DFG Deutsche Forschungsgemeinschaft
Priority Programme 1962

An Example Comparing the Standard and Modified Augmented Lagrangian Methods

Christian Kanzow, Daniel Steck

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-013

received on March 13, 2017

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

An Example Comparing the Standard and Modified Augmented Lagrangian Methods*

Christian Kanzow† Daniel Steck†

February 22, 2017

Abstract

We consider the well-known augmented Lagrangian method for constrained optimization and compare its classical variant to a modified counterpart which uses safeguarded multiplier estimates. In particular, we give a brief overview of the theoretical properties of both methods, focusing on both feasibility and optimality of limit points. Finally, we give an example which illustrates the advantage of the modified method and incidentally shows that some of the assumptions used for convergence of the classical method cannot be relaxed.

1 Introduction

The purpose of this report is to compare two variants of the well-known augmented Lagrangian method (ALM), also known as the multiplier-penalty method or simply method of multipliers. Methods of this type essentially come in two flavours. On the one hand, there is the "classical" ALM [4, 9, 13, 16] which goes back to [10, 15]. On the other hand, modified ALMs [1, 2, 6, 7, 8] which seek to alleviate some of the weaknesses of the classical methods have surfaced in recent years. These methods go back to [1, 5]; note that a similar method was used in [14] for the analysis of quasi-variational inequalities.

On the following pages, we give an overview of the two methods, and refer to them as the *standard ALM* and *modified ALM*, respectively. We also give convergence theorems for both methods (some of these are just taken from the literature). The ultimate purpose of this report is to give a fairly simple example which demonstrates the benefits of the modified ALM when compared to its classical counterpart.

For a better comparison, we have attempted to put the algorithms into a unified framework. For our purposes, this is a finite dimensional optimization problem with

*This research was supported by the German Research Foundation (DFG) within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962) under grant number KA 1296/24-1.

†University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; kanzow@mathematik.uni-wuerzburg.de, daniel.steck@mathematik.uni-wuerzburg.de

inequality constraints. More precisely, let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be given functions, and consider the problem defined by

$$\min\ f(x) \quad \text{s.t.} \quad g(x) \le 0. \tag{1}$$

It is possible to make this framework more general, for instance, by including equality constraints, additional constraint functions which are not penalized, or even considering infinite-dimensional problems. Moreover, augmented Lagrangian methods have also been extended to problem classes which are inherently more complex, such as generalized Nash equilibrium problems [11] and quasi-variational inequalities [12]. However, for our comparison of the two ALMs, we have decided to remain within the framework (1) because it is fairly simple and suffices for a discussion of the algorithmic differences of the two methods. Moreover, one might argue that optimization problems are both the historical origin and the key application of augmented Lagrangian methods. Hence, it makes sense to discuss the applicability and performance of such methods for precisely this problem class.

It is important to note that convergence theorems and properties of ALMs usually come in multiple flavours as well. These occur naturally because ALMs generate a sequence of penalized subproblems, and one has to clarify in which manner these are solved. The two most prominent choices in this regard are global minimization and finding stationary points. Here, we focus on the latter for its practical relevance and because global minimization is infeasible if the underlying problem is non-convex.

This report is organized as follows. In Section 2, we start with some preliminary definitions. The subsections 2.1 and 2.2 are dedicated to the standard and modified ALMs, respectively, and we give (or recall) convergence theorems for each of these methods. In Section 3 and its subsections, we give an example and discuss the results of the standard and modified ALMs, both from a theoretical and practical point of view. We conclude with some final remarks in Section 4.

Notation: The gradient of the continuously differentiable objective function f is denoted by ∇f, whereas the symbol $\nabla g(x)$ stands for the transposed Jacobian of the constraint function g at a given point x. For a mapping of two block variables, say $L(x, \lambda)$, we write $\nabla_x L(x, \lambda)$ to indicate the derivative with respect to the x-variables only. Given any vector z, we use the abbreviation z_+ for $\max\{0, z\}$, where the maximum is taken component-wise. Finally, throughout this note, $\|z\|$ denotes the Euclidean norm of a vector z of appropriate dimension.

