
Exact Augmented Lagrangian Duality for Mixed Integer Linear Programming

Mohammad Javad Feizollahi · Shabbir
Ahmed · Andy Sun

Submitted: August 13, 2015

Abstract We investigate the augmented Lagrangian dual (ALD) for mixed integer linear programming (MIP) problems. ALD modifies the classical Lagrangian dual by appending a nonlinear penalty function on the violation of the dualized constraints in order to reduce the duality gap. We first provide a primal characterization for ALD for MIPs and prove that ALD is able to asymptotically achieve zero duality gap when the weight on the penalty function is allowed to go to infinity. This provides an alternative characterization and proof of a recent result in Boland and Eberhard [7, Proposition 3]. We further show that, under some mild conditions, ALD using any norm as the augmenting function is able to close the duality gap of an MIP with a *finite* penalty coefficient. This generalizes the result in [7, Corollary 1] from pure integer programming problems with bounded feasible region to general MIPs. We also present an example where ALD with a quadratic augmenting function is not able to close the duality gap for any finite penalty coefficient.

Keywords mixed integer linear programming; lagrangian duality; penalty functions

M. J. Feizollahi, Corresponding author
Robinson College of Business, Georgia State University, Atlanta, GA 30303, USA.
Tel.: +1-404-413-7046
E-mail: mfeizollahi@gsu.edu,

S. Ahmed
School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.
E-mail: sahmed@isye.gatech.edu

A. Sun
School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.
E-mail: andy.sun@isye.gatech.edu

1 Introduction

We consider the general mixed integer (linear) programming (MIP) problem

$$z^{\text{IP}} := \inf\{\mathbf{c}^\top \mathbf{x} \mid \mathbf{Ax} = \mathbf{b}, \mathbf{x} \in X\}, \quad (1)$$

and its augmented Lagrangian dual (ALD)

$$z_\rho^{\text{LD}^+} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^n} \inf_{\mathbf{x} \in X} \{\mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) + \rho\psi(\mathbf{b} - \mathbf{Ax})\},$$

where X is a mixed integer linear set, ρ is a given positive scalar, and $\psi(\cdot)$ is an *augmenting function* with $\psi(\mathbf{0}) = 0$ and $\psi(\mathbf{u}) > 0$ for all $\mathbf{u} \neq \mathbf{0}$. Here, $\mathbf{Ax} = \mathbf{b}$ are the complicating constraints, and relaxing these makes the remaining problem easier.

In contrast to the convex setting, for nonconvex optimization problems, a non-zero duality gap may exist when certain constraints are relaxed by using classical Lagrangian dual (LD). ALD modifies classical LD by appending a nonlinear penalty on the violation of the dualized constraints. The resulting ALD problem then involves dual functions which are not necessarily affine as in LD, and may be capable of penetrating possible ‘dents’ in the value function (or perturbation function) thereby reducing the duality gap [26]. Depending on the properties of the value function of the underlying optimization problem, various different forms of ALD approaches have been introduced (cf. [1, 9, 10, 11, 12, 18, 19, 22, 23, 25, 26, 27, 28, 29, 30, 32, 34]). Under certain conditions, a *zero duality gap* can be reached asymptotically by increasing the coefficient on penalty function to infinity [32]. In some cases, the duality gap can be closed with a large enough *finite* value of the penalty coefficient. In this case, we say that the corresponding ALD involves *exact penalization* or is *exact*. Rockafellar [25] and Bertsekas [2] used convex quadratic augmenting functions. Burke [13, 14] used norms as convex augmenting functions. For these cases, necessary and sufficient conditions for exact penalization are provided in [13, 14, 26] which we will review in Section 2. For some classes of non-convex optimization problems, the duality gap cannot be closed by using convex augmenting functions. For these problems, more general forms of ALD are needed. For example, *level-bounded* augmenting functions were used in [18] rather than convex ones. The works in [29] and [30] used a family of augmenting functions with the *almost peak at zero* property, which includes the augmenting functions in [18] as special cases. Note that the class of augmenting functions in [29] and [30] are generalizations of convex augmenting functions in [26]. A weaker peak at zero property was considered in [22]. A more general form of peak at zero property was investigated in [32] to provide a unified nonlinear ALD. Using abstract convexity, ALD was studied in [12] and [9] in Banach and Hausdorff topological spaces, respectively.

In this paper, we consider non-negative level bounded augmenting functions in ALD for solving MIPs. Because of the non-convexity in MIP (1), a non-zero duality gap may exist [33], that is $z^{\text{IP}} - z_\rho^{\text{LD}^+} > 0$. Recently, Boland

and Eberhard [7] showed that in ALD for MIPs, with a specific class of nonnegative convex augmenting functions, $\lim_{\rho \rightarrow \infty} z_{\rho}^{\text{LD}^+} = z^{\text{IP}}$. They also proved that if X is a finite set (e.g. a bounded pure IP), then there exists a finite penalty coefficient which closes the duality gap. We significantly generalize the results of [7]. In particular, our contributions are as follows:

1. We first provide a primal characterization for the ALD of an MIP. This is an alternative characterization to the one provided in [7, Theorem 1]. Using this characterization, the ALD of an MIP can be viewed as a traditional LD in a lifted space.
2. We give an alternative proof for the asymptotic zero duality gap property of ALD for MIPs when the penalty coefficient is allowed to go to infinity. This was first proved in [7, Proposition 3].
3. We prove that ALD using any norm as the augmenting function with a sufficiently large but finite penalty coefficient closes the duality gap for general MIPs. This generalizes the result in [7, Corollary 1] from the case of pure integer programming with a bounded feasible region to general MIPs with unbounded feasible regions.
4. Using our primal characterization, we also present an example where ALD with a quadratic augmenting function is not able to close the duality gap for any finite penalty coefficient.

The paper is organized as follows. Section 2 provides definitions and surveys existing results on Lagrangian relaxation and augmented Lagrangian relaxation of general nonlinear optimization problems and specifically of MIPs. Section 3 presents a primal characterization of the ALD of a general MIP and the zero duality gap property when the penalty coefficient is allowed to go to infinity. Section 4 proves that under mild conditions the ALD achieves zero duality gap using any norm as the augmenting function with a finite penalty coefficient.

2 Preliminaries

Let \mathbb{R} , \mathbb{Q} , and \mathbb{Z} denote the sets of real, rational, and integer numbers, respectively. For any vector \mathbf{a} and matrix \mathbf{A} with finite dimensions, denote their transpose by \mathbf{a}^{\top} and \mathbf{A}^{\top} , respectively. For any set $S \subseteq \mathbb{R}^n$, let $\mathbf{conv}(S)$, $\mathbf{ri}(S)$ and $\mathbf{cl}(S)$ denote the convex hull, relative interior, and closure of the set S , respectively. Moreover, let $\mathbf{diam}(S) := \sup\{\|\mathbf{u} - \mathbf{v}\|_{\infty} : \mathbf{u} \in S, \mathbf{v} \in S\}$ denote the diameter of set S , where $\|\cdot\|_{\infty}$ is the l_{∞} norm.

Let $\mathbf{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}$ be the vector of decision variables, where n_1 and n_2 are numbers of integer and continuous variables, respectively, and $n := n_1 + n_2$. For given $\mathbf{c} \in \mathbb{Q}^n$, $\mathbf{b} \in \mathbb{Q}^m$, and $\mathbf{A} \in \mathbb{Q}^{m \times n}$, consider the general MIP problem (1),

$$z^{\text{IP}} := \inf\{\mathbf{c}^{\top} \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in X\},$$

where m is the number of *complicating* or *coupling constraints*, $\mathbf{A}\mathbf{x} = \mathbf{b}$. The case with $n_2 = 0$ is called a pure IP, while for a MIP we have $n_2 \geq 1$ and

$n_1 \geq 1$. Denote the LP relaxation of z^{IP} in problem (1) by z^{LP} . We consider MIP problems that satisfy the following assumption.

Assumption 1 For the MIP (1) we have the following:

- a) X is a mixed integer linear set given by $X = \{\mathbf{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} : \mathbf{E}\mathbf{x} \leq \mathbf{f}\}$ for some $\mathbf{E} \in \mathbb{Q}^{\bar{m} \times n}$ and $\mathbf{f} \in \mathbb{Q}^{\bar{m}}$, where \bar{m} is the number of the inequality constraints in the definition of X . The problem data \mathbf{A} , \mathbf{b} , \mathbf{c} , \mathbf{E} , and \mathbf{f} all have rational entries, and without loss of generality, we can assume that they are integral.
- b) Problem (1) is feasible and its optimal value is bounded.

Usually problem (1) is taken to be structured so that X includes integrality constraints, simple bounds on variables, and other simple constraints.

Remark 1 Note that under Assumption 1-a, $\text{conv}(X)$ and $\text{conv}(\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in X\})$ are rational polyhedra by Meyer's theorem [20]. By Assumption 1 (rationality of input data and boundedness of z^{IP}), the value of the LP relaxation of MIP (1) is bounded [5], i.e. $-\infty < z^{\text{LP}} \leq z^{\text{IP}} < \infty$. Let $\bar{\boldsymbol{\lambda}}_{\text{LP}}$ be a rational optimal vector of dual variables for $\mathbf{A}\mathbf{x} = \mathbf{b}$ in the LP relaxation of (1). Moreover, z^{IP} is attainable and the inf in the objective function of (1) can be replaced by min. That is, there exists an *optimal solution* \mathbf{x}^* of problem (1) such that $\mathbf{x}^* \in X$, $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ and $z^{\text{IP}} = \mathbf{c}^\top \mathbf{x}^*$.

It is worth mentioning that the equality relation in $\mathbf{A}\mathbf{x} = \mathbf{b}$ does not impose any restriction on the type of these constraints. Because any inequality can be replaced by an equality constraint with a new non-negative slack variable. The non-negativity of the introduced variable can be absorbed in X . Moreover, in the case of a pure IP, this slack variable will automatically be an integer variable following Assumption 1-a.

Definition 1 (Value function). The *value function* for problem (1) is defined as

$$p(\mathbf{u}) := \inf\{\mathbf{c}^\top \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} + \mathbf{u}, \mathbf{x} \in X\}. \quad (2)$$

Note that $p(\mathbf{0}) = z^{\text{IP}}$. The value function is a very important tool for the theoretical examination of constrained optimization problems [29]. The properties of the value functions for IPs and MIPs were studied in [4, 5, 6, 21, 24]. For an MIP problem with rational data, the value function is lower semicontinuous [21] and piecewise polyhedral with finitely many pieces in any bounded set [4].

