Exact Augmented Lagrangian Duality for Nonconvex Mixed-Integer Nonlinear Optimization

Henri Lefebvre, Martin Schmidt

Abstract. In the context of mixed-integer nonlinear problems (MINLPs), it is well-known that strong duality does not hold in general if the standard Lagrangian dual is used. Hence, we consider the augmented Lagrangian dual (ALD), which adds a nonlinear penalty function to the classic Lagrangian function. For this setup, we study conditions under which the ALD leads to a zero duality gap for general MINLPs. In particular, under mild assumptions and for a large class of penalty functions, we show that the ALD yields zero duality gaps if the penalty parameter goes to infinity. If the penalty function is a norm, we also show that the ALD leads to zero duality gaps for a finite penalty parameter. Moreover, we show that such a finite penalty parameter can be computed in polynomial time in the mixed-integer linear case. This generalizes the recent results on linearly constrained mixed-integer problems by Bhardwaj et al. (2024), Boland and Eberhard (2014), Feizollahi et al. (2016), and Gu et al. (2020).

1. Introduction

We study the mixed-integer nonlinear problem (MINLP)

$$
\begin{aligned}
z^* := \inf_x \ & f(x) && \text{(1a)}\\
\text{s.t. } & Ax = b, && \text{(1b)}\\
& g(x) \le 0, && \text{(1c)}\\
& x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{n_2} && \text{(1d)}
\end{aligned}
$$

for which $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ is a given matrix and vector, $f : \mathbb{R}^n \to \mathbb{R}$ as well as $g : \mathbb{R}^n \to \mathbb{R}^\ell$ are given functions with $n := n_1 + n_2$. Additionally, we let $X := \{x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{n_2} : g(x) \le 0\}$ and consider dualizing the linear constraints in (1b).

Under some regularity conditions, it is well known that convex problems have a zero duality gap if the classic Lagrangian dual is considered; see e.g., Bertsekas et al. (2003). Unfortunately, it is also known that this result does not generalize to nonconvex problems like Problem (1). Also note that, even in the case in which f and g are convex functions, the integrality requirements in (1d) lead to a disconnected feasible region and, thus, Problem (1) is still a nonconvex problem. Therefore, we focus on the augmented Lagrangian dual (ALD), which, as its name suggests, augments the classic Lagrangian dual with a nonlinear penalization of some constraint violation. To this end, we consider the augmented Lagrangian relaxation (ALR) given by

$$
z_\rho^{\text{LR+}}(\lambda) := \inf_{x\in X} \ c^\top x + \lambda^\top (Ax - b) + \rho\psi(Ax - b), \tag{2}
$$

Date: July 16, 2024.
2020 *Mathematics Subject Classification.* 90-XX, 90C46, 90C11.
Key words and phrases. Mixed-integer nonlinear optimization, Augmented Lagrangian duality, Exact penalty methods.

where $\rho > 0$ is a given penalty parameter and ψ is a penalty function satisfying $\psi(0) = 0$ and $\psi(z) > 0$ if $z \neq 0$; see e.g., Rockafellar and Wets (1998). The ALD is then defined as the best lower bound which can be obtained by (2), i.e., the ALD is

$$z_\rho^{\mathrm{LD}+} := \sup_{\lambda \in \mathbb{R}^m} z_\rho^{\mathrm{LR}+}(\lambda). \tag{3}$$

It is clear that the ALD reduces to the classic Lagrangian dual for $\rho = 0$. Moreover, the inequalities

$$z_\rho^{\mathrm{LR}+}(\bar{\lambda}) \leq z_\rho^{\mathrm{LD}+} \leq z^* \tag{4}$$

trivially hold for all $\bar{\lambda} \in \mathbb{R}^m$.

We are primarily interested in conditions under which the ALD can close the duality gap between the primal (1) and its dual (3). Throughout this paper, we denote this duality gap by $\gamma_\rho := z^* - z_\rho^{\mathrm{LD}+}$ and say that the dual is strong (or that strong duality holds) if and only if $\gamma_\rho = 0$ holds for some $\rho \in \mathbb{R}_{\geq 0} \cup \{\infty\}$. As detailed in the literature review below in Section 1.1, recent works on ALD for mixed-integer problems seem to indicate that ALD is a good candidate for a strong dual in this setting (Bhardwaj et al. 2024; Boland and Eberhard 2014; Feizollahi et al. 2016; Gu et al. 2020). Additionally, an interesting question is whether a finite penalty parameter exists so that $\gamma_\rho = 0$ and, if so, under what conditions. We call such a penalty parameter an exact penalty parameter.

Definition 1 (Exact Penalty Parameter)**.** *A penalty parameter* $\rho > 0$ *is called an exact penalty parameter if and only if* $\gamma_\rho = 0$ *and* $\rho < \infty$*.*

While this definition is not in line with classic exactness notions in nonlinear continuous optimization (see, e.g., Bertsekas (2016)), it is in line with the papers mentioned above that deal with the mixed-integer case.

1.1. **Literature Review.** There is a vast literature on ALD approaches for continuous nonconvex problems. For example, in Rockafellar (1974), the authors consider general nonconvex problems and show that the duality gap can be closed asymptotically, i.e., by driving the penalty parameter to infinity, using the ALD if the penalty function is a quadratic function. The paper also discusses the existence of an exact penalty parameter provided that some stability conditions hold on the problem's value function under small perturbations. It should be noted that a nonsmooth penalty function is often required to obtain general exact penalty parameters; see, e.g., Chapter 11.K in Rockafellar and Wets (1998). For instance, penalty functions that are norms are used in Burke (1991). In Huang and Yang (2003), the convexity assumption of the penalty function is replaced by a level-boundedness assumption. In Burachik et al. (2017), the authors study the ALD in the context of semi-infinite programming problems.

More recently, ALD approaches for mixed-integer problems gained interest in the research community. In Boland and Eberhard (2014), mixed-integer linear problems (MILPs) are considered and it is shown that a zero duality gap can be reached asymptotically using the ALD with a certain class of convex penalty functions. Additionally, the authors show that an exact penalty parameter exists in the pure integer case. In Feizollahi et al. (2016), a more general class of penalty functions is considered for MILPs. In particular, they show that the ALD can asymptotically close the duality gap if the penalty function is a general level-bounded augmenting function, which does not necessarily need to be convex. Moreover, they show under mild assumptions that an exact penalty parameter exists if the penalty function is a norm. Mixed-integer convex-quadratic problems (MIQPs) are studied in Gu et al. (2020) and similar results are given. Moreover, the authors also discuss the existence of a "small" exact penalty parameter with norm penalty functions. More precisely, they show that an exact penalty parameter exists whose bit-encoding length is bounded above by a polynomial in the bit-encoding length of the MIQP instance. Finally, in Bhardwaj et al. (2024), the authors consider linearly constrained problems with a convex objective function and generalize previous results to this setting. Another noteworthy contribution is the quantification of the finite penalty parameter for the ALD with a norm penalty function.

1.2. **Main Contributions.** Our main contributions are as follows.

(i) We consider nonconvex MINLPs and show that the ALD asymptotically closes the duality gap under mild assumptions.
(ii) For any finite penalty parameter, we derive bounds on the duality gap between the primal problem (1) and its augmented Lagrangian relaxation (2).
(iii) If the penalty function is a norm, we show the existence of an exact penalty parameter for general MINLPs under mild conditions.
(iv) Additionally, we also show that the sets of optimal solutions of (1) and (2) are equal for a sufficiently large but finite penalty parameter if norm penalty functions are used.
(v) Finally, in the MILP case, we further show that such an exact penalty parameter with a bit-encoding length being polynomial in the size of the input data of the problem can be computed in polynomial time (in the input data of the problem, and for fixed dimension).

Let us highlight that contributions (i) and (iii) generalize previous results from the recent literature, in particular, Boland and Eberhard (2014) (Proposition 3 and Corollary 1), Feizollahi et al. (2016) (Theorem 2 and 4), Gu et al. (2020) (Theorem 10 and 11) and Bhardwaj et al. (2024) (Proposition 4.16 and Theorem 3.3). Our proof techniques could be considered conceptually simpler than those employed in the previous literature since they rely on standard results from convex analysis.

Contribution (ii) generalizes Proposition 4.2 of Gu et al. (2020) from the mixed-integer convex setting with norm penalty functions to general MINLPs with norm-like penalty functions, which include the sharp and the proximal augmented Lagrangian dual; see Section 2. Moreover, Contribution (iv) generalizes Proposition 1 of Feizollahi et al. (2016) from the MILP setting to the MINLP setting. Next, Contribution (v) can be related to Theorem 11 of Gu et al. (2020), which shows the existence of a (complexity-wise) small exact penalty parameter for convex MIQPs. Our result goes beyond this result by showing that such a parameter can be computed in polynomial time for MILPs. We also point out that the result can be easily extended to convex MIQPs.

Finally, the contributions of the current paper and the recent literature on ALD for mixed-integer problems is summarized in Table 1. The first column corresponds to the class of problems considered in a given contribution.[1] Then, "Asymptotic" indicates a positive result on asymptotically closed duality gaps. The "Exactness" column corresponds to an existence result on finite penalty parameters closing the duality gap. The columns "Poly. size" and "Poly. time" indicate an existence result of such a penalty parameter of (complexity-wise) small size, which can be computed in polynomial time. Finally, the "Opt. set" column indicates an existence result for exact penalty parameters such that the optimal sets of the primal problem and its ALR (and hence its ALD) are the same.

The rest of the paper is organized as follows. In Section 2, we formalize and discuss our main assumptions. In Section 3, we show that the ALD closes the duality

[1]ILP: Integer linear problem, MILP: Mixed-integer linear problem, MIQP: Convex mixed-integer quadratic problems, MICP: Mixed-integer convex problem with linear constraints, MINLP: Mixed-integer nonlinear problem.

Table 1. State-of-the-Art of ALD for Mixed-Integer Problems

	Asymptotic	Exactness	Poly. size	Poly. time	Opt. set
ILP (Boland and Eberhard 2014)	✓	✓	⇑	⇑	⇑
MILP (Feizollahi et al. 2016)	✓	✓	⇑	✓	✓
MIQP (Gu et al. 2020)	✓	✓	✓	✓*	⇑
MICP (Bhardwaj et al. 2024)	✓	✓			⇑
MINLP (our paper)	✓	✓			✓

*Checkmarks in white cells refer to contributions of the paper cited in the first column while those in gray cells are contributions of the current paper. An upward arrow indicates a result that was not shown in the corresponding paper but which is a consequence of a later contribution. The asterisk * indicates that the result is only shown for the MILP case but that it easily extends to convex MIQPs.*

gap if the penalty parameter goes to infinity. This result holds for general penalty functions under some mild assumptions. In Section 4, we derive bounds on the duality gap between the primal and its augmented Lagrangian dual. We also show some convergence rate for the sharp and proximal Lagrangian penalty function in the MILP case. In Section 5, we give sufficient conditions for the existence of finite penalty parameters closing the duality gap under the assumption that the penalty function is a norm.

2. Assumptions

We now state our main assumptions.

Assumption 1 (Compactness)**.** *The set X is nonempty and compact. Alternatively, we may assume that there exists a set $\mathcal{E} := \{q + P\zeta : \|\zeta\|_E \leq 1\}$, for some norm $\|\cdot\|_E$, such that at least one solution x^* to Problem (1) satisfies $x^* \in \mathcal{E}$ and that the function g is closed on $X \cap \mathcal{E}$.*

Assumption 2 (Convex Objective Function)**.** *The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is a convex and differentiable function.*

Assumption 3 (Penalty Function)**.** *The penalty function $\psi : \mathbb{R}^m \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is*

(i) closed, i.e., $\operatorname{epi}(\psi)$ is a closed set;
(ii) continuous on $\operatorname{dom}(\psi)$, i.e., $\lim_{u\to u^*} \psi(u) = \psi(u^*)$;
(iii) positive definite, i.e., $\psi(u) > 0$ for all $u \neq 0$ and $\psi(0) = 0$.

Moreover, ψ is assumed to have a nonempty effective domain, i.e., $\operatorname{dom}(\psi) \neq \emptyset$.