2 Preliminaries

Recall that we are dealing with the optimization problem (1). Since we are ultimately interested in KKT-type conditions, we assume that f, g are continuously differentiable on $\mathbb{R}^n$. Moreover, for $\rho > 0$, $\lambda \ge 0$, we define the augmented Lagrangian

$$L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \left\| \left(g(x) + \frac{\lambda}{\rho} \right)_+ \right\|^2. \tag{2}$$

It is easily seen that, like f and g, the function L_ρ is continuously differentiable on $\mathbb{R}^n$. Its gradient is given by

$$\nabla_x L_\rho(x,\lambda) = \nabla f(x) + \nabla g(x)(\lambda + \rho g(x))_+, \tag{3}$$

which is in fact the main motivation for the classical Hestenes-Powell multiplier updating scheme.

For our analysis, we will need certain constraint qualifications. The linear independence and Mangasarian-Fromovitz constraint qualifications (LICQ and MFCQ, respectively) are fairly standard and, hence, we do not give their definitions here. Instead, we focus on two other conditions: the extended MFCQ (EMFCQ) and the constant positive linear dependence condition (CPLD), whose definitions are given below. Note that we call a collection of vectors $v_1, \ldots, v_k$ *positively linearly dependent* if the system $\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$, $\lambda \geq 0$, has a nontrivial solution. Otherwise, the vectors are called *positively linearly independent.*

Definition 2.1. *Let $\bar{x} \in \mathbb{R}^n$ be a given point. We say that*

(a) EMFCQ *holds in $\bar{x}$ if the set of gradients $\nabla g_i(\bar{x})$ with $g_i(\bar{x}) \geq 0$ is positively linearly independent.*

(b) CPLD *holds in $\bar{x}$ if, for every $I \subseteq \{i \mid g_i(\bar{x}) = 0\}$ such that the vectors $\nabla g_i(\bar{x})$ $(i \in I)$ are positively linearly dependent, there is a neighbourhood of $\bar{x}$ where the gradients $\nabla g_i(x)$ $(i \in I)$ are linearly dependent.*

It is well-known and easy to verify that EMFCQ boils down to MFCQ for feasible points, and that CPLD is weaker than MFCQ. Moreover, using a standard theorem of the alternative, EMFCQ is equivalent to the existence of a $d \in \mathbb{R}^n$ such that

$$g_i(\bar{x}) \geq 0 \quad \Longrightarrow \quad \nabla g_i(\bar{x})^T d < 0 \tag{4}$$

for all $i \in \{1, \ldots, m\}$.

Note that some subsequent results may hold under weaker assumptions than CPLD or EMFCQ. For instance, there are certain relaxed versions of CPLD [3] which can be used in a similar manner as CPLD. However, for the sake of simplicity, we have decided to remain with the conditions above. Note also that at least one of the aforementioned relaxations of CPLD is in fact equivalent to CPLD for our setting.

2.1 The Standard Method

Here, we give a fairly straightforward version of the standard ALM. Recall that L_ρ is the augmented Lagrangian from (2) and that the optimization problem has inequality constraints only.

Algorithm 2.2. *(Standard ALM)*

(S.0) Let $(x^0, \lambda^0) \in \mathbb{R}^{n+m}$, $\rho_0 > 0$, $\gamma > 1$, $\tau \in (0,1)$, and set $k := 0$.

