INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA

# ON GENERAL AUGMENTED LAGRANGIANS AND A MODIFIED SUBGRADIENT ALGORITHM


Doctoral thesis by

**Jefferson Divino Gonçalves de Melo**

Supervisor: Dr. Alfredo Noel Iusem

Co-supervisor: Dr. Regina Sandra Burachik


November 2009

*To my wife Leila, my daughter Yasmim*

*and all my family*

## Abstract

In this thesis we study a modified subgradient algorithm applied to the dual problem generated by augmented Lagrangians. We consider an optimization problem with equality constraints and study an exact version of the algorithm with a sharp Lagrangian in finite dimensional spaces. An inexact version of the algorithm is extended to infinite dimensional spaces and we apply it to a dual problem of an extended real-valued optimization problem. The dual problem is constructed via augmented Lagrangians which include sharp Lagrangian as a particular case. The sequences generated by these algorithms converge to a dual solution when the dual optimal solution set is nonempty. They have the property that all accumulation points of a primal sequence, obtained without extra cost, are primal solutions. We relate the convergence properties of these modified subgradient algorithms to differentiability of the dual function at a dual solution, and exact penalty property of these augmented Lagrangians. In the second part of this thesis, we propose and analyze a general augmented Lagrangian function, which includes several augmented Lagrangians considered in the literature. In this more general setting, we study a zero duality gap property, exact penalization and convergence of a sub-optimal path related to the dual problem.

**Keywords**: nonsmooth optimization, reflexive Banach spaces, sharp Lagrangian, general augmented Lagrangians, dual problem, modified subgradient algorithm, primal convergence, exact penalization, Hausdorff topological spaces.

## Agradecimentos

Várias pessoas contribuíram para que eu pudesse chegar até aqui. Começo então agradecendo aos meus familiares, que sempre deram apoio aos meus estudos e aos quais devo tudo o que sou. Ao meu irmão querido, também agradeço pela amizade e carinho. Agradeço a minha esposa Leila pela compreensão, apoio e carinho, e a minha filha Yasmim que está sempre iluminando minha vida e me motiva ainda mais a buscar meus objetivos.

Aos meus queridos orientadores Professores Alfredo N. Iusem e Regina S. Burachik, meus sinceros agradecimentos pelos ensinamentos, motivações e competência demonstrados ao longo desses anos. Não há palavras pra descrever o quanto vocês são especiais na minha vida.

Ao Professor Luis Román agradeço pelos ensinamentos e incentivos durante a graduação, os quais foram fundamentais para a continuação dos meus estudos.

Agradeço aos Professores Orizon e Geci pelo apoio, motivação e amizade.

Agradeço também aos membros da banca pelos comentários e sugestões.

A todos os amigos e colegas fica o meu muito obrigado por tudo. Em especial agradeço à Aldo Zang, Alysson Tobias, Artur Menezes, Acir Junior, Bom Kim, Bruno Souza, Carlos Melara, Cristina, Dalva, Fábio Júlio, Ives, Jean Carlos, Maicon Alves, Michel Molina, José Yunier, Roger Behling, Ruan e Zulu, pela amizade, pelas conversas matemáticas e pelos momentos de descontração proporcionados.

Gostaria de agradecer a todos os professores e demais funcionários do IMPA que participaram direta ou indiretamente da minha formação. Em especial, ao Professor Fernando Codá pelos ensinamentos e amizade. Meus agradecimentos aos funcionários da coordenação de ensino, pela eficiência e pela simpática forma que sempre me receberam.

Ao CNPq agradeço pelo apoio financeiro.

I would like to express my gratitude for all the staffs of the School of Mathematics and Statistics at the University of South Australia, specially Jorge Aarão, Regina Burachik, Maureen Coton and Yalcin Kaya, for their help and friendship. A substantial material of this thesis was obtained during my stay in this institute, which provided me a motivating environment and a good infrastructure.

Thank you all my friends: Ali, Andrew, Muntazan, Mustafa, Nahid, Essie, Paul, Arlene and Edmund. Finally, I am very grateful for the friendship and kindness of the Leverenz: John, Caroline and Renae.

# Contents

# Introduction

Duality is a very useful tool in optimization. The duality theory obtained through the ordinary (classical) Lagrangian and its use for convex optimization problems is well-known. However, when the primal problem is not convex, a duality gap (the difference between the optimal primal value and the optimal dual value) may exist when the ordinary Lagrangian is used, compromising the applicability of many methods. This justifies the quest for other kind of Lagrangians which are able to provide algorithms for solving a broader family of constrained optimization problems, including nonconvex ones.

The duality theory by means of a Lagrange-type function is used to transform the (primal) constrained optimization problem into a sequence of unconstrained subproblems, in such a way that information of the constraints is incorporated on this Lagrange-type function. This reduction is effective when the subproblems are much easier to solve than the original problem and the Lagrange-type function is able to provide zero duality gap property. The so-called augmented Lagrangian methods are among the best known methods for reduction to unconstrained optimization. Augmented Lagrangian methods for equality constrained optimization problems were introduced independently by Hestenes [25] and Powell [47]. Its extension to inequality constrained problems started with [19] and [50, 51]. An extensive study of augmented Lagrangian can be found in [3, 6, 7, 9, 30, 52, 53] and references therein. Many others Lagrange-type functions have been proposed and analyzed in the literature, see e.g. [7, 22, 35, 55, 57, 67].

Rockafellar and Wets in [54, Chapter 11] considered an optimization problem of minimizing a (possibly) nonsmooth and nonconvex extended real-valued function in a finite dimensional space. They proposed and analyzed a dual problem constructed via augmented Lagrangians with convex augmenting functions. This primal-dual scheme provides zero duality gap properties under mild assumptions [54, Theorems 11.59]. An exact penalty representation was defined and a criterion for such a property was presented [54, Theorem 11.61]. Recently, this duality approach has been studied in a more general setting. In Huang and Yang [26] the convexity assumption on the augmenting function, which is assumed by Rockafellar and Wets [54], is relaxed to a level boundedness assumption. They studied zero duality gap property, exact penalty representation and convergence of optimal paths in a unified approach. Rubinov et al. [58] considered augmented Lagrangians constructed by a family of augmenting functions with a peak at zero, and related the zero duality gap property with the lower semicontinuity of the perturbation function. Penot [43] and Zhou and Yang [70] considered duality via augmented Lagrangian, studied different growth conditions on the perturbation function and related them with zero duality gap property. Nedic and Ozdaglar [37] considered a geometric primal-dual approach and studied a zero duality gap property for augmented Lagrangian with a convex augmenting function. This geometric viewpoint was extended to augmented Lagrangian with some general augmenting function in [38]. An inequality constrained optimization problem is considered in [65, 63], where a unified approach for duality with some Lagrange-type functions, which include some augmented Lagrangians, is analyzed. Many efforts have been devoted to augmented Lagrangians with a valley at zero property on the augmenting function [17, 27, 58, 56, 68, 69, 70, 71]. Duality via general Lagrange-type functions for an optimization problem with general constraints can be found in [18, 64]. Augmented Lagrangian type functions constructed via auxiliary coupling functions have been considered and extensively studied in the literature [17, 44, 56, 58, 65, 71]. Penot and Rubinov [44] investigated the relationship between the Lagrangian multipliers and a generalized subdifferential of the perturbation function in ordered spaces. In Wang et al. [65, section 3.1], an augmented Lagrangian type function is studied via an auxiliary coupling function, and a valley at zero type property for the derivative of the coupling function with respect to the penalty parameter is proposed. Abstract convex analysis has demonstrated to be a natural language for constructing a duality theory for nonconvex optimization problems [55, 56, 61]. The issue of studying duality theory in the setting of augmented Lagrangian type functions with abstract convexity tools is reinforced in the works of Nedić et al. [41], Rubinov et al. [58], Penot and Rubinov [44] and most recently with the work of Burachik and Rubinov [17].

Another issue to consider in any duality theory is how to solve the dual problem. It is well known that in the classical Lagrangian duality the dual problem is a convex problem (i.e., the dual problem is the maximization of a concave function over a convex set), regardless of the properties of the primal problem. Also, in the process of evaluating the dual function, we obtain "without extra cost" some subgradient information. In the case of the duality constructed via augmented Lagrangian these results still hold. As a consequence, subgradient type methods are natural methods to solving the dual problem.

Subgradient methods are used to solve nonsmooth convex optimization problems. They were introduced by N. Z. Shor in the middle sixties and extensively studied since then. The work of Polyak [46] was fundamental for the development of this method. Subgradient methods are extensions of the gradient method to the convex nonsmooth optimization. One of the main difference between these methods is that the direction opposite to a subgradient is not, in general, a descent direction, as is the case for differentiable functions. We mention that in some very special situations it is still possible to obtain a subgradient such that its opposite direction is a descent direction, see for example [5]. Subgradient methods are slow methods, but they are very simple and their study can help to understand and devise better algorithms, so that they still attract the attention of many researchers. An extensive literature on subgradient methods and their variants can be found in [1, 2, 4, 10, 21, 32, 36, 46, 60] and references therein.

One of the main applications of subgradient methods appears in the context of classical Lagrangian duality, see for instance [39, 40] and references therein. In this context, we obtain a subgradient when solving the subproblems, which can be used to devise algorithms for solving the dual problem. In general, this information is not enough to obtain a descent direction for a subgradient algorithm. Indeed, the subgradient method has another property which makes its convergence possible, namely that the $k+1-$th iteration of a subgradient algorithm is closer to the solution set than the $k$-th iteration. In some variants of subgradient methods this latter property is satisfied partially, that is, we can compare and control the distance of iterates of the algorithm to the dual solution set. We mention that Féjer convergence analysis is a suitable tool for analyzing convergence of subgradient methods [1]. As commented before, the dual problem generated by augmented Lagrangian methods [26, 53, 54] is convex, and thus subgradient methods and its variants are suitable procedures for solving it. Among the augmented Lagrangians proposed by Rockafellar and Wets [54], we mention the "sharp Lagrangian", generated by taking as augmenting function the Euclidean norm.

Modified subgradient algorithms (MSg) have been considered for the dual problem constructed via sharp Lagrangian in finite dimensional spaces [12, 13, 14, 16, 23]. Gasimov in [23] considered a nonlinear optimization problem with equality constraints, and proposed a modified subgradient algorithm for solving the dual problem constructed via a sharp Lagrangian. A deflection in the parameter direction ensures that the dual values are strictly increasing. This increasing property makes the modified subgradient algorithm very attractive, since subgradient methods, as commented above, in general do not have this property. Dual convergence results were obtained and numerical experiments were presented in [23] to illustrate the behavior of the algorithm. In [12], Burachik et al. improved the results of [23] by relaxing the stepsize rule, and presented an example for which the algorithm may fail to achieve primal convergence. An auxiliary sequence, with an extra cost, was considered, and a primal convergence result was obtained for this sequence. In [14] an inexact version of the algorithm with a variant of the stepsize rule considered in [12] is proposed and analyzed. Similar results to those of the exact version were obtained. A version of this modified subgradient algorithm is considered in [24] for solving the dual problem via augmented Lagrangians where the primal problem has a single constraint. The idea of these modified subgradient algorithms was carried out to propose an update rule for a penalty method [13]. The applicability of these algorithms (exact and inexact versions) is better when the primal optimal value is known, or at least a good estimate, see Section 1.6 and [12, Eqs. 16, 23 and Section 5.1] and also [14, Corollary 5.1]. In some problems, even an approximate optimal value is both very hard to obtain and expensive. Therefore, it is desirable to look for a different stepsize rule, unrelated to the optimal value, and such that the convergence properties of the algorithm are preserved. It is also important to ensure convergence of the primal sequence which is generated during the process of the algorithm without any extra cost. This goals are achieved with the methods introduced in this thesis.

This thesis is divided in two parts: in the first part we investigate a modified subgradient algorithm for the dual problem generated by augmented Lagrangians, and in the second part we study general augmented Lagrangians. We now outline the thesis and present with details what we have done.

# Outline of the Thesis

In Chapter 1 we recall some preliminaries materials on Lagrange-type functions and subgradient algorithms. In particular, we present some existing results on augmented Lagrangians and a modified subgradient algorithm with sharp Lagrangian.

In Chapter 2 we consider the same modified subgradient algorithm as [12, 23], but we propose a very simple stepsize rule. With this rule we get rid of the dependence on the optimal value. We obtain dual convergence results as in [12, 23]. We also show that all accumulation points of the primal sequence generated by the algorithm are primal solutions, and thus no auxiliary sequence (as required in [12]) is needed. This primal convergence is ensured even if the dual optimal set is empty. The latter result is very important, because, in general, it is impossible to know a priori whether the dual problem has optimal solutions. We also show that if there exists a dual solution, then it is possible to consider larger stepsizes which ensure that after a finite number of iterations of the algorithm both primal and dual optimal solutions are reached.

The algorithm analyzed in Chapter 2 considers the dual problem induced by the sharp Lagrangian in finite dimensional spaces and assumes exact solution of the subproblems, which is too strong a requirement in many situations, specially in the context of nonsmooth and nonconvex optimization. In Chapter 3 we address this issue by developing an infinite dimensional modified subgradient method which accepts an inexact solution of the subproblems, and can be applied to duality schemes induced by a wide family of augmented Lagrangians in Banach spaces.

We consider a primal problem of minimizing an extended real-valued function (possibly nonconvex and nondifferentiable) in a reflexive Banach space. A duality scheme is considered via augmented Lagrangian functions which include the sharp Lagrangian as a particular case (see Example 3.1.1).

Our dual variables belong to a Hilbert space. Such duality schemes are suitable for solving constrained optimization problems in which the image space of the constraint function is a Hilbert space, see e.g. [28, 29, 62, 71] and references therein. Development and analysis of a given algorithm in infinite dimensional spaces gives a deeper insight into its properties. This issue has also practical interest since usually the performance of numerical algorithms in finite dimensions are closely related to the infinite dimensional performance, see for example [62] and references therein.

We propose a *parameterized* inexact modified subgradient algorithm for solving the dual problem. For this purpose we use a dualizing parameterization function (a function $f(\cdot,\cdot)$ such that $f(\cdot,0)=\varphi(\cdot)$, where $\varphi(\cdot)$ is the primal function). In order to ensure a monotone improvement of the dual values, we consider an augmenting function (not necessarily convex) similar to the one used in [24] (see assumption $(A_0)$ below). We prove that validity of $(A_0)$ is necessary for having a monotone increase of the dual values, see Proposition 3.2.18. This algorithm extends in several ways the one considered in Chapter 2. First, we extend the (finite dimensional) MSg to a reflexive Banach space. Second, our method accepts dual variables in a finite or infinite dimensional Hilbert spaces; as commented above this can have some advantages. Third, the convergence analysis of MSg assumes exact solution of the subproblems, while in Chapter 3 we establish convergence accepting inexact iterates, which is in fact the actual situation in computational implementations. Fourth, the sharp Lagrangian is just a particular case of a wide family of augmented Lagrangians that are considered in this chapter (see Example 3.1.1 and assumption $(A_0)$). Fifth, we consider in our analysis a level-boundedness assumption on the dualizing parameterization function (Definition 3.1.2) which is weaker than the compactness assumption used in Chapter 2 and [12, 14, 16, 23, 24].

We show that, in our more general setting, our algorithm generates a dual sequence strongly convergent to a dual solution when the dual solution set is nonempty. The primal sequence converges in the sense that all its weak accumulation points are primal solutions, even when the dual solution set is empty. We also analyze a stepsize selection rule which ensures that when the dual solution set is nonempty, approximate primal and dual solutions are obtained after a finite number of iterations of the algorithm.

In Chapter 4 we consider the same framework as in Chapter 3. We relate some properties of a penalty mapping to the differentiability of the dual function at a solution, and the convergence of modified subgradient algorithms with augmented Lagrangians in infinite dimensional spaces. In this chapter we characterize the primal convergence property of modified subgradient algorithm. We show that if the algorithm generates a dual sequence convergent to a dual solution, then the dual function is differentiable at this solution if and only if the primal sequence generated by the algorithm has all the accumulation points in the primal solution set. This result is also equivalent to strong exactness of a penalty mapping defined in this chapter.

In the second part of the thesis, Chapter 5, we consider a primal problem of minimizing an extended real-valued function in a Hausdorff topological space. A main tool in our analysis is abstract convexity, which recently became a natural language to investigate duality schemes via augmented Lagrangian type functions, see for example [17, 41, 44, 56, 58]. With abstract convexity tools, we propose and analyze a duality scheme induced by a general augmented Lagrangian function. We consider a valley at zero type property on the coupling (augmenting) function, which generalizes the valley at zero type property proposed in the related literature (e.g., [65, Section 3.1], [17] and references therein). In order to obtain our results, we demand continuity assumptions at a fixed base point instead of at the whole space, the latter being a standard assumption in the literature (see, e.g.,[17]). We show that our duality scheme has a zero duality gap property. A sub-optimal path related to the dual problem is considered, and we prove that all its cluster points are primal solutions. A criterion for exact penalization was presented by Rockafellar and Wets in [54, Theorem 11.61]. This criterion has been generalized, for instance, by Burachik and Rubinov [17], Huang and Yang [26] and Zhou and Yang [71]. We also extend this criterion to our general setting. Since no linearity on the augmented Lagrangian is assumed, this allows us to consider our primal-dual scheme in Hausdorff topological spaces. The main motivation for working in the framework of Hausdorff topological spaces is to develop a duality theory that can encompass different settings found in the literature, such as metric spaces (see e.g., [18, 58, 42, 64]) and Banach spaces with the weak topology (see e.g., [17, 69, 71]), which in general are not metrizable. It is also worthwhile to note that the general augmented Lagrangian studied here, for which the valley at zero type property is assumed directly at the coupling function $\rho$ (see Section 5.1), has not been considered in the literature even in finite dimensional spaces.

