On the Augmented Lagrangian Dual for Integer Programming

N.L. Boland and A. C. Eberhard*

Abstract

We consider the augmented Lagrangian dual for integer programming, and provide a primal characterization of the resulting bound. As a corollary, we obtain proof that the augmented Lagrangian is a strong dual for integer programming. We are able to show that the penalty parameter applied to the augmented Lagrangian term may be placed at a fixed, large value and still obtain strong duality for pure integer programs.

1 Introduction

Duality has long been a powerful tool in *Integer Programming* (IP). Several dual problems have been proposed for IP (see, e.g. Nemhauser and Wolsey [28] Chapter II.3), including the Lagrangian dual [12], the surrogate dual [13], the subadditive dual [15, 18, 19], and more recently, the inference dual [17]. The Lagrangian dual has given rise to a wealth of algorithms, both as a heuristic and through the provision of bounds for use in exact methods such as branch-and-bound. Surrogate duality has been less well explored, but methods have been developed for solving surrogate duals, and for embedding them in methods for solving the primal IP, notably those of Karwan and Rardin [20, 21]. The value of the inference and subadditive duals appears to have been primarily for their use in sensitivity analysis and the study of the value function for IP [3, 32]. One surprising new development is the application of Z-transforms and complex variable techniques by Lasserre [23, 24] to study the value function. Very recently, algorithms for solving the subadditive dual for both linear and nonlinear IP have been developed [22, 26].

One reason for the prevalence of Lagrangian duality has been its tractability. However this advantage has a concomitant downside: it does not, in general, provide a strong bound. As noted in [21], the surrogate dual bound cannot be worse than the Lagrangian dual bound obtained through relaxation of the same set of constraints, and can be better. This is verified in other studies [16, 14, 11]. Strong duality in a form that allows practical, computationally useful methods to estimate bounds for IP has been elusive, although the recent work of Klabjan [22] and Li et al. [26] offers great promise.

By contrast, for nonlinear, nonsmooth optimisation there have been a number of strong duality schemes proposed in the last 10 years. The *augmented* Lagrangian dual has been of particular interest in this area. In convex optimisation, algorithms to solve the augmented Lagrangian dual were found to be more robust, and converged under less stringent assumptions, than their standard Lagrangian dual predecessors [29]. Early application of the augmented Lagrangian in IP [30] was aimed at producing better dual bounds at the root node of the branch-and-bound tree. Li and Sun [25] report some more recent progress with strong duality schemes. In other lines of research, the potential for their application has been observed but not extensively studied. Theory cast in a very general framework may be found in the work of Burachik and Rubinov [5, 6], which requires some standard assumptions for unstructured nonlinear problems that are less natural for IPs. The approach of [5] and [6] provides justification for the penalty methods applied to small nonlinear IPs in [7].

One of the challenges for applying the augmented dual in IP computational methods is that the augmented Lagrangian term destroys the natural separability properties of the Lagrangian dual. However this can be overcome by the use of an approach known as the *alternating direction method of multipliers* [4, 9], which is enjoying recent attention in the literature. Thus the augmented Lagrangian dual for IP warrants further attention.

In this paper, we contribute to the theory of the augmented Lagrangian dual for IP so as to provide insight into how it obtains better bounds than the standard Lagrangian dual. Our main result is a

*Email addresses of the authors: Natashia.Boland@newcastle.edu.au, andy.eb@rmit.edu.au
This research was supported by the ARC Discovery grant no. DP0987445.

primal characterization of the augmented Lagrangian dual value which in some sense mirrors that given in [28] Chapter II.3.6 as Theorem 6.2 (originally proved in [12]) for the standard Lagrangian dual. The primal description of the standard Lagrangian dual is maximization of the original objective over the intersection of the set of points satisfying the relaxed constraints and the convex hull of points satisfying the constraints that are not relaxed, including integrality. In other words, the action of the standard Lagrangian dual is to convexify the integer points feasible for the constraints that are not relaxed. (This is expressed more precisely in equation (1) below.) We show that the augmented Lagrangian dual convexifies only a subset of these points, convexifying the integer points that satisfy the constraints not relaxed *and* that do not violate the constraints that *are* relaxed by too much, within a parameterized framework controlling a measure of the degree of violation of the relaxed constraints. The fact that the augmented Lagrangian term should act to control the violation of the relaxed constraints is intuitively clear: here we provide a rigorous formulation showing how that occurs.

A motivation for considering this direction of research arises from its links to standard Lagrangian methods and their role in IP. For many well-known applications, such as airline crew scheduling and cutting stock problems, the standard Lagrangian dual yields a bound with very small duality gap, and is solved in its primal form via the technique of column generation — undoubtedly one of the success stories of integer programming [8]. Since Theorem 6.2 (Chapter II.3.6, [28]) underpins the use of column generation as a technique for solving the standard Lagrangian dual, it is our hope that providing an analogous result for the augmented Lagrangian dual will stimulate new methods for the latter. Furthermore, column generation is an active subject of current research, with new ideas such as stabilization and dynamic constraint aggregation [27, 10, 2] emerging in recent years, aimed at accelerating convergence. With the augmented Lagrangian providing greater stability and better convergence properties than standard Lagrangian approaches in convex optimization, the idea that the augmented Lagrangian could also lead to convergence improvements in IP is an intriguing possibility.

As a corollary of our primal characterization, we are able to derive strong duality for the augmented Lagrangian dual. In the case of pure integer programs with bounded domains, we also present a stronger result, showing that the zero duality gap is attained for finite values of the augmented penalty term parameter. We demonstrate that strong duality for the augmented Lagrangian dual for IP can also be deduced from the results of [5], but the derivation is somewhat indirect, and does not yield any of the other results we give here.