(S.1) If (x^k, λ^k) *is a KKT point of the problem: STOP.*

(S.2) Compute an approximate solution x^{k+1} *of the problem*

$$\min \ L_{\rho_k}(x, \lambda^k). \tag{5}$$

(S.3) Set $\lambda^{k+1} := \left(\lambda^k + \rho_k g(x^{k+1})\right)_+$ *and*

$$V^{k+1} = \left\| \min\left\{ -g(x^{k+1}), \frac{\lambda^k}{\rho_k} \right\} \right\|. \tag{6}$$

If $k = 0$ *or* $V^{k+1} \leq \tau V^k$*, set* $\rho_{k+1} := \rho_k$*. Otherwise, set* $\rho_{k+1} := \gamma\rho_k$*.*

(S.4) Set $k \leftarrow k+1$ *and go to (S.1).*

The test function in (6) arises from an inherent slack variable transformation which is often used to define the augmented Lagrangian method for inequality constrained problems. Note also that, for formal reasons, we have given the case $k = 0$ specific treatment in Step 3 since (6) only defines V^k for $k \geq 1$ and V^0 is undefined.

Note that we have left the term "approximate solution" unspecified in Step 2. As mentioned in the introduction, multiple choices can be made for the solution process of the subproblems, e.g. one could look for global minima or stationary points. In this report, we will only consider the latter case. More precisely, we assume that $L'_{\rho_k}(x^{k+1}, \lambda^k) \to 0$. Using (3), it is easy to see that

$$\nabla_x L_{\rho_k}(x^{k+1}, \lambda^k) = \nabla f(x^{k+1}) + \nabla g(x^{k+1})\lambda^{k+1}. \tag{7}$$

We now turn to two convergence theorems for the standard ALM. Note that we implicitly assume that the method generates an infinite sequence (x^k). More convergence results using stronger assumptions can be found in [4, 9].

Theorem 2.3. *Let* (x^k) *be generated by Algorithm 2.2, and assume that*

$$x^{k+1} \to \bar{x} \quad \text{and} \quad \nabla_x L_{\rho_k}(x^{k+1}, \lambda^k) \to 0. \tag{8}$$

If $\bar{x}$ *is feasible and CPLD holds in* $\bar{x}$*, then* $\bar{x}$ *is a KKT point of the problem.*

Proof. The result essentially follows by applying [7, Thm. 3.6]. To this end, we need to verify that $\min\{-g(x^{k+1}), \lambda^{k+1}\} \to 0$. This is obvious whenever (ρ_k) stays bounded, cf. (6). Hence consider the case where $\rho_k \to \infty$, and recall that $g(\bar{x}) \leq 0$. If i is an index with $g_i(\bar{x}) < 0$, then the multiplier updating scheme implies $\lambda_i^{k+1} = 0$ for all sufficiently large k. This completes the proof. □

The above theorem does not contain any information about the attainment of feasibility. Since the augmented Lagrangian method is, at its heart, a penalty method, the achievement of feasibility is paramount to the success of the algorithm. The following result contains some information in this direction.

Theorem 2.4. *If (8) holds and $\bar{x}$ satisfies EMFCQ, then $\bar{x}$ is feasible and CPLD holds in $\bar{x}$. In particular, the requirements of Theorem 2.3 are satisfied.*

Proof. Note that, for feasible points, EMFCQ implies CPLD. If (ρ_k) is bounded, then $V^{k+1} \to 0$ and $\bar{x}$ is feasible. Now, let $\rho_k \to \infty$. We first show that (λ^{k+1}) is bounded. If this is not the case, then, subsequencing if necessary, (7) implies that

$$\nabla g(x^{k+1}) \frac{\lambda^{k+1}}{\|\lambda^{k+1}\|} \to 0$$

Denoting by $\bar{\lambda}$ a limit point of $(\lambda^{k+1}/\|\lambda^{k+1}\|)$, it follows that $\nabla g(\bar{x})\bar{\lambda} = 0$. Moreover, for each index i with $g_i(\bar{x}) < 0$, the multiplier updating rule implies $\lambda_i^{k+1} = 0$ for sufficiently large k, and hence $\bar{\lambda}_i = 0$. Now, let d be the vector from (4). Then

