, and J. Yarkony. Tracking objects with higher order interactions via delayed column generation. In *Proc. 20th International Conference on Artificial Intelligence and Statistics*, pages 1132–1140, Fort Lauderdale, Florida, 2017.
- S. Wang, A. Ihler, K. Kording, and J. Yarkony. Accelerating dynamic programs via nested benders decomposition with application to multi-person pose estimation. In *Proc. 15th European Conference on Computer Vision*, pages 652–666, Munich, Germany, 2018.
- Z. Yang, F. Chu, and H. Chen. A cut-and-solve based algorithm for the single-source capacitated facility location problem. *European Journal of Operational Research*, 221(3):521–532, 2012.
- J. Yarkony and C. Fowlkes. Planar ultrametrics for image segmentation. In *Proc. 28th Advances in Neural Information Processing Systems*, pages 64–72, Montreal, Quebec, 2015.
- J. Yarkony, A. Ihler, and C. Fowlkes. Fast planar correlation clustering for image segmentation. In *Proc. 12th European Conference on Computer Vision*, pages 1169–1176, Florence, Italy, 2012.
- J. Yarkony, Y. Adulyasak, M. Singh, and G. Desaulniers. Data association via set packing for computer vision applications. In *Les Cahiers du GERAD*, Montreal, Quebec, 2019.
- C. A. Zetina, I. A. Contreras, and J. Cordeau. Exact algorithms based on benders decomposition for multicommodity uncapacitated fixed-charge network design. *Computers & OR*, 111:311–324, 2019. doi: 10.1016/j.cor.2019.07.007.