2 Lagrangian and Augmented Lagrangian Duals

The standard Lagrangian dual in integer programming is obtained as follows. Let the mixed integer linear program (IP) be defined by $z^{IP} := \sup\{cx \mid Ax = b, x \in X\}$, with A an $m \times n$ matrix ($m \leq n$) having full row rank, where X includes integrality constraints, simple bounds on variable, other simple constraints and is assumed non-empty and closed. We assume that the value z^{IP} and the value of its linear programming relaxation z^{LP} are finite. Form the Lagrangian relaxation

$$z^{LR}(\lambda) := \sup_{x \in X} \left[cx + \lambda(b - Ax)\right],$$

and the associated Lagrangian dual $z^{LD} := \inf_\lambda z^{LR}(\lambda)$. Clearly $z^{LR}(\lambda)$ provides an upper bound on (IP), ($z^{LR}(\lambda) \geq z^{IP}$ for all λ), and z^{LD} provides the best such bound and $z^{LD} \leq z^{LP}$.

The Lagrangian dual has an elegant primal characterization [12, 28]:

$$z^{LD} = \sup\{cx \mid Ax = b, x \in \operatorname{conv}(X)\}, \tag{1}$$

where $\operatorname{conv}(C)$ denotes the smallest convex set containing C. So if X does not have the integrality property[1], the bound from the Lagrangian dual may be better than that obtained from the LP relaxation of (IP).

Let us consider the augmented Lagrangian relaxation

$$z_\sigma^{LR+}(\lambda) := \sup_{x \in X} \left[cx + \lambda(b - Ax) - \sigma\rho\,(b - Ax)\right], \tag{2}$$

[1]The set X satisfies the "integrality property" if ignoring the integrality constraints in the definition of X yields $\operatorname{conv}(X)$.

where $\rho : \mathbf{R}^m \to \mathbf{R}$ is a penalty term such as a norm or the square of the standard 2-norm and $\sigma \geq 0$ is treated as a fixed parameter.

Assumption: Assume ρ is of the form $\rho(x) = \psi(\|x\|)$ for some norm $\|\cdot\|$ on $\mathbf{R}^m$, where $\psi : \mathbf{R}_+ \to \mathbf{R}_+$ is a convex, monotonic increasing function for which $\psi(0) = 0$ and there is a $0 < \delta < 1$ such that

$$\liminf_{|a|\to\infty} \frac{\psi(a)}{|a|^{1+\delta}} \geq \delta > 0 \tag{3}$$

and $\operatorname{diam}\{a \mid \psi(a) \leq \delta\} \downarrow 0$ as $\delta \downarrow 0$.

Clearly $z_\sigma^{LR+}(\lambda) \leq z^{LR}(\lambda)$ for $\sigma \geq 0$. It is obvious that for any $\sigma \geq 0$, $z_\sigma^{LR+}(\lambda)$ provides an upper bound on z^{IP} for any λ, and $z_\sigma^{LD+} := \inf_\lambda z_\sigma^{LR+}(\lambda) \leq z^{LD}$ is the best such bound. In Theorem 1 below, we provide a primal expression for z_σ^{LD+} which provides insight into how the augmented Lagrangian dual can yield a strictly better bound than that of the standard Lagrangian. As we shall subsequently see, as σ increases, the former bound becomes exact.

Recall that the recession directions of a convex set $C \subseteq \mathbf{R}^n$ are given by $0^+C := \{d \in \mathbf{R}^n \mid C + d \subseteq C\}$. We need to use the minimax theorem of Sion [31].

Proposition 1 ([31]) *Let G, H be convex subsets of $\mathbf{R}^n$ with H closed and bounded. Let $f : G \times H \to \mathbf{R}$ with $f(g, \cdot)$ upper semi-continuous and quasi-concave[2] on H for every $g \in G$. If $f(\cdot, h)$ is lower semi-continuous and quasi-convex on G for every $h \in H$, then*

$$\inf_{g\in G} \max_{h\in H} f(g,h) = \max_{h\in H} \inf_{g\in G} f(g,h).$$

Theorem 1 *Suppose $-\infty < z^{IP} \leq z^{LP} < +\infty$ (and hence $z_\sigma^{LD+} < +\infty$). Then for any $\sigma > 0$*

$$z_\sigma^{LD+} = \max_{\kappa\geq 0} \left[\sup\{cx \mid Ax = b,\ x \in \operatorname{conv}\ (X \cap \{y \mid \|b - Ay\| \leq \kappa\})\} - \sigma\psi(\kappa)\right]. \tag{4}$$

The maximum in $\kappa \geq 0$ is attained by a finite value $\kappa(\sigma)$, where $\kappa(\sigma) \downarrow 0$ as $\sigma \uparrow \infty$. Moreover for any $\delta > 0$ there is a σ_δ such that for all $\sigma > \sigma_\delta$ we have $\kappa(\sigma) \leq \delta$.

Proof. We may write

$$\begin{aligned} z_\sigma^{LR+}(\lambda) &= \sup_{x\in X} \sup_{\kappa\geq 0} \{cx + \lambda(b - Ax) - \sigma\psi(\kappa) \mid \|b - Ax\| \leq \kappa\} \\ &= \sup_{\kappa\geq 0} \sup_{x\in X} \{cx + \lambda(b - Ax) - \sigma\psi(\kappa) \mid \|b - Ax\| \leq \kappa\} \\ &= \sup_{\kappa\geq 0} \left[\sup\{(c - \lambda A)x \mid x \in \operatorname{conv}\ (X \cap \{y : \|b - Ay\| \leq \kappa\})\} - \sigma\psi(\kappa)\right] + \lambda b \end{aligned} \tag{5}$$

because the maximum of a linear functional over a set equals its maximum over its convex hull. Note that as $z_\sigma^{LD+} < +\infty$ there exist λ for which $z_\sigma^{LR+}(\lambda) < +\infty$.