$$0 = d^T \nabla g(\bar{x})\bar{\lambda} = \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x})^T d = \sum_{g_i(\bar{x}) \geq 0} \bar{\lambda}_i g_i(\bar{x})^T d,$$

which implies $\bar{\lambda}_i = 0$ for all i with $g_i(\bar{x}) \geq 0$. Hence, $\bar{\lambda} = 0$, a contradiction. Thus, (λ^{k+1}) is bounded. But

$$\nabla f(x^{k+1}) + \nabla g(x^{k+1})(\lambda^k + \rho_k g(x^{k+1}))_+ \to 0$$

in view of (8). Dividing both sides by ρ_k and omitting some zero sequences, we obtain $\nabla g(x^{k+1}) g_+(x^{k+1}) \to 0$. Using once again the vector d from EMFCQ, it follows that

$$0 = d^T \nabla g(\bar{x}) g_+(\bar{x}) = (\nabla g(\bar{x})^T d)^T g_+(\bar{x}),$$

which implies $g_+(\bar{x}) = 0$. Hence, $\bar{x}$ is feasible. □

Let us mention explicitly that both of the above theorems require the convergence of the whole sequence (x^k). If $\bar{x}$ is only a limit point of (x^k), the assertions above do not hold, cf. the example in Section 3.

2.2 The Modified Method

We now turn to a discussion of the modified ALM.

Algorithm 2.5. *(Modified ALM)*

(S.0) Let $(x^0, \lambda^0) \in \mathbb{R}^{n+m}$, $\rho_0 > 0$, $u^{\max} \geq 0$, $\gamma > 1$, $\tau \in (0,1)$, and set $k := 0$.

(S.1) If (x^k, λ^k) is a KKT point of the problem: STOP.

(S.2) Choose $u^k \in [0, u^{\max}]^m$ and compute an approximate solution x^{k+1} of the problem

$$\min \; L_{\rho_k}(x, u^k). \tag{9}$$

(S.3) *Set* $\lambda^{k+1} := \big(u^k + \rho_k g(x^{k+1})\big)_+$ *and*

$$V^{k+1} = \left\| \min\left\{ -g(x^{k+1}), \frac{u^k}{\rho_k} \right\} \right\|. \tag{10}$$

If $k = 0$ *or* $V^{k+1} \leq \tau V^k$, *set* $\rho_{k+1} := \rho_k$. *Otherwise, set* $\rho_{k+1} := \gamma\rho_k$.

(S.4) *Set* $k \leftarrow k+1$ *and go to (S.1).*

As with the standard ALM (Algorithm 2.2), the term "approximate solution" in Step 2 can refer to either global minima or stationary points. In the following discussion, we only deal with the latter, cf. the assumptions made in Theorem 2.6. Convergence results dealing with global minimization can be found, e.g. in [6, 7].

The key difference between Algorithm 2.5 and the standard ALM is the use of the bounded sequence (u^k) in certain places where the standard ALM uses the (possibly unbounded) sequence of multiplier estimates (λ^k). Note that, despite the boundedness of (u^k), the corresponding sequence (λ^k) generated by Algorithm 2.5 might still be unbounded. We further stress that Algorithm 2.5 allows some freedom in the choice of the sequence (u^k). In the extreme case, we may take $u^k := 0$ for all k, in which case Algorithm 2.5 essentially boils down to the classical penalty method. A more natural and numerically often more successful way is to take u^k as the projection of λ^k onto the box $[0, u^{\max}]^m$.

We now give a first convergence theorem for the modified ALM. This result is very similar to Theorem 2.3 but differs in the sense that only a subsequence of (x^{k+1}) needs to converge to $\bar{x}$.