# Notations

$\mathbb{R}^n$, the $n$-dimensional Euclidean space,

$B_X[0, r]$, closed ball of $X$ with radius $r$,

$\partial g$, subdifferential (superdifferential ) of a convex (concave) function $g$,

$X^*$, the topological dual of a Banach space $X$,

$\|\cdot\|$, the norm in a Banach space

$\langle \cdot, \cdot \rangle$, the scalar product of a Hilbert space,

$\bar{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$,

$\mathbb{R}_{++} := (0, +\infty)$, $\mathbb{R}_+ := [0, \infty)$, $\mathbb{R}_{+\infty} = \mathbb{R} \cup \{+\infty\}$, $\mathbb{R}_{-\infty} = \mathbb{R} \cup \{-\infty\}$,

$m_p$, the primal optimal value,

$m_d$, the dual optimal value,

lsc, abbreviation for lower semicontinuous function

usc, abbreviation for upper semicontinuous function

# Chapter 1

# Background and preliminaries

In this chapter we present some preliminary materials on a primal-dual scheme via augmented Lagrangians. We recall the classical augmented Lagrangian and penalty function methods. We present different Lagrange-type functions. We recall some materials on abstract convexity. We consider a duality approach which is used to solving an extended real-valued optimization problem [54, Chapter 11]. In the second part of this chapter we recall subgradient method and present the previous materials on a modified subgradient algorithms applied to a dual problem generated via sharp Lagrangian [12, 14, 23]. We also discuss the difference between our work and the previous literature on the subject.

In this chapter we denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^m$.

## 1.1 Classical augmented Lagrangian and penalty functions

We recall in this section some preliminary results on penalty methods and augmented Lagrangian. A complete material can be found in several books on optimization, for instance [6, 8].

Consider the following equality constrained optimization problem:

$$\text{minimize } f(x) \quad \text{s.t.} \quad x \text{ in } K, \;\; h(x) = 0, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ e $K \subset \mathbb{R}^n$. In this formulation we assume that any possible complicate constraints are contained in the equality ones. The set $K$ is usually a box or $K = \mathbb{R}^n$.

The penalty methods attempt to solve the problem (1.1) via a sequence (finite or not) of easier unconstrained optimization problems. In order to do so, the infeasible points are penalized, that is a high cost is given for points which are not in the feasible region ($\{x \in K : h(x) = 0\}$ ). We describe next this procedure:

Consider a penalty function $P : \mathbb{R}^n \to \mathbb{R}_+$ such that $P(x) = 0$ if and only if $h(x) = 0$. Let $\{c_k\}$ be a sequence of penalty parameter (possible satisfying $c_k \uparrow +\infty$). For each $k$, find $x^k$ solution of the following subproblem:

$$\text{minimize } f(x) + c_k P(x) \quad \text{s.t.} \quad x \text{ in } K. \tag{1.2}$$

Under mild assumptions, the penalty method generates a sequence such that all its limit points are primal solutions. It is necessary in many situations to assume that the sequence of penalty parameters $\{c_k\}$ goes to infinity.

Several penalty functions have been proposed in the literature, and we shall consider just some basic ones. The first one is the quadratic penalty function, which is given by $P(x) = \frac{1}{2}\|h(x)\|^2$. This penalty function inherits differentiability properties of the problem (1.1). However it is necessary to increase the parameter $c_k$ to infinity in most cases, causing ill-conditioning of the subproblems (1.2). Some usual nondifferentiable penalty functions are $P(x) = \sum_{j=1}^{m} |h_j(x)|$ and $P(x) = \max_{j=1,\dots,m} \{|h_j(x)|\}$. These penalty functions have the properties that, under mild assumptions, it is sufficient to solve only one subproblem in order to solve or to obtain significant information of the problem (1.1), see [6, 8]. We mention that although only one subproblem needs to be solved, it is not possible to know a priori how large an initial penalty parameter must be. Therefore these methods still give rise to ill-conditioning. It is also important to note that since the penalty function is nondifferentiable, it destroys any differentiability property of the constrained problem. The quadratic penalty approach can be improved to the following situation, which consist of combining the classical penalty approach with the traditional Lagrangian function $L(x, \lambda) = f(x) + \langle h(x), \lambda \rangle$. More precisely, the subproblems become:

$$\text{minimize } f(x) + \langle h(x), \lambda^k \rangle + \frac{c_k}{2} \|h(x)\|^2 \quad \text{s.t.} \quad x \text{ in } K. \tag{1.3}$$

where $\{\lambda^k\}$ is a bounded sequence. We mention that the optimum value of the subproblem above can be equal to $-\infty$ even if the primal problem has a finite optimal value. Such a situation is avoided, for example, when $K$ is a compact set, and the functions involved are continuous. We present next two basic results related to this method. Let $L_c(x, \lambda) := f(x) + \langle h(x), \lambda \rangle + \frac{c}{2}\|h(x)\|^2$.

**Proposition 1.1.1.** *Assume that $f$ and $h$ are continuous functions, that $K$ is a closed set, and that the constraint set $\{x \in K : h(x) = 0\}$ is nonempty. For $k = 0, 1, ...$, let $x^k$ be a global minimum of the problem* (1.3)*, where $\{\lambda^k\}$ is bounded, $0 < c_k < c_{k+1}$ for all $k$, and $c_k \to \infty$. Then every limit point of the sequence $\{x^k\}$ is a global minimum of the original problem* (1.1)*.*

*Proof.* For the proof see [8, Proposition 4.2.1]. □

**Proposition 1.1.2.** *Assume that $K = \mathbb{R}^n$, and $f$ and $h$ are continuously differentiable. For $k = 0, 1, ...$, let $x^k$ satisfying*

$$\|\nabla_x L_{c_k}(x^k, \lambda^k)\| \leq \varepsilon_k,$$

*where $\{\lambda^k\}$ is bounded, $0 < c_k < c_{k+1}$ for all $k$, and $c_k \to \infty$, and $\varepsilon_k \geq 0$ for all $k$, and $\varepsilon_k \to 0$. Assume that a subsequence $\{x^{k_j}\}$ converges to a vector $x^*$ such that $\nabla h(x^*)$ has rank $m$. Then*

$$\{\lambda^{k_j} + c_{k_j} h(x^{k_j})\} \quad \textit{converges to} \ \lambda^*,$$

*where $\lambda^*$ is a vector satisfying, together with $x^*$, the first order necessary conditions:*

$$\nabla f(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(x^*) = 0, \quad h(x^*) = 0.$$

*Proof.* For the proof see [8, Proposition 4.2.2]. □

Proposition 1.1.2 suggests an updating rule for $\{\lambda_k\}$. Indeed, given a sequence of parameters $\{c_k\}$ and $\lambda^0 \in \mathbb{R}^m$, the multipliers method consists of generating a sequence $\{\lambda^k\}$ through the formula

$$\lambda^{k+1} = \lambda^k + c_k h(x^k),$$

where $x_k$ is a solution of the subproblem (1.3).

This method overcomes or attenuates some of the difficulties of the quadratic penalty method, see [7].

In the first part of this thesis we shall consider a sharp Lagrangian, which uses $P(x) = \|h(x)\|$. This augmented Lagrangian is more suitable for nondifferentiable problems, because of the nondifferentiability of the penalty function $\|\cdot\|$ at 0.

An inequality constrained optimization problem:

$$\text{minimize } f(x) \quad \text{s.t.} \quad g(x) \leq 0, \tag{1.4}$$

where $g : \mathbb{R}^n \to \mathbb{R}^m$, can be treated as an equality one by augmenting the domain, introducing additional variables: $h_j(x, z) = g_j(x) + z_j^2$ or by considering $h_j(x) = \max\{0, g_j(x)\}$, for $j = 1, ..., m$. We mention however that in some situations it is interesting to tackle the problem directly without modifying it. Rockafellar in the seventies proposed and analyzed the following quadratic augmented Lagrangian for inequality constrained optimization problem:

$$l(x, \mu, c) = f(x) + \frac{1}{2c} \sum_{j=1}^{m} \{[\max\{0, \mu_j + c g_j(x)\}]^2 - \mu_j^2\}.$$

In this context the subproblems are

$$\text{minimize } \ l(x, \mu^k, c_k) \ \text{ s.t. } \ x \in \mathbb{R}^n. \tag{1.5}$$

The multipliers method for inequality constraints is similar to the one for equality constraints. The iteration formula to updating the multipliers is $\mu_j^{k+1} = \max\{0, \mu_j^k + c_k g_j(x^k)\}$, where $x^k$ is a solution of (1.5). Several results concerning the quadratic augmented Lagrangian and the multipliers method can be found in the literature, for instance [7, 8, 51, 52, 53]. A detailed review on augmented Lagrangian for inequalities constraints can be found in [53].

## 1.2 Classical primal dual scheme

The classical (linear) Lagrange function for problem (1.4) is given by

$$L(x, \mu) = f(x) + \langle g(x), \mu \rangle.$$

The dual problem via the Lagrange function is

$$\max q(\mu) \quad \text{s.t.} \quad \mu \geq 0,$$

where the dual function $q : \mathbb{R}^m \to \mathbb{R}$ is defined by $q(\mu) = \inf_{x \in K} L(x, \mu)$.

This dual problem is suitable specially when the primal problem (1.4) is convex. For nonconvex optimization problem a positive duality gap can hold. We recall that the duality gap is the difference between the optimal values of the primal and the dual problems, that is,

$$\inf_{x \in K : g(x) \leq 0} f(x) - \sup_{\mu \geq 0} q(\mu).$$

It is straightforward to verify that weak duality holds:

$$\sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in K : g(x) \leq 0} f(x).$$

Therefore the duality gap is nonnegative. When the duality gap is zero, it is possible to solve the primal problem (1.4) via the dual problem, and this approach is suitable in several situations, see [8, Chapter 5]. When the duality gap is positive, this duality approach is not very successful, specially because many methods are efficient only when zero duality gap holds. We mention however, that the use of Lagrange function for nonconvex problem is also interesting in some particular situations, for example in integer programming. In order to deal with nonconvex problems, an alternative is the use of an augmented Lagrangian function (Section 1.1) instead of the Lagrange function. Another approach is the use of nonlinear Lagrange-type function, briefly described next.

## 1.3 Nonlinear Lagrange-type function

In this section we shall give only a basic description of a nonlinear Lagrange-type function; a complete material can be found in [55, 56]. We consider nonlinear Lagrange-type function generated by increasing positive homogeneous functions.

A function $p : \mathbb{R}^m \to \mathbb{R}$ is called increasing positive homogeneous (IPH) iff

*a*) $p(x) \geq p(y)$ when $x \geq y$, where $x \geq y$ means $x_j \geq y_j$ for $j = 1, ..., m$;

*b*) $p(tx) = tp(x)$ for all $t > 0$.

We assume the following properties on the IPH functions:

*i*) There exist positive numbers $a_0, a_1, ..., a_m$ with $a_0 = 1$ such that

$$p(y) \geq \max_{0 \leq j \leq m} a_i y_i, \quad \text{for all } y = (y_0, ..., y_m) \text{ with } y_0 \in \mathbb{R}_+;$$

*ii*) for $y_0 \in \mathbb{R}_+$ it holds that $p(y_0, 0, ..., 0) = y_0$.

We mention that the linear IPH function ($p(y) = \sum_{j=0}^{m} y_j$) does not satisfy these properties. Some examples of IPH functions satisfying properties ($i$) and ($ii$) are:

$p(y) := \max\{y_0, a_1y_1, ..., a_my_m\}$, where $a_1, ..., a_m \in \mathbb{R}_+$;

$$p_k(y_0, ..., y_m) = \left(\sum_{j=0}^{m} \max\{y_j, 0\}^k\right)^{\frac{1}{k}} \quad 0 < k < +\infty.$$

Consider the inequality constrained optimization problem (1.4). Assume that each of the functions of the problem is continuous, and $\inf_{x\in K} f(x) > 0$.

The function $\mathcal{L}$ defined by

$$\mathcal{L}(x, d) = p(f(x), d_1g_1(x), ..., d_mg_m(x)), \quad x \in \mathbb{R}^m, \quad d_j \in \mathbb{R}_+ \ \forall \ j = 1, ..., m,$$

is called an extended Lagrange function for the problem (1.4). The dual function corresponding to $p$ is defined by $q_p(d) = \inf_{x\in K} \mathcal{L}(x, d)$. The dual problem is

$$\text{maximize} \quad q_p(d) \ \text{ s.t. } \ d \in \mathbb{R}^m_+. \tag{1.6}$$

For the proof of the next theorem see [55, Theorem 4.1].

**Theorem 1.3.1.** *Assume that the feasible set of the problem* (1.4) *is a compact set or (if $K$ is unbounded) the function $f$ satisfies* $\lim\limits_{x\in K, x\to\infty} f(x) = +\infty$. *Then there is no duality gap between the primal problem* (1.4) *and its dual* (1.6)*, constructed via extended Lagrange function with properties* (*i*) *and* (*ii*).

## 1.4 Augmented Lagrangian with convex augmenting functions

In this section, we present some of the main results on the primal-dual scheme via augmented Lagrangian, proposed and analyzed by Rockafellar and Wets in [54, Chapter 11].

We consider the following primal problem.

$$\text{minimize } \varphi(x) \quad \text{s.t. } \ x \in \mathbb{R}^n, \tag{1.7}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}_{+\infty}$ is a proper, lsc function. A *dualizing parameterization* for $\varphi$ is a function $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$, such that $\phi(x, 0) = \varphi(x)$ for all $x \in \mathbb{R}^n$. We recall that the perturbation function of the primal problem is defined as

$$\beta(z) = \inf_x \phi(x, z),$$

where $\phi$ is the parameterization function.

The augmented Lagrangian $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ is defined as

$$\ell(x, y, c) := \inf_{z \in \mathbb{R}^m} \{\phi(x, z) - \langle y, z \rangle + c\sigma(z)\},$$

where $\sigma : \mathbb{R}^m \to \mathbb{R}_{+\infty}$ is a lsc and convex augmenting function such that

$$\operatorname*{Argmin}_x \sigma(x) = 0 \quad \text{and} \quad \sigma(0) = 0.$$

The dual function $\psi : \mathbb{R}^m \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ induced by the augmented Lagrangian $\ell$ is defined by

$$\psi(y, c) = \inf_{x \in \mathbb{R}^n} \ell(x, y, c).$$

The dual problem is given by

$$\text{maximize } \psi(y, c) \quad \text{s.t.} \quad (y, c) \in \mathbb{R}^m \times \mathbb{R}_+. \tag{1.8}$$

Denote by $m_p$ and $m_d$ the optimal primal and dual values respectively.

The next definition was introduced in [54, Definition 1.16].

**Definition 1.4.1.** *A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$, is said to be* level-bounded in $x$ locally uniform in $z$, *if for all $\bar{z} \in \mathbb{R}^m$ and for all $\alpha \in \mathbb{R}$, there exist an open neighborhood $V \subset \mathbb{R}^m$ of $\bar{z}$, and a bounded set $B \subset \mathbb{R}^n$, such that*

$$L_{V,f}(\alpha) := \{x \in \mathbb{R}^n : f(x, z) \leq \alpha\} \subset B, \quad \textit{for all } z \in V.$$

Next we summarize some basic results concerning the primal problem (1.7) and its dual (1.8).

**Proposition 1.4.1.** *Consider the primal problem* (1.7) *and the dual problem* (1.8)*. The following statements hold.*

*(i) The dual function $\psi$ is a concave and upper semicontinuous function.*

*(ii) If $r \geq c$ then $\psi(y, r) \geq \psi(y, c)$ for all $y \in \mathbb{R}^m$. In particular, if $(y, c)$ is a dual optimal solution, then also $(y, r)$ is a dual optimal solution.*

*Proof.* Item (i) follows from the fact that $\psi$ is the infimum of affine functions. Item (ii) is a consequence of the fact that $\sigma$ is nonnegative. □

The next theorem guarantees that there is no duality gap for the primal-dual pair (1.7)-(1.8).

**Theorem 1.4.2.** *Consider the primal problem* (1.7) *and its dual problem* (1.8)*. Assume that the dualizing parameterization function $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ for the primal function $\varphi$ is proper, lsc and level bounded in $x$ locally uniform in $z$. Suppose that there exists some $(y, r) \in \mathbb{R}^m \times \mathbb{R}_+$, such that*

$$\inf_{(x,z)} \{\phi(x, z) - \langle y, z\rangle + r\sigma(z)\} > -\infty.$$

*Then zero duality gap holds, i.e. $m_p = m_d$.*

*Proof.* See for instance, [54, Theorem 11.59]. For the proof of this result in a more general setting, see Theorem 5.2.3 in Chapter 5. □

Another important property related to this augmented Lagrangian is the exact penalty representation. This concept has been generalized to more general contexts [17, 26, 71]; see also Chapter 5.

**Definition 1.4.2.** *Consider the primal and dual problems* (1.7)-(1.8)*. A vector $\bar{y} \in \mathbb{R}^m$ is said to* support an exact penalty representation for problem (1.7) *if there exists $\bar{r} \in \mathbb{R}_+$ such that for any $r > \bar{r}$,*

*$E_1$) $\inf_x l(x, \overline{y}, r) = \inf_x \varphi(x)$;*

$E_2$) $\operatorname*{Argmin}_x \varphi(x) = \operatorname*{Argmin}_x l(x, \overline{y}, r)$.