Let $\mathcal{T}(\kappa, \lambda) := \sup\{(c - \lambda A)x \mid x \in K(\kappa)\}$ where $K(\kappa) := \operatorname{conv}(X \cap \{y \mid \|b - Ay\| \leq \kappa\})$. As $z^{IP} := \sup\{cx \mid x \in X \text{ and } Ax = b\}$ is finite we have $cx \leq 0$ for all $x \in \ker(A) \cap 0^+ \operatorname{conv}(X)$. Furthermore we may decompose any $x \in K(\kappa)$ into an orthogonal sum $x = x_1 + x_2$ with $x_1 \in \ker(A) \cap 0^+ \operatorname{conv}(X)$, $x_2 \in \mathrm{R}(A^T)$ (the range of A^T) and $\|b - Ax_2\| \leq \kappa$ as is implied by the theory of pseudo-inverses. Thus it follows that

$$\mathcal{T}(\kappa, \lambda) \leq \sup\left\{(c - \lambda A)\, P_{\mathrm{R}(A^T)}(x) \mid x \in K(\kappa)\right\}, \tag{6}$$

where $P_{R(A^T)}(x)$ denotes the projection of x onto the range of A^T (which is orthogonal to $\ker(A)$). One may easily show that $C(\kappa) := \{P_{\mathrm{R}(A^T)}(x) | x \in K(\kappa)\}$ is closed and bounded, hence by well-known theory [1], $\kappa \mapsto C(\kappa)$ is an upper semi-continuous multifunction.

As $x_2 \in R(A^T)$, it must be that $\gamma\|x_2\| \leq \|Ax_2\|$ for some γ, where (since A has full row rank) $\gamma := \frac{1}{\|A^T(AA^T)^{-1}\|}$ is allowed. Thus we have the bound $\mathcal{T}(\kappa, \lambda) \leq (\frac{1}{\gamma}\|c\| + \|\lambda\|)(\kappa + \|b\|)$, growing linearly in κ. Thus (5) is bounded below by z_σ^{LD+} (as $z_\sigma^{LD+} = \inf_\lambda z_\sigma^{LR+}(\lambda)$) and above by the difference $(\frac{1}{\gamma}\|c\| + \|\lambda\|)(\kappa + \|b\|) - \sigma\psi(\kappa)$ so if κ solves (5) we have

$$z_\sigma^{LD+} \leq (\frac{1}{\gamma}\|c\| + \|\lambda\|)(\kappa + \|b\|) - \sigma\psi(\kappa).$$

[2] A function $x \mapsto f(x)$ is quasi-convex if $\{x \mid f(x) \leq \alpha\}$ is convex for all α. A function $x \mapsto f(x)$ is quasi-concave if $-f$ is quasi-convex.

Hence a contradiction follows from the coercivity of $\psi(\cdot)$ for any fixed σ and λ if κ assumed to be larger than a constant $B(\lambda)$ (which is bounded for bounded λ). Note that $B(\lambda)$ is also σ-dependent, but it is non-increasing for increasing σ, thus under the assumption that $\sigma \geq \bar{\sigma} > 0$ (any $\bar{\sigma} > 0$) this dependence can be ignored.

Define $F(\kappa, \lambda) := \mathcal{T}(\kappa, \lambda) + \lambda b = \sup\{cx + \lambda(b - Ax) \mid x \in K(\kappa)\}$. Then (as $cz \leq 0$ for all $z \in \ker(A) \cap 0^+ \operatorname{conv}(X)$)

$$-\infty < z^{IP} \leq F(\kappa, \lambda) \leq \sup\{cx + \lambda(b - Ax) \mid x \in C(\kappa)\} < +\infty, \tag{7}$$

where the upper bound is finite due to the local uniform boundedness of $\kappa \mapsto C(\kappa)$. Thus $F(\kappa, \lambda)$ is finite valued and moreover we claim $\kappa \mapsto F(\kappa, \lambda)$ is upper semi-continuous. As this function is monotonically non-decreasing we only have to consider a sequence $\kappa_p \downarrow \bar{\kappa}$ and show that $\lim_p F(\kappa_p, \lambda) \leq F(\bar{\kappa}, \lambda)$. Let $\varepsilon_p \downarrow 0$ and let $x_p \in K(\kappa_p)$ achieve the supremum to within ε_p. Decompose x_p into $x_p^1 \in \ker(A) \cap 0^+ \operatorname{conv}(X)$ and $x_p^2 \in C(\kappa_p)$. Using the upper semi-continuity of $C(\cdot)$ we may take a convergent subsequence (and after re-numbering) we have $x_p^2 \to x^2 \in C(\bar{\kappa})$. If $\{x_p^1\}$ is also bounded we may also extract a convergent subsequence (and after re-numbering) have $x_p^1 \to x^1$ with $x := x^1 + x^2 \in K(\bar{\kappa})$. Then $\lim_p F(\kappa_p, \lambda) = cx + \lambda(b - Ax) \leq F(\bar{\kappa}, \lambda)$. When $\{x_p^1\}$ is unbounded, by (7) we have $F(\cdot, \lambda)$ locally uniformly bounded so

$$0 = \lim_p \frac{1}{\|x_p^1\|}(cx_p + \lambda(b - Ax_p)) = \lim_p (c(\frac{x_p^1 + x_p^2}{\|x_p^1\|}) + \frac{1}{\|x_p^1\|}\lambda(b - Ax_p^2)) = \lim_p c(\frac{x_p^1}{\|x_p^1\|}).$$