Theorem 2.6. *Let* (x^k) *be generated by Algorithm 2.2, and assume that*

$$x^{k+1} \to_K \bar{x} \quad \textit{and} \quad \nabla_x L_{\rho_k}(x^{k+1}, u^k) \to_K 0 \tag{11}$$

on some subset $K \subseteq \mathbb{N}$. *Then* $\bar{x}$ *is a stationary point of* $\|g_+(x)\|^2$ *and, if* $\bar{x}$ *is feasible and satisfies CPLD, it is a KKT point of the optimization problem.*

Proof. This is just a summary of [7, Thm. 6.2 and 6.3]. □

Note that, as opposed to Theorem 2.3, the above result includes a feasibility assertion about the limit point $\bar{x}$ which does not require any assumptions (apart from the continuous differentiability of f and g, of course). The main reason why the modified ALM admits a result of this type is that the sequence u^k/ρ_k always converges to zero if $\rho_k \to \infty$. It follows that the minimization of the augmented Lagrangian essentially reduces to the minimization of $\|g_+(x)\|^2$ if ρ is large enough. This property does not hold for the standard ALM. In fact, we will see in Section 3 that the latter faces severe problems if the sequence λ^k/ρ_k (for the standard ALM) remains bounded away from zero.

We now give a second result akin to Theorem 2.4.

Theorem 2.7. *If* $\nabla\|g_+(\bar{x})\|^2 = 0$ *and EMFCQ holds in* $\bar{x}$, *then* $\bar{x}$ *is feasible and CPLD holds in* $\bar{x}$.

Proof. Clearly, we only need to show feasibility. Let $d \in \mathbb{R}^n$ be the vector from EMFCQ. Using $\nabla \|g_+(x)\|^2 = \nabla g(x) g_+(x)$, it follows that

$$0 = d^T \nabla g(\bar{x}) g_+(\bar{x}) = (\nabla g(\bar{x})^T d)^T g_+(\bar{x}).$$

This implies $g_+(\bar{x}) = 0$. Hence, $\bar{x}$ is feasible. □

As mentioned before, the key difference between the assumptions used in the above theorems and the corresponding results for the standard ALM is the fact that $\bar{x}$ only needs to be a limit point of (x^k).

At first glance, the theoretical advantage which the modified ALM possesses in comparison to the standard ALM may not seem significant. However, it should be noted that the stationarity of $\bar{x}$ with respect to the infeasibility measure $\|g_+(x)\|^2$ has many interesting counterparts for other problem classes, e.g. for generalized Nash equilibrium problems [11] and quasi-variational inequalities [12]. Moreover, the resulting properties typically lend themselves to a simple problem-specific analysis which often guarantees the feasibility of $\bar{x}$ without any additional assumptions. We refer the reader to [12] in particular for further reading.

3 An Example

This section is the main contribution of this note. It provides an example where both Algorithms 2.2 and 2.5 generate a sequence of stationary points (in fact, local minimizers) of the corresponding augmented Lagrangian subproblems in such a way that the sequence has two stationary points, one of which is infeasible and violates basically any constraint qualification and can therefore not be expected to be a KKT point of the underyling optimization problem, whereas the other accumulation point is feasible (though different for both methods), satisfies essentially all constraint qualifications, and is therefore necessarily a KKT point of the underlying optimization problem for the modified ALM in view of Theorem 2.7, whereas it does not correspond to a stationary point for the standard ALM. Note that, in view of Theorem 2.3, the example has to be constructed in such a way that it has at least two accumulation points, or one accumulation point together with another subsequence which is unbounded.

Now, let $n = m = 1$ and consider the optimization problem given by

$$\min \ x \quad \text{s.t.} \quad 1 - x^3 \le 0. \tag{12}$$

In other words, we have $f(x) = x$ and $g(x) = 1 - x^3$. It is easy to see that $\bar{x} := 1$ is the unique solution of this optimization problem; moreover, an easy calculation shows that $(\bar{x}, \bar{\lambda}) := (1, 1/3)$ is the only KKT point. A key point in our following analysis is the fact that g has a stationary point at $x = 0$, and that this point is not feasible. Note that this example is easy in the sense that both the objective function and the feasible set are convex, though the representation of the convex feasible set is nonconvex.