Together with this concept of exact penalization, Rockafellar and Wets proved the following criterion.

**Theorem 1.4.3.** *[54, Theorem 11.61] Let $\beta : \mathbb{R}^m \to \bar{R}$ be the perturbation function. With the same assumption of Theorem 1.4.2 the following assertions are equivalent:*

*i) There exist an open neighborhood $V \subset Z$ of $0$ and $\hat{r} > 0$ such that*

$$\beta(z) \geq \beta(0) + \langle \overline{y}, z \rangle - \hat{r}\sigma(z) \ \textit{ for all } \ z \in V;$$

*ii) the vector $\overline{y}$ supports an exact penalty representation for problem* (1.7).

The above considered scheme includes a wide family of augmented Lagrangians. Indeed, many specific augmented Lagrangians can be constructed by choosing different convex augmenting function. We describe now two particular augmented Lagrangians. The first one is constructed by choosing the augmenting function $\sigma(z) = \frac{1}{2}\|z\|^2$ and it is called a proximal (or quadratic) Lagrangian. The second one is the sharp Lagrangian, which is constructed by taking the augmenting function $\sigma(z) = \|z\|$. These examples are discussed in [54, Examples 11.57 e 11.58].

**Remark 1.4.4.** In [54], the sharp Lagrangian function is generated by taking, as the augmenting function, any norm function in the Euclidean space. However, in this thesis we use this notation for the case of the Euclidean norm, that is, $\sigma(z) = (\sum_{j=1}^{m} |z_j|^2)^{\frac{1}{2}}$ for all $z \in \mathbb{R}^m$, which guaranteed some additional properties of the resulting modified subgradient algorithm. For example, if we consider the infinity norm $\sigma(z) = \max_{j=1,\ldots,m} |z_j|$, then the modified subgradient algorithm may fail to have the increasing property proved in [14, 23] and assumed in [12, 16] (see example 3.29 in Chapter 3). We mention, parenthetically, that the 1-norm, $\sigma(z) = \sum_{j=1}^{m} |z_j|$, could also be used. We also remark that in Chapter 3 we study the modified subgradient algorithm applied to the dual problem constructed with a much more general family of augmented Lagrangians. In particular the algorithm can be modified so that it admits the infinity norm as augmenting function (see Remark 3.2.19 in Chapter 3).

The next example show that the augmented Lagrangians considered in this section generalizes the classical augmented Lagrangian for constrained optimization problems, which is presented in Section 1.1.

**Example 1.4.5.** Consider the equality constrained optimization problem (1.1). For $z \in \mathbb{R}^m$, define

$$\Omega(z) := K \cap \{x : h(x) = z\}.$$

Recall that given $A \subset \mathbb{R}^n$, the indicator function $\delta_A : \mathbb{R}^n \to \mathbb{R}_{+\infty}$ is defined as

$$\delta_A(x) = 0, \text{ if } x \in A; \text{and } \delta_A(x) = \infty \text{ otherwise.}$$

Let

$$\varphi(x) := \begin{cases} f(x) & x \in \Omega(0), \\ \infty & \text{otherwise.} \end{cases}$$

Consider now a dualizing parameterization function $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ given by $\phi(x, z) = f(x) + \delta_{\Omega(z)}(x)$. It is easy to see that $\phi(x, 0) = \varphi(x)$ for all $x \in \mathbb{R}^n$. In this situation we obtain

$$\ell(x, y, r) = f(x) - \langle y, h(x) \rangle + r\sigma(h(x)) + \delta_K(x)$$

In particular taking $\sigma(z) = \frac{1}{2}\|z\|^2$ we obtain

$$\ell(x, y, r) = f(x) - \langle y, h(x) \rangle + \frac{r}{2}\|h(x)\|^2 + \delta_K(x), \tag{1.9}$$

which is the classical quadratic augmented Lagrangian plus the indicator function of $K$. In particular, taking $K = \mathbb{R}^n$ we obtain

$$\ell(x, y, r) = f(x) - \langle y, h(x) \rangle + \frac{r}{2}\|h(x)\|^2.$$

In the case of the constrained optimization problem above, if we consider the sharp Lagrangian ($\sigma(\cdot) = \|\cdot\|$), we obtain

$$\ell(x, y, r) = f(x) - \langle y, h(x)\rangle + r\|h(x)\| + \delta_K(x). \tag{1.10}$$

In Chapter 2 we study a modified subgradient algorithm applied to the dual problem constructed by sharp Lagrangian, where the primal problem is (1.1).

## 1.5 Abstract convexity

In the second part of the thesis we use abstract convexity for analyzing our primal-dual scheme. In order to better understand our results we recall here some basic definitions and results of abstract convexity. A detailed material can be found in [55, 56].

**Definition 1.5.1.** *Consider a nonempty set $Z$. Let $H$ be a nonempty set of (elementary) functions $h : Z \to \mathbb{R}$. A function $f : Z \to \mathbb{R}$ is called abstract convex with respect to $H$ (or $H$-convex) if there exists a set $U \subset H$ such that $f$ is the upper envelop of this set:*

$$f(z) = \sup\{h(z) : h \in U\} \text{ for all } z \in Z.$$

**Example 1.5.1.** Let $Z = \mathbb{R}^n$ and take $H := \{\langle\cdot, w\rangle + c : c \in \mathbb{R}, w \in \mathbb{R}^n\}$ (the set of all affine functions defined in $\mathbb{R}^n$). In this context a function $f$ is $H$-convex if and only if $f$ is lower semicontinuous and convex (in the classical definition); see [48, Theorem 12.1]. If $H$ is the set of all linear functions defined in $\mathbb{R}^n$ then $f$ is $H$-convex iff it is a lower semicontinuous sublinear function.

In the following we present this concept of abstract convexity involving a coupling function. Let $(Z, \Omega)$ be a pair of nonempty sets, and take a coupling function $\rho : Z \times \Omega \to \mathbb{R}$. Consider the following set of elementary functions defined in $Z$:

$$H_\rho := \{\rho(\cdot, w) + c : (w, c) \in \Omega \times \mathbb{R}\}.$$

Given a function $g : Z \to \mathbb{R}_{+\infty}$, the support of $g$ with respect to $H_\rho$ is defined by

$$\mathrm{Supp}(g, H_\rho) := \{h \in H_\rho : h \leq g\},$$

where $h \leq g$ means $h(z) \leq g(z)$ for all $z \in Z$. The $H_\rho$-convex hull of $g$ is defined as

$$\mathrm{co}_{H_\rho} g(z) := \sup\{h(z) : h \in \mathrm{Supp}(g, H_\rho)\} \quad \text{for each} \quad z \in Z.$$

The function $g$ is called abstract convex (or $H_\rho$-convex) at $\bar{z}$ iff

$$g(\bar{z}) = \mathrm{co}_{H_\rho} g(\bar{z}).$$

A function $g$ is abstract convex (with respect to $H_\rho$) if and only if it is abstract convex at all $z \in Z$.

**Definition 1.5.2.** *Consider a function $g : Z \to \bar{\mathbb{R}}$. Given $\varepsilon \geq 0$, we say that $w \in \Omega$ is an $\varepsilon-$*abstract subgradient of *$g$ at $\overline{z}$ (with respect to $\rho$) if*

$$g(z) - \rho(z, w) \geq g(\overline{z}) - \rho(\overline{z}, w) - \varepsilon \quad \textit{for all } z \in Z. \tag{1.11}$$

*The set of $\varepsilon-$abstract subgradients of $g$ at $\overline{z}$, denoted by $\partial_{\rho,\varepsilon} g(\overline{z})$, is the* $\varepsilon-$abstract subdifferential of *$g$ at $\overline{z}$ with respect to the coupling function $\rho$. The $0-$abstract subdifferential at $\bar{z}$ is denoted by $\partial_\rho g(\overline{z})$, and is called* abstract subdifferential.

The Fenchel-Moreau conjugate and biconjugate functions of $g$ with respect to the coupling function $\rho$ are defined, respectively, by

$$g^\rho(w) = \sup_{z \in Z}\{\rho(z, w) - g(z)\}$$

and

$$g^{\rho\rho}(z) = \sup_{w \in \Omega}\{\rho(z, w) - g^\rho(w)\}.$$

It follows directly from this definition that

$$g^{\rho\rho}(z) \le g(z) \quad \text{for all} \quad z \in Z. \tag{1.12}$$

### 1.5.1 Abstract Lagrangian

Now we present a general duality scheme in the framework of abstract convex analysis. Let $X, Z$ be general nonempty sets (usually with some topological structure). Consider the primal problem:

$$\min \phi(x) \quad \text{s.t.} \quad x \in X. \tag{1.13}$$

Assume that an element of $Z$, denoted by 0, satisfies $\rho(0, w) = 0$ for all $w \in \Omega$. Let $f : X \times Z \to \bar{\mathbb{R}}$ be a parameterization function for problem (1.13), that is, $f(x, 0) = \phi(x)$. The perturbation function is $\beta(z) = \inf_x f(x, z)$. The abstract Lagrangian related to $\rho$ is defined by

$$\ell(x, w) = \inf_z \{f(x, z) - \rho(z, w)\},$$

the dual function $q : \Omega \to \bar{\mathbb{R}}$ is given by

$$q(w) = \inf_x \ell(x, w),$$

and the dual problem is stated as

$$\max q(w) \quad \text{s.t.} \quad w \in \Omega.$$

Observing that

$$-\beta^{\rho}(w) = \inf_{z \in Z} \{\beta(z) - \rho(z, w)\} = q(w),$$

it follows that

$$\beta^{\rho\rho}(0) = \sup_{w \in \Omega} \{\rho(0, w) - \beta^{\rho}(w)\} = \sup_{w \in \Omega} q(w),$$

where we used $\rho(0, w) = 0$. In particular, weak and strong duality are rewritten, respectively, as

$$\beta^{\rho\rho}(0) \leq \beta(0) \quad \text{and} \quad \beta^{\rho\rho}(0) = \beta(0).$$

In this context strong duality is related to abstract convexity at 0 of the function $\beta$ with respect to the family of functions $H_\rho$.

Let $Z = \mathbb{R}^m, \Omega = \mathbb{R}^m \times \mathbb{R}_+$ and $X = \mathbb{R}^n$. Consider a convex augmenting function $\sigma$, and take $\rho(z, y, c) = \langle y, z\rangle - c\sigma(z)$. In this context the abstract Lagrangian is an augmented Lagrangian with convex augmenting function presented in Section 1.4.

In the next proposition we relate $\partial_\rho\beta(0)$, $\mathrm{Supp}(\beta, H_\rho)$, and $\mathrm{dom}\,\beta^\rho$ to the dual function $q$.

**Proposition 1.5.2.** *Take $\bar{w} \in \Omega$. Then*

*i) $\bar{w} \in \partial_\rho\beta(0)$ if and only if $q(\bar{w}) = \beta(0)$;*

*ii) $\bar{w} \in \mathrm{dom}\,\beta^\rho$ if and only if there exists $\bar{c} \in \mathbb{R}$ such that $q(\bar{w}) \geq \bar{c}$, which in turn is equivalent to $\rho(\cdot, \bar{w}) + \bar{c} \in \mathrm{Supp}(\beta, H_\rho)$.*

*Proof.* Since weak duality holds, (i) follows from the following equivalences:

$$\begin{aligned} \bar{w} \in \partial_\rho\beta(0) \quad &\Leftrightarrow \quad \beta(z) - \rho(z, \bar{w}) \geq \beta(0) \quad (\forall\, z \in Z) \\ &\Leftrightarrow \quad \inf_z\{\beta(z) - \rho(z, \bar{w})\} \geq \beta(0) \\ &\Leftrightarrow \quad q(\bar{w}) \geq \beta(0). \end{aligned}$$

Since $q(\bar{w}) = -\beta^\rho(\bar{w})$, (ii) follows from the following equivalences:

$$\begin{aligned} \rho(\cdot, \bar{w}) + \bar{c} \in \mathrm{Supp}(\beta, H_\rho) \quad &\Leftrightarrow \quad \beta(z) \geq \ \rho(z, \bar{w}) + \bar{c} \quad (\forall\, z \in Z) \\ &\Leftrightarrow \quad \beta(z) - \rho(z, \bar{w}) \geq \bar{c} \quad (\forall\, z \in Z) \\ &\Leftrightarrow \quad \inf_z\{\beta(z) - \rho(z, \bar{w})\} \geq \bar{c} \\ &\Leftrightarrow \quad q(\bar{w}) \geq \bar{c}. \end{aligned}$$

$\square$

We describe now two approaches for studying augmented Lagrangians type functions. The first one was proposed in [17] and the second one was proposed in [66].

Let $X, Z$ be Banach spaces. Consider a function $s : \mathbb{R}^2 \to \mathbb{R}$ such that $s(0,0) = 0$ and for every $a \in \mathbb{R}$ and $b_1 \geq b_2$, it satisfies

$$s(a, b_1) - s(a, b_2) \geq \psi(b_1 - b_2), \tag{1.14}$$

where $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing function such that $\psi(0) = 0$ and $\psi$ is coercive, that is, $\lim_{t\to\infty} \psi(t) = \infty$.

Let $\{\nu_\eta\}_{\eta\in U_1}$ be a family of upper semicontinuous functions defined in $Z$ satisfying

$$\nu_\eta(0) = 0 \text{ for all } \eta \in U_1, \tag{1.15}$$

and consider a family of augmenting functions $\{\sigma_\mu\}_{\mu\in U_2}$ defined in $Z$ and having a valley at zero property, that is, for every $\mu \in U_2$, $\sigma_\mu : Z \to \mathbb{R}_+$ is proper, w-lsc and satisfies

$$\sigma_\mu(0) = 0 \quad \text{and} \quad \inf_{z\in V^c} \sigma_\mu(z) > 0, \tag{1.16}$$

for every open neighborhood $V \subset Z$ of 0.

The coupling function $\rho$ considered in [17] is given by

$$\rho(z, (\eta, \mu), r) = s(\nu_\eta(z), -r\sigma_\mu(z)).$$

In this context the abstract Lagrangian is

$$l(x, w) = l(x, \eta, \mu, r) = \inf_{z\in Z}\{f(x, z) - s(\nu_\eta(z), -r\sigma_\mu(z))\}$$

The corresponding dual function is $q(\eta, \mu, r) = \inf_x l(x, \eta, \mu, r)$ and the dual problem is stated as

$$\max q(\eta, \mu, r) \ \text{ s.t. } \ (\eta, \mu, r) \in U_1 \times U_2 \times \mathbb{R}_+.$$

In order to describe the second approach studied in [66], consider a coupling function $p : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$ such that $p(z, y, \cdot)$ is differentiable.

The dual scheme studied in [66] is constructed via the following nonlinear Lagrangian type function (taking $p = -\rho$)

$$l(x, y, r) = \inf_z \{f(x, z) + p(y, z, r)\},$$

where the coupling function $p$ has the following valley at zero type property:

There exists $\alpha \in [0, 1)$ such that, for every $\varepsilon > 0$ and $y \in \mathbb{R}^m$,

$$\inf_{\|u\| \geq \varepsilon, \tau \geq \varepsilon} \tau^\alpha p'_r(u, y, \tau) > 0.$$

This property was introduced in [65, Section 3.1] for an inequality constrained problem in finite dimensional spaces.

In [17, 66] the authors study the zero duality gap property, exact penalty representations and optimal paths. We present now the latter concept.

Recall that the calculation of the dual function leads to the following problem:

$$\inf\{f(x, z) - \rho(z, y, r) : (x, z) \in X \times Z\}. \tag{1.17}$$

**Definition 1.5.3.** *Let $I \subset \mathbb{R}_+$ be unbounded above, and for each $r \in I$ take $\varepsilon_r \geq 0$. The set $\{(x_r, z_r)\}_{r \in I} \subset X \times Z$ is called a* sub-optimal path of problem (1.17) *if*

$$f(x_r, z_r) - \rho(z_r, y, r) \leq q(y, r) + \varepsilon_r \tag{1.18}$$

*for all $r \in I$. When $(x_r, z_r)$ satisfies* (1.18) *with $\varepsilon_r = 0$ for all $r \in I$, the set $\{(x_r, z_r)\}_{r \in I}$ is called an* optimal path*.*

We mention that it is usual to consider $I = \mathbb{R}_+$. In [17] the authors studied only optimal paths (i.e., the minimization (1.18) is carried out exactly).

## 1.6 Modified subgradient algorithm-MSg

In this section we recall some preliminary material on a modified subgradient algorithm proposed by Gasimov [23] for solving the dual problem generated by sharp Lagrangian. Gasimov considered a deflected direction which guarantees that the algorithm is an ascent method. It is well known that subgradient method, in general, when applied to a maximization of a concave function over a convex set, is not an ascent method. The results of Gasimov was improved in [12] and an inexact version of MSg was proposed in [14]. Before presenting the modified subgradient algorithm we recall a general subgradient method.

Consider the optimization problem

$$\text{minimize } g(x) \quad \text{s.t. } x \in \mathbb{R}^n, \tag{1.19}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Denote by $\bar{g} := \inf_{x \in \mathbb{R}^n} g(x)$.