Then $d := \lim_p \frac{x_p^1}{\|x_p^1\|} \in \ker(A) \cap 0^+ \operatorname{conv}(X)$ (taking subsequences if necessary) and $cd = 0$. By construction for all $\varepsilon > 0$ there exists $y_\varepsilon \in C(\bar{\kappa})$ with $\|y_\varepsilon - x^2\| \leq \varepsilon$, $cy_\varepsilon + \varepsilon \geq cx^2$ and $z_\varepsilon := y_\varepsilon + \alpha_\varepsilon d \in K(\bar{\kappa})$, for some $\alpha_\varepsilon > 0$. Then we have

$$\begin{aligned}\lim_p F(\kappa_p, \lambda) &= \lim_p (cx_p^1 + cx_p^2 + \lambda(b - Ax_p^2)) \leq cx^2 + \lambda(b - Ax^2) \leq cy_\varepsilon + \lambda(b - Ay_\varepsilon) + \varepsilon(1 + \|\lambda A\|) \\ &= cz_\varepsilon + \lambda(b - Az_\varepsilon) + \varepsilon(1 + \|\lambda A\|) \leq F(\bar{\kappa}, \lambda) + \varepsilon(1 + \|\lambda A\|).\end{aligned}$$

As ε is arbitrary we have $\lim_p F(\kappa_p, \lambda) \leq F(\bar{\kappa}, \lambda)$ in both cases, establishing upper semi-continuity.

Consider the mapping

$$B \mapsto Z(B) := \inf_\lambda \max_{\kappa \in [0,B]} [F(\kappa, \lambda) - \sigma\psi(\kappa)]. \tag{8}$$

which is monotonically non-decreasing and bounded above as $z_\sigma^{LD+} < +\infty$. Now as $\kappa \mapsto F(\kappa, \lambda)$ is monotone, it is quasi-concave, its level sets being unbounded intervals. Hence $\kappa \mapsto F(\kappa, \lambda) - \sigma\psi(\kappa)$ is also quasi-concave. It is evident that $\lambda \mapsto F(\kappa, \lambda)$ is convex for fixed κ. Thus we may apply the Sion minimax theorem. By the previous analysis for every $\varepsilon > 0$ there exists λ_ε attaining the infimum to within ε and a $B_\varepsilon = B(\lambda_\varepsilon)$ such that

$$\begin{aligned} z_\sigma^{LD+} + \varepsilon = Z(+\infty) + \varepsilon &> \sup_{\kappa \geq 0} [F(\kappa, \lambda_\varepsilon) - \sigma\psi(\kappa)] = \max_{\kappa \in [0, B_\varepsilon]} [F(\kappa, \lambda_\varepsilon) - \sigma\psi(\kappa)] \\ &\geq \max_{\kappa \in [0, B_\varepsilon]} \inf_\lambda [F(\kappa, \lambda) - \sigma\psi(\kappa)] = \inf_\lambda \max_{\kappa \in [0, B_\varepsilon]} [F(\kappa, \lambda) - \sigma\psi(\kappa)] = Z(B_\varepsilon) \geq Z(0) = z^{IP}, \end{aligned} \tag{9}$$

where the last equalities follow from an application of the Sion minimax theorem. The maximum in (9) is attained by κ_ε due to the upper semi-continuity of F in κ. Suppose there does not exists a $B > 0$ such that $Z(B) = Z(+\infty)$. Then we necessarily have $\|\lambda_\varepsilon\| \to \infty$ (to assume otherwise results in a bound on κ) and for any $\kappa_\varepsilon \in [0, B_\varepsilon]$ attaining the maximum in (9) we must have $\kappa_\varepsilon \to +\infty$ as $\varepsilon \downarrow 0$. Note that (9) implies

$$z_\sigma^{LD+} + \varepsilon \geq \inf_\lambda F(\kappa_\varepsilon, \lambda) - \sigma\psi(\kappa_\varepsilon) \geq z^{IP}.$$

As $\sigma > 0$ and $\psi(\kappa_\varepsilon) \to +\infty$ we have $\inf_\lambda F(\kappa_\varepsilon, \lambda) \to +\infty$ and the following contradiction ensues:

$$+\infty > z^{LP} \geq z^{LD} = \inf_\lambda \sup\{cx + \lambda(b - Ax) \mid x \in K(+\infty)\} \geq \lim_{\kappa_\varepsilon \to \infty} \inf_\lambda F(\kappa_\varepsilon, \lambda) = +\infty.$$

Applying the Sion minimax theorem we obtain for a fixed but sufficiently large $B > 0$

$$\begin{aligned}\inf_\lambda \max_{\kappa \geq 0} [F(\kappa, \lambda) - \sigma\psi(\kappa)] &= \inf_\lambda \max_{\kappa \in [0,B]} [F(\kappa, \lambda) - \sigma\psi(\kappa)] = \max_{\kappa \in [0,B]} \inf_\lambda [F(\kappa, \lambda) - \sigma\psi(\kappa)] \\ &\leq \max_{\kappa \geq 0} \inf_\lambda [F(\kappa, \lambda) - \sigma\psi(\kappa)] \leq \inf_\lambda \max_{\kappa \geq 0} [F(\kappa, \lambda) - \sigma\psi(\kappa)]\end{aligned}$$

forcing equality. Finally (for $B > 0$ sufficiently large)