3.1 The Standard Method

Let us consider the standard augmented Lagrangian method applied to this problem. The subsequent analysis is fairly general and only assumes (mainly for the sake of convenience) that $\rho_0 > 1/3$ and $\lambda^0 \le 1/3$.

It is easily seen that, for all $\lambda \ge 0$ and $\rho > 0$, the function $L_\rho(\cdot, \lambda)$ is coercive on $\mathbb{R}$. Moreover, using the formula

$$L'_\rho(x,\lambda) = f'(x) + g'(x)\big(\lambda + \rho g(x)\big)_+$$

for the derivative of L_ρ, we obtain $L'_\rho(0,\lambda) = 1$ and $L'_\rho(1,\lambda) = 1 - 3\lambda$. It follows that L_ρ always attains a local minimum in $(-\infty, 0)$ and, if $\lambda > 1/3$, it attains another local minimum in $(1, +\infty)$. Let $(x^k)_{k\ge1}$ be a sequence of such local minimizers such that

- for k odd, x^k is the largest local minimizer in $(-\infty, 0)$,
- for k even, x^k is the smallest local minimizer in $(1, +\infty)$.

If k is odd, we have $x^k < 0$ and $g(x^k) > 1$. It follows that

$$\lambda^k = \big(\lambda^{k-1} + \rho_{k-1} g(x^k)\big)_+ \ge \rho_{k-1}. \tag{13}$$

Since $\rho_{k-1} > 1/3$, we conclude that x^{k+1} is well-defined. Another property of the sequence (x^k) is boundedness.

Lemma 3.1. $x^k \in [-1, 2]$ *for all* $k \ge 1$.

Proof. If k is odd, then $x^k < 0$. Moreover, $L'_{\rho_{k-1}}(-1, \lambda^{k-1}) = 1 - 3(\lambda^{k-1} + 2\rho_{k-1})_+ < 0$, which implies $x^k > -1$ since x^k is supposed to be the largest local minimum in $(-\infty, 0)$.

Before showing that $x^k \le 2$ for k even, we need some information about the multiplier sequence (λ^k). First, if $k > 1$ is even, then $0 = L'_{\rho_{k-1}}(x^k, \lambda^{k-1}) = 1 - 3(x^k)^2\lambda^k$, which implies $\lambda^k = 1/(3(x^k)^2) \le 1/3$. By our assumption on λ^0, this assertion also holds for $k = 0$. Hence, using $\rho_0 > 1/3$ and $x^k \ge -1$ for k odd, it follows that

$$\lambda^k = \big(\lambda^{k-1} + \rho_{k-1} g(x^k)\big)_+ \le \frac{1}{3} + 2\rho_{k-1} \le 3\rho_k,$$

again for k odd. We now use this inequality to prove $x^k \le 2$ for k even. To this end, let $k > 1$ be even, and note that

$$L'_{\rho_{k-1}}(2, \lambda^{k-1}) = 1 - 12(\lambda^{k-1} - 7\rho_{k-1})_+ = 1,$$

since $k-1$ is odd. Hence, the definition of x^k as the smallest local minimum in $(1, +\infty)$ implies $x^k < 2$. □

The boundedness of (x^k) implies that the sequence has at least one limit point in $[-1, 0]$ and one in $[1, 2]$. In particular, we have $\rho_k \to \infty$, for otherwise every limit point of (x^k) would have to be feasible.