Assumption 1 is mild in our setting. Note that compactness of X implies the existence of a set $\mathcal{E}$ satisfying $X \subseteq \mathcal{E}$. We briefly show that the second part of the assumption is weaker than compactness of X. For example, consider a convex MIQP with unbounded feasible region and rational data. Then, by Lemma 4 in Del Pia et al. (2016), any solution x^* (if it exists) satisfies $\|x^*\|_2 \leq M$ for some $M > 0$ of reasonable size, i.e., whose bit-encoding length is bounded from above by a polynomial of the bit-encoding length of the input data. Moreover, all the constraints of an MIQP (other than integrality requirements) are linear and, hence, closed. Thus, convex MIQPs with an unbounded feasible region satisfy Assumption 1 despite X not being compact.

We highlight that Assumption 2 is made without loss of generality since, if the objective function would not be differentiable or if it would be nonconvex, one could resort to an epigraph reformulation to move the objective function f into the constraints g by using one additional variable, i.e., one could consider

$$\min\{t : t \geq f(x), Ax = b, x \in X\} = \min\{f(x) : Ax = b, x \in X\} = z^*.$$

Note that since f is differentiable, hence continuous, and X is nonempty and compact, by the theorem of Weierstraß, $\max_{x\in X} f(x)$ and $\min_{x\in X} f(x)$ exist. Thus, t is bounded.

Finally, Assumption 3 defines the class of penalty functions under consideration and is similar to those in Bhardwaj et al. (2024), Feizollahi et al. (2016), and Gu et al. (2020). We now give two well-known examples of penalty functions satisfying Assumption 3; see, e.g., Rockafellar and Wets (1998).

Definition 2 (Sharp Lagrangian)**.** *Let $\psi = \|\cdot\|$ with $\|\cdot\|$ being any norm. Then, the ALR (2) is called a sharp Lagrangian.*

Definition 3 (Proximal Lagrangian)**.** *Let $\psi = \frac{1}{2}\|\cdot\|_2^2$. Then, the ALR (2) is called a proximal Lagrangian.*

We close this section with some notation that is used throughout the rest of this paper.

Notations. For a given set W, let $\operatorname{int}(W)$ denote the interior of W, $\operatorname{ri}(W)$ denote its relative interior and $\operatorname{conv}(W)$ denote its convex hull, i.e., the smallest convex set such that $W \subseteq \operatorname{conv}(W)$. When clear from the context, we let $\operatorname{proj}_x(W)$ denote the projection of W onto the variables x. For a given function $\zeta : \mathbb{R}^n \to \mathbb{R}$, we denote by $\operatorname{vex}_W(\zeta)$ its convex envelope, i.e., the function such that $\operatorname{epi}_W(\operatorname{vex}_W(\zeta)) = \operatorname{conv}(\operatorname{epi}_W(\zeta))$ where $\operatorname{epi}(\cdot)$ denotes the epigraph over W. The domain of ζ is defined as $\operatorname{dom}(\zeta) := \{z \in \mathbb{R}^n : \zeta(z) < \infty\}$. For a given norm $\|\cdot\|$, its dual norm is denoted by $\|\cdot\|_*$.

Throughout this paper, we use the notation $x = (x_1, x_2)$ where x_1 is the vector made of the n_1 first components of x. We naturally extend this notation to matrices and vectors that multiply x, e.g., $Ax = A_1x_1 + A_2x_2$. Moreover, we may write $f(x) = f(x_1, x_2)$ instead of $f((x_1, x_2))$. We define the sets $X_1 := \operatorname{proj}_{x_1}(X)$ and $X_2 := \operatorname{proj}_{x_2}(X)$.

3. Asymptotic Zero Duality Gap

In this section, we show that the duality gap γ_ρ goes to zero as the penalty parameter ρ goes to infinity. This result holds for quite general penalty functions ψ satisfying Assumption 3. In particular, we highlight that ψ does not need to be nonsmooth or convex for this result to hold. We start with a lemma.

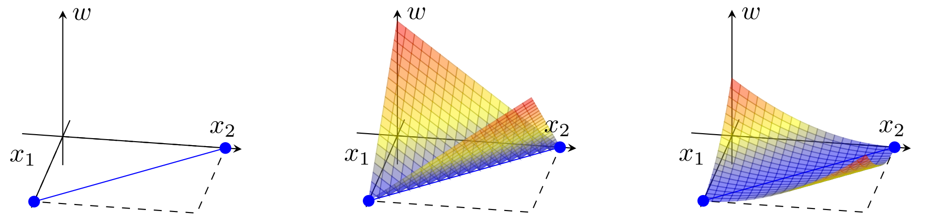
Lemma 1. *Let $\delta \in (0, \infty]$ and let $\tilde{X}_\delta := \{(x, w) \in X \times [0, \delta] : \psi(Ax - b) \leq w\}$. Then,*

$$\operatorname{conv}(\tilde{X}_\delta \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0})) = \operatorname{conv}(\tilde{X}_\delta) \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0}). \tag{5}$$

Proof. The inclusion from left to right is trivial. To show the other direction, let $(x, w) \in \operatorname{conv}(\tilde{X}_\delta)$ be such that $w \leq 0$ and assume that $(x, w) \notin \operatorname{conv}(\tilde{X}_\delta \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0}))$. Note that $w \leq 0$ and $\psi(Ax-b) \leq w$ implies that $w = 0$. Now, by Carathéodory's theorem, since $(x, w) \in \operatorname{conv}(\tilde{X}_\delta)$, there exists $\alpha_1, \ldots, \alpha_{n+2} \geq 0$ and $(\bar{x}^1, \bar{w}^1), \ldots, (\bar{x}^{n+1}, \bar{w}^{n+2}) \in \tilde{X}$ such that $\sum_{k=1}^{n+2} \alpha_k = 1$ and $(x, w) = \sum_{k=1}^{n+2} \alpha_k (\bar{x}^k, \bar{w}^k)$. Note that there must exist a $\bar{k} \in \{1, \ldots, n+2\}$ such that $\alpha_{\bar{k}} > 0$ and $\bar{w}^{\bar{k}} > 0$ since, otherwise, this would show that $(x, w) \in \operatorname{conv}(\tilde{X}_\delta \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0}))$. This, however, is impossible since we then have

$$0 = w = \sum_{k=1}^{n+2} \alpha_k \bar{w}^k = \underbrace{\sum_{k:\bar{w}^k=0} \alpha_k \bar{w}^k}_{=0} + \underbrace{\sum_{k:\bar{w}^k>0} \alpha_k \bar{w}^k}_{>0} > 0. \qquad \square$$

We illustrate this lemma with a small example.



(A) $\operatorname{conv}(X \cap \{x : Ax = b\})$ (B) $\operatorname{conv}(\tilde{X}_\delta)$, $\psi = \|\cdot\|_1$ (C) $\operatorname{conv}(\tilde{X}_\delta)$, $\psi = \frac{1}{2}\|\cdot\|_2^2$

FIGURE 1. Illustration of Lemma 1 and Example 1

Example 1. *Let a mixed-integer linear feasible region be defined as*

$$F := \{x \in \{0,1\}^2 : e^\top x = 1\}.$$

We consider dualizing the constraint "$e^\top x = 1$". Hence, we set $X = \{0,1\}^2$, $A = e^\top$ and $b = 1$. Clearly, $(0,1)$ and $(1,0)$ are the only feasible points. The convex hull of F is given by $\operatorname{conv}(F) = \{x \in [0,1]^2 : e^\top x = 1\}$ and is depicted in Figure 1a. In Figure 1b and Figure 1c, the set $\tilde{X}_\delta$ is depicted for the sharp Lagrangian ($\psi = \|\cdot\|_1$) and the proximal Lagrangian ($\psi = \frac{1}{2}\|\cdot\|_2^2$). It is easily seen that fixing $w = 0$ reduces $\operatorname{conv}(\tilde{X}_\delta)$ to $\operatorname{conv}(F)$ (in a higher dimensional space), i.e., (5) holds.

In the next theorem, we exploit Lemma 1 to show that the best lower bound, which can be obtained by the augmented Lagrangian relaxation $z^{\mathrm{LR}+}$, evaluates to z^*, independent of the choice of λ.

Theorem 2. *For all $\bar{\lambda} \in \mathbb{R}^m$ it holds*

$$z^* = \sup_{\rho \in \mathbb{R}_{\geq 0}} z_\rho^{LR+}(\bar{\lambda}).$$

Proof. Let us define δ as the maximum penalization for violating the constraints "$Ax = b$" by a point $x \in X$, i.e., let $\delta := \max\{\psi(Ax - b) : x \in X \cap \operatorname{dom}(\psi(A(\cdot) - b))\}$. Note that, by Assumption 3, ψ is continuous and that $\operatorname{dom}(\psi(A(\cdot) - b))$ is closed. Moreover, by Assumption 1, X is compact. Thus, δ is finite. Hence,

$$\begin{aligned} z_\rho^{\mathrm{LR}+}(\bar{\lambda}) &= \min_{x \in X} \; f(x) + \bar{\lambda}^\top (Ax - b) + \rho\psi(Ax - b) \\ &= \min_{(x,w) \in \tilde{X}_\delta} \; f(x) + \bar{\lambda}^\top (Ax - b) + \rho w. \end{aligned}$$

We also have that ψ is closed. Thus, $\tilde{X}_\delta$ is compact. Therefore, by Proposition 2.4 in Tardella (2004), it holds

$$z_\rho^{\mathrm{LR}+}(\bar{\lambda}) = \min_{(x,w) \in \operatorname{conv}(\tilde{X}_\delta)} \; \operatorname{vex}_X(f)(x) + \bar{\lambda}^\top (Ax - b) + \rho w.$$

Note that we used $\operatorname{proj}_x(\tilde{X}_\delta) = X$ to argue that $\operatorname{vex}_{\tilde{X}_\delta}((x,w) \mapsto f(x))(x,w) = \operatorname{vex}_X(f)(x)$. Now, by Theorem 1 in Perchet and Vigeral (2015), the following equalities hold:

$$\begin{aligned} \sup_{\rho \in \mathbb{R}_{\geq 0}} z_\rho^{\mathrm{LR}+}(\bar{\lambda}) &= \sup_{\rho \in \mathbb{R}_{\geq 0}} \min_{(x,w) \in \operatorname{conv}(\tilde{X}_\delta)} \; \operatorname{vex}_X(f)(x) + \bar{\lambda}^\top (Ax - b) + \rho w \\ &= \min_{(x,w) \in \operatorname{conv}(\tilde{X}_\delta)} \sup_{\rho \in \mathbb{R}_{\geq 0}} \; \operatorname{vex}_X(f)(x) + \bar{\lambda}^\top (Ax - b) + \rho w. \end{aligned}$$

By Assumption 1 and Inequality (4), it must be that $w \leq 0$ holds in any optimal point of the last minimization problem. Thus, it also equals

$$\min_{(x,0) \in \operatorname{conv}(\tilde{X}_\delta)} \; \operatorname{vex}_X(f)(x) + \bar{\lambda}^\top (Ax - b),$$

which, by Lemma 1, is also equal to

$$\min_{(x,0)\in\operatorname{conv}(\tilde{X}_\delta\cap(\mathbb{R}^n\times\mathbb{R}_{\le 0}))} \operatorname{vex}_X(f)(x)+\bar{\lambda}^\top(Ax-b).$$

Now, applying Proposition 2.4 from Tardella (2004) again, we obtain

$$\min_{(x,0)\in\tilde{X}_\delta} f(x)+\bar{\lambda}^\top(Ax-b).$$

The proof is achieved by noticing that $(x,0)\in\tilde{X}_\delta$ implies $\psi(Ax-b)=0$, which, in turn, implies $Ax=b$ and, thus, $\bar{\lambda}^\top(Ax-b)=0$. □

A direct consequence of Theorem 2 is the following limit result, which states that the augmented Lagrangian relaxation converges to z^* as ρ approaches infinity.

