Operations Research Letters 45 (2017) 598–603

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

An example comparing the standard and safeguarded augmented Lagrangian methods

Christian Kanzow *, Daniel Steck

University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany

A R T I C L E I N F O

Article history:
Received 6 March 2017
Received in revised form 19 September 2017
Accepted 19 September 2017
Available online 27 September 2017

Keywords:
Augmented Lagrangian method
Nonlinear programming
Multiplier safeguarding
Counterexample
Global convergence

A B S T R A C T

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 [1,2,6–8] 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 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 [12] and quasi-variational inequalities [11]. 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. Sections 2.1 and 2.2 are dedicated to the standard and modified ALMs, respectively, and we give (or

* Corresponding author.
E-mail addresses: kanzow@mathematik.uni-wuerzburg.de (C. Kanzow), daniel.steck@mathematik.uni-wuerzburg.de (D. Steck).

https://doi.org/10.1016/j.orl.2017.09.005

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 $\rho > 0, \lambda \geq 0$, 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 $\lambda_1 v_1 + \cdots + \lambda_k v_k = 0, \lambda \geq 0$, 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 $\rho_k \to \infty$.