On the other hand, (13) implies that $\lambda^k/\rho_k \geq \gamma^{-1}$ for odd k. Define

$$\hat{x} = \left(1 + \frac{1}{2\gamma}\right)^{1/3},$$

which yields $g(\hat{x}) = -1/(2\gamma)$. It follows that, for all $x \in [1, \hat{x}]$ and k even,

$$\begin{aligned} L'_{\rho_{k-1}}(x, \lambda^{k-1}) &= 1 - 3\rho_{k-1}x^2 \left(g(x) + \frac{\lambda^{k-1}}{\rho_{k-1}}\right)_+ \\ &\leq 1 - 3\rho_{k-1}\left(\frac{1}{\gamma} - \frac{1}{2\gamma}\right)_+ \\ &= 1 - \frac{3}{2\gamma}\rho_{k-1} < 0 \end{aligned}$$

for sufficiently large values of ρ_{k-1}. Since $\rho_k \to \infty$, we conclude that $x^k > \hat{x}$ for sufficiently large (even) k. In particular, any accumulation point of (x^k) in $[1, 2]$ is strictly greater than 1. But none of these accumulation points correspond to a KKT point of the optimization problem (12).

3.2 The Modified Method

We now consider the modified method applied to problem (12). For the sake of convenience, we will again make certain assumptions on the algorithmic parameters. That is, we assume $\rho_0 > 1/3$ and $\lambda^0 \leq 1/3$. Moreover, we define u^k as the projection of λ^k onto the interval $[0, u^{\max}]$, where $u^{\max} > 1/3$, cf. the comments after the statement of Algorithm 2.5.

These assumptions allow us to compare the algorithm fairly easily to the standard ALM. In particular, we can choose (x^k) as in Section 3.1, and the proof of Lemma 3.1 can be carried over as well.

Lemma 3.2. $x^k \in [-1, 2]$ *for all* $k \geq 1$.

Proof. The proof is virtually identical to that of Lemma 3.1. Note that $u^k \leq \lambda^k$ for all k; hence, any upper bound for λ^k automatically translates to one for u^k. □

As with the standard ALM, it follows that the sequence generated by the modified ALM has at least two limit points, one in $[-1, 0]$ and one in $[1, 2]$. Using standard convergence theorems, e.g. Theorem 2.6 or [7, Thm. 6.2 and 6.3], we know that

- every limit point of (x^k) is a stationary point of $\|g_+(x)\|^2$, and
- every feasible limit point of (x^k) where g satisfies CPLD is a KKT point.

Clearly, the interval $[-1, 0]$ consists only of infeasible points. However, the point $x = 0$ is the only point in this interval which is a stationary point of $\|g_+\|^2$. Hence, the subsequence of (x^k) consisting of odd k must converge to $x = 0$. On the other hand, the interval $[1, 2]$ consists entirely of feasible points, and CPLD (in fact, LICQ) holds at every one of these points. Hence, the subsequence of (x^k) consisting of even k converges to $x = 1$ which is the solution (and only stationary point) of the optimization problem (12).

3.3 Numerical Results

Here, we give some numerical results illustrating the practical behaviour of the two methods. We chose the parameters

$$x^0 = -1, \quad \lambda^0 = 0, \quad \rho_0 = 1, \quad \tau = 0.1, \quad \gamma = 2, \quad u^{\max} = 10^4.$$

The subproblems are solved with the MATLAB® function `fminunc` and a tolerance of 10^{-8}. The overall stopping criterion is

$$|f'(x) + \lambda g'(x)| \le 10^{-4} \quad \text{and} \quad |\min\{-g(x), \lambda\}| \le 10^{-4}.$$

Table 1 shows the iterates generated by both algorithms.

	Standard Method			Modified Method		
k	x^k	λ^k	ρ_k	x^k	λ^k	ρ_k
1	-1.000000	0.000000	2^0	-1.000000	0.000000	2^0
2	-0.537207	1.155034	2^0	-0.537207	1.155034	2^0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	1.144714	$2.543809e-01$	2^{15}	1.144714	$2.543809e-01$	2^{15}
19	-0.003189	$3.276826e+04$	2^{16}	-0.003189	$3.276826e+04$	2^{16}
20	1.144714	$2.549944e-01$	2^{17}	1.048473	$3.032242e-01$	2^{17}
21	-0.001595	$1.310723e+05$	2^{18}	-0.001595	$1.310723e+05$	2^{18}
22	1.144714	$2.543809e-01$	2^{19}	1.012557	$3.251171e-01$	2^{18}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	0.000000	$1.374390e+11$	2^{38}	-0.000050	$1.342177e+08$	2^{28}
42	2.100000	0.000000	2^{39}	1.000012	$3.333247e-01$	2^{28}

Table 1: Numerical results for the optimization problem (12).