Theorem 3. *For all $\bar{\lambda}\in\mathbb{R}^m$ it holds*

$$z^*=\lim_{\rho\to\infty} z_\rho^{LR+}(\bar{\lambda}).$$

Proof. Let $\bar{\lambda}\in\mathbb{R}^m$ be fixed and define $\pi_{\bar{\lambda}}:\rho\mapsto z_\rho^{\mathrm{LR}+}(\bar{\lambda})$. Then, $\pi_{\bar{\lambda}}$ is a continuous function because it can be expressed as a minimum of affine functions. Moreover, it is non-decreasing since, for all ρ_1 and ρ_2 with $\rho_1\le\rho_2$, it holds

$$f(x)+\bar{\lambda}^\top(Ax-b)+\rho_1\psi(Ax-b)\le f(x)+\bar{\lambda}^\top(Ax-b)+\rho_2\psi(Ax-b) \tag{7}$$

for all $x\in X$ (since $\psi\ge 0$), implying that $z_{\rho_1}^{\mathrm{LR}+}(\bar{\lambda})\le z_{\rho_2}^{\mathrm{LR}+}(\bar{\lambda})$, i.e., $\pi_{\bar{\lambda}}(\rho_1)\le\pi_{\bar{\lambda}}(\rho_2)$. By Inequality (4), $\pi_{\bar{\lambda}}$ is bounded by z^*. Thus, the limit exists and coincides with the supremum. By Theorem 2, the supremum is exactly z^*. □

From this last theorem, one easily concludes that the duality gap between the primal problem (1) and the augmented Lagrangian dual converges to zero as the penalty parameter ρ goes to infinity. This is established in the next corollary, which directly follows from Theorem 3 and Inequality (4).

Corollary 1. *Let $\gamma_\rho := z^*-z_\rho^{LD+}$. It holds*

$$\lim_{\rho\to\infty}\gamma_\rho=0.$$

4. Gap Guarantees for Finite Penalty Parameters

In the previous section, we show that an infinite penalty parameter can close the duality gap between the augmented Lagrangian dual and the primal problem (1). In this section, we study the case in which the penalty parameter ρ is chosen to be finite and derive guarantees on the quality of the lower bound provided by the augmented Lagrangian dual. This is particularly interesting since Feizollahi et al. (2016) show, even in the MILP setting, that the proximal Lagrangian dual cannot guarantee to close the duality gap with a finite parameter. However, they do not provide bounds for the gap for finite penalty parameters.

4.1. A Perturbed Value Function.

We start by introducing the perturbed value function $p:\mathbb{R}^m\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ defined by

$$p(\lambda,u):=\min_{x,w}\ \operatorname{vex}_X(f)(x)+\lambda^\top(Ax-b) \tag{8a}$$

$$\text{s.t. } w\le u, \tag{8b}$$

$$(x,w)\in\operatorname{conv}(\tilde{X}_\delta), \tag{8c}$$

where $\operatorname{vex}_X(f)$ denotes the convex envelope of f on X. We make the following remark.

Remark 1. *For all $\lambda\in\mathbb{R}^m$, it holds $z^*=p(\lambda,0)$.*

Proof. This directly follows from Lemma 1, since

$$p(\lambda, 0) = \min_{x,w} \ \mathrm{vex}_X(f)(x) + \lambda^\top(Ax - b) \quad \text{s.t. } (x, w) \in \mathrm{conv}(\tilde{X}_\delta \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0})).$$

By Tardella (2004), we have

$$p(\lambda, 0) = \min_{x,w} \ f(x) + \lambda^\top(Ax - b) \quad \text{s.t. } (x, w) \in \tilde{X}_\delta \cap (\mathbb{R}^n \times \mathbb{R}_{\leq 0}),$$

from which we derive $\psi(Ax - b) \leq 0$, which implies $Ax = b$ and $\lambda^\top(Ax - b) = 0$. □

We now state a theorem that puts into relation the perturbed value function $p(\lambda, u)$ and the augmented Lagrangian relaxation $z_\rho^{\mathrm{LR+}}(\lambda)$.

Theorem 4. *Suppose that there exists a point $\tilde{x} \in \mathrm{ri}(\mathrm{conv}(X))$ such that $A\tilde{x} = b$ holds and let $\rho^*(u)$ denote an optimal Lagrange multiplier of Constraint (8b). Then, for all fixed $\bar{\lambda} \in \mathbb{R}^m$, for all $u > 0$, and for all $\rho \geq \rho^*(u)$, it holds*

$$p(\bar{\lambda}, u) \leq z_\rho^{LR+}(\bar{\lambda}) \leq z^*. \tag{11}$$

Proof. Let $\lambda \in \mathbb{R}^m$ and $u > 0$ be fixed. We first show that the optimization problem in (8) is strictly feasible, i.e., there exists $(\hat{x}, \hat{w}) \in \mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta))$ such that $w \leq u$. By assumption, there exists $\tilde{x} \in \mathrm{ri}(\mathrm{conv}(X))$ such that $A\tilde{x} = b$ holds, which implies that $(\tilde{x}, 0) \in \mathrm{conv}(\tilde{X}_\delta)$. Note that $\mathrm{proj}_w(\tilde{X}_\delta) \subseteq \mathbb{R}_{\geq 0}$. Thus, it cannot be that $(\tilde{x}, 0) \in \mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta))$. Yet, we show that there exists $(\bar{x}, \bar{w}) \in \mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta))$. From this, it will be easy to show that there exists an $\alpha \in (0, 1)$ such that $(\hat{x}, \hat{w}) := \alpha(\tilde{x}, 0) + (1 - \alpha)(\bar{x}, \bar{w})$ with $\hat{w} \leq u$ and $(\hat{x}, \hat{w}) \in \mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta))$. To show the existence of $(\bar{x}, \bar{w})$ it suffices to take $\bar{x} = \tilde{x}$. Indeed, since $\tilde{x} \in \mathrm{ri}(\mathrm{conv}(X))$ and $\mathrm{proj}_x(\mathrm{conv}(\tilde{X}_\delta)) = \mathrm{conv}(X)$, we have $\tilde{x} \in \mathrm{ri}(\mathrm{proj}_x(\mathrm{conv}(\tilde{X}_\delta))) = \mathrm{proj}_x(\mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta)))$, where the last equality holds by Theorem 6.6 in Rockafellar (1970). By definition of $\tilde{x} \in \mathrm{proj}_x(\mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta)))$, there exists $\bar{w}$ such that $(\bar{x}, \bar{w}) \in \mathrm{ri}(\mathrm{conv}(\tilde{X}_\delta))$.

From this, we conclude that strong duality holds for Problem (8); i.e., there exists a ρ^* with

$$\begin{aligned} p(\lambda, u) &= \max_{\rho \geq 0} \min_{(x,w) \in \mathrm{conv}(\tilde{X}_\delta)} \ \mathrm{vex}_X(f)(x) + \lambda^\top(Ax - b) + \rho(w - u) \\ &= \min_{(x,w) \in \mathrm{conv}(\tilde{X}_\delta)} \ \mathrm{vex}_X(f)(x) + \lambda^\top(Ax - b) + \rho^*(w - u). \end{aligned}$$

Yet, since $u > 0$ and $\rho \geq 0$, it holds

$$\begin{aligned} p(\lambda, u) &\leq \min_{(x,w) \in \mathrm{conv}(\tilde{X}_\delta)} \ \mathrm{vex}_X(f)(x) + \lambda^\top(Ax - b) + \rho^* w \\ &= \min_{x \in X} \ f(x) + \lambda^\top(Ax - b) + \rho^* \psi(Ax - b) \\ &= z_{\rho^*}^{\mathrm{LR+}}(\lambda). \end{aligned}$$

The proof is completed by the simple observation that $z_{\rho^*}^{\mathrm{LR+}}(\lambda) \leq z_\rho^{\mathrm{LR+}}(\lambda)$ holds for all $\rho \geq \rho^*$. □

In the following remark, we highlight that Theorem 4 can also be used to prove Theorem 3 by using the lower-semicontinuity of $p(\bar{\lambda}, \cdot)$ for any fixed $\bar{\lambda} \in \mathbb{R}^m$.

Remark 2. *Let $\bar{\lambda} \in \mathbb{R}^m$ be fixed. Then, $u \mapsto p(\bar{\lambda}, u)$ is a lower semi-continuous function; see Theorem 5.1 in Still (2018). Thus, by definition, it holds*

$$\liminf_{u \to 0} \ p(\bar{\lambda}, u) \geq p(\bar{\lambda}, 0),$$

which, together with Inequality (11) shows that

$$p(\lambda, 0) \leq \liminf_{u \to 0} \ p(\bar{\lambda}, u) \leq \lim_{\rho \to \infty} \ z_\rho^{LR+}(\bar{\lambda}) \leq p(\lambda, 0)$$

holds and $p(\lambda, 0) = z^*$*. Here, we used the fact that* $\pi_{\bar{\lambda}} : \rho \mapsto z_\rho^{LR+}(\bar{\lambda})$ *is continuous so that* $\liminf_{\rho\to\infty} z_\rho^{LR+}(\bar{\lambda}) = \lim_{\rho\to\infty} z_\rho^{LR+}(\bar{\lambda})$.

A clear drawback of Theorem 4 is that the relation between the perturbed problem (8) and the original problem (1) is not straightforward. We therefore aim for a more informative theorem. In the next theorem, we first show that the penalty parameter can be used to control the penalization of an optimal point of (2). To be numerically applied, this theorem requires that the set $\mathcal{E}$ such that $x^* \in X \cap \mathcal{E}$ for at least one solution x^* to Problem (1) is known; see Assumption 1. Before stating the theorem, we give an example in which $\mathcal{E}$ can be computed easily.

Example 5. *Consider the case in which* X *is explicitly bounded, i.e.,* $X \subseteq [\underline{x}, \bar{x}]$*. Then, setting* $q = (\bar{x} + \underline{x})/2$*,* $P = I(\bar{x} - \underline{x})/2$*, and* $\|\cdot\|_E = \|\cdot\|_\infty$ *leads to*

$$X \subseteq [\underline{x}, \bar{x}] = \{q + P\zeta : \|\zeta\|_\infty \leq 1\} =: \mathcal{E}.$$

Theorem 6. *Let* $\bar{\lambda} \in \mathbb{R}^m$ *be fixed and let* x_ρ^* *denote a solution to the augmented Lagrangian relaxation (2) for a fixed penalty parameter* $\rho > 0$ *as well as* $\lambda = \bar{\lambda}$ *and and let* x^* *denote a solution to the primal problem (1). Let* $\varepsilon > 0$ *be given. Then, the inequality*

$$\psi(Ax_\rho^* - b) \leq \varepsilon$$

holds for all $\rho \geq \kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda})/\varepsilon$ *with*

$$\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda}) := \|P^\top \nabla f(x^*)\|_{E_*} + \|P^\top(\nabla f(x^*) + A^\top \bar{\lambda})\|_{E_*} + \bar{\lambda}^\top(b - Aq),$$

where $\|\cdot\|_{E_*}$ *is the dual norm of* $\|\cdot\|_E$*.*

Proof. Let $\bar{\lambda} \in \mathbb{R}^m$ be arbitrary but fixed. For all $\rho \geq 0$, $z_\rho^{\mathrm{LR}+}(\bar{\lambda}) \leq z^*$ holds; see (4). Equivalently,

$$f(x_\rho^*) + \bar{\lambda}^\top(Ax_\rho^* - b) + \rho\psi(Ax_\rho^* - b) \leq f(x^*) \tag{14a}$$

$$\iff \quad \rho\psi(Ax_\rho^* - b) \leq f(x^*) - f(x_\rho^*) + \bar{\lambda}^\top(b - Ax_\rho^*) \tag{14b}$$

$$\implies \quad \rho\psi(Ax_\rho^* - b) \leq \nabla f(x^*)^\top(x^* - x_\rho^*) + \bar{\lambda}^\top(b - Ax_\rho^*). \tag{14c}$$

The last inequality holds by convexity and differentiability of f; see Assumption 2. By Assumption 1, we then have

$$\rho\psi(Ax_\rho^* - b) \leq \nabla f(x^*)^\top x^* - (\nabla f(x^*) + A^\top\bar{\lambda})^\top x_\rho^* + \bar{\lambda}^\top b$$

$$\implies \quad \rho\psi(Ax_\rho^* - b) \leq \max_{x\in\mathcal{E}} \nabla f(x^*)^\top x + \max_{x\in\mathcal{E}} -(\nabla f(x^*) + A^\top\bar{\lambda})^\top x.$$