$$\begin{aligned}
z_\sigma^{LD+} &= \inf_\lambda z_\sigma^{LR+}(\lambda) \\
&= \inf_\lambda \max_{\kappa\in[0,B]} \left[\sup\left\{cx + \lambda(b - Ax) \mid x \in \mathrm{conv}\left(X \cap \{y \mid \|b - Ay\| \le \kappa\}\right)\right\} - \sigma\psi\left(\kappa\right)\right] \\
&= \inf_\lambda \max_{\kappa\in[0,B]} \left[F\left(\kappa, \lambda\right) - \sigma\psi\left(\kappa\right)\right] \\
&= \max_{\kappa\in[0,B]} \left[\inf_\lambda \sup_{x\in X} \left\{cx + \lambda(b - Ax) \mid \|b - Ax\| \le \kappa\right\} - \sigma\psi\left(\kappa\right)\right] \\
&= \max_{\kappa\in[0,B]} \left[\sup\left\{cx \mid Ax = b,\, x \in \mathrm{conv}\,\left(X \cap \{y \mid \|b - Ay\| \le \kappa\}\right)\right\} - \sigma\psi\left(\kappa\right)\right]
\end{aligned}$$

where the last equality comes from the standard argument from the theory of Lagrangian relaxation (see page 328 of [28]). In particular there exists a $\kappa(\sigma) \in [0, B]$ such that

$$\begin{aligned}
z_\sigma^{LD+} &= \sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\left(X \cap \{y \mid \|b - Ay\| \le \kappa(\sigma)\}\right)\} - \sigma\psi\left(\kappa(\sigma)\right) \\
&\le \sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\left(X\right)\} = z^{LD}.
\end{aligned}$$

Now let $G(\kappa) := \sup\left\{cx \mid Ax = b,\, x \in K(\kappa)\right\}$ and note that $\kappa \mapsto G\left(\kappa\right)$ is a non-decreasing function with $G\left(\infty\right) = z^{LD} < +\infty$. As $z_\sigma^{LD+} = \max_{\kappa\in[0,B]}\left[G(\kappa) - \sigma\psi\left(\kappa\right)\right]$ we see that the term $\sigma\psi\left(\kappa\right)$ penalizes large values of $\kappa \ge 0$. Suppose the solution $\kappa(\sigma)$ of $\max_{\kappa\in[0,B]}\left[G(\kappa) - \sigma\psi\left(\kappa\right)\right]$ satisfies $\kappa(\sigma) \ge \delta > 0$ for all $\sigma \ge 0$. Then

$$\begin{aligned}
-\infty < z^{IP} &\le \lim_{\sigma\to\infty} z_\sigma^{LD+} = \lim_{\sigma\to\infty} \max_{\kappa\ge 0}\left[G(\kappa) - \sigma\psi\left(\kappa\right)\right] = \lim_{\sigma\to\infty} \max_{\kappa\ge\delta}\left[G(\kappa) - \sigma\psi\left(\kappa\right)\right] \\
&\le \lim_{\sigma\to\infty} \max_{\kappa\ge\delta}\left[G(\infty) - \sigma\psi\left(\kappa\right)\right] = \lim_{\sigma\to\infty} \left[z^{LD} - \sigma\psi\left(\delta\right)\right] = -\infty,
\end{aligned}$$

which is a contradiction. Thus the solution $\kappa(\sigma)$ tends to zero as σ tends to infinity.

More can be said about the behavior of $\kappa\left(\sigma\right)$ using the property $\mathrm{diam}\,\mathrm{lev}_\delta\,\psi \downarrow 0$ as $\delta \downarrow 0$, where $\mathrm{lev}_\delta\,\psi := \{\kappa \mid \psi\left(\kappa\right) \le \delta\}$. As the value function $\max_{\kappa\in[0,B]}\left[G(\kappa) - \sigma\psi\left(\kappa\right)\right]$ is monotonically non-increasing in σ and bounded below by z^{IP} we know that the solution $\kappa\left(\sigma\right)$ must be contained in a level set

$$\left\{\kappa \in [0, B] \mid G(\kappa) - \sigma\psi\left(\kappa\right) \ge z^{IP}\right\} \quad \text{for all} \quad \sigma \ge \bar{\sigma} > 0. \tag{10}$$

Note that $G\left(\cdot\right)$ is dominated by the following function (as $cz \le 0$ for all $z \in \ker\left(A\right) \cap 0^+\,\mathrm{conv}\left(X\right)$):

$$G(\kappa) \le H\left(\kappa\right) := \sup\left\{cx \mid Ax = b,\, x \in C(\kappa)\right\}.$$

By the Berge maximum principle [1] we have $\kappa \mapsto H\left(\kappa\right)$ upper semi-continuous. Let M be the maximum of $H\left(\kappa\right)$ over $\kappa \in [0, B]$. Then by (10) we have

$$\kappa\left(\sigma\right) \in \left\{\kappa \in [0, B] \mid \frac{M - z^{IP}}{\sigma} \ge \psi\left(\kappa\right)\right\}.$$

Consequently for any $\delta > 0$, there is a σ_δ such that for all $\sigma > \sigma_\delta$, $\kappa\left(\sigma\right) \le \mathrm{diam}\,\mathrm{lev}_\delta\,\psi \le \delta$. ■

We also consider the "full" augmented Lagrangian dual $z^{LD*} = \inf_{\sigma\ge 0,\lambda} z_\sigma^{LR+}(\lambda)$.