Definition 2 (Lagrangian relaxation and dual). For a given Lagrange multiplier vector $\boldsymbol{\lambda} \in \mathbb{R}^m$, the corresponding Lagrangian relaxation (LR) of (1) is given as

$$z^{\text{LR}}(\boldsymbol{\lambda}) := \inf_{\mathbf{x} \in X} \left\{ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) \right\}, \quad (3)$$

and the associated Lagrangian dual (LD) is

$$z^{\text{LD}} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} z^{\text{LR}}(\boldsymbol{\lambda}). \quad (4)$$

A well known primal characterization of LD is given by [17] as

$$z^{\text{LD}} = \inf_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} \mid \mathbf{Ax} = \mathbf{b}, \mathbf{x} \in \text{conv}(X) \}. \quad (5)$$

Remark 2 Note that by rationality of the input data in Assumption 1, z^{LD} is attainable and \inf in the objective function of (5) can be replaced by \min .

Definition 3 (Augmented Lagrangian relaxation and dual). The augmented Lagrangian relaxation (ALR) of (1) has the following form [26]:

$$z_\rho^{\text{LR}+}(\boldsymbol{\lambda}) := \inf_{\mathbf{x} \in X} \{ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) + \rho \psi(\mathbf{b} - \mathbf{Ax}) \}. \quad (6)$$

Here, $\rho > 0$ is a fixed given parameter called *penalty coefficient* and ψ is an augmenting function. In this paper, unless explicitly mentioned, we assume that ψ satisfies the following assumption.

Assumption 2 $\psi : \mathbb{R}^m \rightarrow \mathbb{R}_+$ is a proper, nonnegative, lower semicontinuous, and level-bounded augmenting function, that is $\psi(\mathbf{0}) = 0$, $\psi(\mathbf{u}) > 0$ for all $\mathbf{u} \neq \mathbf{0}$, $\text{diam}\{\mathbf{u} \mid \psi(\mathbf{u}) \leq \delta\} < +\infty$ for all $\delta > 0$. Moreover $\lim_{\delta \downarrow 0} \text{diam}\{\mathbf{u} \mid \psi(\mathbf{u}) \leq \delta\} = 0$.

Note that non-negative convex augmenting functions satisfy Assumption 2. The augmented Lagrangian dual (ALD) is as follows.

$$z_\rho^{\text{LD}+} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} z_\rho^{\text{LR}+}(\boldsymbol{\lambda}). \quad (7)$$

Since the augmenting function $\psi(\cdot)$ is nonnegative, $z_\rho^{\text{LD}+}$ is a non-decreasing function of ρ . Moreover, since $\rho \psi(\mathbf{b} - \mathbf{Ax}) \geq 0$ we have $z^{\text{LR}}(\boldsymbol{\lambda}) \leq z_\rho^{\text{LR}+}(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\rho > 0$. Therefore, $z^{\text{LD}} \leq z_\rho^{\text{LD}+}$ for any $\rho > 0$. Moreover

$$\begin{aligned} z_\rho^{\text{LR}+}(\boldsymbol{\lambda}) &\leq \inf_{\mathbf{x} \in X: \mathbf{Ax}=\mathbf{b}} \left\{ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) + \rho \psi(\mathbf{b} - \mathbf{Ax}) \right\} \\ &= \inf_{\mathbf{x} \in X: \mathbf{Ax}=\mathbf{b}} \mathbf{c}^\top \mathbf{x} \\ &= z^{\text{IP}}, \end{aligned}$$

where the inequality holds because $\{\mathbf{x} \in X : \mathbf{Ax} = \mathbf{b}\} \subseteq X$. The first equality follows from $\psi(\mathbf{0}) = 0$. The second equation holds by definition of z^{IP} . Thus, $z_\rho^{\text{LD}+} \leq z^{\text{IP}}$. In summary, for all $\rho > 0$,

$$-\infty < z^{\text{LP}} \leq z^{\text{LD}} \leq z_\rho^{\text{LD}+} \leq z^{\text{IP}} < +\infty, \quad (8)$$

where the strict inequalities in the upper and lower bounds hold from Assumption 1.

2.1 Exact Penalty Representation

Definition 4 (Exact penalty representation [26, Definition 11.60]). A dual vector $\bar{\lambda}$ is said to support an *exact penalty representation* for problem (1) if, for all ρ sufficiently large, $z^{\text{IP}} = z_{\rho}^{\text{LR}+}(\bar{\lambda})$ and

$$\operatorname{argmin}_{\mathbf{x} \in X: \mathbf{Ax}=\mathbf{b}} \mathbf{c}^{\top} \mathbf{x} = \operatorname{argmin}_{\mathbf{x} \in X} \{ \mathbf{c}^{\top} \mathbf{x} + \bar{\lambda}^{\top} (\mathbf{b} - \mathbf{Ax}) + \rho \psi(\mathbf{b} - \mathbf{Ax}) \}.$$

The smallest ρ which satisfies this property is called the *adequate penalty threshold*. A criterion for the exact penalty representation presented in [26] is as follows. Suppose that $z_{\rho}^{\text{LR}+}(\bar{\lambda}) > -\infty$ for at least one $(\bar{\lambda}, \rho) \in \mathbb{R}^m \times (0, \infty)$. Then, a vector $\bar{\lambda}$ supports an exact penalty representation for problem (1) if and only if there exist an open neighborhood $W \subset \mathbb{R}^m$ of $\mathbf{0}$ and a scalar $\hat{\rho} > 0$ such that

$$p(\mathbf{u}) \geq p(\mathbf{0}) - \bar{\lambda}^{\top} \mathbf{u} - \hat{\rho} \psi(\mathbf{u}), \quad \forall \mathbf{u} \in W.$$

If $z_{\rho}^{\text{LD}+} = z^{\text{IP}}$ for some $\rho > 0$, then ALR (6) can recover a primal solution for the MIP problem (1).

Proposition 1 *Suppose Assumption 1 holds and $z^{\text{IP}} = z_{\hat{\rho}}^{\text{LD}+} = z_{\hat{\rho}}^{\text{LR}+}(\bar{\lambda})$ for some finite $\hat{\rho} > 0$ and $\bar{\lambda} \in \mathbb{R}^m$. Then, any optimal solution of ALR (6) with $\lambda = \bar{\lambda}$ and $\rho = \rho^* > \hat{\rho}$ is an optimal solution of the MIP problem (1), and vice versa. That is, $\bar{\lambda}$ supports an exact penalty representation for the MIP problem (1).*

Proof Let ρ^* be any scalar such that $\rho^* > \hat{\rho}$. Let $\bar{\mathbf{x}}$ be an optimal solution of MIP problem (1) (the existence of an optimal solution for problem (1) is guaranteed under Assumption 1). Then, it holds that $\bar{\mathbf{x}} \in X$, $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$, and $\mathbf{c}^{\top} \bar{\mathbf{x}} = z^{\text{IP}}$. Thus,

$$\mathbf{c}^{\top} \bar{\mathbf{x}} + \bar{\lambda}^{\top} (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) + \rho^* \psi(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = \mathbf{c}^{\top} \bar{\mathbf{x}} = z^{\text{IP}} = z_{\rho^*}^{\text{LR}+}(\bar{\lambda}).$$

where the last equality follows from the facts that $z_{\hat{\rho}}^{\text{LR}+}(\bar{\lambda}) \leq z_{\rho^*}^{\text{LR}+}(\bar{\lambda}) \leq z^{\text{IP}}$ and $z^{\text{IP}} = z_{\rho^*}^{\text{LR}+}(\bar{\lambda})$. Therefore, $\bar{\mathbf{x}}$ solves ALR (6) with ρ^* and $\bar{\lambda}$. Moreover, it shows that the optimality is achieved for this case of ALR (6).

To prove the other side, let $\mathbf{x}^* \in X$ be any optimal solution of ALR (6) with ρ^* and $\bar{\lambda}$, i.e. $\mathbf{c}^{\top} \mathbf{x}^* = z_{\rho^*}^{\text{LR}+}(\bar{\lambda})$. We claim that \mathbf{x}^* solves problem (1), i.e. $\mathbf{x}^* \in X$, $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ and $\mathbf{c}^{\top} \mathbf{x}^* = z^{\text{IP}}$. Note that as a feasible solution of ALR (6), \mathbf{x}^* belongs to X . Assume by contradiction $\mathbf{A}\mathbf{x}^* \neq \mathbf{b}$. Then, $\psi(\mathbf{b} - \mathbf{A}\mathbf{x}^*) > 0$ and therefore

$$\hat{\rho} \psi(\mathbf{b} - \mathbf{A}\mathbf{x}^*) < \rho^* \psi(\mathbf{b} - \mathbf{A}\mathbf{x}^*). \quad (9)$$

Moreover,

$$\begin{aligned} z^{\text{IP}} = z_{\hat{\rho}}^{\text{LD}+} = z_{\hat{\rho}}^{\text{LR}+}(\bar{\lambda}) &\leq \mathbf{c}^{\top} \mathbf{x}^* + \bar{\lambda}^{\top} (\mathbf{b} - \mathbf{A}\mathbf{x}^*) + \hat{\rho} \psi(\mathbf{b} - \mathbf{A}\mathbf{x}^*) \\ &< \mathbf{c}^{\top} \mathbf{x}^* + \bar{\lambda}^{\top} (\mathbf{b} - \mathbf{A}\mathbf{x}^*) + \rho^* \psi(\mathbf{b} - \mathbf{A}\mathbf{x}^*) \\ &= z_{\rho^*}^{\text{LR}+}(\bar{\lambda}), \end{aligned} \quad (10)$$

which contradicts $z_{\rho^*}^{\text{LR}+}(\bar{\boldsymbol{\lambda}})$ being a lower bound for z^{IP} . Therefore, $\mathbf{Ax}^* = \mathbf{b}$. Note that in (10) the equality relations hold by assumption, the first inequality holds by definition of $z_{\rho^*}^{\text{LR}+}(\bar{\boldsymbol{\lambda}})$, and the strict inequality follows from (9). Furthermore,

$$\begin{aligned} z^{\text{IP}} &= z_{\rho^*}^{\text{LR}+}(\bar{\boldsymbol{\lambda}}) \\ &\leq z_{\rho^*}^{\text{LR}+}(\bar{\boldsymbol{\lambda}}) = \mathbf{c}^\top \mathbf{x}^* + \bar{\boldsymbol{\lambda}}^\top (\mathbf{b} - \mathbf{Ax}^*) + \rho^* \psi(\mathbf{b} - \mathbf{Ax}^*) = \mathbf{c}^\top \mathbf{x}^* \quad (11) \\ &\leq z^{\text{IP}}, \end{aligned}$$

where the first two equalities hold by assumption and the third equality follows from $\mathbf{Ax}^* = \mathbf{b}$. Therefore, $\mathbf{c}^\top \mathbf{x}^* = z^{\text{IP}}$ which completes the proof. \square

Two important cases of ALR are the proximal and sharp Lagrangian. Next, we present their definitions, and necessary and sufficient conditions for supporting an exact penalty representation in these cases.