Now, by definition of dual norms, for all $\pi \in \mathbb{R}^n$, it holds

$$\max_{x\in\mathcal{E}} \ \pi^\top x = \max_{\|\zeta\|_E \leq 1} \ \pi^\top(q + P\zeta) = q^\top\pi + \|P^\top\pi\|_{E_*}.$$

Thus, with $\pi = \nabla f(x^*)$ and $\pi = -(\nabla f(x^*) + A^\top\bar{\lambda})$, we conclude

$$\max_{x\in\mathcal{E}} \nabla f(x^*)^\top x = \nabla f(x^*)^\top q + \|P^\top \nabla f(x^*)\|_{E*},$$

$$\max_{x\in\mathcal{E}} -(\nabla f(x^*) + A^\top\bar{\lambda})^\top x = -(\nabla f(x^*) + A^\top\bar{\lambda})^\top q + \|P^\top(\nabla f(x^*) + A^\top\bar{\lambda})^\top\|_{E*}.$$

Combining these results with (14c), we obtain

$$\rho\psi(Ax_\rho^* - b) \leq \underbrace{\|P^\top\nabla f(x^*)\|_{E_*} + \|P^\top(\nabla f(x^*) + A^\top\bar{\lambda})\|_{E_*} + \bar{\lambda}^\top(b - A^\top q)}_{=:\kappa_1(\mathcal{E},A,b,\nabla f(x^*),\bar{\lambda})}.$$

The proof is achieved by the implications

$$\begin{aligned} & \rho \geq \frac{1}{\varepsilon}\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda}) \\ \implies \quad & \rho\psi(Ax^*_\rho - b) \geq \frac{1}{\varepsilon}\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda})\psi(Ax^*_\rho - b) \\ \implies \quad & \kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda}) \geq \frac{1}{\varepsilon}\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda})\psi(Ax^*_\rho - b) \\ \implies \quad & \psi(Ax^*_\rho - b) \leq \varepsilon. \end{aligned}$$

□

Remark 3. *Since f is a convex function, it is Lipschitz continuous over any compact set. Hence, it holds $\|\nabla f(x^*)\|_{E*} \leq L$ with $L > 0$ being the Lipschitz constant of f over X w.r.t. $\|\cdot\|_{E*}$. Hence, it is not necessary to know x^* for Theorem 6 to be applied.*

In the next section, we show how Theorem 6 can be further exploited to relate solutions to the augmented Lagrangian relaxation (2) to those of a more "natural" perturbed problem compared to (8).

4.2. **Norm-Like Penalty Functions.** Theorem 6 is particularly interesting for cases in which ψ has some relation to a given norm. In such a case, it is possible to relate solutions to the augmented Lagrangian relaxation (2) to those of the perturbed problem

$$\tilde{z}^*_{\varepsilon,\|\cdot\|} := \min_x \; f(x) \tag{15a}$$
$$\text{s.t. } \|Ax - b\| \leq \varepsilon, \tag{15b}$$
$$x \in X. \tag{15c}$$

Arguably, the relation between the perturbed problem (15) and the original problem (1) is much more natural compared to the relation between (8) and (1).

In the next definition, we formalize the concept of norm-like penalty functions. Note that such functions have already been used by Boland et al. (2012) in the study of the feasibility pump heuristic (Fischetti et al. 2005) for mixed-integer problems under the name of "integer compatible regularization function". More recently, they are also used by Fabiani et al. (2022) in the context of mixed-integer Nash equilibria.

Definition 4 (Norm-Like Penalty Functions)**.** *Let $\|\cdot\|_\varphi$ be a given norm. A penalty function ψ is called $\|\cdot\|_\varphi$-norm-like if and only if there exists a bijection $\varphi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for all $\varepsilon \geq 0$ and any $x \in \mathbb{R}^n$, it holds*

$$\psi(x) \leq \varepsilon \implies \|x\|_\varphi \leq \varphi(\varepsilon).$$

Obviously, any norm is a norm-like penalty function. We now show another example of a norm-like penalty function, namely, proximal functions. This example is particularly relevant since Feizollahi et al. (2016) shows that, even in the mixed-integer linear setting, such a penalty function cannot guarantee to close the duality gap.

Example 7 (Proximal Penalty Function)**.** *Assume $\psi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$. Clearly, we have that $\frac{1}{2}\|x\|_2^2 \leq \varepsilon$ implies $\|x\|_2 \leq \sqrt{2\varepsilon}$ for all $\varepsilon \geq 0$. Thus, the proximal Lagrangian penalty function is a norm-like penalty function.*

A direct consequence of Theorem 6 is that the penalty parameter ρ can be used to control the violation of the dualized constraints "$Ax = b$" by optimal points of the augmented Lagrangian relaxation (2), provided that ψ is norm-like. Hence, we get the two following corollaries.

Corollary 2. *Assume that ψ is $\|\cdot\|_\varphi$-norm-like. Let $\bar\lambda \in \mathbb{R}^m$ be fixed and let x^*_ρ denote an optimal point of (2) for a fixed penalty parameter $\rho > 0$ and $\lambda = \bar\lambda$. Let $\varepsilon > 0$. Then, the inequality*

$$\tilde z_{\varepsilon,\|\cdot\|_\varphi} \le z_\rho^{LR+}(\bar\lambda) \le z^*$$

holds for all $\rho \ge \kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar\lambda)/\varphi^{-1}(\varepsilon)$.

Proof. According to Theorem 6, ρ is such that $\psi(Ax^*_\rho - b) \le \varphi^{-1}(\varepsilon)$ holds. Since ψ is $\|\cdot\|_\varphi$-norm-like, it holds $\|Ax^*_\rho - b\|_\varphi \le \varphi(\varphi^{-1}(\varepsilon)) = \varepsilon$. Thus, x^*_ρ is feasible for (15). □

Corollary 3. *Assume that ψ is $\|\cdot\|_\varphi$-norm-like. Let $\bar\lambda \in \mathbb{R}^m$ be fixed and let x^*_ρ denote an optimal point of (2) for a fixed penalty parameter $\rho > 0$ and $\lambda = \bar\lambda$. Let $\varepsilon > 0$. Then, the inequality*

$$\tilde z_{\varepsilon,\|\cdot\|_\infty} \le z_\rho^{LR+}(\bar\lambda) \le z^*$$

holds for all $\rho \ge \kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar\lambda)/\varphi^{-1}(C_\varphi\varepsilon)$ with C_φ being a constant such that $C_\varphi\|\cdot\|_\infty \le \|\cdot\|_\varphi$ holds.

Proof. According to Corollary 2, ρ is such that $\tilde z_{C_\varphi\varepsilon,\|\cdot\|_\varphi} \le z_\rho^{\mathrm{LR}+}(\bar\lambda)$ holds. Moreover,

$$\{x \in X : \|Ax - b\|_\varphi \le C_\varphi\varepsilon\} \subseteq \{x \in X : \|Ax - b\|_\infty \le \varepsilon\},$$

which shows $\tilde z_{C_\varphi\varepsilon,\|\cdot\|_\varphi} \ge \tilde z_{\varepsilon,\|\cdot\|_\infty}$. □

Although Corollary 2 and 3 are very simple, they are key to proving the next theorem. Assuming ψ is $\|\cdot\|_\varphi$-norm-like and that (1) is a linear problem if x_2 is temporarily held fixed and appears in the right-hand side, we show that φ can be used to bound the distance between $z_\rho^{\mathrm{LR}+}(\lambda)$ and z^*.

Theorem 8. *Assume that ψ is $\|\cdot\|_\varphi$-norm-like. Let $c \in \mathbb{Z}^n$ and $T \in \mathbb{Z}^{\ell\times n}$ be a given vector, resp., matrix and let $\tilde g_2 : \mathbb{R}^{n_2} \to \mathbb{R}^\ell$. Suppose $f(x) = c^\top x = c_1^\top x_1 + c_2^\top x_2$ and $g(x_1, x_2) = T_1x_1 + \tilde g_2(x_2)$. Then, for all $\rho > 0$, the inequality*

$$z^* - \frac{1}{C_\varphi}\kappa_2(\mathcal{E}, C_E, A, b, T_1, \theta)\varphi(\kappa_1(\mathcal{E}, A, b, c, \bar\lambda)/\rho) \le z_\rho^{LR+}(\bar\lambda) \le z^*$$

holds for some constant $\kappa_2(\mathcal{E}, C_E, A, b, T_1, \theta)$ with C_E such that $\|x\|_\infty \le C_E\|x\|_E$ holds for all $x \in \mathbb{R}^m$ and $\theta := \max\{\|\tilde g_2(x_2)\|_\infty : x_2 \in X_2\}$.

Proof. Let $\rho > 0$ be fixed. Note that $\nabla f(x^*) = c$. According to Corollary 3, ρ guarantees that $\tilde z^*_{\tilde\varepsilon,\|\cdot\|_\infty} \le z_\rho^{\mathrm{LR}+}(\bar\lambda)$ with $\tilde\varepsilon := \varphi(\kappa_1(\mathcal{E}, A, b, c, \bar\lambda)/\rho)/C_\varphi$. To ease readability, we temporarily drop the subscript $\|\cdot\|_\infty$. We now adapt a proof by Beck et al. (2023) to show that there exists κ' such that $\tilde z^*_{\tilde\varepsilon} \ge z^* - \tilde\varepsilon\kappa'$. To this end, let us use the notation $\tilde z_\varepsilon(x_2)$ for

$$\tilde z_\varepsilon(x_2) := c_2^\top x_2 + \min_{x_1} c_1^\top x_1 \tag{16a}$$

$$\text{s.t. } \|A_1x_1 + A_2x_2 - b\|_\infty \le \varepsilon, \tag{16b}$$

$$T_1x_1 \le -\tilde g(x_2) \tag{16c}$$

for some $x_2 \in X_2$. Note that $\tilde z^*_\varepsilon = \min_{x_2\in X_2} \tilde z_\varepsilon(x_2)$ and $\tilde z_0 = z^*$. Thus, it suffices to show that there exists κ' such that $\tilde z_{\tilde\varepsilon}(x_2) \ge \tilde z_0(x_2) - \tilde\varepsilon\kappa'$ for all $x_2 \in X_2$, because then, this implies

$$\begin{aligned} & \tilde z_{\tilde\varepsilon}(x_2) \ge \tilde z_0(x_2) - \tilde\varepsilon\kappa' \quad \forall x_2 \in X_2 \\ \Longrightarrow\quad & \min_{x_2\in X_2} \tilde z_{\tilde\varepsilon}(x_2) \ge \min_{x_2\in X_2} \{\tilde z_0(x_2)\} - \tilde\varepsilon\kappa' \\ \Longleftrightarrow\quad & \tilde z_{\tilde\varepsilon} \ge z^* - \tilde\varepsilon\kappa'. \end{aligned}$$

Now, (16) can be formulated as the linear program

$$\begin{aligned} c_2^\top x_2 + \min_{x_1} \ & c_1^\top x_1 \\ \text{s.t. } & A_1 x_1 \le \varepsilon e + b - A_2 x_2, \\ & -A_1 x_1 \le \varepsilon e - b + A_2 x_2, \\ & T_1 x_1 \le -\tilde{g}(x_2). \end{aligned}$$

Following the argument in the proof of Lemma 1 in Beck et al. (2023), we consider a solution $(\alpha^*, \beta^*, \gamma^*)$ to the LP dual of (16) with $\varepsilon = 0$, i.e., let $(\alpha^*, \beta^*, \gamma^*)$ be a solution to

$$\begin{aligned} \max_{\alpha,\beta,\gamma} \ & (\alpha - \beta)^\top (b - A_2 x_2) - \gamma^\top \tilde{g}_2(x_2) \\ \text{s.t. } & A_1^\top (\alpha - \beta) + T_1^\top \gamma = c_1, \\ & \alpha, \beta, \gamma \le 0. \end{aligned}$$