Recall that the subdifferential of a convex function $g$ at $x_0 \in \text{dom}g$ is defined by

$$\partial g(x_0) := \{v : g(x) \geq g(x_0) + \langle v, x - x_0 \rangle, \text{ for all } x \in \mathbb{R}^n\}.$$

An element of $\partial g(x_0)$ is a subgradient of $g$ at $x_0$. The superdifferential of a concave function $q$ at $y_0 \in \text{dom}\, q$ is the set

$$\partial q(y_0) := \{-v : v \text{ is a subgradient of the convex function } -q \text{ at } y_0\}.$$

The subgradient method is an iterative method of the form

$$x_{k+1} = x_k - s_k u_k, \tag{1.20}$$

where $x_0$ is an arbitrary initial point, $u_k \in \partial g(x_k)$ and $s_k \geq 0$ is a stepsize parameter. If at some iteration-$j$, the subgradient direction is 0 then the method stops. In such a situation, it follows from the subgradient inequality that $x_j$ is a solution of the problem (1.19). Therefore for the convergence results of the method is assumed that $u_k \neq 0$ for all $k$. Some well known stepsize rules for the subgradient method are:

- Diminishing stepsize rule:

  $\{s_k \|u_k\|\}$ converges to zero, and $\sum_k s_k = \infty$;

- Polyak's (or dynamic) stepsize rule:

$$s_k = \gamma_k \frac{g(x_k) - \bar{g}}{\|u_k\|^2}, \quad \text{with } 0 < \delta \leq \gamma_k \leq 2 - \delta.$$

It is well known that the diminishing stepsize rule, although simple, it makes the method slow in general. One of the reasons is that no information during the process of the algorithm is used. However, in some applications, it is very difficult and expensive to obtain informations of the problem at hand, for example in large scale optimization problems. Therefore for such problems, simple rules as constant ($s_k = s$ for all $k$, with $s \in \mathbb{R}_+$) and diminishing stepsizes are convenient and many strategies are used in order to overcome some drawbacks of the method [36, 39] and references therein. The Polyak's stepsize rule is better than diminishing stepsize when one has some"a priori" knowledge of the optimal value. In practice, in the case that the optimal value is not known, some strategies can be considered in such a way that significant convergence results of the method are obtained [32, 46]. One approach proposed in the literature in order to avoid the optimal value used in the Polyak's stepsize rule, is a variable target value [32]. Many variants and strategies to improve subgradient methods have been proposed and analyzed in the literature, see [11, 32, 36] and references therein. One of the main variant of subgradient methods is the"bundle method". This method is the main tool for solving a nonsmooth convex optimization problem and much research has been done in order to improve or adapt it to some particular applications. We will not deal with bundle methods in this thesis; we refer the reader interested on this method to [10, 31, 34] and references therein.

It is usual to consider a normalized direction in the subgradient method:

$$x_{k+1} = x_k - s_k \frac{u_k}{\|u_k\|}, \tag{1.21}$$

where $u_k \in \partial g(x_k)$ and $u_k \neq 0$. In the following, we state some of the well known convergence results for subgradient methods.

**Proposition 1.6.1.** *Consider the optimization problem* (1.19).

*a)* *The sequence* $\{x_k\}$ *generated by the subgradient method* (1.20) *satisfies* $\liminf_{k\to\infty} g(x_k) = \bar{g}$.

*b)* *Assume that the solution set of the problem* (1.19) *is nonempty, and the stepsize satisfy* $\sum_{k=1}^{\infty} s_k = \infty$ *and* $\sum_{k=1}^{\infty} s_k^2 < \infty$. *Then* $\{x_k\}$ *generated by the subgradient method* (1.21) *converges to a solution of* (1.19).

*Proof.* See [33, Propositions 1.2 and 5.1]. □

**Proposition 1.6.2.** *Assume that the solution set of the problem* (1.19) *is nonempty. Then the sequence* $\{x_k\}$ *generated by the subgradient method* (1.20) *with Polyak's stepsize rule converges to a solution of* (1.19).

*Proof.* See [11, Theorem 2.4]. □

In the following we state the modified subgradient algorithm. This algorithm is used to solve the dual problem of an equality constrained optimization problem (1.1). It was proposed by Gasimov [23]. We use the notation

$$A(y,c) := \operatorname*{Argmin}_{x\in K} \ell(x,y,c), \tag{1.22}$$

where $\ell(x,y,r) = f(x) - \langle \lambda, h(x)\rangle + c\|h(x)\|$ is the sharp Lagrangian.

The dual function is given by $q(y,c) = \min_{x\in K} \ell(x,y,c)$. The dual problem is

$$\max q(y,c) \ \text{ s.t. } \ (y,c) \in \mathbb{R}^m \times \mathbb{R}_+.$$

We mention that the set $A(y,c)$ is nonempty for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, because the function $\ell(\cdot,y,c)$ is lower semicontinuous for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, and $K$ is a compact set.

### Modified Subgradient Algorithm-MSg:

**Step** 0. Choose $(y_0, c_0) \in \mathbb{R}^m \times \mathbb{R}_+$, $\{\alpha_k\} \subset \mathbb{R}_{++}$, and set $k := 0$.

**Step** 1 (**Subproblem and stopping criterion)**

*a*) Find $x_k \in A(y_k, c_k)$,

*b*) if $h(x_k) = 0$ stop,

*c*) if $h(x_k) \neq 0$, go to **Step** 2.

**Step** 2 (**Stepsize selection and update of dual variable)**

Take $s_k > 0$ a stepsize and update the dual variables,

$y_{k+1} := y_k - s_k h(x_k)$,

$c_{k+1} := c_k + (1 + \alpha_k) s_k \|h(x_k)\|$.

Set $k = k + 1$ and go to **Step** 1.

In [23], Gasimov showed that MSg generates a monotonic increasing sequence of dual values, that is, $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$. The stepsize selection rule considered in [23] is a dynamic stepsize, more precisely, $s_k$ is chosen as

$$s_k = \frac{\bar{q} - q_k}{5\|h(x_k)\|^2}, \tag{1.23}$$

where $\bar{q}$ is the dual optimal value and $q_k = q(y_k, c_k)$. [23] showed that $\{q_k\}$ converges monotonically to $\bar{q}$ when the dual optimal solution is nonempty. The sequence $\{\epsilon_k\}$ with $\epsilon_k := \alpha_k s_k$, is chosen satisfying $\epsilon_k < s_k$ for all $k$. In [12] the authors improved the results of [23] relaxing the choice for the stepsize and proving that MSg generates a dual sequence which converges to a dual solution when the dual solution set is nonempty. An example presented in [12] shows that the primal sequence $\{x_k\}$ generated by MSg can converge to a point which is not a primal solution. In order to solve this inconvenient, they propose an auxiliary sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k \in A(y_k, c_k + \beta)$ for all $k$ and some $\beta > 0$, and establish optimality of all accumulation points of this sequence. However this auxiliary sequence has an extra cost. We show, in a more general setting, in Proposition 4.3.5 that this auxiliary sequence has finite termination, that is, there exists $\bar{k}$ such that $\tilde{x}_{\bar{k}}$ is a primal solution.

For further use and for the sake of completeness, we prove now two results related to MSg. The first one establishes the relation between the minimization implicit in $A(y,c)$ (defined in (1.22)) and the superdifferential $\partial q(y,c)$. The second one shows that MSg is an ascent method. The proofs of these results are very similar to the ones given in [23]. In Chapter 3 we prove these results in a more general setting.

**Theorem 1.6.3.** *The following results hold for* MSg.
*a) If $\hat{x} \in A(\hat{y},\hat{c})$, then $(-h(\hat{x}), \|h(\hat{x})\|) \in \partial q(\hat{y},\hat{c})$.*

*b) If the sequence $\{\alpha_k\}$ is bounded, then* MSg *generates a dual bounded sequence $\{(y_k, c_k)\}$ if and only if $\sum_k s_k \|h(x_k)\| < +\infty$.*

*c) If* MSg *stops at the $k-$*th *iteration, then $x_k$ is an optimal primal solution, and $(y_k, c_k)$ is an optimal dual solution.*

*Proof.* a) For all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_{++}$ we have

$$\begin{aligned} q(y,c) &= \min_{x\in K}\{f(x) - \langle h(x), y\rangle + c\|h(x)\|\} \\ &\leq f(\hat{x}) - \langle h(\hat{x}), y\rangle + c\|h(\hat{x})\| \\ &= f(\hat{x}) - \langle h(\hat{x}), \hat{y}\rangle + \hat{c}\|h(\hat{x})\| + \langle -h(\hat{x}), y-\hat{y}\rangle + (c-\hat{c})\|h(\hat{x})\|. \end{aligned} \tag{1.24}$$

Using that $\hat{x} \in A(\hat{y},\hat{c})$ in (1.24), we obtain

$$\begin{aligned} q(y,c) &\leq q(\hat{y},\hat{c}) + \langle -h(\hat{x}), y-\hat{y}\rangle + (c-\hat{c})\ \|h(\hat{x})\| \\ &= q(\hat{y},\hat{c}) + \langle (-h(\hat{x}), \|h(\hat{x})\|), (y,c) - (\hat{y},\hat{c})\rangle. \end{aligned}$$

That is, $(-h(\hat{x}), \|h(\hat{x})\|) \in \partial q(\hat{y},\hat{c})$.
b) Since $\{\alpha_k\}$ is bounded, the equivalence follows from the expressions:

$$\|y_{k+1} - y_0\| \leq \sum_{j=0}^{k} \|y_{j+1} - y_j\| = \sum_{j=0}^{k} s_j \|h(x_j)\|, \tag{1.25}$$

$$c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1) s_j \|h(x_j)\|. \tag{1.26}$$

c) Example 1.4.5 shows that the sharp Lagrangian is an augmented Lagrangian with convex augmenting function. Therefore, it follows from Theorem 1.4.2 that $m_p = m_d$. If MSg stops at iteration $k$, then $h(x_k) = 0$. Hence

$$m_d \;=\; m_p \leq f(x_k) = f(x_k) - \langle y_k, 0\rangle + c_k\|0\| = q(y_k, c_k) \leq m_p,$$

which implies that $m_d = q(y_k, c_k)$, and $f(x_k) = m_p$. That is to say, $x_k$ is an optimal primal solution, and $(y_k, c_k)$ is an optimal dual solution. The theorem is proved. □

**Theorem 1.6.4.** *Consider the primal problem* (1.1). *Assume that $f$ is a lower semicontinuous function, $h$ is a continuous function and $K$ is a compact set. Let $\{(y_k, c_k)\}$ be the sequence generated by* MSg. *If $(y_k, c_k)$ is not a dual solution, then $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$.*

*Proof.* The assumptions imply that for any $(y_j, c_j)$ there exists $x_j \in A(y_j, c_j)$, that is $q(y_j, c_j) = f(x_j) - \langle h(x_j), y_j\rangle + c_j\|h(x_j)\|$. Let $\epsilon_k := \alpha_k s_k$. Therefore

$$\begin{aligned}
q(y_{k+1}, c_{k+1}) &= \min_{x\in K}\{f(x) - \langle h(x), y_{k+1}\rangle + c_{k+1}\|h(x)\|\} \\
&= f(x_{k+1}) - \langle h(x_{k+1}), y_{k+1}\rangle + c_{k+1}\|h(x_{k+1})\| \\
&= f(x_{k+1}) - \langle h(x_{k+1}), y_k\rangle + c_k\|h(x_{k+1})\| + \\
&\quad (s_k + \epsilon_k)\|h(x_k)\|\|h(x_{k+1})\| + s_k\langle h(x_{k+1}), h(x_k)\rangle \\
&\geq q(y_k, c_k) + (s_k + \epsilon_k)\|h(x_k)\|\|h(x_{k+1})\| + s_k\langle h(x_{k+1}), h(x_k)\rangle.
\end{aligned}$$

Using Cauchy-Schwarz inequality we obtain

$$q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k) + \epsilon_k\|h(x_k)\|\|h(x_{k+1})\|. \tag{1.27}$$

If $h(x_{k+1}) = 0$, then $(y_{k+1}, c_{k+1})$ is a dual solution, by Theorem 1.6.3 ($c$). In particular $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$, because $(y_k, c_k)$ is not a dual solution, by assumption. Therefore assume that $h(x_{k+1}) \neq 0$. Since $(y_k, c_k)$ is not a dual solution, it follows that $h(x_k) \neq 0$. From (1.27) we obtain that

$$q(y_{k+1}, c_{k+1}) > q(y_k, c_k),$$

because $\epsilon_k > 0$ and $h(x_k), h(x_{k+1}) \neq 0$. The result follows. □

**Remark 1.6.5.** We mention that the results in Theorems 1.6.3, 1.6.4 do not depend on the choice of the stepsize $s_k$.

The next theorem is one of the main convergence results of MSg presented in [12]. For its proof, see [12, Theorem 10].

**Theorem 1.6.6.** *Consider the equality constrained problem* (1.1)*, where $f$ and $h$ are continuous function and $K$ is a compact set. Assume that the stepsize $s_k$ in MSg satisfies*

$$\frac{\eta(\bar{q} - q_k)}{\|h(x_k)\|^2} < s_k < \frac{2(\bar{q} - q_k)}{\|h(x_k)\|^2} \quad \textit{with}\ \eta \in (0, 2).$$

*Suppose also that the dual optimal solution is nonempty. Then:*

*i) the dual sequence $\{(y_k, c_k)\}$ converges to a dual solution;*

*ii) the sequence $\{q(y_k, c_k)\}$ converges to the optimal dual value;*

*iii) all accumulation points of the auxiliary sequence $\{\tilde{x}_k\}$ are solutions of the primal problem* (1.1)*.*

In [14] an inexact version of MSg was proposed and analyzed. They allowed that the subproblems are solved inexactly. More precisely, the subproblems are solved with possibly some errors $r_k$ satisfying

$$\ell(x_k, y_k, c_k) \leq q(y_k, c_k) + r_k. \tag{1.28}$$

[14] assumes the following conditions on the stepsize:

$a_1$) $s_k \geq \dfrac{\eta(\bar{q} - q_k) + \theta r_k}{\|h(x_k)\|^2}$ for some $\theta, \eta > 0$, and for all $k$;

$a_2$) the sequence $\{s_k \|h(x_k)\|\}$ is bounded.

With this stepsize rule, similar results of the exact version studied in [12] are obtained in [14]. We denote by IMSg the inexact version of MSg. One of their main convergence results is:

**Theorem 1.6.7.** *[14, Theorem 4.2] Assume that the dual optimal solution set is nonempty. Let $\{x_k\}$ and $\{(y_k, c_k)\}$ be generated by IMSg satisfying (1.28). Suppose that conditions $(a_1)$ and $(a_2)$ hold. Then the sequence $\{(y_k, c_k)\}$ generated by IMSg converges to a dual solution. In particular, $\{q_k\}$ tends to $\bar{q}$ and $\{r_k\}$ tends to 0. Moreover, if $\{x_k\}$ has an accumulation point $\bar{x}$ such that $h(\bar{x}) = 0$, then $\bar{x}$ is a primal solution.*

We emphasize that Theorem 1.6.6 and Theorem 1.6.7 hold in finite dimensional spaces and assume that the dual optimal solution set is nonempty. In the next chapter we propose a simple stepsize selection rule for MSg with sharp Lagrangian. We show that our stepsize rule ensures primal convergence even when the optimal dual solution is empty. In Chapter 3 we propose an inexact version of this algorithm for a broader class of augmented Lagrangians in infinite dimensional space. We obtain similar results as the ones obtained for the sharp Lagrangian in Chapter 2.

## 1.7 Subdifferential of the maximum of convex functions

For completing this chapter we present a characterization of the subdifferential of the maximum of convex functions. We use the following notation: $X$ is a Banach space and $X^*$ its topological dual. Given a set $C \subset X^*$, we denote its closure in the weak $^*$ topology by $\overline{C}^{w*}$. We denote the convex hull of $C$ by

$$co(C) := \left\{ \sum_{j=1}^{k} \alpha_j c_j \mid \sum_{j=1}^{k} \alpha_j = 1, c_j \in C \text{ and } k \in \mathbb{N} \right\}.$$

Consider a nonempty set $S$ and convex functions $f_s : X \to \mathbb{R}$ with $s \in S$. Let $f$ be a convex function given by

$$f(x) = \max_{s \in S} f_s(x), \quad x \in X. \tag{1.29}$$

We set

$$S(\bar{x}) := \{s \in S \mid f_s(\bar{x}) = f(\bar{x})\}.$$

**Theorem 1.7.1.** *Let $S$ be a compact Hausdorff space. For any $s \in S$, let $f_s : X \to \mathbb{R}$ be convex on $X$ and continuous at $\bar{x} \in X$. Assume further that there exists a neighborhood $U$ of $\bar{x}$ such that for every $z \in U$, the functional $s \mapsto f_s(z)$ is upper semicontinuous on $S$. Then the function $f : X \to \mathbb{R}$ defined by* (1.29) *satisfies*

$$\partial f(\bar{x}) = \overline{\mathrm{co}}^{w*}\left(\bigcup_{s \in S(\bar{x})} \partial f_s(\bar{x})\right).$$

*Proof.* The proof of this result can be found in [59, Proposition 4.5.2]. $\square$

# Chapter 2

# Modified subgradient algorithm with sharp Lagrangian

In this chapter we consider an optimization problem with equality constraints. We construct a dual problem via a sharp Lagrangian. In order to solve the dual problem, we use a modified subgradient algorithm. We propose two stepsize rules. The first one guarantees that all the accumulation points of a primal sequence generated by the algorithm, without extra cost, are primal solutions. The dual sequence converges to a dual solution when the dual solution set is nonempty. The second stepsize rule assures that the algorithm finds a primal-dual solution after a finite number of iterations when the dual solution set is nonempty. In Chapter 3 we will present a generalization of the results of this chapter in several directions (like inexact solution of the subproblems and infinite dimensional spaces). We have chosen to keep this chapter for the sake of clarity of the exposition, and also because have been useful as an introduction to the more involved material of Chapter 3.