Proposition 2

$$z^{LD*} = \sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\left(X \cap \{y \mid \|b - Ay\| \le 0\}\right)\} = z^{IP}.$$

Proof. From the previous result we have

$$\begin{aligned}
z^{IP} &\le z^{LD*} := \inf_{\sigma\ge 0,\lambda} z_\sigma^{LR+}(\lambda) \\
&= \min_{\sigma\ge 0} \inf_\lambda z_\sigma^{LR+}(\lambda) = \min_{\sigma\ge 0} \max_{\kappa\ge 0} \left[\sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\,\left(X \cap \{y \mid \|b - Ay\| \le \kappa\}\right)\} - \sigma\psi\left(\kappa\right)\right] \\
&\le \max_{\kappa\ge 0} \left[\sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\,\left(X \cap \{y \mid \|b - Ay\| \le \kappa\}\right)\} - \sigma\psi\left(\kappa\right)\right] \quad \text{for all } \sigma \ge 0 \\
&\to \sup\{cx \mid Ax = b,\, x \in \mathrm{conv}\,\left(X \cap \{y \mid Ay = b\}\right)\} \quad \text{as } \sigma \to \infty \\
&= \sup\{cx \mid x \in \mathrm{conv}\,\left(X \cap \{y \mid Ay = b\}\right)\} = \sup\{cx \mid x \in X \text{ and } Ax = b\} = z^{IP}.
\end{aligned}$$

■

We now make the observation that when X constrains us to optimize over a finite (but large) set of discrete elements then the parameter σ need not go to infinity.

Corollary 1 *Suppose that X has a finite number of elements (i.e. $X = \{x_1, \ldots . x_M\}$), in which case we say X is* finite. *Then there exist positive constants $\underline{\kappa}$ and Δ (dependent only on the constraint data in (IP)) such that for all $\sigma \geq \frac{\|c\|\Delta}{\psi(\underline{\kappa})}$, taking $\kappa(\sigma) = 0$ will suffice, and hence $z^{IP} = z_\sigma^{LD+}$.*

Proof. Recall $K(\kappa) := \text{conv}\ (X \cap \{y \mid \|b - Ay\| \leq \kappa\})$ and $K(0) = \text{conv}\ (X \cap \{y \mid Ay = b\})$, the convex hull of the (IP) feasible set. Let $E(\kappa)$ denote the set of extreme points of $\{y \mid Ay = b\} \cap K(\kappa)$, which must be a bounded polyhedron if X is finite. Hence $E(\kappa)$ is also finite. Clearly under the condition that X is finite, for all κ sufficiently small, say below $\underline{\kappa} > 0$, it must be that $K(\kappa) = K(0)$, and hence $E(\kappa) = E(0)$. There must also be a second threshold, $\hat{\underline{\kappa}} \geq \underline{\kappa}$, so that for $\kappa < \hat{\underline{\kappa}}$, it must be that $E(\kappa) = E(0)$, and otherwise, for all $\kappa \geq \hat{\underline{\kappa}}$, $E(\kappa) \neq E(0)$[3]. In the latter case the distance from any point in $E(\kappa) \setminus E(0)$ to $K(0)$ is bounded below by a constant, which we denote by Δ. So if $d\,(x, K(0)) := \min\{\|x - x'\| \mid x' \in K(0)\}$, then we define

$$\Delta := \min\{d\,(x, K(0)) \mid x \in \cup_{\kappa \geq \hat{\underline{\kappa}}} E(\kappa) \setminus E(0)\} > 0.$$

Define

$$\mathcal{K}(\sigma) := \arg\max_{\kappa \geq 0} \left[\max\{cx \mid Ax = b,\, x \in \text{conv}\ (X \cap \{y \mid \|b - Ay\| \leq \kappa\})\} - \sigma\psi\,(\kappa)\right].$$

Let $\sigma > 0$ be such that taking $\kappa(\sigma) = 0$ will *not* suffice, i.e. such that $0 \notin \mathcal{K}(\sigma)$, and let x^* be an optimal solution of the (IP). Then for any $\kappa^* \in \mathcal{K}(\sigma)$, it must be that $\kappa^* \geq \underline{\kappa}$ and

$$x^* \notin \arg\max\{cx \mid Ax = b,\, x \in \text{conv}\ (X \cap \{y \mid \|b - Ay\| \leq \kappa^*\})\}.$$

Let

$$y^* \in \arg\max\{cx \mid Ax = b,\, x \in \text{conv}\ (X \cap \{y \mid \|b - Ay\| \leq \kappa^*\})\}$$

be chosen to be an extreme point solution, i.e. in $E(\kappa^*)$ (possible since the objective is linear, and $\{y \mid Ay = b\} \cap K(\kappa^*)$ is a bounded polyhedron). Clearly $y^* \notin E(0)$ and $\|y^* - x^*\| \geq \Delta$. Furthermore

$$\begin{aligned} & cy^* - \sigma\psi(\kappa^*) > cx^* - \sigma\psi(0) \\ \Rightarrow\ & \sigma\psi(\kappa^*) < c(y^* - x^*) \leq \|c\|\,\|y^* - x^*\| \\ \Rightarrow\ & \sigma\psi(\underline{\kappa}) < \|c\|\,\Delta, \end{aligned}$$

since ψ is monotonically increasing, and thus $\sigma < \frac{\|c\|\Delta}{\psi(\underline{\kappa})}$. We conclude that for any $\sigma \geq \frac{\|c\|\Delta}{\psi(\underline{\kappa})}$, it must be that $0 \in \mathcal{K}(\sigma)$, i.e. we may take $\kappa(\sigma) = 0$, and the result follows immediately from Theorem 1.