Moreover, let $x^{\tilde{\varepsilon}}$ denote a solution to (16) with $\varepsilon = \tilde{\varepsilon}$. Then, for all fixed $x_2 \in V$, the following holds:

$$\begin{aligned} \tilde{z}_{\tilde{\varepsilon}}(x_2) &= c_1^\top x_1^{\tilde{\varepsilon}} + c_2^\top x_2 \\ &= (A_1^\top (\alpha^* - \beta^*) + T_1^\top \gamma^*)^\top x_1^{\tilde{\varepsilon}} + c_2^\top x_2 \\ &= (\alpha^* - \beta^*)^\top (A_1 x_1^{\tilde{\varepsilon}}) + {\gamma^*}^\top (T_1 x_1^{\tilde{\varepsilon}}) + c_2^\top x_2 \\ &\ge (\alpha^* - \beta^*)^\top (b - A_2 x_2 + \tilde{\varepsilon} e) - {\gamma^*}^\top \tilde{g}_2(x_2) + c_2^\top x_2 \\ &= \tilde{z}_0(x_2) + \tilde{\varepsilon}(\|\beta^*\|_1 - \|\alpha^*\|_1) \\ &\ge \tilde{z}_0(x_2) - \tilde{\varepsilon}\|\alpha^*\|_1. \end{aligned}$$

Thus, we achieve the proof by bounding $\|\alpha^*\|_1$. Note that $(\alpha^*, \beta^*, \gamma^*)$ is a solution to a linear problem, which, w.l.o.g, is an extreme point of its feasible region. By Lemma 4 in Buchheim (2023), it holds

$$\|\alpha^*\|_1 \le (2m + \ell)! \max\{\|b - A_2 x_2\|_\infty, \|\tilde{g}_2(x_2)\|_\infty\} \max\{\|A_1\|_\infty, \|T_1\|_\infty\}^{2m+\ell-1}.$$

Then, it remains to argue that $\|b - A_2 x_2\|_\infty \le \|b\|_\infty + \|A_1\|_\infty \|x_2\|_\infty$ and, since $X_2 \subseteq \operatorname{proj}_{x_2}(\mathcal{E})$, that

$$\|x_2\|_\infty \le C_E \|x_2\|_E \le C_E \left(\max_{x_2' \in \operatorname{proj}_{x_2}(\mathcal{E})} \|x_2'\|_E \right) \le C_E \left(\|q\|_E + \|P\|_E \right)$$

holds, where we again use that $\|q + P x_2\|_E \le \|q\|_E + \|P\|_E \|x_2\|_E$ and $\|x_2\|_E \le 1$ hold for all $x_2 \in \operatorname{proj}_{x_2}(\mathcal{E})$. Additionally, for all $x_2 \in X_2$,

$$\|\tilde{g}_2(x_2)\|_\infty \le \max_{x_2 \in X_2} \|\tilde{g}(x_2)\|_\infty =: \theta.$$

This shows that $\|\alpha^*\|_1 \le \kappa_2(\mathcal{E}, C_E, A, b, T, \theta)$. □

Remark 4. *In Theorem 8, let us further assume that* $\tilde{g}_2(x_2) = T_2 x_2 + r$. *Then,* $\kappa_2(\mathcal{E}, C_E, A, b, T_1, \theta)$ *can be replaced by* $\kappa_2'(\mathcal{E}, C_E, A, b, T, r)$ *with* $T = (T_1, T_2)$.

Proof. It holds

$$\theta = \max_{x_2 \in X_2} \|T_2 x_2 + r\|_\infty \le \|r\|_\infty + \|T_2\|_\infty \left(\max_{x \in \operatorname{proj}_{x_2}(\mathcal{E})} \|x\|_\infty \right). \qquad \square$$

Applying Theorem 8 to norms and to the proximal penalty function leads to the following examples, which are generalizations of Theorem 2 in Feizollahi et al. (2016) for the MILP setting.

Example 9 (Approximation Guarantees)**.** *Assume that* $f(x) = c^\top x$ *and* $g(x) = Tx + r$ *for some rational vectors and matrices* c*,* T*, and* r*. Let* $\bar{\lambda} \in \mathbb{R}^m$ *be fixed. Then, for all* $\rho > 0$*, the following holds.*

- *If* $\psi = \|\cdot\|$*, then*
$$z^* - o\left(\frac{1}{\rho}\right) \leq z_\rho^{LR+}(\bar{\lambda}) \leq z^*.$$
- *If* $\psi = \frac{1}{2}\|\cdot\|_2^2$*, then*
$$z^* - o\left(\frac{1}{\sqrt{\rho}}\right) \leq z_\rho^{LR+}(\bar{\lambda}) \leq z^*.$$

Note that the square root is due to the square root appearing in Example 7.

5. Exact Penalty Parameter

In this section, we study conditions for the existence of an exact penalty parameter. Before doing so, we show that, if a finite penalty parameter exists such that the value of the primal problem (1) and the one of its augmented Lagrangian relaxation (2) are equal, then one also exists so that they share the same set of (global) optimal points. Hence, all our existence theorems for an exact penalty parameter extend to existence of such parameters so that Problem (1) and (2) are fully equivalent.

Theorem 10. *Let* S^* *denote the set of optimal points of the primal problem* (1) *and let* $S^*_\rho(\lambda)$ *denote the set of optimal points of the augmented Lagrangian relaxation* (2)*. Let* $\lambda \in \mathbb{R}^m$ *be arbitrary but fixed and assume that there exists a finite exact penalty parameter* ρ*, i.e., assume that there exists* $\rho < \infty$ *such that* $z^* = z_\rho^{LR+}(\lambda)$*, then, for any* $\rho' > \rho$*, it holds*
$$S^* = S^*_{\rho'}(\lambda).$$
This, in particular, implies
$$z^* = z_{\rho'}^{LR+}(\lambda).$$

Proof. Let ρ be an exact penalty parameter and let $\rho' > \rho$. We first show $S^*_{\rho'}(\lambda) \subseteq S^*$. To this end, consider a point $x^+ \in S^*_{\rho'}(\lambda)$. Note that it is sufficient to show that x^+ satisfies $Ax^+ = b$ since, then, it follows that
$$z_\rho^{\mathrm{LR}+}(\bar{\lambda}) = f(x^+) + \underbrace{\bar{\lambda}^\top(Ax^+ - b) + \rho'\psi(Ax^+ - b)}_{=0} = f(x^+),$$
which implies
$$\begin{aligned} & f(x^+) \leq f(x) + \bar{\lambda}^\top(Ax - b) + \rho'\psi(Ax - b) \quad \text{for all } x \in X \\ \implies\ & f(x^+) \leq f(x) \quad \text{for all } x \in X \text{ with } Ax = b \\ \iff\ & x^+ \text{ is a solution to (1).} \end{aligned}$$
Hence, for the sake of contradiction, suppose that $Ax^+ \neq b$, i.e., $\psi(Ax^+ - b) > 0$. It then holds
$$\begin{aligned} z^* = z_\rho^{\mathrm{LR}+}(\lambda) &\leq f(x^+) + \lambda^\top(Ax^+ - b) + \rho\psi(Ax^+ - b) \\ &< f(x^+) + \lambda^\top(Ax^+ - b) + \rho'\psi(Ax^+ - b) \\ &= z_{\rho'}^{\mathrm{LR}+}(\lambda). \end{aligned}$$
The first inequality holds by $x^+ \in X$ (since $x^+ \in S^*_{\rho'}(\lambda)$) which shows that x^+ a feasible point of Problem (2) with penalty parameter ρ. The strict inequality follows from $\psi(Ax^+ - b) > 0$ and $\rho < \rho'$. Hence, we reach the contradiction that $z^* < z_{\rho'}^{\mathrm{LR}+}(\lambda)$ despite $z_{\rho'}^{\mathrm{LR}+}(\lambda)$ being a lower bound on z^*; see Inequality (4). Hence, it must be that $Ax^+ = b$.

We end the proof by showing $S^* \subseteq S^*_{\rho'}(\bar{\lambda})$. To this end, let $x^* \in S^*$ and assume, for the sake of contradiction, that $x^* \notin S^*_{\rho'}(\bar{\lambda})$. Then, for all $x^+ \in S^*_{\rho'}(\bar{\lambda})$,

$$f(x^+) + \bar{\lambda}^\top(Ax^+ - b) + \rho'\psi(Ax^+ - b) < f(x^*) + \bar{\lambda}^\top(Ax^* - b) + \rho'\psi(Ax^* - b).$$

Yet, as shown above, both x^+ and x^* are feasible for (1). Thus, we conclude that $f(x^+) < f(x^*)$, which contradicts $x^* \in S^*$. Hence, $S^* \subseteq S^*_{\rho'}(\bar{\lambda})$. □

5.1. **General Penalty Functions.** In the next theorem, we give a sufficient condition for the existence of a finite exact penalty parameter. This result holds for the general class of penalty functions satisfying Assumption 3 and generalizes Theorem 3 of Feizollahi et al. (2016) from the mixed-integer linear to the mixed-integer nonlinear setting.

Theorem 11. *Let Assumptions 1–3 hold. Assume that for all $x \in X$ such that $Ax \neq b$ it holds $\psi(Ax - b) \geq \delta > 0$. Then, for all $\bar{\lambda} \in \mathbb{R}^m$, there exists a finite penalty parameter ρ such that $z^* = z^{LR+}_\rho(\bar{\lambda})$.*

Proof. Let $\rho \geq 2\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda})/\delta$ be fixed and let $x^*_\rho \in S^*_\rho(\bar{\lambda})$. With the same argument as in the proof of Theorem 10, it suffices to show that x^*_ρ satisfies $Ax^*_\rho = b$ to conclude that $x^*_\rho \in S^*$ and, thus, that $z^* = z^{\mathrm{LR}+}(\bar{\lambda})$. By Theorem 6, ρ is sufficiently large to guarantee that $\psi(Ax^*_\rho - b) \leq \delta/2$. Yet, assuming that $Ax^*_\rho \neq b$ leads to the contradiction

$$0 < \delta \leq \psi(Ax^*_\rho - b) \leq \delta/2. \quad (20)$$

Hence, it must be that $Ax^*_\rho = b$, which ends the proof. □

A clear consequence of this theorem is that a finite penalty parameter closing the duality gap always exists for integer nonlinear problems. Hence, the next corollary generalizes Corollary 1 from Boland and Eberhard (2014) to nonlinearly constrained integer problems.

Corollary 4. *Let Assumption 1–3 hold and assume that $n_1 = 0$, i.e., Problem (1) is an integer problem. Then, for all $\bar{\lambda} \in \mathbb{R}^m$, there exists a finite ρ such that $z^* = z^{LR+}_\rho(\bar{\lambda})$.*

Proof. Under the stated assumption, the set $\{x \in X : Ax \neq b\}$ is compact and ψ is continuous, hence, $\min\{\psi(Ax - b) : x \in X, Ax \neq b\}$ exists. Positive definiteness of ψ shows the existence of a δ as required by Theorem 11. □

While Theorem 11 applies to general penalty functions, it requires the existence a positive minimal penalization of violated constraints. This is a rather strong assumption. Moreover, even checking that such a condition is satisfied seems to be challenging. Corollary 4 exhibits a class of problems for which it is easy to ensure the existence of such minimal violation $\delta > 0$. Yet, this class of problems is far from covering the general class of problems that can be cast as Problem (1). Note that this is, in fact, necessary since Feizollahi et al. (2016) shows that the ALD equipped with the proximal Lagrangian penalty function fails to close the duality gap on some mixed-integer linear instances; see Proposition 7 in Feizollahi et al. (2016). Since the proximal Lagrangian penalty function satisfies Assumption 3, one cannot hope for a stronger result with such a generality. In the next section, we specialize the penalty function ψ to be a norm and derive weaker sufficient conditions for exact penalty parameters to exist.

Before moving to the next section, we show that the existence of a finite exact penalty parameter for a given penalty function ψ can be used to show the existence of a finite exact penalty parameter with a different penalty function.