2.2 Proximal Lagrangian

Definition 5 (Proximal Lagrangian). An ALR generated with the augmenting function $\psi(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2$ is called a proximal Lagrangian.

Definition 6 (Proximal subgradient [26, Definition 8.45]). A vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ is called a proximal subgradient of a function $f : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ at $\bar{\mathbf{u}}$, a point where $f(\bar{\mathbf{u}})$ is finite, if there exist $\rho > 0$ and $\delta > 0$ such that

$$f(\mathbf{u}) \geq f(\bar{\mathbf{u}}) - \boldsymbol{\lambda}^\top (\mathbf{u} - \bar{\mathbf{u}}) - \frac{1}{2}\rho\|\mathbf{u} - \bar{\mathbf{u}}\|_2^2, \quad \forall \mathbf{u} \text{ s.t. } \|\mathbf{u} - \bar{\mathbf{u}}\|_2 \leq \delta.$$

The existence of a proximal subgradient at $\bar{\mathbf{u}}$ corresponds to the existence of a ‘local quadratic support’ to f at $\bar{\mathbf{u}}$.

In a proximal Lagrangian, suppose that there exists $(\boldsymbol{\lambda}, \rho) \in \mathbb{R}^n \times (0, \infty)$ such that $z_{\rho}^{\text{LR}+}(\boldsymbol{\lambda}) > -\infty$. Then, a necessary and sufficient condition for a vector $\bar{\boldsymbol{\lambda}}$ to support an exact penalty representation is that $\bar{\boldsymbol{\lambda}}$ is a proximal subgradient of the value function $p(\mathbf{u})$ at $\mathbf{u} = \mathbf{0}$ [26].

2.3 Sharp Lagrangian

Definition 7 (Sharp Lagrangian). An ALR which uses a norm as an augmenting function, i.e. $\psi(\mathbf{u}) = \|\mathbf{u}\|$, is called a sharp Lagrangian.

Definition 8 (Calmness [26, Ch. 8.F]). A function $f : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ is calm at $\bar{\mathbf{u}}$ from below with modulus $\kappa \in \mathbb{R}_+$ if $f(\bar{\mathbf{u}})$ is finite and on some open neighborhood V of $\bar{\mathbf{u}}$, one has

$$f(\mathbf{u}) \geq f(\bar{\mathbf{u}}) - \kappa\|\mathbf{u} - \bar{\mathbf{u}}\|, \quad \forall \mathbf{u} \in V.$$

Consider a function f which is not calm at $\bar{\mathbf{u}}$ from below. Then, a small shift in \mathbf{u} can produce a proportionally unbounded downward shift in f . Calmness is a basic regularity condition under which we can study the sensitivity properties of certain variational systems [15].

In the sharp Lagrangian, suppose that $z_\rho^{\text{LR}+}(\mathbf{0}) > -\infty$ for some $\rho \in (0, \infty)$. Then, a necessary and sufficient condition for the vector $\bar{\boldsymbol{\lambda}} = \mathbf{0}$ to support an exact penalty representation is that the value function $p(\mathbf{u})$ is calm from below at $\mathbf{u} = \mathbf{0}$ [13, 14, 26].

2.4 ALD for MIPs

For the MIP problem (1), under some technical assumptions, Boland and Eberhard [7] showed that the duality gap for ALD, $z_\rho^{\text{LD}+} - z^{\text{IP}}$, goes to zero as the penalty coefficient ρ goes to infinity.

Proposition 2 [7, Proposition 3] *Suppose ψ is of the form $\psi(\mathbf{u}) = \phi(\|\mathbf{u}\|)$ for some norm $\|\cdot\|$ in \mathbb{R}^m where $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a convex, monotonically increasing function for which $\phi(0) = 0$ and there exists $\delta > 0$ for which*

$$\liminf_{a \rightarrow +\infty} \frac{\phi(a)}{a} \geq \delta > 0$$

with $\text{diam}\{\mathbf{a} \mid \phi(\mathbf{a}) \leq \delta\} \downarrow 0$ as $\delta \downarrow 0$. Moreover, at least one of the following conditions holds: 1) The solution set of the LP relaxation of problem (1) does not contain a lineality space. 2) The matrices \mathbf{A} and \mathbf{D} have rational entries and the norm $\|\cdot\|$ used in the definition of ψ is the l_∞ norm. 3) $\text{conv}(X)$ is bounded. Then

$$z^{\text{LD}^*} := \sup_{\rho > 0} z_\rho^{\text{LD}+} = \lim_{\rho \rightarrow \infty} z_\rho^{\text{LD}+} = z^{\text{IP}}.$$

Boland and Eberhard [7] also showed that if X is a finite set of discrete elements then ρ does not need to go to infinity to close duality gap.

Corollary 1 [7, Corollary 1] *Suppose X is a finite set and assumptions in Proposition 2 hold. Then, there exists a ρ^* with $0 < \rho^* < \infty$ such that $z_{\rho^*}^{\text{LD}+} = z^{\text{IP}}$.*

3 Zero Duality Gap with ALD

In this section, we first present a primal characterization of the ALD for MIPs. Then, we prove that strong duality holds for ALD of general MIPs when the penalty coefficient is allowed to go to infinity. Our primal characterization and the strong duality result hold for a general, not necessarily convex augmenting function, satisfying Assumption 2. We also discuss the relation of our results to the recent results in [7].

3.1 A Primal Characterization of ALD

Similar to the equivalence of (4) and (5) for the LD, we can give a primal characterization for the ALD problem (7). The key observation is that (7) can be viewed as an LD of a problem in a lifted space. Then, the primal characterization follows from strong duality in convex optimization with usual regularity conditions.

Let us first find the primal problem for the ALD problem (7).

$$\begin{aligned} z_\rho^{\text{LD}^+} &= \sup_{\lambda \in \mathbb{R}^m} \inf_{\mathbf{x} \in X} \{ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho\psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \} \\ &= \sup_{\lambda \in \mathbb{R}^m} \inf_{\mathbf{x} \in X, \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \omega} \{ \mathbf{c}^\top \mathbf{x} + \rho\omega + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) \} \end{aligned} \quad (12)$$

$$= \sup_{\lambda \in \mathbb{R}^m} \inf_{\mathbf{x}, \omega} \{ \mathbf{c}^\top \mathbf{x} + \rho\omega + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : (\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi) \}, \quad (13)$$

where S_ψ denotes the feasible region of the inf problem in (12), i.e.

$$S_\psi := \{ (\mathbf{x}, \omega) \in \mathbb{R}^{n+1} : \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \omega, \mathbf{x} \in X \}, \quad (14)$$

and (13) holds because the objective function in (12) is linear. Now switching the sup and inf in (13), we have the dual problem of (13) given as

$$\begin{aligned} \hat{z}_\rho^{\text{LD}^+} &:= \inf_{(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi)} \sup_{\lambda \in \mathbb{R}^m} \{ \mathbf{c}^\top \mathbf{x} + \rho\omega + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) \} \\ &= \inf_{\mathbf{x}, \omega} \{ \mathbf{c}^\top \mathbf{x} + \rho\omega : \mathbf{A}\mathbf{x} = \mathbf{b}, (\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi) \}. \end{aligned} \quad (15)$$

Theorem 1 below shows that, under a mild regularity condition, strong duality holds between (13) and (15), i.e. $z_\rho^{\text{LD}^+} = \hat{z}_\rho^{\text{LD}^+}$. Note that (15) only involves primal variables \mathbf{x}, ω . Therefore, this gives a primal characterization of the ALD problem (13). To prove this result, we need a few simple propositions and a nonlinear Farkas lemma.

Proposition 3 $\text{Proj}_x(\mathbf{conv}(S_\psi)) = \mathbf{conv}(X)$.

Proof For any $(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi)$, there exist $\mathbf{x}^i \in X$ and $\psi(\mathbf{b} - \mathbf{A}\mathbf{x}^i) \leq \omega^i$ for $i = 1, \dots, n+2$ so that $\mathbf{x} = \sum_{i=1}^{n+2} \lambda_i \mathbf{x}^i$, $\omega = \sum_{i=1}^{n+2} \lambda_i \omega^i$, and $\sum_{i=1}^{n+2} \lambda_i = 1$ with $\lambda_i \geq 0$ for all $i = 1, \dots, n+2$ (by Caratheodory's Theorem). Clearly, $\mathbf{x} \in \mathbf{conv}(X)$, which shows $\text{Proj}_x(\mathbf{conv}(S_\psi)) \subseteq \mathbf{conv}(X)$.

For the other direction, take any $\mathbf{x} \in \mathbf{conv}(X)$. Then \mathbf{x} can be written as $\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}^i$ for each $\mathbf{x}^i \in X$ and λ_i 's form a convex combination. Let $\omega_i := \psi(\mathbf{b} - \mathbf{A}\mathbf{x}^i)$ and $\omega := \sum_i \lambda_i \omega_i$. Then, for each i , $(\mathbf{x}^i, \omega^i) \in S_\psi$, and $(\mathbf{x}, \omega) = \sum_i \lambda_i (\mathbf{x}^i, \omega^i)$. Therefore, $(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi)$, i.e. $\mathbf{x} \in \text{Proj}_x(\mathbf{conv}(S_\psi))$. This completes the proof. \square

Proposition 4 Let S be a nonempty convex set in \mathbb{R}^{n+1} . Then $\mathbf{ri}(\text{Proj}_x(S)) = \text{Proj}_x(\mathbf{ri}(S))$.

This follows from the well-known fact $\mathbf{ri}(A(S)) = A(\mathbf{ri}(S))$, where A is a linear transformation and S is a convex set. See e.g., [3].