■

3 Comparison with Previous Duality Results

As noted in the introduction, there now exists a rich theory of duality for very general continuous optimisation problems [5, 6]. As integrality can be modeled using concave constraints or penalties it is possible to use these results to deduce strong dual for IPs. In this paper, we have not so far (other than in Corollary 1) made the assumptions necessary to ensuring attainment of optimal solutions; they were not required for Theorem 1 nor for Proposition 2. The attainment of solutions appears to be the assumption that has to be made in order to connect the approach we use to those of [5] and [6]. Let us consider the main result of [5]. The general problem (P) is defined to be:

$$v(P) := \max_{x \in X_0} \quad f_0\,(x)$$

where X_0 is a closed set (that models the constraints) and $f_0 : \mathbf{R}^n \to \mathbf{R}$ is a continuous function. In [5, Example 3] this set is chosen as $X_0 := \{x \in \mathbf{R}^n \mid \varphi\,(x) = 0\}$ and allows the inclusion of concave

[3]In the special case that $E(\kappa) = E(0)$ for all κ, it is obvious from Theorem 1 that $\kappa(\sigma) = 0$ will suffice for *all* σ.

constraints to enforce the integrality condition. It has been long recognized in IP that continuous optimization over concave constraints can be used to represent the integrality condition. Thus binary problems and all IPs can be modelled within this framework.

Continuing the derivation of [5], define $q(\lambda) := \max_{x\in\mathbf{R}^n} L(x,\lambda)$ for a given "Lagrangian function" $L : \mathbf{R}^n \times \Lambda \to \mathbf{R}$. Then the dual (D) problem of (P) is defined to be

$$v(D) := \min_{\lambda\in\Lambda} q(\lambda).$$

To apply this to $z^{IP} = \sup\{cx : Ax = b,\ x \in X\}$ where the integrality constraints are in X, take $\varphi(x) = (Ax - b, d(x,X))$ so that

$$X_0 := \{x \in \mathbf{R}^n \mid Ax = b,\ d(x,X) = 0\} = \{x \in X \mid Ax = b\},$$

where $d(x,Y) := \inf\{||x-y|| \mid y \in Y\}$ for any set $Y \subseteq \mathbf{R}^n$.

Strong duality is established in [5] under the following set of assumptions.

(H_1): $f_0(x) = L(x,\lambda)$ for all $x \in X_0$ and $\lambda \in \Lambda$.

$(H_2(\Lambda_0))$: There exists a subset $\Lambda_0 \subseteq \Lambda$ such that for all $\alpha > v(P)$ and $\delta > 0$

$$\inf_{\lambda\in\Lambda_0}\left[\sup_{\substack{x\in X\\ d(x,X_0)\geq\delta}} L(x,\lambda)\right] < \alpha.$$

$(H_3(\Lambda_0))$: For the same subset $\Lambda_0 \subseteq \Lambda$ we have

$$f_0(x) \geq L(x,\lambda) \quad \text{for all } x \in X,\ \lambda \in \Lambda_0.$$

(A_1): For all $\alpha < v(P)$, the level set $\{z \in \mathbf{R}^n \mid f_0(z) \geq \alpha\}$ is compact.

The condition (A_1) is an unnatural assumption in IP, but if one assumes that the IP attains an optimal solution then one may introduce into X_0 a box constraint on the variables that is large enough to contain this optimal solution. Then (A_1) may be replaced by an assumption that X_0 is compact, i.e. for large enough L we have

$$X_0 = \{x \in X \mid Ax = b,\ -L \leq x_i \leq L, \text{ for } i = 1,\ldots,n\}.$$

As noted earlier, we did not need assurance of the attainment of a solution to obtain our results in Theorem 1 and Proposition 2, however it appears to be needed here in order to apply the result of [5]. Continuing with this assumption one can see that the Lagrangian function

$$L(x,(\mu,\sigma)) := cx + \mu(b - Ax) - \sigma\rho(b - Ax)$$

along with the sets

$$\Lambda := \{\lambda = (\mu,\sigma) \mid \mu \in \mathbf{R}^m \text{ and } \sigma \geq 0\} \quad \text{and} \quad \Lambda_0 := \{0\} \times \mathbf{R}_+$$

satisfy the assumptions (H_1)–(A_1) as follows.

1. (H_1): For $x_0 \in X_0$, $L(x_0,(\mu,\sigma)) = cx_0 = f_0(x_0)$.
2. $(H_3(\Lambda_0))$: For all $x \in X$, $(\mu,\sigma) \in \Lambda_0$ we have
$$\begin{aligned} L(x,(\mu,\sigma)) &:= cx + 0(b - Ax) - \sigma\rho(b - Ax) \\ &= cx - \sigma\rho(b - Ax) \leq cx = f_0(x) \end{aligned}$$
3. $(H_2(\Lambda_0))$: This may now be established exactly as in [5, Example 3]. Note that (A_1) is critical for this argument and can be replaced by the assumption of compactness of X_0.

On application of [5, Theorem 2.2] we have $v(P) = v(D)$ or

$$\begin{aligned} z^{LD*} &= \inf_{\sigma\geq 0,\lambda} z_\sigma^{LR+}(\lambda) = \inf_{\sigma\geq 0,\lambda} \max_{x\in X} [cx + \mu(b - Ax) - \sigma\rho(b - Ax)] \\ &= \max_{x\in X_0} cx = \max\{cx \mid Ax \leq b,\ x \in X\} = z^{IP} \end{aligned}$$

where the last equality hold under the assumption that the optimal solution $\bar{x}$ exists and

$$\bar{x} \in \{x \mid -L < x_i < L, \text{ for } i = 1,\ldots,n\}.$$

The intermediate results we have established, in particular, Theorem 1, are not deducible in this way and as a consequence neither is Corollary 1.