Theorem 12. *Let Assumption 1–3 hold. Let $\bar{\lambda} \in \mathbb{R}^m$ be fixed and assume that there exists $\rho < \infty$ such that $z^* = z_\rho^{LR+}(\bar{\lambda})$, i.e., there exists a finite exact penalty parameter for the ALR equipped with ψ. Let ψ' be a penalty function satisfying Assumption 3 and assume that, for every $x \in X$ with $Ax = b$, there exists a neighborhood $\mathcal{N}(x)$ satisfying the following two conditions:*

(i) there exists $\gamma > 0$ such that, for all $x' \in X \cap \mathcal{N}(x)$ with $Ax' \neq b$, it holds

$$\psi'(Ax' - b) \geq \gamma \psi(Ax' - b);$$

(ii) there exists $\delta > 0$ such that for all $x' \in X \backslash \mathcal{N}(x)$ with $Ax' \neq b$, it holds

$$\psi'(Ax' - b) \geq \delta.$$

Then, there exists a finite penalty parameter for the ALR equipped with ψ'.

Proof. To avoid ambiguity, we let $x_\rho^*(\psi)$ denote an optimal point of the ALR (2) equipped with the penalty function ψ. Similarly, we let $z_\rho^{\mathrm{LR}+}(\bar{\lambda}, \psi)$ denote its objective function. Let now

$$\rho' > \max \left\{ \frac{2\kappa_1(\mathcal{E}, A, b, \nabla f(x^*), \bar{\lambda})}{\delta}, \frac{\rho}{\gamma} \right\}.$$

Following the argument in the proof of Theorem 11, it is sufficient to show that $x_{\rho'}^*(\psi')$ satisfies $Ax_{\rho'}^*(\psi') = b$ to conclude $z^* = z_{\rho'}^{\mathrm{LR}+}(\bar{\lambda}, \psi')$. Hence, for the sake of contradiction, assume that $x_{\rho'}^*(\psi')$ is such that $Ax_{\rho'}^*(\psi') \neq b$. We have two cases.

1. Suppose $x_{\rho'}^*(\psi') \in \mathcal{N}(x)$ for some $x \in X$ satisfying $Ax = b$. It holds

$$\begin{aligned} z^* &= z_\rho^{\mathrm{LR}+}(\bar{\lambda}, \psi) \\ &\leq f(x_{\rho'}^*(\psi')) + \bar{\lambda}^\top (Ax_{\rho'}^*(\psi') - b) + \rho\psi(Ax_{\rho'}^*(\psi') - b) \\ &\leq f(x_{\rho'}^*(\psi')) + \bar{\lambda}^\top (Ax_{\rho'}^*(\psi') - b) + \frac{\rho}{\gamma}\psi'(Ax_{\rho'}^*(\psi') - b) \\ &< f(x_{\rho'}^*(\psi')) + \bar{\lambda}^\top (Ax_{\rho'}^*(\psi') - b) + \rho'\psi'(Ax_{\rho'}^*(\psi') - b) \\ &= z_{\rho'}^*(\bar{\lambda}, \psi'). \end{aligned}$$

Hence, we reach the contradiction that the primal problem (1) is strictly upper bounded by its ALR equipped with the penalty function ψ'; see Inequality (4).

2. Suppose $x_{\rho'}^*(\psi') \notin \mathcal{N}(x)$ for all $x \in X$ satisfying $Ax = b$. By assumption, it holds $\psi'(Ax_{\rho'}^*(\psi') - b) \geq \delta > 0$. Using Theorem 6, we conclude that ρ is sufficiently large to guarantee $\psi'(Ax_{\rho'}^*(\psi') - b) \leq \delta/2$. Hence, we reach the contradiction

$$0 < \delta \leq \psi'(Ax_{\rho'}^*(\psi') - b) \leq \delta/2.$$

All in all, we conclude that $x_{\rho'}^*(\psi')$ satisfies $Ax_{\rho'}^*(\psi') = b$ and, thus, the theorem holds. □

Let us note that Theorem 12 is a generalization of Theorem 5 from Feizollahi et al. (2016) in two ways. First, we consider general MINLPs instead of MILPs. Second, we consider two general penalty functions ψ and ψ' instead of a general penalty function ψ and the infinity norm. Later, we will specialize this result in the context of norms; see Lemma 2.

5.2. **Sharp Lagrangian.** Throughout this section, we make the assumption that the penalty function is a norm. Note that this assumption is stronger than Assumption 3. For instance, the proximal penalty function fulfills Assumption 3 but is not a norm. However, it is clear that any norm satisfies the properties of Assumption 3.

Assumption 4. $\psi = \|\cdot\|$ *for some norm* $\|\cdot\|$.

To simplify our proofs, we first show that the existence of an exact penalty parameter with $\psi = \|\cdot\|_\infty$ is enough to show the existence of an exact penalty parameter for any other norm $\|\cdot\|$. Even more, we show that this also implies an existence result for the nonconvex penalty function $\|\cdot\|^r$ with $0 < r \leq 1$.

Lemma 2. *Assume that there exists an exact penalty parameter for $\psi = \|.\|_\infty$. Then, there exists an exact penalty parameter for $\psi = \|\cdot\|^r$ with $\|\cdot\|$ being any norm and $r \in (0, 1]$.*

Proof. By the equivalence of norms in finite-dimensional spaces, there exists $\gamma > 0$ such that $\|\cdot\| \geq \gamma \|\cdot\|_\infty$. Then, for any $x \in X$ with $Ax = b$, there exists a neighborhood $\mathcal{N}(x)$ such that, for all $x' \in X \cap \mathcal{N}(x)$ with $Ax' \neq b$, it holds $\|Ax' - b\| \leq 1$ and, for all $x' \in X \backslash \mathcal{N}(x)$ with $Ax' \neq b$, $\|Ax' - b\| \geq 1$ holds. Hence, for all $x' \in X \cap \mathcal{N}(x)$, it holds

$$\gamma \|Ax' - b\|_\infty \leq \|Ax' - b\| \leq \|Ax' - b\|^r.$$

Moreover, for all $x \in X \backslash \mathcal{N}(x)$ with $Ax' \neq b$, we have $\|Ax' - b\| \geq 1$ and, thus, $\|Ax' - b\|^r \geq 1$. Hence, by Theorem 12, the claimed result holds. □

Note that the same result has been frequently used in the recent literature with $r = 1$; see, e.g., Feizollahi et al. (2016) and Gu et al. (2020). In the next lemma, we also show that focusing on the case $\lambda = 0$ is sufficient.

Lemma 3. *Assume that there exists a finite ρ^* such that $z^* = z^{LR+}_{\rho^*}(0)$ with $\psi = \|\cdot\|_\infty$. Then, for all $\bar{\lambda} \in \mathbb{R}^m$,*

$$z^* = z^{LR+}_{\sqrt{m}\|\bar{\lambda}\|_2 + \rho^*}(\bar{\lambda}).$$

Proof. Let $\bar{\lambda} \in \mathbb{R}^m$ be fixed and assume that there exists a finite ρ^* such that $z^* = z^{\mathrm{LR}+}_{\rho^*}(0)$. Then, by the Cauchy–Schwarz inequality and using $\|\cdot\|_2 \leq \sqrt{n}\|\cdot\|_\infty$, we get

$$\begin{aligned}
z^* &= \min_{x \in X} \; c^\top x + \rho^* \psi(Ax - b) \\
&= \min_{x \in X} \; c^\top x + \bar{\lambda}^\top (Ax - b) - \bar{\lambda}^\top (Ax - b) + \rho^* \psi(Ax - b) \\
&\leq \min_{x \in X} \; c^\top x + \bar{\lambda}^\top (Ax - b) + \|\bar{\lambda}\|_2 \|Ax - b\|_2 + \rho^* \psi(Ax - b) \\
&\leq \min_{x \in X} \; c^\top x + \bar{\lambda}^\top (Ax - b) + \sqrt{m}\|\bar{\lambda}\|_2 \|Ax - b\|_\infty + \rho^* \psi(Ax - b) \\
&= \min_{x \in X} \; c^\top x + \bar{\lambda}^\top (Ax - b) + (\sqrt{m}\|\bar{\lambda}\|_2 + \rho^*) \psi(Ax - b) \\
&= z^{\mathrm{LR}+}_{\sqrt{m}\|\bar{\lambda}\|_2 + \rho^*}(\bar{\lambda}).
\end{aligned}$$

The proof is achieved by Inequality (4). □

We are now ready to state conditions under which the existence of an exact penalty parameter is guaranteed for the primal problem (1). We do so in the following sections. The first section is dedicated to convex MINLPs, i.e., we assume that all complicating variables are integer. Then, we specialize this result for MILPs and show that such an exact penalty parameter can be computed in polynomial time. Finally, in the last section, we study general MINLPs.

5.2.1. *Convex MINLPs.* In this section, we consider convex MINLPs. That is, we assume that Problem (1) is a convex problem if x_2 is fixed. In this case, we show that an exact penalty parameter always exists under mild conditions. To this end, we need the following additional assumption.

Assumption 5 (Slater's Condition)**.** *Let $N \subseteq \{1, \dots, \ell\}$ denote the set of indices such that g_i is nonlinear. For all $x_2 \in X_2$, there exists $x_1 \in \mathbb{R}^{n_1}$ such that $A_1 x_1 = b - A_2 x_2$, $g_i(x_1, x_2) < 0$ for all $i \in N$ and $g_i(x_1, x_2) \leq 0$ for all $i \notin N$.*

Assumption 5 states that Slater's condition hold for the primal problem (1) if the complicating variables x_2 are fixed. Note that these conditions are immediately satisfied by the class of problems previously studied in the literature and is not restrictive in this sense. For instance, Assumption 5 holds for MILPs (Feizollahi et al. 2016), convex MIQPs (Gu et al. 2020), and linearly constrained convex problems (Bhardwaj et al. 2024).

Before we state the theorem, we introduce the notation $z_\rho^{\mathrm{LR+}}(\lambda, x_2)$ for

$$\begin{aligned} z_\rho^{\mathrm{LR+}}(\lambda, x_2) := \lambda^\top (A_2 x_2 - b) + \min_{x_1}\ & f(x_1, x_2) + \lambda^\top A_1 x_1 + \rho\psi(Ax - b) \\ \text{s.t. }\ & g(x_1, x_2) \leq 0, \\ & x_1 \in \mathbb{R}^{n_1} \end{aligned}$$

with $x_2 \in X_2$ and $\lambda \in \mathbb{R}^m$. Similarly, we define $z^*(x_2)$ as Problem (1) for a fixed $x_2 \in X_2$.

Theorem 13. *Let Assumptions 1–5. For all $\bar{\lambda} \in \mathbb{R}^m$, there exists $\rho^* < \infty$ such that*

$$z^* = z_{\rho^*}^{LR+}(\bar{\lambda}).$$

Proof. By Lemma 2 and 3, it is sufficient to consider $\psi = \|\cdot\|_\infty$ and $\bar{\lambda} = 0$. We first show that, for all $x_2 \in X_2$, there exists $\rho^*(\bar{\lambda}, x_2) < \infty$ such that

$$z^*(x_2) = z_{\rho^*(\bar{\lambda}, x_2)}^{\mathrm{LR+}}(\bar{\lambda}, x_2).$$

To this end, let $\hat{x}_2 \in X_2$ be fixed. Note that, by Assumption 5, Slater's Condition is satisfied for the convex problem

$$\begin{aligned} z^*(\hat{x}_2) := \min_{x_1, w}\ & f(x_1, \hat{x}_2) \\ \text{s.t. }\ & A_1 x_1 = b - A_2 \hat{x}_2, \\ & g(x_1, \hat{x}_2) \leq 0, \\ & x_1 \in \mathbb{R}^{n_1}. \end{aligned}$$

Thus, its optimal value is equal to the one of its dual and it holds

$$z^*(\hat{x}_2) = \max_{\alpha \in \mathbb{R}^m, \beta \geq 0}\ \min_{x_1}\ f(x_1, \hat{x}_2) + \beta^\top g(x_1, \hat{x}_2) + \alpha^\top (A_1 x_1 + A_2 \hat{x}_2 - b). \tag{22}$$

Moreover, the maximum is attained at some point $(\alpha^*(\hat{x}_2), \beta^*(\hat{x}_2)) \in \mathbb{R}^m \times \mathbb{R}^\ell_{\geq 0}$.