Proposition 5 *There exists $\mathbf{x} \in \mathbf{ri}(\mathbf{conv}(X))$ and $\mathbf{Ax} = \mathbf{b}$ if and only if Problem (15) has a feasible point in $\mathbf{ri}(\mathbf{conv}(S_\psi))$.*

Proof If (15) has a feasible point $(\bar{\mathbf{x}}, \bar{\omega})$ in $\mathbf{ri}(\mathbf{conv}(S_\psi))$, then $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ and $\bar{\mathbf{x}} \in \text{Proj}_x(\mathbf{ri}(\mathbf{conv}(S_\psi)))$. by Proposition 4, $\bar{\mathbf{x}} \in \mathbf{ri}(\text{Proj}_x(\mathbf{conv}(S_\psi)))$. By Proposition 3, we have $\bar{\mathbf{x}} \in \mathbf{ri}(\mathbf{conv}(X))$.

For the other direction, take any $\bar{\mathbf{x}} \in \mathbf{ri}(\mathbf{conv}(X))$ and $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$. By Proposition 3, we have $\bar{\mathbf{x}} \in \mathbf{ri}(\text{Proj}_x(\mathbf{conv}(S_\psi)))$. By Proposition 4, then we know that $\bar{\mathbf{x}} \in \text{Proj}_x(\mathbf{ri}(\mathbf{conv}(S_\psi)))$, i.e. there exists $(\bar{\mathbf{x}}, \bar{\omega}) \in \mathbf{ri}(\mathbf{conv}(S_\psi))$ and $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$. \square

Lemma 1 (Nonlinear Farkas' Lemma (Prop. 3.5.4, [3])) *Let C be a nonempty convex subset of \mathbb{R}^n , and let $f : C \rightarrow \mathbb{R}$ and $g_j : C \rightarrow \mathbb{R}$, for $j = 1, \dots, r$ be convex functions. Consider the set F given by $F = \{\mathbf{x} \in C : g(\mathbf{x}) \leq 0\}$, where $g(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_r(\mathbf{x}))$, and assume that $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in F$. Consider the subset Q^* of \mathbb{R}^r given by*

$$Q^* = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^r : \boldsymbol{\lambda} \geq \mathbf{0}, f(\mathbf{x}) + \boldsymbol{\lambda}^\top g(\mathbf{x}) \geq 0, \forall \mathbf{x} \in C \right\}.$$

Then, Q^ is nonempty if the functions g_j for $j = 1, \dots, r$ are affine, and F contains a relative interior point of C .*

Next, we present the primal characterization of the ALD problem (7) as following theorem.

Theorem 1 *If there exists $\mathbf{x} \in \mathbf{ri}(\mathbf{conv}(X))$ such that $\mathbf{Ax} = \mathbf{b}$, and $z_\rho^{LD+} > -\infty$, then for all $\rho > 0$,*

$$z_\rho^{LD+} = \inf_{\mathbf{x}, \omega} \mathbf{c}^\top \mathbf{x} + \rho \omega \quad (16a)$$

$$\text{s.t. } \mathbf{Ax} = \mathbf{b} \quad (16b)$$

$$(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi). \quad (16c)$$

Proof Essentially, we want to show that strong duality holds between the primal and dual pair of convex programs (13) and (15), i.e. $\hat{z}_\rho^{LD+} = z_\rho^{LD+}$ (recall that \hat{z}_ρ^{LD+} is defined in (15)). By Proposition 5, if there exists $\mathbf{x} \in \mathbf{ri}(\mathbf{conv}(X))$ and $\mathbf{Ax} = \mathbf{b}$, then (16) has a feasible point in $\mathbf{ri}(\mathbf{conv}(S_\psi))$. To apply the nonlinear Farkas' lemma, we first rewrite the linear equality constraints in (16b) as linear inequalities $\tilde{\mathbf{A}}\mathbf{x} \leq \tilde{\mathbf{b}}$ with $\tilde{\mathbf{A}} = [\mathbf{A}^\top, -\mathbf{A}^\top]^\top$ and $\tilde{\mathbf{b}} = [\mathbf{b}^\top, -\mathbf{b}^\top]^\top$; we can also subtract \hat{z}_ρ^{LD+} from the objective function of (16a) so the new optimal value is zero. Furthermore, denote the feasible region of (16) as

$$F := \left\{ (\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi) : \tilde{\mathbf{A}}\mathbf{x} \leq \tilde{\mathbf{b}} \right\}.$$

Since F contains a point in the relative interior of $\mathbf{conv}(S_\psi)$, by Lemma 1, we know that there exists a multiplier vector $\boldsymbol{\lambda}^* \leq \mathbf{0}$ such that

$$\mathbf{c}^\top \mathbf{x} + \rho \omega - \hat{z}_\rho^{LD+} + (\boldsymbol{\lambda}^*)^\top (\tilde{\mathbf{b}} - \tilde{\mathbf{A}}\mathbf{x}) \geq 0, \quad \forall (\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi).$$

From this, we obtain

$$\begin{aligned} & \inf_{(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi)} \mathbf{c}^\top \mathbf{x} + \rho\omega + (\boldsymbol{\lambda}^*)^\top (\tilde{\mathbf{b}} - \tilde{\mathbf{A}}\mathbf{x}) \geq \hat{z}_\rho^{LD+} \\ \Rightarrow z_\rho^{LD+} &= \sup_{\lambda \leq \mathbf{0}} \inf_{(\mathbf{x}, \omega) \in \mathbf{conv}(S_\psi)} \mathbf{c}^\top \mathbf{x} + \rho\omega + \boldsymbol{\lambda}^\top (\tilde{\mathbf{b}} - \tilde{\mathbf{A}}\mathbf{x}) \geq \hat{z}_\rho^{LD+}. \end{aligned}$$

By the weak duality between (13) and (15), we already have $z_\rho^{LD+} \leq \hat{z}_\rho^{LD+}$, therefore, this shows that $z_\rho^{LD+} = \hat{z}_\rho^{LD+}$ for all $\rho > 0$. \square

Remark 3 A similar primal characterization of (7) is given in [7, Theorem 1]. In particular, the primal characterization in [7] has the following form

$$z_\rho^{LD+} = \min_{\hat{\omega} > 0} \left\{ \rho\hat{\omega} + \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in X_\psi(\hat{\omega}) \} \right\}, \quad (17)$$

where $X_\psi(\hat{\omega}) := \mathbf{conv}(\{\mathbf{x} \in \mathbb{R}^n : \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \hat{\omega}, \mathbf{x} \in X\})$. Note that (17) first minimizes over $\hat{\omega}$ then over \mathbf{x} , whereas the primal characterization obtained in (16) minimizes \mathbf{x}, ω jointly. Of course, (16) can also be written in this order as

$$z_\rho^{LD+} = \min_{\hat{\omega} > 0} \left\{ \rho\hat{\omega} + \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in X'_\psi(\hat{\omega}) \} \right\}, \quad (18)$$

where $X'_\psi(\hat{\omega}) := \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}, \hat{\omega}) \in \mathbf{conv}(S_\psi)\}$. The difference between (17) and (16) is more clear if we rewrite the sets $X_\psi(\hat{\omega})$ and $X'_\psi(\hat{\omega})$ as follows,

$$\begin{aligned} X_\psi(\hat{\omega}) &= \text{Proj}_x \left(\mathbf{conv}(S_\psi \cap \{(\mathbf{x}, \omega) : \omega = \hat{\omega}\}) \right) \\ X'_\psi(\hat{\omega}) &= \text{Proj}_x \left(\mathbf{conv}(S_\psi) \cap \{(\mathbf{x}, \omega) : \omega = \hat{\omega}\} \right). \end{aligned} \quad (19)$$

From this, we can see $X_\psi(\hat{\omega}) \subseteq X'_\psi(\hat{\omega})$. In this sense, (17) provides a stronger characterization than (18), when the joint minimization over (\mathbf{x}, ω) is split out in the order of ω and \mathbf{x} . In fact, the proof in [7] that established (17) is quite involved. The difficulty exactly lies in characterizing the properties of the optimal objective value of the inner minimization in (17) as a single variable function in ω . In comparison, our primal characterization (16) bypasses this difficulty by only looking at the joint minimization problem. It seems that this insight to view the ALD as a traditional LD problem in a lifted space is new, which makes the derivation of (16) quite simpler. Our primal characterization also requires less assumptions than (17). In particular, (17) requires that the augmenting function is convex in a particular form and at least one of the three assumptions stated in Proposition 2 hold, whereas *our primal characterization works for both convex and non-convex augmenting functions*, and the relative interior condition in Theorem 1 is a rather mild regularity condition. In addition, as we will show now, Assumptions 1 and 2 are enough to prove the zero duality gap result for ALD of general MIPs. A similar result is also proved in [7, Proposition 3] through their characterization (17), again under more restricted conditions.

3.2 Zero Duality Gap for MIPs

From the primal characterization (16) we can see the $z_\rho^{\text{LD}+}$ is a non-decreasing function of ρ . Since $z_\rho^{\text{LD}+}$ is upper bounded by z^{IP} , therefore we have

$$-\infty < z^{\text{LD}*} := \sup_{\rho > 0} z_\rho^{\text{LD}+} = \lim_{\rho \rightarrow +\infty} z_\rho^{\text{LD}+} \leq z^{\text{IP}}.$$

We want to show that in fact $z^{\text{LD}*} = z^{\text{IP}}$. Recall that $\bar{\lambda}_{\text{LP}}$ is defined as a rational optimal vector of dual variables for $\mathbf{Ax} = \mathbf{b}$ in the LP relaxation of problem (1).

Proposition 6 *Suppose Assumptions 1 and 2 hold. For given $\rho > 0$ and $\epsilon > 0$, define $\omega_{\rho,\epsilon}^*$ as*

$$\begin{aligned} \omega_{\rho,\epsilon}^* &:= \inf_{\mathbf{x}, \omega} \omega \\ \text{s.t. } \mathbf{x} &\in X, \\ \psi(\mathbf{b} - \mathbf{Ax}) &\leq \omega, \\ \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho\omega - z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) &\leq \epsilon. \end{aligned} \quad (20)$$

Then, the limit $\omega_\rho^* := \lim_{\epsilon \downarrow 0} \omega_{\rho,\epsilon}^*$ exists and $\lim_{\rho \rightarrow +\infty} \omega_\rho^* = 0$.