This chapter is organized as follows. We start by recalling the primal dual problem via sharp Lagrangians. In Section 2.2 we state and analyze a first version of the modified subgradient algorithm. In this section we also establish the main results of this chapter Theorem 2.2.6 and Theorem 2.2.7.

In Section 2.3 we present a stepsize selection rule which ensures that the sequence generated by the algorithm reaches a primal-dual solution after a finite number of iterations. By means of a simple example, we show in Section 2.4 that the stepsize used in [23] may not have primal convergence, while our stepsize produces a primal solution in a finite number of steps. The issue of finite convergence is discussed in detail in Remark 2.4.4 at the end of Section 2.4.

The results of this chapter is accepted for publication, see [16].

## 2.1 Introduction

We consider the nonlinear (primal) optimization problem

$$\text{minimize } f(x) \quad \text{s.t.} \quad x \text{ in } K, \;\; h(x) = 0, \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, is lower semicontinuous, $h : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $K \subset \mathbb{R}^n$ is compact. We consider the sharp Lagrangian

$$L(x, y, c) := f(x) - \langle y, h(x)\rangle + c\|h(x)\|. \tag{2.1}$$

Associated with the sharp Lagrangian we consider the dual function $q : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ defined by

$$q(y, c) = \inf_{x \in K} L(x, y, c)$$

and the dual augmented problem given by

$$\text{maximize } \; q(y, c) \; \text{ s.t } \; (y, c) \text{ in } \; \mathbb{R}^m \times \mathbb{R}_+. \tag{D}$$

The primal and dual optimal values are denoted, respectively, by $m_p$ and $m_d$. Example 1.4.5 shows that the sharp Lagrangian (2.1) is an augmented Lagrangian with convex augmenting function, presented in Section 1.4. Therefore, it follows from Theorem 1.4.2 that zero duality gap property holds for the primal and dual problems $(P)$ and $(D)$, that is $m_p = m_d$.

The fact that in our approach the penalty parameter $c$ is a dual variable, together with the use of a sharp Lagrangian, has some interesting consequences on the structure of the dual solution set $D_*$, improving upon the result: $(y, c) \in D_* \Rightarrow (y, t) \in D_*$, for all $t > s$; see Proposition 1.4.1(ii).

**Proposition 2.1.1.** *Take $(y^*, c^*) \in D_*$, $\rho > 0$, and define*

$$\Delta_\rho = \{(y, c) : \|y - y^*\| \leq \rho, c \geq c^* + \rho\}.$$

*Then $\Delta_\rho \subset D_*$ for all $\rho > 0$.*

*Proof.* Take $(y, c) \in \Delta_\rho$. By assumption $q(y^*, c^*) = m_d$. Therefore,

$$\begin{aligned} q(y,c) &= \inf_{x \in K} \{f(x) - \langle h(x), y\rangle + c\|h(x)\|\} \\ &= \inf_{x \in K} \{f(x) - \langle h(x), y^*\rangle + c^*\|h(x)\| + \\ &\qquad (c - c^*)\|h(x)\| - \langle h(x), y - y^*\rangle\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^*\rangle + c^*\|h(x)\| + (c - c^* - \|y - y^*\|)\|h(x)\|\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^*\rangle + c^*\|h(x)\| + (c - c^* - \rho)\|h(x)\|\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^*\rangle + c^*\|h(x)\|\} = q(y^*, c^*) = m_d, \end{aligned} \tag{2.2}$$

where we used Cauchy-Schwarz inequality in the first inequality, the fact that $\|y - y^*\| \leq \rho$ in the second one, and the fact that $c \geq c^* + \rho$ in the third one. We conclude from (2.2) that $q(y, c) = m_d$ and so $(y, c) \in D_*$, proving that $\Delta_\rho \subset D_*$. □

**Corollary 2.1.2.** *If $(y^*, c^*) \in D_*$ then $\{(0, c) : c \geq c^* + \|y^*\|\} \subset D_*$.*

*Proof.* The proof follows from Proposition 2.1.1, taking $\rho = \|y^*\|$. □

We recall that given a concave function $q : \mathbb{R}^p \to \mathbb{R}$, the superdifferential of $q$ at $y_0 \in \mathbb{R}^p$ is the set $\partial q(y_0)$ defined by

$$\partial q(y_0) := \{z \in \mathbb{R}^p : q(y) \leq q(y_0) + \langle z, y - y_0\rangle \;\; \forall y \in \mathbb{R}^p\}.$$

We mention that the set $\partial q(y_0)$ is called *subdifferential* in [12, 14, 23]. Since $q(\cdot)$ is concave, we prefer our notation in order to avoid any confusion between the set above and the subdifferential of a convex function, where the inequality is reversed. However, we still call the algorithm a subgradient algorithm, instead of a supergradient algorithm. Consider the following set

$$A(y,c) = \{x \in K \subset \mathbb{R}^n : L(x,y,c) = q(y,c)\}. \tag{2.3}$$

Note that $A(y,c) = \mathrm{Argmin}_{x\in K} L(x,y,c)$. Since K is compact, $f$ is a lsc function and $h$ is a continuous function, we have that $L(\cdot,y,c)$ is a lsc function for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, and $A(y,c)$ is nonempty for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$. Thus, we have $q(y,c) > -\infty$, for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, and also $m_P > -\infty$. In particular, since by Proposition 1.4.1(ii) the dual function $q$ is concave, we conclude also that $q$ is continuous (note that $q$ can be extended in a natural way to $\mathbb{R}^m \times \mathbb{R}$, preserving its concavity). Next we state our first version of the Modified Subgradient Algorithm (MSg-1).

## 2.2 Algorithm 1

**Step** 0. Choose $(y_0, c_0) \in \mathbb{R}^m \times \mathbb{R}_+$, and exogenous parameters, $\{\alpha_k\} \subset \mathbb{R}_{++}$. Also fix $\beta \geq \eta > 0$. Set $k := 0$.

**Step** 1 (**Subproblem and stopping criterion**)

*a*) Find $x_k \in A(y_k, c_k)$,

*b*) if $h(x_k) = 0$ stop,

*c*) if $h(x_k) \neq 0$, go to **Step** 2.

**Step** 2 (**Stepsize selection and update of dual variable**)

Set $\eta_k := \min\{\eta, \|h(x_k)\|\}$, $\beta_k := \max\{\beta, \|h(x_k)\|\}$;

choose $s_k$ in $[\eta_k, \beta_k]$ and update the variables,

$y_{k+1} := y_k - s_k h(x_k)$,

$c_{k+1} := c_k + (1+\alpha_k)s_k\|h(x_k)\|$.

Set $k = k+1$ and go to **Step** 1.

**Remark 2.2.1.** Note that $[\eta,\beta] \subset [\eta_k,\beta_k]$. In particular, if we consider $\eta=\beta$ then we see that constant stepsizes ($s_k=\eta$, for all $k$) are acceptable. Another simple choice for the stepsize is $s_k := \|h(x_k)\|$.

**Remark 2.2.2.** The parameter $\epsilon_k := \alpha_k s_k$ (which "modifies" the classical subgradient step) was proposed by Gasimov in [23]. It ensures that the dual values are strictly increasing. It is well known that *pure* subgradient methods (i.e., when $\alpha_k = 0$ for all $k$) in general do not have this property. This is a special characteristic of this modified subgradient algorithm. The stepsize selection rule given above has not been considered in [12, 14, 23, 24]. In all these references, some knowledge of the optimal value is required (see, for instance Section 1.6 in Chapter 1).

It follows from Theorem 1.6.3 that if MSg-1 stops at iteration $k$ (that is, $h(x_k)=0$), then $x_k$ is a primal solution and $(y_k,c_k)$ is a dual solution. Therefore, from now on we assume that $h(x_k)\neq 0$ for all $k$, which means that the algorithm produces an infinite primal-dual sequence.

The next result provides an estimate which is essential for proving our main result. We will use in the sequel the following notation: $q_k := q(y_k,c_k)$, $\bar{q} := m_d$.

**Lemma 2.2.3.** *The following estimate is satisfied for all $k\geq 1$,*

$$\max\{q_0 + (\sum_{j=0}^{k-1}\alpha_j s_j\|h(x_j)\|)\|h(x_k)\|, f(x_k) - \langle y_0, h(x_k)\rangle\} \leq q_k. \tag{2.4}$$

*Proof.* It is easy to see that $y_k = y_0 - \sum_{j=0}^{k-1} s_j h(x_j)$ . Therefore we have

$$\begin{aligned}\langle y_k, h(x_k)\rangle &= \langle y_0, h(x_k)\rangle - \sum_{j=0}^{k-1} s_j\langle h(x_j), h(x_k)\rangle \\ &\leq \langle y_0, h(x_k)\rangle + \sum_{j=0}^{k-1} s_j\|h(x_j)\|\|h(x_k)\|,\end{aligned}$$

using Cauchy-Schwarz inequality. Hence

$$\begin{aligned} q_k &= f(x_k) - \langle y_k, h(x_k)\rangle + c_k\|h(x_k)\| \\ &\geq f(x_k) - \langle y_0, h(x_k)\rangle + (c_k - \sum_{j=0}^{k-1} s_j\|h(x_j)\|)\|h(x_k)\|. \end{aligned} \tag{2.5}$$

On the other hand, a simple manipulation in (1.26) gives

$$c_k - \sum_{j=0}^{k-1} s_j\|h(x_j)\| = c_0 + \sum_{j=0}^{k-1} \alpha_j s_j\|h(x_j)\| \geq 0. \tag{2.6}$$

Using (2.6) in (2.5) we obtain

$$\begin{aligned} q_k &\geq f(x_k) - \langle y_0, h(x_k)\rangle + c_0\|h(x_k)\| + \sum_{j=0}^{k-1} \alpha_j s_j\|h(x_j)\|\|h(x_k)\| \\ &\geq \max\{f(x_k) - \langle y_0, h(x_k)\rangle, q_0 + \sum_{j=0}^{k-1} \alpha_j s_j\|h(x_j)\|\|h(x_k)\|\}. \end{aligned}$$

The result follows. □

**Lemma 2.2.4.** *Consider the sequence $\{(y_k, c_k)\}$ generated by* MSg-1*. If $\{c_k\}$ is bounded then $\{y_k\}$ is bounded. If the dual optimal set is nonempty, the converse of the last statement also holds.*

*Proof.* The first statement follows directly from (1.25) and (1.26). For proving the second statement, suppose that $\{y_k\}$ is bounded and take a dual solution $(\bar{y}, \bar{c})$. The supergradient inequality yields

$$q(\bar{y}, \bar{c}) \leq q(y_k, c_k) - \langle h(x_k), \bar{y} - y_k\rangle + (\bar{c} - c_k)\|h(x_k)\|,$$

and therefore

$$c_k \leq \frac{q(y_k, c_k) - q(\bar{y}, \bar{c})}{\|h(x_k)\|} + \|y_k - \bar{y}\| + \bar{c},$$

using Cauchy Schwarz inequality. Since $(\bar{y}, \bar{c})$ is a dual solution we get that

$$c_k \leq \|y_k - \bar{y}\| + \bar{c}.$$

Therefore, since $\{y_k\}$ is bounded, we conclude that $\{c_k\}$ is bounded. The result follows. □

Next, we establish convergence for a stepsize which is more general than the one used in Step 2 of MSg-1. Indeed, we prove convergence for stepsizes $s_k \in [\eta_k, \bar{\beta}_k]$ for all $k$, where $\bar{\beta}_k \geq \beta_k$. More precisely, we make the following assumption.

$(\mathbf{A_1})$ : There exist $\hat{k} > 0$ such that

$$\eta_k \leq s_k \leq \beta_k + \frac{2(\bar{q} - q_k)}{\|h(x_k)\|^2} =: \bar{\beta}_k \quad \text{for all } k > \hat{k}.$$

**Remark 2.2.5.** At least from the theoretical point of view, the step $(A_1)$ is an improvement over the stepsizes used in [12, 23]. Indeed, the step $(A_1)$ ensures primal and dual convergence, while in [12, 23] only dual convergence results hold, and primal convergence is proved only for an auxiliary primal sequence in [12]. In fact in [12, Example 1], the authors presented an optimization problem for which the MSg with their stepsize rule produces a primal sequence convergent to a point which is not a primal solution. We will see later on that the step $(A_1)$ guarantees that all the accumulation points of the primal sequence generated by MSg-1 are primal solutions (see Theorem 2.2.6). The dual optimal value considered in Assumption $(A_1)$ is just for enlarging the interval where the stepsizes can be chosen. It is clear that the interval $[\eta_k, \beta_k]$ considered at iteration $k$ in Step 2 of MSg-1, is contained in the interval $[\eta_k, \bar{\beta}_k]$, and $s_k$ can be chosen in $[\eta_k, \beta_k]$ without the knowledge of the dual optimal value.

From now on we take $\{\alpha_k\} \subset (0, \alpha)$ for some $\alpha > 0$. Next we establish our main convergence results.

**Theorem 2.2.6.** *If the dual optimal set is nonempty then the following statements hold.*

*i) Algorithm* MSg-1 *generates a bounded dual sequence.*

*ii)* $\{h(x_k)\}$ *converges to zero and* $\{q_k\}$ *converges to* $\bar{q}$.

*iii) All accumulation points of* $\{(y_k, c_k)\}$ *are dual solutions.*

*iv) All accumulation points of $\{x_k\}$ are primal solutions.*

*Proof.* For proving $(i)$, note that if $\{c_k\}$ is bounded then $\{(y_k, c_k)\}$ is bounded by Lemma 2.2.4. Thus it suffices to prove that $\{c_k\}$ is bounded. Suppose, for the sake of contradiction, that $\{c_k\}$ is unbounded. By monotonicity of $\{c_k\}$ we have that

$$\lim_{k\to\infty} c_k = \infty. \tag{2.7}$$

Observe that by continuity of $h$ and compactness of $K$, we have that

$$\sup_k \|h(x_k)\| := b < \infty,$$

in particular $\{\beta_k\}$ is bounded. Consider $\hat{\beta}$ such that $\beta_k \le \hat{\beta}$ for all $k$. Take $\hat{k}$ as in Assumption $(A_1)$. In view of (2.7), there exists $k_0 > \hat{k}$ such that $c_k \ge \bar{c} + \dfrac{\hat{\beta} b}{2}$, for all $k \ge k_0$. Take $(\bar{y}, \bar{c}) \in D_*$. Denote $d_j = \|\bar{y} - y_j\|$ for all $j$. For each $k \ge k_0$ we can write

$$\begin{aligned} d_{k+1}^2 &= \|\bar{y} - (y_k - s_k h(x_k))\|^2 \\ &= d_k^2 + s_k^2 \|h(x_k)\|^2 + 2 s_k \langle \bar{y} - y_k, h(x_k)\rangle \\ &\le d_k^2 + s_k^2 \|h(x_k)\|^2 + 2 s_k \left[q_k - \bar{q} + \|h(x_k)\|(\bar{c} - c_k)\right] \end{aligned} \tag{2.8}$$

using the update of the dual variables in the first equality, and the super-gradient inequality in the inequality. Rearranging the right-hand side of the expression above, and using Assumption $(A_1)$, we obtain

$$\begin{aligned} d_{k+1}^2 &\le d_k^2 + s_k[s_k \|h(x_k)\|^2 + 2(q_k - \bar{q})] + 2 s_k \|h(x_k)\|(\bar{c} - c_k) \\ &\le d_k^2 + s_k \beta_k \|h(x_k)\|^2 + 2 s_k \|h(x_k)\|(\bar{c} - c_k) \\ &= d_k^2 + s_k \|h(x_k)\|(\beta_k \|h(x_k)\| + 2\bar{c} - 2c_k) \\ &\le d_k^2 + s_k \|h(x_k)\|(\hat{\beta} b + 2\bar{c} - 2c_k), \end{aligned} \tag{2.9}$$

using the definition of $b$ in the last inequality. Therefore,

$$\|\bar{y} - y_{k+1}\|^2 \le \|\bar{y} - y_k\|^2 + s_k \|h(x_k)\|(\hat{\beta} b + 2\bar{c} - 2c_k). \tag{2.10}$$

The expression between parentheses in (2.10) is negative by definition of $k_0$. Hence, we obtain

$$\|y_{k+1} - \bar{y}\| \leq \|y_k - \bar{y}\| \leq \|y_{k_0} - \bar{y}\|, \text{ for all } k \geq k_0. \tag{2.11}$$

Thus, $\{y_k\}$ is bounded, and by Lemma 2.2.4 we conclude that $\{(y_k, c_k)\}$ is bounded, in contradiction with (2.7), and hence $(i)$ holds. Moreover, we have that $\sum_k s_k \|h(x_k)\| < \infty$, by Theorem 1.6.3$(b)$. In particular $\{s_k \|h(x_k)\|\}$ converges to zero. On the other hand, by the first inequality in assumption $(A_1)$ we have that

$$s_k \|h(x_k)\| \geq \eta_k \|h(x_k)\| > 0 \quad \text{for all } k \geq \hat{k}, \tag{2.12}$$

where $\eta_k = \min\{\eta, \|h(x_k)\|\}$. Hence we obtain from (2.12) that $\{h(x_k)\}$ converges to zero. We are going to prove $(ii)$ and $(iv)$ simultaneously. Since $\{x_k\} \subset K$ and $K$ is compact, $\{x_k\}$ is bounded. Take an accumulation point $\bar{x}$ of $\{x_k\}$. Suppose that $\{x_{k_j}\}$ converges to $\bar{x}$. By lower semi-continuity of $f$, and Lemma 2.2.3, we obtain

$$f(\bar{x}) \leq \liminf_j \left(f(x_{k_j}) - \langle y_0, h(x_{k_j})\rangle\right) \leq \liminf_j q_{k_j} \leq \bar{q} = m_p, \tag{2.13}$$

using also the fact that $\{h(x_{k_j})\}$ converges to zero. On the other hand, by continuity of $h$ we have that $h(\bar{x}) = 0$. Therefore, we conclude from (2.13) that $\bar{x}$ is a primal solution. In particular, all inequalities in (2.13) are equalities and then $\liminf q_{k_j} = \bar{q}$. Since $\{q_k\}$ is increasing by Theorem 1.6.4, we get that $\{q_k\}$ converges to $\bar{q}$, and we have thus proved $(ii)$ and $(iv)$. For proving $(iii)$, take a subsequence $\{(y_{k_j}, c_{k_j})\}_j$ converging for some $(\hat{y}, \hat{c})$. By upper-semicontinuity of $q$ (Proposition 1.4.1(i)), we get

$$q(\hat{y}, \hat{c}) \geq \limsup_j q(y_{k_j}, c_{k_j}) = \lim_j q_{k_j} = \bar{q},$$

using the fact that $\{q_k\}$ converges to $\bar{q}$ by $(ii)$. Hence we have that $(\hat{y}, \hat{c})$ is a dual solution. This proves $(iii)$, and the theorem follows. □

Theorem 2.2.6 presents convergence results for the primal and dual sequences generated by Algorithm MSg-1 assuming the existence of an optimal dual solution. The next theorem ensures convergence of the primal sequence generated by MSg-1 even when the dual solution set is empty. This is very important, because in general, it is not possible to know "a priori" whether the dual solution set is nonempty. Also, in our dual formulation, which includes the penalty parameter $c$ as a dual variable, optimal dual solutions exist only when the problem admits exact penalization and some problems fail to enjoy this property.