Now, consider the reformulation of $z_\rho^{\mathrm{LR+}}(\lambda, \hat{x}_2)$ in which "$\|A_1 x_1 + A_2 \hat{x}_2 - b\|_\infty \leq w$" has been linearized, i.e., consider

$$\begin{aligned} z_\rho^{\mathrm{LR+}}(\lambda, \hat{x}_2) = \min_{x_1, w}\ & f(x_1, \hat{x}_2) + \rho w \\ \text{s.t. }\ & A_1 x_1 \leq we + b - A_2 \hat{x}_2, \\ & -A_1 x_1 \leq we - b + A_2 \hat{x}_2, \\ & g(x_1, \hat{x}_2) \leq 0, \\ & x_1 \in \mathbb{R}^{n_1}. \end{aligned}$$

By strong duality, it holds

$$\begin{aligned} z_\rho^{\mathrm{LR+}}(\lambda, x_2) = \max_{\alpha^-, \alpha^+, \beta \geq 0}\ \min_{x_1 \in X_1, w \geq 0}\ & f(x_1, \hat{x}_2) + \rho w + \beta^\top g(x_1, \hat{x}_2) \\ & + (\alpha^+ - \alpha^-)^\top (A_1 x_1 + A_2 \hat{x}_2 - b) - (\alpha^+ + \alpha^-)^\top e w. \end{aligned}$$

With the change of variables $\alpha = \alpha^+ - \alpha^-$, we obtain

$$z_\rho^{\mathrm{LR}+}(\lambda, \hat{x}_2) = \max_{\alpha \in \mathbb{R}^m, \alpha^-, \beta \geq 0} \ \min_{x_1 \in X_1} \ f(x_1, \hat{x}_2) + \beta^\top g(x_1, \hat{x}_2) + \alpha^\top (A_1 x_1 + A_2 \hat{x}_2 - b) + \min_{w \geq 0} (\rho - e^\top(\alpha + 2\alpha^-)) w.$$

Looking at the last minimization problem, it follows that

$$z_\rho^{\mathrm{LR}+}(\lambda, \hat{x}_2) = \max_{\substack{e^\top(\alpha + 2\alpha^-) \leq \rho, \\ \alpha \in \mathbb{R}^m, \alpha^-, \beta \geq 0}} \ \min_{x_1 \in X_1} \ f(x_1, \hat{x}_2) + \beta^\top g(x_1, \hat{x}_2) + \alpha^\top (A_1 x_1 + A_2 \hat{x}_2 - b).$$

Observe that α^- can be fixed to zero since it only appears in a $\leq$-constraint with a nonnegative coefficient. Thus, for a sufficiently large ρ, this last optimization problem is exactly the dual Problem (22). Clearly, $\rho = \|\alpha^*(\hat{x}_2)\|_1$ is finite and large enough so that $z_\rho^{\mathrm{LR}+}(\lambda, \hat{x}_2) = z^*(\hat{x}_2)$. Thus, for all $x_2 \in X_2$, there exists an exact penalty parameter $\rho^*(\hat{x}_2, \lambda)$.

Now, consider $\rho^* = \max_{x_2 \in X_2} \ \rho^*(\lambda, x_2)$, which is a finite maximum of finite numbers since X_2 is discrete and bounded ($X_2 = \mathrm{proj}_{x_2}(X) \subset \mathbb{Z}^{n_2}$ with X bounded) and $\rho^*(\lambda, x_2) < \infty$ for all $x_2 \in X_2$. Hence, $\rho^* < \infty$. For all $x_2 \in X_2$, it then follows that

$$z^*(x_2) = z_{\rho^*(\lambda, x_2)}^{\mathrm{LR}+}(\lambda, \hat{x}_2) \leq z_{\rho^*}^{\mathrm{LR}+}(\lambda, \hat{x}_2).$$

Taking the minimum over X_2 leads to

$$z^* \leq z_{\rho^*}^{\mathrm{LR}+}(\lambda).$$

The proof is achieved by Inequality (4). □

5.2.2. *MILPs.* In this section, we show that an exact penalty parameter can be computed in polynomial time if the primal problem (1) is a MILP. We start with a lemma showing that the second part of Assumption 1 is fulfilled if the problem is feasible and bounded from below.

Lemma 4. *Let $\tilde{A} \in \mathbb{Z}^{\tilde{m} \times \tilde{n}}$ with $\tilde{m} \leq \tilde{n}$, $\tilde{b} \in \mathbb{Z}^{\tilde{m}}$, and $\tilde{c} \in \mathbb{Z}^{\tilde{n}}$ be integer matrices and vectors and consider the MILP*

$$\min_x \ \tilde{c}^\top x \tag{23a}$$

$$\text{s.t. } \tilde{A} x = \tilde{b}, \tag{23b}$$

$$x \in \mathbb{R}_{\geq 0}^{\tilde{n}_1} \times \mathbb{Z}_{\geq 0}^{\tilde{n}_2}, \tag{23c}$$

with $\tilde{n} = \tilde{n}_1 + \tilde{n}_2$. Assume that (23) is feasible and bounded from below. Then, there exists a solution x^* to Problem (23) satisfying $\|x^*\|_\infty \leq M$ for some finite $M > 0$. Moreover, there exists a polynomial-time algorithm (in the input data and for fixed dimension $(\tilde{m}, \tilde{n})$) computing M with input data $\tilde{A}$ and $\tilde{b}$.

Proof. Since (23) is feasible, bounded from below, and has rational entries, its continuous relaxation is feasible and bounded from below. Let $\tilde{x}^*$ denote a solution to the continuous relaxation of Problem (23). Without loss of generality, $\tilde{x}^*$ is an extreme point of $\{x \in \mathbb{R}_{\geq 0}^{\tilde{n}} : \tilde{A}x = \tilde{b}\}$. By Lemma 4 in Buchheim (2023), it holds

$$\|\tilde{x}^*\|_\infty \leq \tilde{m}! \|\tilde{b}\|_\infty \|\tilde{A}\|_\infty^{\tilde{m}-1}. \tag{24}$$

Now, by Theorem 17.2 in Schrijver (1998), there exists an optimal solution x^* to (23) which is such that

$$\|x^* - \tilde{x}^*\|_\infty \leq \tilde{n} \Delta, \tag{25}$$

with Δ an upper bound on each sub-determinant of $\tilde{A}$. Now let $\bar{A}$ denote any square sub-matrix of $\tilde{A}$ of size k. It holds

$$\det(\bar{A}) = \sum_{\sigma \in \mathcal{S}_k} \operatorname{sgn}(\sigma) \prod_{j=1}^{k} (\bar{A})_{j,\sigma(j)} \leq \sum_{\sigma \in \mathcal{S}_k} \prod_{j=1}^{k} \left|(\bar{A})_{j,\sigma(j)}\right| \leq k! \|\bar{A}\|_\infty^k.$$

Thus, it holds $\Delta \leq \tilde{m}! \|\tilde{A}\|_\infty^{\tilde{m}}$. By the reverse triangular inequality, it then follows that

$$\|x^*\|_\infty \leq \|x^* - \tilde{x}^*\|_\infty + \|\tilde{x}^*\|_\infty \leq \tilde{n}\Delta + \|\tilde{x}^*\|_\infty.$$

Using (24) and (25), it follows that

$$\|x^*\|_\infty \leq \tilde{n}\tilde{m}! \|\tilde{A}\|_\infty^{\tilde{m}} + \tilde{m}! \|\tilde{b}\|_\infty \|\tilde{A}\|_\infty^{\tilde{m}-1}.$$

Clearly, the constant in the right-hand side can be computed in polynomial time (in $\tilde{A}$ and $\tilde{b}$ and for fixed dimension $(\tilde{m}, \tilde{n})$). □

We now state the complexity result.

Theorem 14. *Let $\psi = \|\cdot\|_\infty$. Assume that Problem (1) is feasible and bounded from below and that $f(x) = c^\top x = c_1^\top x_1 + c_2^\top x_2$, and $g(x) = Tx + r = T_1 x_1 + T_2 x_2 + r$ for some matrix $T \in \mathbb{Z}^{\ell \times n}$ and some vectors $c \in \mathbb{Z}^n$, $r \in \mathbb{Z}^\ell$. Then, for all $\bar{\lambda} \in \mathbb{Q}^m$, there exists a polynomial-time algorithm computing a finite ρ such that $z^* = z_\rho^{LR+}(\bar{\lambda})$ with input data A, b, c, T, r and for fixed dimension (ℓ, m, n).*

Proof. By Lemma 4, since A, b, c, r, and T are integral matrices, there exists an $M > 0$, which can be computed in polynomial time such that $\|x^*\|_\infty \leq M$ for at least one solution to the primal problem (1). Hence, Assumption 1 is satisfied. Assumption 2 is also fulfilled as f is linear. Since assuming $\psi = \|\cdot\|_\infty$ is stronger than Assumption 4 and 3, they are readily satisfied. Again, since all constraints are linear, Assumption 5 is also satisfied. Finally, note that, by Lemma 3, it suffices to consider the case $\bar{\lambda} = 0$ since $\|\bar{\lambda}\|_\infty$ can be computed in polynomial time and $\sqrt{m} \leq m$. Thus, we consider $\bar{\lambda} = 0$.

Following the proof of Theorem 13, let $\hat{x}_2 \in X_2$ be fixed. We show that a solution $\alpha^*(\hat{x}_2)$ to the dual problem (22) exists such that $\|\alpha^*(\hat{x}_2)\|_\infty \leq \bar{\alpha}$ for all $\hat{x}_2 \in X_2$ and $\bar{\alpha}$ is a constant, which can be computed in polynomial time with input data A, b, c, T, r and for fixed dimension (ℓ, m, n). Now, the dual can be expressed as

$$\begin{aligned} c_2^\top \hat{x}_2 + \max_{\alpha,\beta} \; & (b - A_2\hat{x}_2)^\top \alpha - (r + T_2\hat{x}_2)^\top \beta \\ \text{s.t. } & A_1^\top \alpha + T_1^\top \beta = c_1, \\ & \beta \leq 0. \end{aligned}$$

Since $\hat{x}_2$ is fixed, this problem is an LP with rational entries and, w.l.o.g., there exists a solution $\alpha^*(\hat{x}_2)$ which is an extreme point of the feasible region. By Lemma 4 in Buchheim (2023), it follows that there exists α^* such that

$$\|\alpha^*(\hat{x}_2)\|_\infty \leq n_1! \|c_1\|_\infty \max\{\|A_1\|_\infty, \|T_1\|_\infty\}^{n_1+1} =: \bar{\alpha}.$$

By equivalence of norms, we conclude that $\|\alpha^*(\hat{x}_2)\|_1 \leq n_1\bar{\alpha}$. By the same argument as in the proof of Theorem 13, it follows that any ρ with $\rho \geq n_1\bar{\alpha} \geq \max_{x_2 \in X_2} \|\alpha^*(x_2)\|_1$ is an exact penalty parameter for Problem (1). Clearly, $n\bar{\alpha}$ can be computed in polynomial time. □

5.2.3. *General MINLPs.* We now turn to general MINLPs and give a sufficient condition for the existence of a finite exact penalty parameter. Note that Theorem 11 and Corollary 4 already show that a finite exact penalty parameter exists if all variables in Problem (1) are integer. Thus, we may now assume that some variables are continuous, i.e., $n_1 \geq 1$.

Following the idea of the proof of Theorem 13, it is sufficient to show that a finite exact penalty parameter exists for Problem (1) if all integer variables are fixed. Then, it suffices to consider the maximum of these finite exact penalty parameters over the set of all possible fixations (i.e., over X_2) to obtain a finite exact penalty parameter for the complete problem (1). Moreover, in virtue of Lemma 2 and Lemma 3, it is sufficient to consider the existence of a finite penalty parameter in the case $\psi = \|\cdot\|_\infty$ and $\lambda = 0$. Hence, we first need a theorem on exact penalization for nonlinear problems. The next theorem is due to Di Pillo and Grippo (1989).