Proof From definition and finiteness of $z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$, and feasibility of problem (1), we know that problem (20) is feasible. For all $\rho > 0$ and $\epsilon > 0$, nonnegativity of ψ implies $\omega_{\rho,\epsilon}^* \geq 0$. Moreover, the first and third constraints in (20) imply

$$\begin{aligned} \omega_{\rho,\epsilon}^* &\leq \frac{1}{\rho} (z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) + \epsilon - \mathbf{c}^\top \mathbf{x} - \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax})), \text{ for some } \mathbf{x} \in X \\ &\leq \frac{1}{\rho} (z^{\text{IP}} + \epsilon - z^{\text{LP}}), \end{aligned} \quad (21)$$

where the second inequality follows from the facts that $z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) \leq z^{\text{IP}}$ and $z^{\text{LP}} \leq \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax})$, for all $\mathbf{x} \in X$. By taking limits $\epsilon \downarrow 0$ on both sides of (21) we have

$$0 \leq \omega_\rho^* = \lim_{\epsilon \downarrow 0} \omega_{\rho,\epsilon}^* \leq \lim_{\epsilon \downarrow 0} \frac{1}{\rho} (z^{\text{IP}} + \epsilon - z^{\text{LP}}) = \frac{1}{\rho} (z^{\text{IP}} - z^{\text{LP}}) \quad (22)$$

Note that $\omega_{\rho,\epsilon}^*$ is non-decreasing as $\epsilon \downarrow 0$. Moreover, $\omega_{\rho,\epsilon}^*$ is upper bounded. Then, $\lim_{\epsilon \downarrow 0} \omega_{\rho,\epsilon}^*$ exists. By taking limits $\rho \rightarrow +\infty$ on both sides of (22) we have

$$\lim_{\rho \rightarrow +\infty} \omega_\rho^* = 0. \quad \square$$

Lemma 2 Consider ω_ρ^* as described in Proposition 6. Let us define $\tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$ as follows:

$$\begin{aligned} \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) &:= \inf_{\mathbf{x}, \omega} \{ \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho\omega \} \\ &\text{s.t. } \mathbf{x} \in X, \\ &\psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \omega, \\ &(1 - \delta)\omega_\rho^* \leq \omega \leq (1 + \delta)\omega_\rho^*. \end{aligned} \quad (23)$$

Then,

$$\begin{aligned} z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) &= \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) \\ &\geq \inf_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho(1 - \delta)\omega_\rho^* \} \\ &\text{s.t. } \mathbf{x} \in X, \\ &\psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq (1 + \delta)\omega_\rho^*, \\ &\geq \inf_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) \} \\ &\text{s.t. } \mathbf{x} \in X, \\ &\psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq (1 + \delta)\omega_\rho^*, \end{aligned} \quad (24)$$

for any $0 < \delta < 1$.

Proof Clearly, $z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) \leq \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$, due to the last constraint in (23). Let $\alpha_\rho := \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) - z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$. Assume by contradiction, $z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) < \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$ or equivalently $\alpha_\rho > 0$. Then, for all (\mathbf{x}, ω) satisfying constraints of (23) it holds

$$\mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho\omega \geq \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) = z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) + \alpha_\rho,$$

which implies (\mathbf{x}, ω) is infeasible for problem (20) if $0 < \epsilon < \alpha_\rho$. Therefore, $\omega_{\rho, \epsilon}^* \notin ((1 - \delta)\omega_\rho^*, (1 + \delta)\omega_\rho^*)$ for $0 < \epsilon < \alpha_\rho$, which contradicts with $\omega_\rho^* = \lim_{\epsilon \downarrow 0} \omega_{\rho, \epsilon}^*$. Therefore, $z_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) = \tilde{z}_\rho^{\text{LR}+}(\bar{\lambda}_{\text{LP}})$. Inequalities in (24) hold, because $\rho\omega \geq \rho(1 - \delta)\omega_\rho^* \geq 0$ and $\psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq (1 + \delta)\omega_\rho^*$, for all (\mathbf{x}, ω) satisfying constraints of (23). \square

Theorem 2 Suppose Assumptions 1 and 2 hold. Then, $\sup_{\rho > 0} z_\rho^{\text{LD}+} = z^{\text{IP}}$.

Proof Following (8), it is enough to show that $\sup_{\rho>0} z_\rho^{\text{LD}+} \geq z^{\text{IP}}$. Let δ be a given positive scalar in $(0, 1)$. By definition of ALD, we have

$$\begin{aligned} z_\rho^{\text{LD}+} &= \sup_{\lambda \in \mathbb{R}^m} \inf_{\mathbf{x}, \omega} \{ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho\omega : \mathbf{x} \in X, \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \omega \} \\ &\geq \inf_{\mathbf{x}, \omega} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho\omega : \mathbf{x} \in X, \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq \omega \} \\ &\geq \inf_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \psi(\mathbf{b} - \mathbf{A}\mathbf{x}) \leq (1 + \delta)\omega_\rho^* \} \quad (25a) \end{aligned}$$

$$\geq \inf_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq \kappa_\rho \} \quad (25b)$$

$$= \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq \kappa_\rho \} \quad (25c)$$

where $\kappa_\rho := \text{diam}\{\mathbf{u} \mid \psi(\mathbf{u}) \leq (1 + \delta)\omega_\rho^*\} = \sup\{\|\mathbf{u}\|_\infty \mid \psi(\mathbf{u}) \leq (1 + \delta)\omega_\rho^*\}$. Inequality (25a) holds by Lemma 2, and (25b) follows from level boundedness of ψ . Equality (25c) is valid by Assumption 1. By taking limits on both sides of (25b) we have

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} z_\rho^{\text{LD}+} &\geq \lim_{\rho \rightarrow +\infty} \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq \kappa_\rho \} \\ &\geq \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq \lim_{\rho \rightarrow +\infty} \kappa_\rho \} \quad (26a) \end{aligned}$$

$$= \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq 0 \} \quad (26b)$$

$$= \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \in X, \mathbf{A}\mathbf{x} = \mathbf{b} \}$$

$$= \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} : \mathbf{x} \in X, \mathbf{A}\mathbf{x} = \mathbf{b} \}$$

$$= z^{\text{IP}}.$$

where (26a) follows from lower semicontinuity of value functions for MIPs with rational data [21]. Equality (26b) holds by Assumption 2, i.e. $\lim_{\rho \rightarrow +\infty} \kappa_\rho = 0$. This completes the proof. \square

4 Exact Penalty Representation of ALD for MIPs

4.1 Pure IP case

A special case of problem (1) is the *pure IP case*, where all variables are integral, i.e. $n_2 = 0$. Zero duality gap and exact penalty representation using proximal Lagrangian for pure IPs were established in [5, Theorem 1.5]. Boland and Eberhard [7, Corollary 1] proved exact penalty representation for ALD of pure IPs with a bounded feasible region, i.e. X is finite, and the augmenting functions satisfying assumptions in Proposition 2. In this section, we extend this recent result to show exact penalty representation for pure IPs

under weaker assumptions on the augmenting functions (e.g., the augmenting function does not have to be convex) and X may not be necessarily finite.

Theorem 3 *Suppose problem (1) is a pure IP with potentially infinitely many feasible solutions, and Assumptions 1 holds. If*

$$\inf\{\psi(\mathbf{b} - \mathbf{Ax}) : \mathbf{x} \in X, \mathbf{Ax} \neq \mathbf{b}\} \geq \delta > 0 \quad (27)$$

for some strictly positive value of δ , then there exists a finite $\rho^* \in (0, +\infty)$ such that $z_{\rho^*}^{\text{LD}+} = z^{\text{IP}}$.

Proof Following (8), it suffices to find a finite ρ^* such that $z_{\rho^*}^{\text{LD}+} \geq z^{\text{IP}}$. Let $\bar{\rho} > 0$ be any positive penalty coefficient. By assumption, there exists a $\delta > 0$ which satisfies (27). Furthermore, let \mathbf{x}^0 be any arbitrary feasible solution of (1), i.e. $\mathbf{x}^0 \in X$ and $\mathbf{Ax}^0 = \mathbf{b}$. Set $\rho^* = \frac{\mathbf{c}^\top \mathbf{x}^0 - z^{\text{LP}}}{\delta}$. Note that $0 < \rho^* < +\infty$, because $\delta > 0$ and $-\infty < z^{\text{LP}} \leq \mathbf{c}^\top \mathbf{x}^0 < +\infty$. We claim that $z_{\rho^*}^{\text{LD}+} \geq z^{\text{IP}}$. Observe that we have

$$z_{\rho^*}^{\text{LD}+} = \sup_{\lambda} z_{\rho^*}^{\text{LD}+}(\lambda) \geq z_{\rho^*}^{\text{LD}+}(\bar{\lambda}_{\text{LP}}) = \inf_{\mathbf{x} \in X} \{\mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \psi(\mathbf{b} - \mathbf{Ax})\}. \quad (28)$$

There are two cases.

1. For all $\mathbf{x} \in X$ with $\mathbf{Ax} = \mathbf{b}$,

$$\mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \psi(\mathbf{b} - \mathbf{Ax}) = \mathbf{c}^\top \mathbf{x} \geq z^{\text{IP}}. \quad (29)$$

2. For all $\mathbf{x} \in X$ with $\mathbf{Ax} \neq \mathbf{b}$,

$$\begin{aligned} & \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \psi(\mathbf{b} - \mathbf{Ax}) \\ &= \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \left(\frac{\mathbf{c}^\top \mathbf{x}^0 - z^{\text{LP}}}{\delta} \right) \psi(\mathbf{b} - \mathbf{Ax}) \\ &\geq \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + (\mathbf{c}^\top \mathbf{x}^0 - z^{\text{LP}}) \end{aligned} \quad (30a)$$

$$\geq z^{\text{LP}} + (\mathbf{c}^\top \mathbf{x}^0 - z^{\text{LP}}) \quad (30b)$$

$$= \mathbf{c}^\top \mathbf{x}^0$$

$$\geq z^{\text{IP}},$$

where (30a) holds because $-\infty < z^{\text{LP}} \leq z^{\text{IP}} \leq \mathbf{c}^\top \mathbf{x}^0$ and $\psi(\mathbf{b} - \mathbf{Ax}) \geq \delta > 0$ for all $\mathbf{x} \in X$ with $\mathbf{Ax} \neq \mathbf{b}$ by (27); (30b) follows by definition of $\bar{\lambda}_{\text{LP}}$.