**Theorem 2.2.7.** *Assume that* $\hat{\alpha} := \inf_k \alpha_k > 0$. *Then* $\{h(x_k)\}$ *converges to zero and* $\{q_k\}$ *converges to* $\bar{q}$. *Moreover, all accumulation points of the primal sequence* $\{x_k\}$ *are primal solutions.*

*Proof.* By monotonicity of the sequence $\{c_k\}$, either it goes to infinite, or it converges to some $\hat{c}$. In the second case, we have that $\{c_k\}$ is bounded, therefore by Lemma 2.2.4 we get that $\{(y_k, c_k)\}$ is also bounded. Hence repeating the proof of Theorem 2.2.6 $(ii), (iii)$ and $(iv)$ we get that the dual solution set is nonempty (observe that in Theorem 2.2.6 we use the nonemptiness of the dual solution set just for ensuring the boundedness of the dual sequence). Thus, in this case (i.e., when $\{c_k\}$ is bounded) the theorem is proved. So we just need to consider the case in which $\{c_k\}$ goes to infinite. In this case, by Theorem 1.6.3(b), $\sum_j s_j \|h(x_j)\| = \infty$. On the other hand, by Lemma 2.2.3 we obtain that

$$\hat{\alpha} \left( \sum_{j=0}^{k-1} s_j \|h(x_j)\| \right) \|h(x_k)\| \leq \left( \sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\| \right) \|h(x_k)\| \leq q_k - q_0 \leq \bar{q} - q_0. \tag{2.14}$$

Note that $\bar{q} < \infty$ and $\sum_j s_j \|h(x_j)\| = \infty$. Therefore we conclude that $\{h(x_k)\}$ converges to zero. The proof of the remaining statements follows the same steps as in $(ii)$ and $(iv)$ of Theorem 2.2.6. $\square$

The following corollary establishes the equivalence between the boundedness of the dual sequence and the existence of dual solutions. A similar result was obtained in [12, 14].

**Corollary 2.2.8.** *The dual sequence $\{(y_k, c_k)\}$ generated by* MSg-1 *is bounded if and only if the dual optimal solution set is nonempty.*

*Proof.* If the dual optimal set is nonempty, then Theorem 2.2.6 (i) ensures that $\{(y_k, c_k)\}$ is bounded. For proving the converse statement, we just note that in the proof of Theorem 2.2.6, we only use the existence of a dual solution for ensuring boundedness of the dual sequence. Thus, if we assume that the dual sequence is bounded, we can repeat the proof of Theorem 2.2.6($ii$) and ($iii$), and prove that the dual optimal set is nonempty. The result follows. $\square$

In Theorem 2.2.6 we proved that all accumulation points of the dual sequence $\{(y_k, c_k)\}$ generated by MSg-1 are optimal solutions. Since the dual problem is convex and we are applying a subgradient method, we should expect convergence of the whole sequence. The next proposition establishes this result.

**Proposition 2.2.9.** *Take the dual sequence $\{(y_k, c_k)\}$ generated by* MSg-1*, and assume that $0 < \alpha_k < \alpha < \infty$. If $D_*$ is nonempty, then $\{(y_k, c_k)\}$ converges to a dual solution.*

*Proof.* Since $D_*$ is nonempty, it follows from Theorem 2.2.6 that $\{(y_k, c_k)\}$ is bounded. In particular $\{y_k\}$ and $\{c_k\}$ are bounded. Take an accumulation point $(\bar{y}, \bar{c})$ of $\{(y_k, c_k)\}$. It follows that $(\bar{y}, \bar{c})$ belongs to $D_*$ (Theorem 2.2.6($iii$)). Since $\{c_k\}$ is increasing and bounded, it converges to $\bar{c}$. Therefore we just need to prove that $\{y_k\}$ converges to $\bar{y}$. Consider a subsequence $\{y_{k_j}\}_j$ converging to $\bar{y}$. Using the same calculations as in (2.10), we obtain

$$\|\bar{y} - y_{k+1}\|^2 \leq \|\bar{y} - y_k\|^2 + \tilde{\beta} s_k \|h(x_k)\|, \tag{2.15}$$

where $\tilde{\beta} := \hat{\beta}b + 2\bar{c}$, with $b$ and $\hat{\beta}$ as in the proof of Theorem 2.2.6. On the other hand, since $\{(y_k, c_k)\}$ is bounded, we have that $\sum s_k\|h(x_k)\| < \infty$, by Theorem 1.6.3. Therefore, given an arbitrary $\epsilon > 0$, there exists $k_0$ sufficiently large such that

$$\sum_{k>k_0} s_k\|h(x_k)\| < \frac{\epsilon}{2\tilde{\beta}}.$$

Since $\{y_{k_j}\}_j$ converges to $\bar{y}$, there exists $j_0$ such that for each $j \geq j_0$ it follows that $k_j > k_0$ and $\|y_{k_j} - \bar{y}\|^2 < \frac{\epsilon}{2}$. Using (2.15) we obtain, for all $k > k_{j_0}$,

$$\|\bar{y} - y_k\|^2 \leq \|\bar{y} - y_{k_{j_0}}\|^2 + \tilde{\beta} \sum_{l=k_{j_0}}^{k-1} s_l\|h(x_l)\| < \epsilon. \tag{2.16}$$

Since $\epsilon$ is arbitrary, we conclude that $\{y_k\}$ converges to $\bar{y}$. Therefore $\{(y_k, c_k)\}$ converges to $(\bar{y}, \bar{c})$, and the proposition follows. □

## 2.3 Algorithm 2

In this section we present and analyze algorithm MSg-2. This algorithm has a stepsize which ensures finite termination, as long as there exist dual solutions.

**Step** 0. Choose $(y_0, c_0) \in \mathbb{R}^m \times \mathbb{R}_{++}$, and exogenous parameters $\{\alpha_k\} \subset \mathbb{R}_{++}$. Also fix $\bar{\beta} \geq \bar{\eta} > 0$. Set $k := 0$.

**Step** 1 (**Subproblem and Stopping Criterion)**

$a)$ Find $x_k \in A(y_k, c_k)$,

$b)$ if $h(x_k) = 0$ stop,

$c)$ if $h(x_k) \neq 0$, go to **Step** 2.

**Step** 2 (**Step-size Choice and Update of Dual Variables)**

$\eta_k := \frac{\bar{\eta}}{\|h(x_k)\|}$, $\beta_k := \frac{\bar{\beta}}{\|h(x_k)\|}$, and choose $s_k \in [\eta_k, \beta_k]$,

$y_{k+1} := y_k - s_k h(x_k)$,

$c_{k+1} := c_k + (1 + \alpha_k) s_k \|h(x_k)\|$.

Set $k = k + 1$ and go to **Step** 1.

Observe that the only difference between MSg-1 and MSg-2 is the choice of $\beta_k$ and $\eta_k$.

**Theorem 2.3.1.** *Suppose that the dual solution set is nonempty. Consider the* MSg-2 *algorithm. Choose* $\{\alpha_k\} \subset (0, \alpha)$ *for some* $\alpha > 0$. *Then there exists* $\bar{k} > 0$ *such that* $h(x_{\bar{k}}) = 0$, *i.e.* MSg-2 *stops at the* $\bar{k}$*-th iteration. In particular* $x_{\bar{k}}$ *and* $(y_{\bar{k}}, c_{\bar{k}})$ *are primal and dual optimal solutions, respectively.*

*Proof.* We prove first that the dual sequence is bounded. If $c_k \leq \frac{\bar{\beta}}{2} + \bar{c}$ for all $k$, then $\{y_k\}$ is also bounded by Lemma 2.2.4. Assuming, for the sake of contradiction, that $\{c_k\}$ is unbounded, we can repeat the same calculations as in (2.8), (observing that $q_k \leq \bar{q}$), to get

$$\begin{aligned} \|\bar{y} - y_{k+1}\|^2 &\leq \|\bar{y} - y_k\|^2 + s_k\|h(x_k)\|(\beta_k\|h(x_k)\| + 2\bar{c} - 2c_k) \\ &= \|\bar{y} - y_k\|^2 + s_k\|h(x_k)\|(\bar{\beta} + 2\bar{c} - 2c_k). \end{aligned} \tag{2.17}$$

Since $\{c_k\}$ is increasing, there exists $k_0$ such that $c_k > \frac{\bar{\beta}}{2} + \bar{c}$ for all $k > k_0$. Using this estimate in (2.17), we obtain, for all $k > k_0$,

$$\|\bar{y} - y_{k+1}\|^2 \leq \|\bar{y} - y_k\|^2. \tag{2.18}$$

From (2.18) we obtain that $\{y_k\}$ is bounded. Thus, $\{(y_k, c_k)\}$ is bounded by Lemma 2.2.4, contradicting the supposed unboundedness of $\{c_k\}$. Hence, the dual sequence is bounded. Let us prove now that the algorithm has finite termination. If this is not true, we must have $h(x_k) \neq 0$ for all $k$ (note that the algorithm stops at $k$ if and only if $h(x_k) = 0$). Using the definition of the stepsizes, we can write for $k \geq 1$:

$$c_k - c_0 = \sum_{j=0}^{k-1}(c_{j+1} - c_j) = \sum_{j=0}^{k-1}(\epsilon_j + s_j)\|h(x_j)\| \geq \sum_{j=0}^{k-1} s_j\|h(x_j)\| \geq \sum_{j=0}^{k-1} \bar{\eta} = k\bar{\eta},$$

which entails a contradiction with the boundedness of $\{c_k\}$. Thus there exists $\bar{k}$ such that $h(x_{\bar{k}}) = 0$. In view of Remark 1.6.5, the result follows by Theorem 1.6.3(c). □

## 2.4 Final remarks

As established above, our stepsize rules allow us to prove primal convergence, which in principle does not hold for the stepsize rule proposed in [23]. We show next that the improvement over the convergence analysis in [23] is not purely theoretical, but has indeed computational consequences; we do this by exhibiting a simple example where MSg with our stepsize rule finds the primal solution after a finite number of steps, while MSg with the stepsize rule given in [23] stays away from it.

**Example 2.4.1.** Consider the following primal problem:

$$\min f(x) := -x \quad \text{subject to } h(x) := x = 0 \text{ and } x \in [0,1]. \tag{2.19}$$

The sharp Lagrangian related to problem (2.19) is:

$$L(x,y,c) = -x - yx + c|x|,$$

and the dual function is

$$q(y,c) = \min_{x\in[0,1]} L(x,y,c) = \min_{x\in[0,1]} (c-y-1)x.$$

Therefore the dual problem is stated as

$$\max_{(y,c)\in\mathbb{R}\times\mathbb{R}_+} q(y,c).$$

Let $A(y,c) := \operatorname{Argmin}_{x\in[0,1]} L(x,y,c)$. It follows that:

$$\begin{aligned} A(y,c) &= [0,1] && \text{if} \quad c-y=1,\\ A(y,c) &= \{0\} && \text{if} \quad c-y>1,\\ A(y,c) &= \{1\} && \text{if} \quad c-y<1. \end{aligned} \tag{2.20}$$

In the current example we consider an initial point $(y_0, c_0)$ satisfying

$$0 \le c_0 - y_0 < 1.$$

We claim that MSg with our stepsize has finite convergence, that is, there exists $\bar{k}$ such that $x_{\bar{k}} = 0$. In this case we have $c_{\bar{k}} - y_{\bar{k}} \ge 1$, $q(y_{\bar{k}}, c_{\bar{k}}) = 0 = m_p$ (where $m_p$ is the optimal value) and $h(x_{\bar{k}}) = h(0) = 0$. Thus, MSg stops at iteration $\bar{k}$, producing a primal-dual solution. On the other hand, MSg with stepsize as in [23] satisfies $x_k = 1$, $q(y_k, c_k) = c_k - y_k - 1 < 0$ and $h(x_k) = h(1) = 1$, for each $k \ge 0$. Indeed, in the latter situation, the dual update of MSg is

$$\begin{aligned} c_{k+1} &= c_k + (1+\alpha_k)s_k|h(x_k)| = c_k + (1+\alpha_k)s_k\\ y_{k+1} &= y_k - s_k h(x_k) = y_k - s_k. \end{aligned}$$

Hence, taking $\alpha_k = \frac{1}{2}$ yields

$$c_{k+1} - y_{k+1} = c_k - y_k + (2 + \alpha_k)s_k = c_k - y_k + \frac{5}{2}s_k. \qquad (2.21)$$

We recall that the stepsize rule suggested in [23] is

$$s_k = \frac{\bar{q} - q_k}{5\|h(x_k)\|^2},$$

which in the current situation becomes

$$s_k = \frac{0 - (c_k - y_k - 1)}{5} = \frac{1 - (c_k - y_k)}{5}.$$

Applying MSg with the stepsize rule proposed in [23], and supposing that at the $k$-th iteration $c_k - y_k < 1$, we have

$$c_{k+1} - y_{k+1} = \frac{c_k - y_k}{2} + \frac{1}{2} < 1.$$

Thus the primal-dual sequences $\{x_k\}$ and $\{(y_k, c_k)\}$ satisfy $c_k - y_k < 1$ and $x_k = 1$ for all $k$. The dual sequence converges to a dual solution, in agreement with the convergence results in [12, Theorem 10], but the primal sequence does not approach the primal solution $x^* = 0$.

On the other hand, our method overcomes this obstacle, obtaining a primal solution after a finite number of iterations, as we show next. Take $\beta > \eta > 0$. If $c_k - y_k = 1$ for some $k > 0$ and $x_k \neq 0$, then

$$c_{k+1} - y_{k+1} \geq c_k - y_k + \frac{5\min\{\eta, \|h(x_k)\|\}}{2} = 1 + \frac{5\min\{\eta, \|h(x_k)\|\}}{2} > 1.$$

Therefore, it follows from (2.20) that $x_{k+1} = 0$, which is the primal solution. Thus, suppose that $c_k - y_k < 1$ for each $k \leq k_1 := \left\lfloor \frac{2}{5\min\{\eta,1\}} \right\rfloor$. Then the stepsize $s_k$ satisfies

$$s_k \in [\min\{\eta, 1\}, \max\{\beta, 1\}]$$

for each $k \leq k_1$, because $x_k = 1$. In this situation we have, using (2.21),

$$c_{k+1} - y_{k+1} \geq c_k - y_k + \frac{5}{2}\min\{\eta, 1\}.$$

It follows recursively that

$$c_{k_1+1} - y_{k_1+1} \geq c_0 - y_0 + \frac{5(k_1+1)}{2} \min\{\eta, 1\} > 1,$$

using $c_0 - y_0 \geq 0$ and definition of $k_1$. Then $c_{k_1+1} - y_{k_1+1} > 1$, so that, in view of (2.20), MSg stops at iteration $k_1 + 1$, achieving the primal solution $x^* = 0$.

**Remark 2.4.2.** Observe that if we consider in the previous example $s_k = |h(x_k)| = 1$ (one of the simple possible choices for $s_k$ discussed in Remark 2.2.1), then $x_{k+1} = 0$, which is the primal solution.