Theorem 15 (Di Pillo and Grippo (1989), Theorem 4.a)**.** *Consider the general nonlinear problem*

$$\min_x \ \tilde{f}(x) \tag{27a}$$
$$\text{s.t. } \tilde{g}_i(x) \le 0, \quad i = 1, \dots, \tilde{m}, \tag{27b}$$
$$\tilde{h}_i(x) = 0, \quad i = 1, \dots, \tilde{p}, \tag{27c}$$

in which $\tilde{f}, \tilde{g}$, and $\tilde{h}$ are continuously differentiable functions. Let $\tilde{S}^*$ denote the set of global solutions to (27) and assume that the Mangasarian–Fromovitz Constraint Qualification (MFCQ) holds at any $x^* \in \tilde{S}^*$, i.e., assume that $\nabla\tilde{h}_1(x^*), \dots, \nabla\tilde{h}_{\tilde{p}}(x^*)$ are linearly independent and that there exists a direction $d \in \mathbb{R}^{\tilde{n}}$ such that

$$\nabla g_i(x^*)^\top d < 0 \quad i \in I(x^*) := \{i : g_i(x^*) = 0\},$$
$$\nabla h_i(x^*)^\top d = 0 \quad i = 1, \dots, \tilde{p}.$$

Assume that the feasible region $F := \{x \in \mathbb{R}^{\tilde{n}} : \tilde{g}(x) \le 0, \tilde{h}(x) = 0\}$ is nonempty and compact so that there exists a compact set D with $F \subset \operatorname{int}(D)$. Then, there exists a positive $\rho < \infty$ such that $\tilde{S}^*$ coincides with the set of global solutions to the penalized problem

$$\min_{x \in \operatorname{int}(D)} \ \tilde{f}(x) + \rho\left(\|\tilde{g}(x)^+\|_\infty + \|\tilde{h}(x)\|_\infty\right).$$

In the next theorem, we apply the result of Theorem 15 to Problem (1). To this end, we introduce the following assumption.

Assumption 6. *Let $\hat{x}_2 \in X_2$ be given and consider the optimization problem (1) in which variables x_2 have been fixed to $\hat{x}_2$, i.e., consider*

$$z^*(\hat{x}_2) := \min_x \ f(x_1, \hat{x}_2) \tag{28a}$$
$$\text{s.t. } g(x_1, \hat{x}_2) \le 0, \tag{28b}$$
$$A_1 x_1 = b - A_2\hat{x}_2. \tag{28c}$$

Let $S^*(\hat{x}_2)$ denote the set of global solutions to (28). We assume that f and g are continuously differentiable functions and that MFCQ holds at every point in $S^*(\hat{x}_2)$, i.e., A_1 has full row rank and, for all $x_1^* \in S^*(\hat{x}_2)$, there exists a direction $d \in \mathbb{R}^{n_1}$ such that

$$\nabla g_i(x_1^*)^\top d < 0 \quad i \in I(x_1^*, \hat{x}_2) := \{i : g(x_1^*, \hat{x}_2) = 0\},$$
$$A_1^\top d = 0.$$

Theorem 16. *Let Assumption 1–4 and 6 hold and let $\bar{\lambda} \in \mathbb{R}^m$ be arbitrary but fixed. Then, there exists a $\rho^* < \infty$ such that $z^* = z_{\rho^*}(\bar{\lambda})$.*

Proof. As anticipated, it is sufficient to consider $\psi = \|\cdot\|_\infty$ and $\bar{\lambda} = 0$; see Lemma 2 and 3. Now, with $\hat{x}_2 \in X_2$ arbitrary but fixed and using Theorem 15, it is easy to show that a finite $\rho^*(\bar{\lambda}, \hat{x}_2)$ exists such that

$$z^*(\hat{x}_2) = \min_{(x_1, \hat{x}_2) \in \operatorname{int}(\mathcal{E})} f(x_1, \hat{x}_2) + \rho^*(\bar{\lambda}, \hat{x}_2)\left(\|g(x_1, \hat{x}_2)^+\|_\infty + \|A_1x_1 + A_2\hat{x}_2 - b\|_\infty\right)$$

for a properly chosen $\mathcal{E}$ as defined in Assumption 1, potentially enlarging it so that $X \subset \text{int}(\mathcal{E})$. Note that, for any global solution $x_1^* \in S^*(\hat{x}_2)$ it holds $g(x_1^*, \hat{x}_2) \leq 0$ so that

$$\begin{aligned} z^*(\hat{x}_2) = \min_{x_1} \ & f(x_1, \hat{x}_2) + \rho^*(\bar{\lambda}, \hat{x}_2)\|A_1 x_1 + A_2 \hat{x}_2 - b\|_\infty \\ \text{s.t. } & g(x_1, \hat{x}_2) \leq 0, \\ & (x_1, \hat{x}_2) \in \text{int}(\mathcal{E}). \end{aligned}$$

By construction of $\mathcal{E}$, we directly obtain

$$z^*(\hat{x}_2) = \min_{x_1 \in X_1} \ f(x_1, \hat{x}_2) + \rho^*(\bar{\lambda}, \hat{x}_2)\|A_1 x_1 + A_2 \hat{x}_2 - b\|_\infty = z^{\text{LR+}}_{\rho(\bar{\lambda}, x_2)}(\bar{\lambda}).$$

We can now argue as in the proof of Theorem 13. For all $\hat{x}_2 \in X_2$, there exists a finite $\rho^*(\bar{\lambda}, \hat{x}_2)$ such that $z^*(\hat{x}_2) = z^{\text{LR+}}_{\rho^*(\bar{\lambda}, \hat{x}_2)}(\bar{\lambda})$. Hence, we can define $\rho^* = \max_{x_2 \in X_2} \ \rho^*(\bar{\lambda}, x_2)$, which is a finite maximum of finite numbers. Hence, ρ^* is finite and it holds

$$z^*(\hat{x}_2) = z^{\text{LR+}}_{\rho(\bar{\lambda}, x_2)}(\bar{\lambda}) \leq z_{\rho^*}(\bar{\lambda}, \hat{x}_2).$$

Taking the minimum over all $x_2 \in X_2$ leads to

$$z^* \leq z_{\rho^*}(\bar{\lambda}),$$

which, by Inequality (4), ends our proof. □

6. Conclusion

In this paper, we have shown that, under mild assumptions, the ALD equipped with any norm leads to a zero duality gap for nonconvex MINLPs. While this constitutes a generalization of the existing literature (see Table 1), several open questions remain.

For the case of MILPs and MIQPs, we have shown that a (complexity-wise) small exact penalty parameter can be computed in polynomial time. However, while the proof is constructive, the derived penalty parameter is too large to be useful in practical applications. Hence, the existence of a polynomial-time algorithm capable of computing a smaller exact penalty parameter is still an open and practically important question. Related to the previous question, we do not know if computing the smallest exact penalty parameter can be done in polynomial time. We conjecture a negative answer to this question. In the same vein, the existence of an exact penalty parameter of polynomial size is limited to MIQPs and it is still open if such a result can be obtained for a more general class of problems such as, e.g., quadratically constrained quadratic problems.

Acknowledgements

We acknowledge the support by the German Bundesministerium für Bildung und Forschung within the project "RODES" (Förderkennzeichen 05M22UTB).

References

Beck, Y., D. Bienstock, M. Schmidt, and J. Thürauf (2023). "On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level." In: *Journal of Optimization Theory and Applications* 198.1, pp. 428–447. DOI: 10.1007/s10957-023-02238-9.

Bertsekas, D., A. Nedic, and A. Ozdaglar (2003). *Convex Analysis and Optimization.* Athena Scientific.

Bertsekas, D. P. (2016). *Nonlinear Programming.* Athena scientific Belmont.

Bhardwaj, A., V. Narayanan, and A. Pathapati (2024). "Exact Augmented Lagrangian Duality for Mixed Integer Convex Optimization." In: *SIAM Journal on Optimization* 34.2, pp. 1622–1645. DOI: 10.1137/22M1526204.

Boland, N. L. and A. C. Eberhard (2014). "On the augmented Lagrangian dual for integer programming." In: *Mathematical Programming* 150.2, pp. 491–509. DOI: 10.1007/s10107-014-0763-3.

Boland, N. L., A. C. Eberhard, F. Engineer, and A. Tsoukalas (2012). "A New Approach to the Feasibility Pump in Mixed Integer Programming." In: *SIAM Journal on Optimization* 22.3, pp. 831–861. DOI: 10.1137/110823596.

Buchheim, C. (2023). "Bilevel linear optimization belongs to NP and admits polynomial-size KKT-based reformulations." In: *Operations Research Letters* 51.6, pp. 618–622. DOI: 10.1016/j.orl.2023.10.006.

Burachik, R. S., X. Q. Yang, and Y. Y. Zhou (2017). "Existence of Augmented Lagrange Multipliers for Semi-infinite Programming Problems." In: *Journal of Optimization Theory and Applications* 173.2, pp. 471–503. DOI: 10.1007/s10957-017-1091-6.

Burke, J. V. (1991). "An Exact Penalization Viewpoint of Constrained Optimization." In: *SIAM Journal on Control and Optimization* 29.4, pp. 968–998. DOI: 10.1137/0329054.

Del Pia, A., S. S. Dey, and M. Molinaro (2016). "Mixed-integer quadratic programming is in NP." In: *Mathematical Programming* 162.1–2, pp. 225–240. DOI: 10.1007/s10107-016-1036-0.

Di Pillo, G. and L. Grippo (1989). "Exact Penalty Functions in Constrained Optimization." In: *SIAM Journal on Control and Optimization* 27.6, 1333—1360. DOI: 10.1137/0327068.

Fabiani, F., B. Franci, S. Sagratella, M. Schmidt, and M. Staudigl (2022). "Proximal-like algorithms for equilibrium seeking in mixed-integer Nash equilibrium problems." In: *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE. DOI: 10.1109/cdc51059.2022.9993250.

Feizollahi, M. J., S. Ahmed, and A. Sun (2016). "Exact augmented Lagrangian duality for mixed integer linear programming." In: *Mathematical Programming* 161.1–2, pp. 365–387. DOI: 10.1007/s10107-016-1012-8.

Fischetti, M., F. Glover, and A. Lodi (2005). "The feasibility pump." In: *Mathematical Programming* 104.1, pp. 91–104. DOI: 10.1007/s10107-004-0570-3.

Gu, X., S. Ahmed, and S. S. Dey (2020). "Exact Augmented Lagrangian Duality for Mixed Integer Quadratic Programming." In: *SIAM Journal on Optimization* 30.1, pp. 781–797. DOI: 10.1137/19m1271695.

Huang, X. X. and X. Q. Yang (2003). "A Unified Augmented Lagrangian Approach to Duality and Exact Penalization." In: *Mathematics of Operations Research* 28.3, pp. 533–552. URL: http://www.jstor.org/stable/4126984.

Perchet, V. and G. Vigeral (2015). "A Minmax Theorem for Concave-Convex Mappings with no Regularity Assumptions." In: *Journal of Convex Analysis* 22. URL: https://hal.science/hal-00927071/document.

Rockafellar, R. T. (1974). "Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming." In: *SIAM Journal on Control* 12.2, pp. 268–285. DOI: 10.1137/0312021.

Rockafellar, R. T. and R. J. B. Wets (1998). *Variational Analysis*. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-02431-3.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton: Princeton University Press. DOI: doi:10.1515/9781400873173.

Schrijver, A. (1998). *Theory of Linear and Integer Programming*. Wiley Series in Discrete Mathematics & Optimization. Wiley.

Still, G. (2018). "Lectures on parametric optimization: An introduction." In: *Optimization Online*, p. 2. URL: https://optimization-online.org/?p=15154.

Tardella, F. (2004). "On the existence of polyhedral convex envelopes." In: *Nonconvex Optimization and Its Applications*. Springer US, pp. 563–573. DOI: 10.1007/978-1-4613-0251-3_30.

(Henri Lefebvre, Martin Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: henri.lefebvre@uni-trier.de

Email address: martin.schmidt@uni-trier.de