Inequalities (29) and (30) imply

$$\inf_{\mathbf{x} \in X} \left\{ \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \psi(\mathbf{b} - \mathbf{Ax}) \right\} \geq z^{\text{IP}}.$$

Together with (28), we have

$$z_{\rho^*}^{\text{LD}+} \geq z^{\text{IP}}.$$

This completes the proof. \square

Note that for the pure IP case of problem (1) under Assumption 1, any augmenting function defined in Proposition 2 satisfies (27). Even the index function $\mathbb{I} : \mathbb{R}^m \rightarrow \{0, 1\}$ where

$$\mathbb{I}(\mathbf{u}) = \begin{cases} 0 & \text{if } \mathbf{u} = \mathbf{0}, \\ 1 & \text{otherwise} \end{cases}$$

can be used as an augmenting function $\psi(\cdot)$ to satisfy (27).

4.2 MIP case

For a general MIP with both continuous and integer variables, we need more conditions on the augmenting function to have an exact penalty representation. For example, if $\psi(\cdot) = \|\cdot\|_2^2$, i.e. the proximal Lagrangian case, this augmenting function satisfies the assumptions in Proposition 2 as well as (27) when X is a pure integer set. However, for a general MIP, there may not exist a finite $0 < \rho^* < \infty$ such that $z_{\rho^*}^{\text{LD}^+} = z^{\text{IP}}$ under this augmenting function. In this section, we first give an example to show that proximal Lagrangian fails to have an exact penalty representation for a simple MIP in three variables. Then we prove that, when the augmenting function is any norm (but not the squared norm) i.e., for the sharp Lagrangians, the ALD always has an exact penalty representation for general MIPs. Finally, we extend this result to some classes of augmenting functions that are not convex.

4.2.1 Counterexample MIP for Proximal Lagrangian

Proposition 7 *There exists an MIP problem of the form (1) and an augmenting function satisfying assumptions in Proposition 2 such that $z_{\rho}^{\text{LD}^+} < z^{\text{IP}}$ for all finite $\rho > 0$.*

Next, we verify this proposition with a simple example.

Example 1 Consider the following MIP problem, with one binary and two continuous variables.

$$\begin{aligned} z^{\text{IP}} &= \min_{x_1, x_2, x_3} && -x_1 - x_2 \\ \text{s.t.} &&& -x_1 + x_2 = 0 \\ &&& 0 \leq x_1 \leq x_3 \\ &&& 0 \leq x_2 \leq 1 - x_3 \\ &&& x_3 \in \{0, 1\} \end{aligned} \tag{31}$$

The only feasible points for (31) are $(x_1, x_2, x_3) = (0, 0, 0)$ and $(x_1, x_2, x_3) = (0, 0, 1)$ with objective value 0. Then, $z^{\text{IP}} = 0$. Projection of the feasible region of (31) without the constraint $-x_1 + x_2 = 0$ into the space of x_1 and x_2 contains the blue lines in Figure 1. The points satisfying $-x_1 + x_2 = 0$ are depicted by a red line in this space.

We show that in Example 1, for ALD with $\psi(\cdot) = \|\cdot\|_2^2$, $z_\rho^{\text{LD}^+} < 0$ for all $\rho > 0$. From Theorem 1, ALD (15) with $\psi(\cdot) = \|\cdot\|_2^2$ becomes

$$\begin{aligned} z_\rho^{\text{LD}^+} = \inf \quad & -x_1 - x_2 + \rho\omega \\ \text{s.t.} \quad & (x_1, x_2, x_3, \omega) \in \mathbf{conv}(S_2) \\ & -x_1 + x_2 = 0. \end{aligned} \quad (32)$$

where,

$$S_2 := \left\{ (x_1, x_2, x_3, \omega) \in \mathbb{R}^4 : \begin{array}{l} \omega \geq (-x_1 + x_2)^2 \\ 0 \leq x_1 \leq x_3 \\ 0 \leq x_2 \leq 1 - x_3 \\ x_3 \in \{0, 1\} \end{array} \right\}.$$

Consider $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega}) = (0, r, 0, r^2)$ and $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega}) = (r, 0, 1, r^2)$ where $r(\rho) = \min\{1, \frac{1}{2\rho}\}$. Obviously, both $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega})$ and $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega})$ belong to S_2 . Then, $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega}) := \frac{1}{2}(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega}) + \frac{1}{2}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega}) = (\frac{r}{2}, \frac{r}{2}, \frac{1}{2}, r^2)$ belongs to $\mathbf{Conv}(S_2)$. Projection of the points $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega})$, $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega})$ and $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega})$ in the space of x_1 and x_2 can be depicted as points A, B and C, respectively, in Figure 1. Because $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega}) \in \mathbf{conv}(S_2)$ and $-\bar{x}_1 + \bar{x}_2 = 0$, the point $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega})$ is a feasible solution of (32). Therefore,

$$z_\rho^{\text{LD}^+} \leq -\bar{x}_1 - \bar{x}_2 + \rho\bar{\omega} = -r + \rho r^2 \leq \max\left\{-\frac{1}{2}, -\frac{1}{4\rho}\right\} < 0 = z^{\text{IP}}, \quad \forall \rho > 0 \quad (33)$$

which shows $z_\rho^{\text{LD}^+} < z^{\text{IP}}$ for all $\rho > 0$, i.e. there is no finite ρ^* such that $z_{\rho^*}^{\text{LD}^+} = z^{\text{IP}}$. Note that the second inequality in (33) follows from the fact that

$$-r + \rho r^2 = \begin{cases} -1 + \rho \times 1^2 = -1 + \rho \leq -\frac{1}{2}, & \text{if } 0 < \rho < \frac{1}{2} \\ -\frac{1}{2\rho} + \rho \times \left(\frac{1}{2\rho}\right)^2 = -\frac{1}{4\rho}, & \text{if } \frac{1}{2} \leq \rho. \end{cases}$$

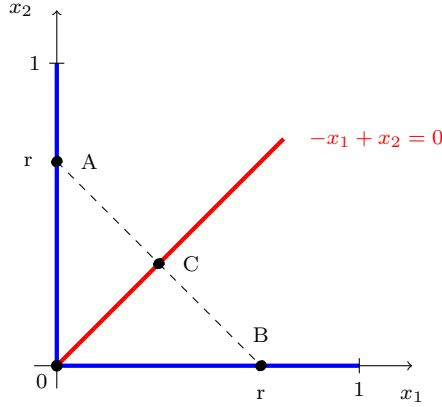


Fig. 1 Projection of the feasible region of Example 1 in the space of x_1 and x_2 .

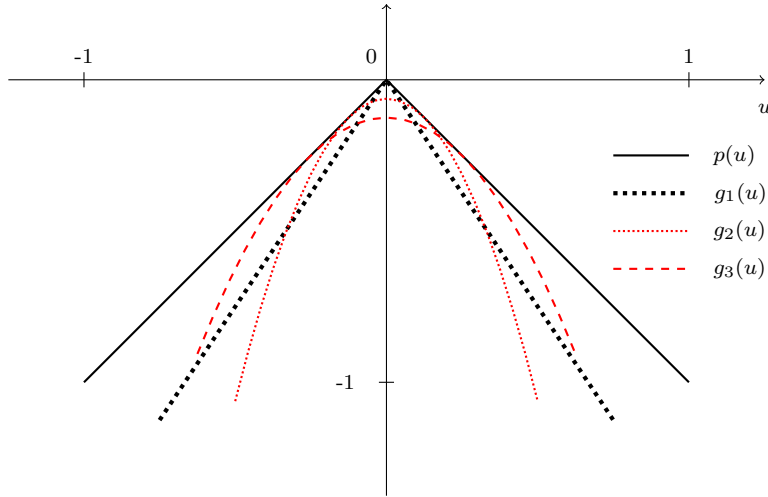


Fig. 2 Value function and some augmenting functions for Example 1

Another point of view for this example is exploring its value function with respect to the first constraint, which is

$$\begin{aligned}
 p(u) = \inf_{x_1, x_2, x_3} & \quad -x_1 - x_2 \\
 \text{s.t.} & \quad -x_1 + x_2 = u \\
 & \quad 0 \leq x_1 \leq x_3 \\
 & \quad 0 \leq x_2 \leq 1 - x_3 \\
 & \quad x_3 \in \{0, 1\}.
 \end{aligned} \tag{34}$$

It is easy to check that $p(u) = -|u|$, for any $u \in [-1, 1]$. Note that the problem (34) is infeasible if $u \notin [-1, 1]$. Consider the supporting functions $g_1(u)$, $g_2(u)$ and $g_3(u)$ which are based on augmenting functions $\psi_1(u) = 1.5|u|$, $\psi_2(u) = u^2$ and $\psi_3(u) = 0.5u^2$, respectively. The value function and these supporting functions are depicted in Figure 2. From this figure, it is clear that $\psi_2(u) = u^2$ can result in a strictly smaller duality gap comparing to $\psi_3(u) = 0.5u^2$. But, no quadratic supporting function to $p(\cdot)$ can reach $p(0) = 0$. However, the sharp augmenting function $\psi_1(u) = 1.5|u|$ closes the duality gap in this example.

Example 1 showed that, for $\psi(\cdot) = \|\cdot\|_2^2$, there may exist MIP problems such that $z_\rho^{\text{LD}^+} < z^{\text{IP}}$, for any finite value of ρ .

4.2.2 Exact ALD with the sharp Lagrangian for MIPs

Next, we show that using any norm as an augmenting function with a sufficiently large penalty coefficient closes the duality gap for the general MIPs. One approach is to verify the calmness condition in Subsection 2.3 for the value function of an MIP. In this section, we provide a self contained proof for this result.

Theorem 4 Consider problem (1) with both integer and continuous variables. Suppose Assumption 1 holds, and $\psi(\cdot) = \|\cdot\|$, where $\|\cdot\|$ is any norm. Then there exists a finite $0 < \rho^* < +\infty$ such that $z_{\rho^*}^{\text{LD}^+} = z^{\text{IP}}$.