Up to iteration 19, the methods perform identically. This is because the bound $u^{\max} = 10^4$ only becomes active when $|\lambda^k| > u^{\max}$, which first occurs in iteration 19. Starting with $k = 20$, the modified method tends to the point $x = 1$ (for even k), while the standard method alternates between 1.144714 and (almost) zero. Moreover, the penalty parameter becomes substantially larger for the standard ALM, which eventually causes the subproblem algorithm to terminate unsuccessfully and return the value 2.1, which is not a solution of the subproblem. Note also that the modified method terminates successfully at $k = 42$ whereas the standard method eventually keeps alternating between the two points 0 and 2.1.

For a better understanding of the different behaviour of the two methods, note that both methods generate an unbounded multiplier sequence. However, a closer look at the ratio u^k/ρ_k shows that this sequence converges to zero for the modified ALM (which is clear since (u^k) is bounded), whereas the corresponding sequence λ^k/ρ_k (for k odd) from the standard ALM appears to converge to 1/2. This agrees with our theoretical analysis in Section 3.1.

4 Final Remarks

We have compared two variants of the well-known augmented Lagrangian method. In particular, we have shown that the modified ALM possesses stronger convergence properties and given an example which demonstrates how the safeguarding of multipliers can salvage convergence in certain cases. This example together with the increasingly rich background of the modified ALM (e.g. when applied to other problem classes) highlight the benefits of the modified method in terms of attaining feasibility, preventing ill-conditioning, and achieving optimality.

We note that our analysis explicitly only deals with convergence results for KKT points. Methods of the augmented Lagrangian type are known to enjoy certain properties with regard to global minimization as well, especially the modified ALM [6, 7]. It might be interesting to consider how the standard ALM behaves when applied to optimization problems for which (i) global minimization of the subproblems is feasible and (ii) the solution of the underlying problem does not satisfy the KKT conditions. However, due to the practical relevance of KKT conditions when dealing with general (nonlinear and non-convex) optimization problems, we have decided not to pursue this avenue in this report.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optim.*, 18(4):1286–1309, 2007.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, 111(1-2, Ser. B):5–32, 2008.

[3] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM J. Optim.*, 22(3):1109–1135, 2012.

[4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.

[5] E. G. Birgin, R. A. Castillo, and J. M. Martínez. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. *Comput. Optim. Appl.*, 31(1):31–55, 2005.

[6] E. G. Birgin, C. A. Floudas, and J. M. Martínez. Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Math. Program.*, 125(1, Ser. A):139–162, 2010.

[7] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.

[8] E. G. Birgin, J. M. Martínez, and L. F. Prudente. Optimality properties of an augmented Lagrangian method on infeasible problems. *Comput. Optim. Appl.*, 60(3):609–631, 2015.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28(2):545–572, 1991.

[10] M. R. Hestenes. Multiplier and gradient methods. *J. Optimization Theory Appl.*, 4:303–320, 1969.

[11] C. Kanzow and D. Steck. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.*, 26(4):2034–2058, 2016.

[12] C. Kanzow and D. Steck. Augmented Lagrangian and exact penalty methods for quasi-variational inequalities. *Technical Report*, Institute of Mathematics, University of Würzburg, June 2016.

[13] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[14] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Comput. Manag. Sci.*, 2(1):21–56, 2005.

[15] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968)*, pages 283–298. Academic Press, London, 1969.

[16] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control*, 12:268–285, 1974.