4 Conclusions

We hope that the results we have presented here will stimulate interest in, and provide insights to support, the development of augmented Lagrangian algorithms for IP. One step in that direction may have already occurred, in the form of column generation stabilization techniques [27]: the primal form of the stabilization allows a controlled violation of the constraints relaxed in a corresponding Lagrangian dual problem. However in these settings, the control occurs outside the convexification procedure: it would be interesting to see how this could be moved inside it, to provide a column generation analogue for the augmented Lagrangian dual.

References

[1] G. Beer (1993). *Topologies on Closed and Closed Convex Sets*, Mathematics and its Applications Vol. 268, Kluwer Acad. Publ.

[2] P. Benchimol, G. Desaulniers and J. Desrosiers, “Stabilized dynamic constraint aggregation for solving set partitioning problems”, *Euro. J. Operational Research* (2012), to appear.

[3] C.E. Blair and R. Jeroslow, “The value function in mixed integer programming 1”, *Discrete Mathematics*, (1977) 19:121-138.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers”, *Foundations and Trends in Machine Learning*, (2010) 3:1-122.

[5] R. Burachik and A. Rubinov, “On the absence of duality gap for Lagrangian-type functions”, *Journal of Industrial and Management Optimisation*, (2005) 1:33-38.

[6] R. Burachik and A. Rubinov, “Abstract convexity and augmented Lagrangians”, *SIAM J. Opt.*, (2007) 18:413-436.

[7] R. Burachik and Y.C. Kaya, “A deflected subgradient algorithm using a general augumented Lagrangian duality with implications on penalty methods”, In: R.S. Burachik and J-C Yao (eds.), *Variational Analysis and Generalized Differentiation in Optimization and Control*, Springer Optim. Appl., 47, Springer, New York, (2010) 109–132.

[8] G. Desaulniers, J. Desrosiers and M.M. Solomon, *Column Generation*, Springer, 2005.

[9] J. Eckstein and D.P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators”, *Math. Progr.*, (1992) 55:293-318.

[10] I. Elhallaoui, A. Metrane, F. Soumis and G. Desaulniers, “Multi-phase dynamic constraint aggregation for set partitioning type problems”, *Math. Progr.*, (2010) 123(2):345-370.

[11] R. D. Galvão, L. G. A. Espejo and B. Boffey, “A comparison of Lagrangian and surrogate relaxations for the maximal covering location problem”, *European Journal of Operational Research*, (2000) 124:377-389.

[12] A.M. Geoffrion, “Lagrangian relaxation for integer programming”, *Math. Progr. Studies*, (1974) 2:82-114.

[13] F. Glover, “A multiphase-dual algorithm for the zero-one integer programming problem”, *Operations Research*, 1965 13:879-919.

[14] F. Glover, “Surrogate constraint duality in mathematical programming”, *Operations Research*, 1975 23:434-451.

[15] R. Gomory, “Some polyhedra related to combinatorial problems”, *Linear Algebra and Applications* (1969) 2:451-558.

[16] H. J. Greenberg and W. P. Pierskalla, “Surrogate mathematical programming”, *Operations Research*, 1970 18:924-939.

[17] J.N. Hooker, “Inference duality as a basis for sensitivity analysis”, in E.C. Freudern (ed.), *Principles and Practice of Constraint Programming-CP96*, LNCS Vol. 1118, Springer, 1996, pp. 224-236.

[18] E. Johnson, “Cyclic groups, cutting planes and shortest path”, in: T. Hu and S.Robinson (Eds.), *Mathematical Programming*, Academic Press, New York, 1973, pp. 185-211.

[19] E. Johnson, *Integer Programming: Facets, Subadditivity, and Duality for Group and Semi-group Problems*, Society for Industrial and Applied Mathematics, 1980.

[20] M. Karwan and R. Rardin, “Surrogate dual multiplier search procedures in integer programming”, *Operations Research*, (1984) 32(1):52-69.

[21] M. Karwan and R. Rardin, “Some relationships between Lagrangian and surrogate duality in integer programming”, *Mathematical Programming*, (1979), 17(1):320-334.

[22] D. Klabjan, “Subadditive approaches in integer programming”, *Euro. J. Oper. Res.*, (2007) 183:525-545.

[23] J. B. Lasserre, “Generating functions and duality for integer programs”, *Discrete Optimisation*, (2005) 2(1):167–187.

[24] J.B. Lasserre, “Integer programming, duality and superadditive functions”, *Contemporary Mathematics*, (2005) 374:139–150.

[25] D. Li and X. Sun, *Nonlinear Integer Programming*, International Series in Operations Resarch & Management, Springer, 2006.

[26] D. Li, J. Wang and X. L. Sun, “Computing exact solution to nonlinear integer programming: convergent Lagrangian and objective level cut method”, *Journal of Global Optimization*, (2007) 39:127-154.

[27] O. du Merle, D. Villeneuve, J. Desrosiers and P. Hansen, “Stabilized column generation”, *Discrete Mathematics*, (1999) 194(1–3):229–237.

[28] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization,* Wiley-Interscience, 1988.

[29] R.T. Rockafellar, “Augmented Lagrangians and applications of the proximal point algorithm in convex optimization”, *Math. Oper. Res.*, (1976) 1:97–116.

[30] M. Sinclair, “Augmented Lagrangean relaxations in general mixed integer programming”, *Quaestiones Math.*, (1978/79) 3:115-146.

[31] M. Sion, “On general minimax theorems”, *Pac. J. Math.*,(1958) 8:171-176.

[32] L. A. Wolsey, “Integer programming duality: price functions and sensitivity analysis”, *Mathematical Programming*, (1981) 20:173-195.