Proof First, let us show the result for $\psi(\cdot) = \|\cdot\|_\infty$. Then, we extend it to any norm by the equivalence of norms in a Euclidean space. Let $\psi(\cdot) = \|\cdot\|_\infty$ and $\mathbf{1}_m$ be an m dimensional vector with all entries equal to 1. Then S_ψ is a polyhedron,

$$\begin{aligned} S_\psi &= \{(\mathbf{x}, \omega) \in \mathbb{R}^{n+1} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \leq \omega, \mathbf{x} \in X\} \\ &= \{(\mathbf{x}, \omega) \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2+1} : -\mathbf{1}_m\omega \leq \mathbf{b} - \mathbf{A}\mathbf{x} \leq \mathbf{1}_m\omega, \mathbf{E}\mathbf{x} \leq \mathbf{f}\}. \end{aligned} \quad (35)$$

Then, by Assumption 1 and due to Meyer's theorem [20], there is a rational polyhedral representation for the set $\text{conv}(S_{\|\cdot\|_\infty}) \cap \{(\mathbf{x}, \omega) \in \mathbb{R}^{n+1} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$. Denote this representation by $\mathbf{H} \begin{bmatrix} \mathbf{x} \\ \omega \end{bmatrix} \geq \mathbf{h}$, where $\mathbf{H} \in \mathbb{Q}^{\hat{m} \times (n+1)}$ and $\mathbf{h} \in \mathbb{Q}^{\hat{m}}$, for some finite integer \hat{m} . Then, by the primal characterization of ALD in Theorem 1, the ALD problem (7) for a given $\rho > 0$ can be written as follows,

$$\begin{aligned} z_\rho^{\text{LD}^+} &= \inf_{\mathbf{x}, \omega} \mathbf{c}^\top \mathbf{x} + \rho\omega \\ \text{s.t. } &\mathbf{H} \begin{bmatrix} \mathbf{x} \\ \omega \end{bmatrix} \geq \mathbf{h}. \end{aligned} \quad (36)$$

Note that, for a given $\rho > 0$, problem (36) is an LP and its dual can be written as follows.

$$\begin{aligned} z_\rho^{\text{DL}^+} &:= \sup_{\mathbf{y}} \mathbf{h}^\top \mathbf{y} \\ \text{s.t. } &\mathbf{H}^\top \mathbf{y} = \begin{bmatrix} \mathbf{c} \\ \rho \end{bmatrix} \\ &\mathbf{y} \geq \mathbf{0}. \end{aligned} \quad (37)$$

Note that $z_\rho^{\text{LD}^+} = z_\rho^{\text{DL}^+}$, since $z_\rho^{\text{LD}^+} > -\infty$ and by strong duality for LPs. We are interested in a finite positive ρ^* such that $z_{\rho^*}^{\text{LD}^+} \geq \mathbf{c}^\top \mathbf{x}^*$, where \mathbf{x}^* is an optimal solution of (1). The existence of such a ρ^* is equivalent to the existence of $(\mathbf{y}^*, \xi^*, \rho^*)$ with $\xi^* = 0$ for the following feasibility problem in (\mathbf{y}, ξ, ρ) ,

$$\begin{aligned} \mathbf{h}^\top \mathbf{y} + \xi &\geq \mathbf{c}^\top \mathbf{x}^* \\ \mathbf{H}^\top \mathbf{y} &= \begin{bmatrix} \mathbf{c} \\ \rho \end{bmatrix} \\ \mathbf{y} &\geq \mathbf{0} \\ \rho &\geq 1 \\ \xi &\geq 0. \end{aligned} \quad (38)$$

Let Ξ be the projection of the feasible set of (38) into the ξ space. Note that by Fourier-Motzkin Elimination, Ξ is itself a polyhedron. Then, Ξ is a closed set.

Consider a sequence $\xi_k \downarrow 0$ as $k \rightarrow +\infty$. Since $z_\rho^{\text{DLD}+} = z_\rho^{\text{LD}+} \leq z^{\text{IP}}$ for any $\rho \geq 0$, and $\psi(\cdot) = \|\cdot\|_\infty$ satisfies Assumption 2, by Theorem 2, $z_\rho^{\text{LD}+} \uparrow z^{\text{IP}} = \mathbf{c}^\top \mathbf{x}^*$ as $\rho \rightarrow \infty$. By closedness of Ξ , $0 \in \Xi$ because $\xi^* = 0$ is a cluster point of Ξ . That is, there exists some \mathbf{y}^* and ρ^* such that $(\mathbf{y}^*, 0, \rho^*)$ is a feasible solution of (38). Therefore, $z_{\rho^*}^{\text{LD}+} \geq z^{\text{IP}}$, which along with $z_{\rho^*}^{\text{LD}+}$ being a lower bound for z^{IP} , we conclude $z_{\rho^*}^{\text{LD}+} = z^{\text{IP}} = \mathbf{c}^\top \mathbf{x}^*$. Note that

$$\begin{aligned}
z^{\text{IP}} &= \sup_{\lambda \in \mathbb{R}^m} \inf_{\mathbf{x} \in X} \mathbf{c}^\top \mathbf{x} + \lambda^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \\
&= \sup_{\lambda \in \mathbb{R}^m} \inf_{(\mathbf{x}, \omega) \in S_{\|\cdot\|_\infty}} \mathbf{c}^\top \mathbf{x} + \lambda^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \omega \\
&= \sup_{\lambda \in \mathbb{R}^m} \inf_{(\mathbf{x}, \omega) \in \mathbf{conv}(S_{\|\cdot\|_\infty})} \mathbf{c}^\top \mathbf{x} + \lambda^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \omega \\
&= \inf_{(\mathbf{x}, \omega) \in \mathbf{conv}(S_{\|\cdot\|_\infty})} \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \omega \\
&= \inf_{(\mathbf{x}, \omega) \in S_{\|\cdot\|_\infty}} \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \omega \\
&= \inf_{\mathbf{x} \in X} \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty
\end{aligned} \tag{39}$$

for some $\bar{\lambda} \in \mathbb{R}^m$, where the second equality follows from definition of $S_{\|\cdot\|_\infty}$ in (35). The third and fifth equations hold because minimizing a linear objective function on a set is equivalent to minimizing it on the convex hull of that set. The fourth equality is valid by strong duality for LPs, because under Assumption 1 and due to Meyer's theorem [20], $\mathbf{conv}(S_{\|\cdot\|_\infty})$ is a rational polyhedron. The last equality follows from definition of $S_{\|\cdot\|_\infty}$ in (35). Then,

$$\mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty \geq z^{\text{IP}}, \quad \forall \mathbf{x} \in X$$

Recall that for any norm $\|\cdot\|$ in finite dimensions there exists $0 < \gamma < 1$ such that $\frac{1}{\gamma} \|\mathbf{u}\| \geq \|\mathbf{u}\|_\infty \geq \gamma \|\mathbf{u}\|$, by the equivalence of norms. Take $\hat{\rho} = \frac{\rho^*}{\gamma}$. Then,

$$\mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \hat{\rho} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \geq \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \rho^* \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_\infty, \quad \forall \mathbf{x} \in X$$

which implies

$$z_{\hat{\rho}}^{\text{LD}+} \geq z_{\hat{\rho}}^{\text{LR}+}(\bar{\lambda}) = \inf_{\mathbf{x} \in X} \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) + \hat{\rho} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \geq z^{\text{IP}} \tag{40}$$

On the other hand, $z_{\hat{\rho}}^{\text{LD}+} \leq z^{\text{IP}}$ by (8). Therefore, $z_{\hat{\rho}}^{\text{LD}+} = z^{\text{IP}}$. \square

Remark 4 Note that $\hat{\rho}$ and $\bar{\lambda}$ in the proof of Theorem 4 satisfy the assumptions in Proposition 1. Therefore, any optimal solution of ALR (6) with $\lambda = \bar{\lambda}$ and $\rho > \hat{\rho}$ solves the MIP problem (1).

Next, we show that the value of $\bar{\lambda}$ in the proof of Theorem 4 really does not matter.

Proposition 8 Consider problem (1) under Assumption 1. Suppose $\psi(\cdot) = \|\cdot\|$, where $\|\cdot\|$ is any norm. For any $\tilde{\lambda} \in \mathbb{R}^m$, there exists a finite $\rho^*(\tilde{\lambda})$ such that $z_{\rho^*}^{\text{LR}+}(\tilde{\lambda}) = z^{\text{IP}}$.

Proof Let $\hat{\rho}$ and $\bar{\lambda}$ be as considered in (40). By the equivalence of norms, there exists $0 < \gamma < 1$ such that $\frac{1}{\gamma}\|\mathbf{u}\| \geq \|\mathbf{u}\|_2 \geq \gamma\|\mathbf{u}\|$. From Cauchy-Schwarz inequality, for all $\mathbf{x} \in X$, it holds

$$\tilde{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) \geq -\|\tilde{\lambda}\|_2 \|\mathbf{b} - \mathbf{Ax}\|_2 \geq -\frac{1}{\gamma}\|\tilde{\lambda}\|_2 \|\mathbf{b} - \mathbf{Ax}\|,$$

$$\bar{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) \leq \|\bar{\lambda}\|_2 \|\mathbf{b} - \mathbf{Ax}\|_2 \leq \gamma\|\bar{\lambda}\|_2 \|\mathbf{b} - \mathbf{Ax}\|,$$

and consequently,

$$\tilde{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) \geq \bar{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) - \left(\frac{1}{\gamma}\|\tilde{\lambda}\|_2 + \gamma\|\bar{\lambda}\|_2 \right) \|\mathbf{b} - \mathbf{Ax}\|. \quad (41)$$

Take $\rho^* = \hat{\rho} + \left(\frac{1}{\gamma}\|\tilde{\lambda}\|_2 + \gamma\|\bar{\lambda}\|_2 \right)$. Then,

$$\mathbf{c}^\top \mathbf{x} + \tilde{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \|\mathbf{b} - \mathbf{Ax}\| \geq \mathbf{c}^\top \mathbf{x} + \bar{\lambda}^\top (\mathbf{b} - \mathbf{Ax}) + \hat{\rho} \|\mathbf{b} - \mathbf{Ax}\|. \quad (42)$$

By taking $\inf_{\mathbf{x} \in X}$ from both sides of (42) and considering (40) it is implied that $z_{\rho^*}^{\text{LR}+}(\tilde{\lambda}) \geq z^{\text{IP}}$. This result along with $z_{\rho^*}^{\text{LR}+}(\tilde{\lambda})$ being a lower bound for z^{IP} , concludes $z_{\rho^*}^{\text{LR}+}(\tilde{\lambda}) = z^{\text{IP}}$. \square

Next, we extend Theorem 4 to a more general class of augmenting functions than norms.