**Remark 2.4.3.** It is also worthwhile to mention that a standard penalty method, with $y_k = 0$ and penalty parameters $c_k$ such that $\lim_{k\to\infty} c_k = \infty$, will exhibit in this example the same behavior as MSg with our stepsize rule, i.e., finite primal convergence.

**Remark 2.4.4.** A finitely convergent algorithm for nonsmooth and nonconvex problems might seem too good to be true, but the point here is that the assumption of existence of optimal dual solutions is stronger than it looks at first sight. Observe that we have included the penalty parameter $c$ among the dual variables, and hence the existence of optimal dual solutions implies in particular the existence of an optimal penalty parameter $c^*$. It is easy to verify that any $c$ larger than such a $c^*$ turns out to be an exact penalty parameter, in the sense of [54, Definition 11.60]. Thus, in our formulation, if optimal dual solutions exist then the problem admits exact penalization. In such a setting, for achieving finite convergence it is enough to have a stepsize selection rule which allows $c_k$ to attain arbitrarily large values. In fact, after establishing that the sequence $\{y_k\}$ is bounded, as is the case for both Algorithms 1 and 2, Proposition 2.1.1 provides an alternative argument for the finite convergence of Algorithm 2, assuming existence of a dual solution $(y^*, c^*)$: if $\rho$ is such that $\|y_k - y^*\| \leq \rho$ for all $k$, then any pair $(y, c)$ with $\|y - y^*\| \leq \rho$, $c \geq c^* + \rho$ belongs to $D_*$ by Proposition 2.1.1, and hence we get $(y_k, c_k) \in D_*$ as soon as $c_k > c^* + \rho$. Once such a value of $k$ is reached, $x_k$ will be an optimal primal solution, because, as commented above, the fact that $x_k$ belongs to $A(y_k, c_k)$, as prescribed in Step 1(a) of Algorithm 2, is equivalent to saying that $x_k$ is an exact minimizer of $L(\cdot, y_k, c_k)$ on $K$.

Similarly to the argument above, one can establish that if $D_* \neq \emptyset$ then the penalty method with $\|\cdot\|$ as penalty function also obtains a primal solution after a finite number of iterations. In this situation $y_k = 0$ for all $k$, and the penalty parameter $c_k$ can be arbitrarily large (see also Corollary 2.1.2).

It should be emphasized, however, that attempting to circumvent the dual updating by guessing the "right" values of $c^*$ and $\rho$ (assuming that it is known in advance that the problem admits exact penalization), does not seem to be in general a good strategy: quite likely one will overshoot the value of the parameters, and then suffer the consequences, in terms of numerical instability, of a too large penalty parameter (of course, this comment applies to any penalty method in the presence of exact penalization; not just to ours). A sensible gradual increase of the penalty parameter, like the updating of $c_k$ in Algorithm 2, is likely to give rise to a better numerical behavior. See also the discussion in [13] on the comparison of the actual numerical behavior of a dual updating similar to ours with a classical penalty method.

A careful study of the numerical behavior of our method, and its comparison with other variants of subgradient techniques, are subjects of our future research.

# Chapter 3

# Inexact MSg with augmented Lagrangian

In this chapter we consider a primal problem of minimizing an extended real-valued function (possibly nonconvex and nondifferentiable) in a reflexive Banach space. A duality scheme is considered via augmented Lagrangian functions, which include the sharp Lagrangian as a particular case (see Example 3.1.1). Our dual variables belong to a Hilbert space. We propose a *parameterized* inexact modified subgradient algorithm for solving the dual problem. We show that this algorithm guarantees monotone increase of the dual values. We show that, in our more general setting, our algorithm generates a dual sequence strongly convergent to a dual solution when the dual optimal solution set is nonempty. The primal sequence converges in the sense that all its weak accumulation points are primal solutions, even when the dual solution set is empty. We also analyze a stepsize rule which ensures that when the dual solution set is nonempty, approximate primal and dual solutions are obtained after a finite number of iterations of the algorithm (see Section 3.2.2). Many of the results presented in Chapter 2 are extended in this chapter to a more general primal-dual scheme in infinite dimensional spaces.

This chapter is organized as follows. In Section 3.1 we describe the setting of our primal and dual problems, and give some basic definitions, assumptions and examples. We also recall in this section some useful facts. In Section 3.2 we consider the inexact modified subgradient algorithm (IMSg) and establish its convergence properties which do not depend on the choice of the stepsize. In Section 3.2.1 we propose a stepsize rule for IMSg and state and prove our main results. In Section 3.2.2 another stepsize rule for IMSg is proposed and we show that, under this stepsize rule, IMSg converges in a finite number of steps. In the last section we compare our algorithm with the algorithms with sharp Lagrangian considered in [12, 16, 14, 23] and in Chapter 2.

## 3.1 Introduction

Let $X$ be a reflexive Banach space and $H$ a Hilbert space. We denote by $\langle \cdot, \cdot \rangle$ the scalar product in $H$, and by $\| \cdot \|$ the norm, where the same notation will be used for the norm both in $X$ and $H$. We consider the optimization problem

$$\min\ \varphi(x)\ \text{ s.t. } x \text{ in } X, \tag{3.1}$$

where the function $\varphi : X \to \mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ is a proper (i.e., $\operatorname{dom} \varphi \neq \emptyset$ and $\varphi > -\infty$) weakly lower semicontinuous (w-lsc) function. We also assume that $\varphi$ has weakly compact sublevel sets. In order to introduce our duality scheme, we consider a *dualizing parameterization* for (3.1), which is a function $f : X \times H \to \bar{R} := \mathbb{R}_{+\infty} \cup \{-\infty\}$ that verifies $f(x, 0) = \varphi(x)$ for all $x \in X$. The perturbation function induced by this dualizing parameterization is the function $\beta : H \to \bar{\mathbb{R}}$ defined by

$$\beta(z) := \inf_{x \in X} f(x, z).$$

Because $\varphi$ is proper, we have $\beta(0) < +\infty$. Next we recall the definition of a level-bounded augmenting function.

**Definition 3.1.1.** *A function $\sigma : H \to \mathbb{R}_{+\infty}$ is said to be a level-bounded augmenting function if it is proper, w-lsc, level-bounded on $H$ and*

$$\sigma(0) = 0 \ \ \textit{and} \ \ \operatorname*{Argmin}_{y} \sigma(y) = \{0\}.$$

The *augmented Lagrangian function* $\ell : X \times H \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ is defined as

$$\ell(x, y, r) := \inf_{z \in H} \{f(x, z) - \langle z, y\rangle + r\sigma(z)\}. \tag{3.2}$$

The dual function $q : H \times \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty\}$ is defined as

$$q(y, r) := \inf_{x \in X} \ell(x, y, r)$$

and therefore the dual problem is stated as

$$\max \ \ q(y, r) \ \ \text{s.t.} \ (y, r) \ \in H \times \mathbb{R}_+. \tag{3.3}$$

Denote by $M_p := \inf_{x \in X} \varphi(x)$ and by $M_d := \sup_{(y,r) \in H \times \mathbb{R}_+} q(y, r)$ the optimal values of the primal and dual problem, respectively. The primal and dual solution sets are denoted by $P_*$ and $D_*$, respectively. Next we show how this framework contains the one studied in Chapter 2. Here we allow the space that contains the image of the constraints to be an infinite dimensional space. The next example is similar to Example 1.4.5.

**Example 3.1.1.** Consider the following equality constrained problem

$$\min \ \psi(x) \ \ \text{s.t.} \ x \in K, h(x) = 0, \tag{3.4}$$

where $h : X \to H$ has a weakly closed graph, i.e., $G(h) := \{(x, h(x)) : x \in K\}$ is weakly closed in $X \times H$, $\psi : X \to \mathbb{R}$ is w-lsc, and $K \subset X$ is weakly closed. We consider the following equivalent unconstrained problem:

$$\min \ \phi(x) := \psi(x) + \delta_V(x), \quad \text{s.t.} \ x \in X,$$

where $V := \{x \in K : h(x) = 0\}$ and $\delta_V(x) = 0$ if $x \in V$, $\delta_V(x) = \infty$ otherwise. Consider the augmenting function given by $\sigma(\cdot) = \|\cdot\|$, and the canonical dualizing parameterization function given by

$$f(x,z) = \begin{cases} \psi(x) & \text{if } x \in K \text{ and } h(x) = z, \\ \infty, & \text{otherwise.} \end{cases}$$

By definition,

$$\ell(x,y,r) = \inf_{z \in H} \{f(x,z) - \langle z, y \rangle + r\sigma(z)\},$$

and then we obtain

$$\ell(x,y,r) = \begin{cases} \psi(x) - \langle y, h(x) \rangle + r\|h(x)\| & \text{if } x \in K, \\ \infty & \text{otherwise,} \end{cases}$$

which generalizes the sharp Lagrangian proposed in [54, Example 11.58] in finite dimensional space. The dual function induced by this Lagrangian is

$$q(y,r) := \inf_{x \in K} \{\psi(x) - \langle y, h(x) \rangle + r\|h(x)\|\},$$

and the dual problem is

$$\max \; q(y,r) \; \text{ s.t. } \; (y,r) \in H \times \mathbb{R}_+.$$

We recall that a modified subgradient algorithm was studied in [12, 16, 14, 23] for the primal-dual scheme described in Example 3.1.1, under the assumptions that $X = \mathbb{R}^n$, $H = \mathbb{R}^m$ and $K$ is a compact set; see Section 1.6.

From now on, in this section, we make the following assumption on the augmenting function.

$(\mathbf{A_0})$ : $\sigma(z) \geq \|z\|$ for all $z \in H$.

**Remark 3.1.2.** An assumption similar to $(A_0)$ was considered in [24], where the primal problem is a constrained optimization problem with a single constraint. [24] considers a dynamic stepsize rule, and results similar to those of [23] are obtained. At the end of this chapter we make some comments on assumption $(A_0)$, in particular we show in Proposition 3.2.18 that $(A_0)$ is a necessary assumption, under mild conditions, for obtaining the increase property of the modified subgradient algorithm.

Next we list some examples of augmenting functions that satisfy $(A_0)$.

i) Let $\sigma_{p,q} : H \to \mathbb{R}$ be defined as

$$\sigma_{p,q}(z) := \begin{cases} \|z\|^p & \text{if } \|z\| \leq 1, \\ \|z\|^q & \text{otherwise,} \end{cases}$$

with $0 < p \leq 1 \leq q$.

ii) Let $H = \mathbb{R}^n$, and $\sigma_k : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\sigma_k(z) := \left( \sum_{i=1}^{n} |z_i|^{\frac{1}{k}} \right)^k, \quad \text{with} \quad k \in \mathbb{N}.$$

The next definition has been considered in [17, Section 5] and it is a natural generalization of Definition 1.4.1.

**Definition 3.1.2.** *A function $f : X \times H \to \bar{\mathbb{R}}$ is said to be* weakly level-compact *if for every $\bar{z} \in H$ and $\alpha \in \mathbb{R}$ there exist a weak neighborhood $V \subset H$ of $\bar{z}$, and a weak compact set $B \subset X$, such that*

$$L_{V,f}(\alpha) := \{x \in X : f(x, z) \leq \alpha\} \subset B \ \textit{ for all } z \in V.$$

**Remark 3.1.3.** If $f$ verifies Definition 3.1.2 then every sequence in $L_{V,f}(\alpha)$ has a weakly convergent subsequence. It is immediate that the canonical dualizing parameterization function considered in Example 3.1.1 is weakly level-compact if $K$ is a weakly compact set. We will only consider dualizing parameterization functions which are proper, w-lsc and weakly level-compact.

Similarly to Proposition 1.4.1, we have the following basic properties of the dual function.

**Proposition 3.1.4.** *i) The dual function $q$ is concave and weakly upper-semicontinuous* (w-usc).
*ii) If $r \geq c$ then $q(y, r) \geq q(y, c)$ for all $y \in H$. In particular, if $(y, c)$ is a dual solution, then also $(y, r)$ is a dual solution for all $r \geq c$.*

*Proof.* Item $(i)$ follows from the fact that $q$ is the infimum of affine functions. Item $(ii)$ follows from the fact that the penalty function $\sigma$ is nonnegative. □

Augmented Lagrangians are special cases of abstract Lagrangians ([17] and [56, Section 5.2]; see Section 1.5.1). The next theorem is a consequence of [17, Proposition 4.1]. It ensures that there is no duality gap between the primal problem (3.1) and its dual problem (3.3).

**Theorem 3.1.5.** *Consider the primal problem* (3.1) *and its dual problem* (3.3). *Assume that the dualizing parameterization function $f : X \times H \to \bar{\mathbb{R}}$ for the primal function $\varphi$ is proper, w-lsc and weakly level-compact. Suppose that there exists some $(y, r) \in H \times \mathbb{R}_+$ such that $q(y, r) > -\infty$. Then zero duality gap holds, i.e. $M_p = M_d$.*

*Proof.* This result has been proved in a more general setting in [17, Proposition 4.1] and [71, Theorem 3.1]. We also emphasize that the result of this theorem is a consequence of Theorem 5.2.3 in Chapter 5. □

**Remark 3.1.6.** The augmented Lagrangians studied in [71] use a weakly continuous function $g : Y \times V \to \mathbb{R}$, where $Y, V$ are reflexive Banach spaces, but the main prototypical of $g$ is $g(y, v) = \langle y, v\rangle$, where $Y = V$ is a Hilbert space. In this case $g$ is not weakly continuous, but we mention that only weakly upper semicontinuity of $g(y, \cdot)$ is required in the proof of [71, Theorem 3.1], and this assumption holds for $g(y, v) = \langle y, v\rangle$.

From now on we assume that the hypotheses of Theorem 3.1.5 are verified. We give next some definitions.

**Definition 3.1.3.** *We say that $x_* \in X$ is an $\epsilon_*$-optimal primal solution if $\varphi(x_*) \le M_p + \epsilon_*$; we say that $(y_*, c_*) \in H \times \mathbb{R}_+$ is an $\epsilon_*$-optimal dual solution if $q(y_*, c_*) \ge M_d - \epsilon_*$.*

For $r \ge 0$ consider the following set

$$A_r(y, c) = \{(x, z) \in X \times H : f(x, z) - \langle z, y\rangle + c\sigma(z) \le q(y, c) + r\}. \quad (3.5)$$

By definition of $q$ and (3.2), we see that $A_r(y,c)$ is nonempty for all $r > 0$ and all $(y,c)$ such that $q(y,c) > -\infty$. Fix $(y,c) \in H \times \mathbb{R}_+$ and define $\Phi_{(y,c)} : X \times H \to \bar{\mathbb{R}}$ as

$$\Phi_{(y,c)}(x,z) = f(x,z) - \langle z, y\rangle + c\sigma(z). \tag{3.6}$$

Observe that computation of an element in $A_r(y,c)$ is tantamount to an approximate unconstrained minimization of $\Phi_{(y,c)}(\cdot,\cdot)$, with tolerance $r$.

## 3.2 Inexact modified subgradient algorithm-IMSg

**Step** 0. Choose $(y_0, c_0) \in H \times \mathbb{R}_+$ such that $q(y_0, c_0) > -\infty$, and exogenous parameters $\epsilon_* > 0$ (a prescribed tolerance), $\delta < 1$, $\{\alpha_k\} \subset (0,\alpha)$ for some $\alpha > 0$, and $\{r_k\} \subset \mathbb{R}_+$ such that $r_k \to 0$. Let $k := 0$.

**Step** 1. (Subproblem and stopping criterion)

*a*) Find $(x_k, z_k) \in A_{r_k}(y_k, c_k)$,

*b*) if $z_k = 0$ and $r_k \leq \epsilon_*$ stop,

*c*) if $z_k = 0$ and $r_k > \epsilon_*$, then $r_k := \delta r_k$ and go to $(a)$,

*d*) if $z_k \neq 0$ go to Step 2.

**Step** 2. (Selection of the stepsize and updating the variables)

Choose a stepsize $s_k > 0$ and update the dual variables,

$y_{k+1} := y_k - s_k z_k,$

$c_{k+1} := c_k + (\alpha_k + 1) s_k \sigma(z_k),$

$k := k + 1$, go to Step 1.

Note that IMSg has the general form of standard augmented Lagrangian methods: in Step 1 the primal variables are updated through the approximate solution of an unconstrained minimization problem (in this case producing the pair $(x, z)$), and then in Step 2 the dual variables $(y_k, c_k)$ is updated through an explicit formula, in this case moving along a direction of dual ascent. Observe also that here the penalty parameters $\{c_k\}$ are considered as variables. The parameters $\{\alpha_k\}$ ensure monotonic increase of the dual values; see Theorem 3.2.3.

First, we present some results which do not depend on the selection of the stepsize. The next proposition establishes the relation between the approximate minimization implicit in $A_r(y, c)$ and the approximate superdifferential $\partial_r q(y, c)$.