Theorem 5 Consider an MIP problem (1) satisfying Assumption 1. Then, there exists a finite $\hat{\rho}$ such that $z_{\hat{\rho}}^{\text{LD}+} = z_{\hat{\rho}}^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) = z^{\text{IP}}$ if ψ is an augmenting function such that

- $\psi(\mathbf{0}) = 0$,
- $\psi(\mathbf{u}) \geq \delta > 0$, for all $\mathbf{u} \notin V$,
- $\psi(\mathbf{u}) \geq \gamma\|\mathbf{u}\|_\infty$, for all $\mathbf{u} \in V$,

for some open neighborhood V of $\mathbf{0}$, and positive scalars $\delta, \gamma > 0$.

Proof From Proposition 8, there exists a finite ρ^* such that $z_{\rho^*}^{\text{LR}+}(\bar{\lambda}_{\text{LP}}) = z^{\text{IP}}$ for $\psi(\cdot) = \|\cdot\|_\infty$. Now, consider the cases where ψ is not a norm but it satisfies the conditions stated above. Take $\hat{\rho} = \max \left\{ \frac{z^{\text{IP}} - z^{\text{LP}}}{\delta}, \frac{\rho^*}{\gamma} \right\}$. There are two cases.

1. For all $\mathbf{x} \in X$ such that $(\mathbf{b} - \mathbf{Ax}) \notin V$, it holds

$$\begin{aligned} \mathbf{c}^\top \mathbf{x} + \bar{\lambda}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \hat{\rho}\psi(\mathbf{b} - \mathbf{Ax}) &\geq z^{\text{LP}} + \hat{\rho}\psi(\mathbf{b} - \mathbf{Ax}) \\ &\geq z^{\text{LP}} + \frac{z^{\text{IP}} - z^{\text{LP}}}{\delta}\psi(\mathbf{b} - \mathbf{Ax}) \\ &\geq z^{\text{LP}} + z^{\text{IP}} - z^{\text{LP}} \\ &\geq z^{\text{IP}} \end{aligned}$$

2. For all $\mathbf{x} \in X$ such that $(\mathbf{b} - \mathbf{Ax}) \in V$, it holds

$$\begin{aligned} \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \hat{\rho} \psi(\mathbf{b} - \mathbf{Ax}) &\geq \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \frac{\rho^*}{\gamma} \psi(\mathbf{b} - \mathbf{Ax}) \\ &\geq \mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \rho^* \|\mathbf{b} - \mathbf{Ax}\|_\infty \\ &\geq z^{\text{IP}}. \end{aligned}$$

Then,

$$\mathbf{c}^\top \mathbf{x} + \bar{\boldsymbol{\lambda}}_{\text{LP}}^\top (\mathbf{b} - \mathbf{Ax}) + \hat{\rho} \psi(\mathbf{b} - \mathbf{Ax}) \geq z^{\text{IP}}, \quad \forall \mathbf{x} \in X, \quad (43)$$

which implies $z_{\hat{\rho}}^{\text{LD}+} \geq z_{\hat{\rho}}^{\text{LR}+}(\bar{\boldsymbol{\lambda}}_{\text{LP}}) \geq z^{\text{IP}}$. This result along with $z_{\hat{\rho}}^{\text{LR}+}(\bar{\boldsymbol{\lambda}}_{\text{LP}})$ being a lower bound for z^{IP} , concludes $z_{\hat{\rho}}^{\text{LR}+}(\tilde{\boldsymbol{\lambda}}) = z^{\text{IP}}$. \square

Remark 5 It is easy to check that for any norm $\|\cdot\|$ and scalar r with $0 < r < 1$, the non-convex function $\psi(\cdot) = \|\cdot\|^r$ satisfies the conditions stated in Theorem 5.

5 Conclusions and final remarks

In this paper we studied ALD for general linear MIP problems. We presented a primal characterization of ALD for MIPs and showed the asymptotic zero duality gap property with non-negative level bounded and not necessarily convex augmenting functions. Moreover, we showed that under some mild assumptions, ALD achieves zero duality gap for general MIPs with a finite penalty coefficient and a general class of augmenting functions. We also show that some augmenting functions such as the squared Euclidean norm are exact in the pure IP cases, but there exists MIP counterexamples for which these augmenting functions may result in a non-zero duality gap for any value of the penalty coefficient.

Solving IP and MIP problems by ALD may have computational advantages over the classical Lagrangian relaxation approaches, since ALD may produce better dual bounds and provide primal solutions. The main drawback of ALR and ALD methods is that the resulting subproblems are not separable because of the nonlinear augmenting functions. To overcome this issue, the alternating direction method of multipliers (ADMM) [8] and related schemes have been developed for convex optimization problems. However, it is not at all clear how to decompose ALD for MIP and more general nonconvex problems and utilize parallel computation. Based on ADMM, a heuristic decomposition method was developed in [16] to solve MIPs arising from electric power network unit commitment problems. In a continuous and non-convex setting, a decomposition approach using ADMM was developed for the AC optimal power flow problem in electric power grid optimization [31]. Further developing theories and algorithms to solve ALD for MIPs and general non-convex optimization problems with decomposition schemes is an important future research direction.

Acknowledgment

We would like to thank Dr. Natasha Boland for the helpful discussions and feedback on our draft. This research has been supported in part by the National Science Foundation Grant 1331426 and the Office of Naval Research Grant N00014-15-1-2078.

References

1. Ben-Tal, A., Zibulevsky, M.: Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal on Optimization* **7**(2), 347–366 (1997)
2. Bertsekas, D.P.: *Constrained optimization and Lagrange multiplier methods*. Academic press (1982)
3. Bertsekas, D.P., Nedić, A., Ozdaglar, A.E.: *Convex analysis and optimization*. Athena Scientific Belmont (2003)
4. Blair, C., Jeroslow, R.: The value function of a mixed integer program: I. *Discrete Mathematics* **19**(2), 121–138 (1977)
5. Blair, C., Jeroslow, R.: The value function of a mixed integer program: II. *Discrete Mathematics* **25**(1), 7–19 (1979)
6. Blair, C., Jeroslow, R.: The value function of an integer program. *Mathematical Programming* **23**(1), 237–273 (1982)
7. Boland, N.L., Eberhard, A.C.: On the augmented Lagrangian dual for integer programming. *Mathematical Programming* **150**(2), 491–509 (2015)
8. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**(1), 1–122 (2011)
9. Burachik, R.S., Iusem, A.N., Melo, J.G.: Duality and exact penalization for general augmented lagrangians. *Journal of optimization theory and applications* **147**(1), 125–140 (2010)
10. Burachik, R.S., Iusem, A.N., Melo, J.G.: The exact penalty map for nonsmooth and nonconvex optimization. *Optimization* **64**(4), 717–738 (2015)
11. Burachik, R.S., Rubinov, A.: On the absence of duality gap for lagrange-type functions. *Journal of Industrial and Management Optimization* **1**(1), 33 (2005)
12. Burachik, R.S., Rubinov, A.: Abstract convexity and augmented Lagrangians. *SIAM Journal on Optimization* **18**(2), 413–436 (2007)
13. Burke, J.V.: Calmness and exact penalization. *SIAM Journal on Control and Optimization* **29**(2), 493–497 (1991)
14. Burke, J.V.: An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization* **29**(4), 968–998 (1991)
15. Clarke, F.H.: A new approach to Lagrange multipliers. *Mathematics of Operations Research* **1**(2), 165–174 (1976)

16. Feizollahi, M.J., Costley, M., Ahmed, S., Grijalva, S.: Large-scale decentralized unit commitment. *International Journal of Electric Power & energy Systems* **73**(1), 97–106 (2015)
17. Geoffrion, A.M.: Lagrangean relaxation for integer programming. *Mathematical Programming Study* **2**, 82–114 (1974)
18. Huang, X., Yang, X.: A unified augmented lagrangian approach to duality and exact penalization. *Mathematics of Operations Research* **28**(3), 533–552 (2003)
19. Luo, Z.Q., Pang, J.S., Ralph, D.: *Mathematical programs with equilibrium constraints*. Cambridge University Press (1996)
20. Meyer, R.R.: On the existence of optimal solutions to integer and mixed-integer programming problems. *Mathematical Programming* **7**(1), 223–235 (1974)
21. Meyer, R.R.: Integer and mixed-integer programming models: general properties. *Journal of Optimization Theory and Applications* **16**(3-4), 191–206 (1975)
22. Nedić, A., Ozdaglar, A.: A geometric framework for nonconvex optimization duality using augmented Lagrangian functions. *Journal of Global Optimization* **40**(4), 545–573 (2008)
23. Nedic, A., Ozdaglar, A.: Separation of nonconvex sets with general augmenting functions. *Mathematics of Operations Research* **33**(3), 587–605 (2008)
24. Ralphs, T.K., Hassanzadeh, A.: On the value function of a mixed integer linear optimization problem and an algorithm for its construction. Tech. rep., Laboratory for Computational Optimization at Lehigh (CORL), Department of Industrial and Systems Engineering, Lehigh University, Technical Report 14T-004 (2014)
25. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control* **12**(2), 268–285 (1974)
26. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*, vol. 317. Springer (1998)
27. Rubinov, A., Yang, X., Bagirov, A.: Penalty functions with a small penalty parameter. *Optimization Methods and Software* **17**(5), 931–964 (2002)
28. Rubinov, A.M., Glover, B.M., Yang, X.: Decreasing functions with applications to penalization. *SIAM Journal on Optimization* **10**(1), 289–313 (1999)
29. Rubinov, A.M., Huang, X., Yang, X.: The zero duality gap property and lower semicontinuity of the perturbation function. *Mathematics of Operations Research* **27**(4), 775–791 (2002)
30. Rubinov, A.M., Yang, X.: *Lagrange-type functions in constrained nonconvex optimization*, vol. 85. Springer Science & Business Media (2003)
31. Sun, X. A., Phan, D., Ghosh, S.: Fully decentralized AC optimal power flow algorithms. *IEEE Power and Energy Society General Meeting*, July 21-25, (2013).

-
32. Wang, C.Y., Yang, X.Q., Yang, X.M.: Nonlinear augmented Lagrangian and duality theory. *Mathematics of Operations Research* **38**(4), 740–760 (2013)
 33. Wolsey, L.A., Nemhauser, G.L.: *Integer and Combinatorial Optimization*. Wiley Series in Discrete Mathematics and Optimization. Wiley (1999)
 34. Yang, X., Huang, X.: A nonlinear Lagrangian approach to constrained optimization problems. *SIAM Journal on Optimization* **11**(4), 1119–1144 (2001)