**Proposition 3.2.1.** *The following facts hold for* IMSg.

*i) For all $r \geq 0$ it holds that if $(\hat{x}, \hat{z}) \in A_r(\hat{y}, \hat{c})$, then $(-\hat{z}, \sigma(\hat{z})) \in \partial_r q(\hat{y}, \hat{c})$.*

*ii) Assume that $(A_0)$ holds. Then* IMSg *generates a bounded dual sequence $\{(y_k, c_k)\}$ if and only if $\sum_k s_k \sigma(z_k) < +\infty$.*

*iii) If* IMSg *stops at iteration k, then $x_k$ is an $\epsilon_*$-optimal primal solution, and $(y_k, c_k)$ is an $\epsilon_*$-optimal dual solution.*

*Proof.* i) For all $(y, c) \in H \times \mathbb{R}_+$ we have:

$$
\begin{aligned}
q(y, c) &= \inf_{(x,z)}\{f(x, z) - \langle z, y\rangle + c\sigma(z)\} \\
&\leq f(\hat{x}, \hat{z}) - \langle \hat{z}, y\rangle + c\sigma(\hat{z}) \qquad (3.7) \\
&= f(\hat{x}, \hat{z}) - \langle \hat{z}, \hat{y}\rangle + \hat{c}\sigma(\hat{z}) + \langle -\hat{z}, y - \hat{y}\rangle + (c - \hat{c})\sigma(\hat{z}).
\end{aligned}
$$

Using that $(\hat{x}, \hat{z}) \in A_r(\hat{y}, \hat{c})$ in (3.7), we obtain

$$
\begin{aligned}
q(y, c) &\leq q(\hat{y}, \hat{c}) + r + \langle -\hat{z}, y - \hat{y}\rangle + (c - \hat{c})\sigma(\hat{z}) \\
&= q(\hat{y}, \hat{c}) + \langle (-\hat{z}, \sigma(\hat{z})), (y, c) - (\hat{y}, \hat{c})\rangle + r.
\end{aligned}
$$

Therefore, $(-\hat{z}, \sigma(\hat{z})) \in \partial_r q(\hat{y}, \hat{c})$.

ii)Using $(A_0)$ and simple manipulations in the definition of $\{y_k\}$, we obtain

$$\|y_{k+1} - y_0\| \leq \sum_{j=0}^{k} \|y_{j+1} - y_j\| = \sum_{j=0}^{k} s_j \|z_j\| \leq \sum_{j=0}^{k} s_j \sigma(z_j). \tag{3.8}$$

On the other hand,

$$c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1) s_j \sigma(z_j). \tag{3.9}$$

Since $\{\alpha_k\}$ is bounded, $(ii)$ follows from (3.8) and (3.9).

For proving $(iii)$, observe that if IMSg stops at iteration $k$, then $z_k = 0$ and $r_k \leq \epsilon_*$. Therefore we have (see Theorem 3.1.5):

$$\begin{aligned} M_d &= M_p \leq \varphi(x_k) = f(x_k, 0) - \langle y_k, 0 \rangle + c_k \sigma(0) \\ &\leq q(y_k, c_k) + r_k \leq q(y_k, c_k) + \epsilon_* \leq M_p + \epsilon_*, \end{aligned}$$

which implies that $M_d \leq q(y_k, c_k) + \epsilon_*$, and $\varphi(x_k) \leq M_p + \epsilon_*$. That is to say, $x_k$ is an $\epsilon_*$-optimal primal solution, and $(y_k, c_k)$ is an $\epsilon_*$-optimal dual solution. □

Next we establish boundedness properties of the sub-level sets of $\Phi_{(y,c)}(\cdot, \cdot)$.

**Lemma 3.2.2.** *Let $(y, c_y) \in H \times \mathbb{R}_+$ be such that $q(y, c_y) > -\infty$. Then for each $w \in H$, $r \geq M_P$ and $c > c_y + \|w - y\|$ the set*

$$L_r(y, c) = \{(x, z) : \Phi_{(y,c)}(x, z) := f(x, z) - \langle z, y \rangle + c\sigma(z) \leq r\}$$

*is nonempty and weakly-compact. In particular, for each $w \in H$ and $c > c_y + \|w - y\|$ there exists some $(\tilde{x}, \tilde{z})$ such that*

$$q(w, c) = f(\tilde{x}, \tilde{z}) - \langle \tilde{z}, y \rangle + c\sigma(\tilde{z}).$$

*Proof.* Since the function $\varphi(\cdot) = f(\cdot, 0)$ is w-lsc and weakly-level compact, there exists a global minimizer $x^*$ of $\varphi$, so that $(x^*, 0) \in L_r(w, c)$ for all $w, c$

and $r \geq M_P$, ensuring that $L_r(w,c)$ is nonempty. For proving that $L_r(w,c)$ is bounded, suppose by contradiction that $L_r(w,c)$ is unbounded for some $w, c, r$ with $c > c_y + \|w - y\|$ and $r \geq M_P$, so that there exists some unbounded sequence $\{(x_k, z_k)\} \subset L_r(w,c)$. Therefore

$$\begin{aligned} r &\geq f(x_k, z_k) - \langle z_k, w\rangle + c\sigma(z_k) \\ &= f(x_k, z_k) - \langle z_k, y\rangle + c_y\sigma(z_k) + \langle z_k, y - w\rangle + (c - c_y)\sigma(z_k) \\ &\geq q(y, c_y) + (c - c_y - \|w - y\|)\|z_k\|, \end{aligned} \tag{3.10}$$

using Cauchy-Schwarz inequality and assumption $(A_0)$ in the inequality. It follows from (3.10) that

$$\|z_k\| \leq \frac{r - q(y, c_y)}{c - c_y - \|w - y\|},$$

and hence $\{z_k\}$ is bounded. Without loss of generality we can assume that the whole sequence $\{z_k\}$ converges weakly to some $\bar{z}$. Since $\{z_k\}$ is bounded and $\sigma(z) \geq 0$ for all $z$, we obtain from (3.10) that

$$f(x_k, z_k) \leq r + \|w\|\|z_k\| \leq \tilde{\alpha} \tag{3.11}$$

for some $\tilde{\alpha} \in \mathbb{R}$. Take a weak compact set $B \subset X$ and a weak neighborhood $V$ of $\bar{z}$ given by the level compactness property of $f$ related to $\bar{z}$ and $\tilde{\alpha}$ (see Definition 3.1.2). We know that there exists $k_0$ such that $z_k \in V$ for all $k > k_0$. Thus $\{x_k\}_{k>k_0} \subset L_{V,f}(\tilde{\alpha}) \subset B$, by (3.11). Therefore $\{x_k\}$ is bounded, and hence $\{(x_k, z_k)\}$ is bounded, which is a contradiction, establishing boundedness of $L_r(w,c)$. Since the function $\Phi_{(w,c)}(\cdot,\cdot)$ given by (3.6) is w-lsc, $L_r(w,c)$ is also weakly-closed, and so $L_r(w,c)$ is weakly-compact, by Banach-Alaoglu theorem. The last assertion of the Lemma is equivalent to

$$(\tilde{x}, \tilde{z}) \in \operatorname{Arg}\min_{(x,z)\in X\times H} \Phi_{(w,c)}(x, z).$$

Indeed, $(\tilde{x}, \tilde{z})$ verifies the inclusion above if and only if

$$q(w, c) = \inf_{(x,z)\in X\times H} \Phi_{(w,c)}(x, z) = \Phi_{(w,c)}(\tilde{x}, \tilde{z}) = f(\tilde{x}, \tilde{z}) - \langle \tilde{z}, w\rangle + c\sigma(\tilde{z}), \tag{3.12}$$

where we use the definitions of $\Phi_{(w,c)}(\cdot,\cdot)$ and $q$. Note also that $X \times H$ is reflexive and $L_r(w,c)$ is a nonempty weakly-compact sub-level set of $\Phi_{(w,c)}(\cdot,\cdot)$, as we have already established. Since $\Phi_{(w,c)}(\cdot,\cdot)$ is w-lsc, we know (see for example [15, Proposition 3.1.15]) that $\Phi_{(w,c)}(\cdot,\cdot)$ attains its minimum $(\tilde{x},\tilde{z})$ on $L_r(w,c)$, which must coincide with the unconstrained minimum. In view of (3.12), we conclude that $(\tilde{x},\tilde{z})$ verifies

$$q(w,c) = f(\tilde{x},\tilde{z}) - \langle \tilde{z}, w\rangle + c\sigma(\tilde{z}).$$

□

The following theorem guarantees a special property of IMSg, which is not verified by the classical subgradient algorithm. It states that IMSg guarantees a monotonic increase of the dual function, generalizing Theorem 1.6.4.

**Theorem 3.2.3.** *Let $\{z_k\}$ and $\{(y_k, c_k)\}$ generated by* IMSg. *If $z_k \neq 0$ and $(y_k, c_k)$ is not a dual solution, then $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$.*

*Proof.* For all $k$ consider $\epsilon_k := \alpha_k s_k$. Using the update rule to the dual variables we have

$$\begin{aligned}
q(y_{k+1}, c_{k+1}) &= \inf_{(x,z)} \{f(x,z) - \langle z, y_{k+1}\rangle + c_{k+1}\sigma(z)\} \\
&= \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + s_k\langle z, z_k\rangle + \\
&\quad [c_k + (\epsilon_k + s_k)\sigma(z_k)]\sigma(z)\} \\
&= \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k\sigma(z_k))\sigma(z) + \\
&\quad (\sigma(z_k)\sigma(z) + \langle z, z_k\rangle)s_k\} \\
&\geq \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k\|z_k\|)\sigma(z) + \\
&\quad (\|z_k\|\|z\| + \langle z, z_k\rangle)s_k\}
\end{aligned}$$

where the inequality follows from $(A_0)$. Now we obtain, using Cauchy-

Schwarz inequality,

$$\begin{aligned} q(y_{k+1}, c_{k+1}) &\geq \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k \|z_k\|)\sigma(z)\} \\ &= q(y_k, c_k + \epsilon_k\|z_k\|) \geq q(y_k, c_k), \end{aligned}$$

where the second inequality follows from Proposition 3.1.4 (ii). Thus we have

$$q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k + \epsilon_k\|z_k\|) \geq q(y_k, c_k) \quad \text{for all } \ k. \tag{3.13}$$

In particular we have $q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k)$ for all k. It follows that $q(y_k, c_k) > -\infty$ for all $k$, because $q_0 = q(y_0, c_0) > -\infty$. Therefore we obtain from Lemma 3.2.2 that, fixing $k$ such that $z_k \neq 0$, there exists $(\tilde{x}, \tilde{z})$ such that

$$q(y_k, c_k + \epsilon_k\|z_k\|) = f(\tilde{x}, \tilde{z}) - \langle \tilde{z}, y_k\rangle + (c_k + \epsilon_k\|z_k\|)\sigma(\tilde{z}). \tag{3.14}$$

If $\tilde{z} = 0$, then we get from (3.14) that

$$q(y_k, c_k + \epsilon_k\|z_k\|) = f(\tilde{x}, 0) = \varphi(\tilde{x}) \geq M_d > q(y_k, c_k),$$

where the last strict inequality follows from the fact that $(y_k, c_k)$ is not a dual solution. Therefore we conclude from (3.13) that $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$. If $\tilde{z} \neq 0$, then, since $z_k \neq 0$ and $\epsilon_k > 0$, we obtain from (3.13) and (3.14) that

$$q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k) + \epsilon_k\|z_k\|\|\tilde{z}\| > q(y_k, c_k),$$

using $(A_0)$ and definition of $q$. The proof is complete. $\square$

**Proposition 3.2.4.** *Consider the dual sequence* $\{(y_k, c_k)\}$ *generated by* IMSg. *The set* $A(y_k, c_k) := \{(x,z) \in X \times H : f(x,z) - \langle z, y_k\rangle + c_k\sigma(z) = q(y_k, c_k)\}$ *is nonempty for each* $k \geq 1$. *In particular, the exact version of* IMSg *is well defined; that is to say, for all* $k \geq 1$ *there exists* $(x_k, z_k) \in X \times H$ *satisfying* $q(y_k, c_k) = f(x_k, z_k) - \langle y_k, z_k\rangle + c_k\sigma(z_k)$.

*Proof.* It follows from inequalities (3.8) and (3.9) that

$$c_k \geq c_0 + \|y_k - y_0\| + \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j) > c_0 + \|y_k - y_0\|.$$

Therefore the result follows from the definition of $A(y_k, c_k)$ and Lemma 3.2.2, because by assumption $(y_0, c_0)$ satisfies $q(y_0, c_0) > -\infty$ (see Step 0 of IMSg).

□

From now on we assume that $z_k \neq 0$ for all $k$. In other words, we assume from now on that the method generates an infinite sequence.

**Lemma 3.2.5.** *Let* $\{(y_k, c_k)\}$ *be the sequence generated by* IMSg *and consider a sequence* $\{z_k\}$ *such that* $(x_k, z_k) \in A_{r_k}(y_k, c_k)$ *for all* $k$. *Then, the sequence* $\{\sigma(z_k)\}$ *is bounded and in particular* $\{z_k\}$ *is bounded.*

*Proof.* From (3.8) and (3.9) we obtain, for all $k \geq 1$,

$$c_k - c_0 \geq \|y_k - y_0\| + \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j) \geq \|y_k - y_0\| + a,$$

for some $a > 0$, ( e.g., we may take $a = \alpha_0 s_0 \sigma(z_0)$). Hence we have

$$\|y_k - y_0\| + c_0 - c_k \leq -a. \tag{3.15}$$

On the other hand, by Proposition 3.2.1, $(-z_k, \sigma(z_k)) \in \partial q_{r_k}(y_k, c_k)$. Thus

$$\begin{aligned} q_0 = q(y_0, c_0) &\leq q(y_k, c_k) + \langle -z_k, y_0 - y_k \rangle + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq M_d + \|z_k\|\|y_k - y_0\| + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq M_d + \sigma(z_k)(\|y_k - y_0\| + c_0 - c_k) + r_k \\ &\leq M_d - a\sigma(z_k) + r_k, \end{aligned}$$

using Cauchy-Schwarz inequality in the second inequality, $(A_0)$ in the third one, and (3.15) in the last one. It follows that

$$q_0 \leq M_d - a\sigma(z_k) + r_k. \tag{3.16}$$

Rewriting (3.16) we obtain

$$\sigma(z_k) \leq \frac{M_d - q_0 + r_k}{a} \leq \frac{M_d - q_0 + \tilde{r}}{a} := b,$$

where $\tilde{r} > 0$ is an upper bound for $\{r_k\}$. Since $\|z_k\| \leq \sigma(z_k) \leq b$ for all $k \geq 1$, the proof is complete. □

From now on we use the notation $q_k := q(y_k, c_k)$ for all $k$, and $\bar{q} := M_d$.

**Lemma 3.2.6.** *Consider the sequences $\{(x_k, z_k)\}$, $\{(y_k, c_k)\}$ generated by* IMSg *algorithm.*

*a) The following estimates hold for all $k \geq 1$,*

$$\begin{aligned} f(x_k, z_k) - \langle z_k, y_0 \rangle &\leq q_k + r_k, \text{ and} &(3.17)\\ \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) &\leq q_k - q_0 + r_k. &(3.18)\end{aligned}$$

*b) Assume that the dual solution set $D_*$ is nonempty. If $(\bar{y}, \bar{c}) \in D_*$ then for all $k$,*

$$\|y_{k+1} - \bar{y}\|^2 \leq \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left(\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right). \tag{3.19}$$

*Proof.* From the update formula for $\{(y_k, c_k)\}$ we have

$$y_k = y_0 - \sum_{j=0}^{k-1} s_j z_j \text{ and } c_k = c_0 + \sum_{j=0}^{k-1}(1 + \alpha_j)s_j\sigma(z_j). \tag{3.20}$$

Hence

$$\langle y_k, z_k \rangle = \langle y_0, z_k \rangle - \sum_{j=0}^{k-1} s_j \langle z_j, z_k \rangle.$$

By Cauchy Schwarz inequality and $(A_0)$,

$$\langle y_k, z_k \rangle \leq \langle y_0, z_k \rangle + \sum_{j=0}^{k-1} s_j \sigma(z_j)\sigma(z_k).$$

Using the expression for $c_k$ given in (3.20) in the inequality above, we obtain

$$\langle y_k, z_k\rangle \leq \langle y_0, z_k\rangle - c_0\sigma(z_k) - \sum_{j=0}^{k-1}\alpha_j s_j \sigma(z_j)\sigma(z_k) + c_k\sigma(z_k).$$

Adding $f(x_k, z_k)$ to both sides of this inequality, and observing that $\sigma \geq 0$, we have, after some simple algebra,

$$\begin{aligned} f(x_k, z_k) - \langle y_0, z_k\rangle &\leq f(x_k, z_k) - \langle y_0, z_k\rangle + c_0\sigma(z_k) + \sum_{j=0}^{k-1}\alpha_j s_j\sigma(z_j)\sigma(z_k) \\ &\leq f(x_k, z_k) - \langle y_k, z_k\rangle + c_k\sigma(z_k) \leq q_k + r_k. \end{aligned}$$

From these inequalities, using the definition of $q$, we obtain

$$f(x_k, z_k) - \langle y_0, z_k\rangle \leq q_k + r_k \text{ and } q_0 + \sum_{j=0}^{k-1}\alpha_j s_j\sigma(z_j)\sigma(z_k) \leq q_k + r_k,$$

which are the statements of (a). For proving (b), take $(\bar{y}, \bar{c}) \in D_*$. For all $k$ we have

$$\begin{aligned} \|y_{k+1} - \bar{y}\|^2 &= \|y_k - s_k z_k - \bar{y}\|^2 \\ &= \|y_k - \bar{y}\|^2 + s_k^2\|z_k\|^2 + 2s_k\langle \bar{y} - y_k, z_k\rangle \\ &\leq \|y_k - \bar{y}\|^2 + s_k^2\sigma(z_k)^2 + 2s_k(q_k - \bar{q} + r_k + \sigma(z_k)(\bar{c} - c_k)) \\ &= \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left(\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right) \end{aligned}$$

where the inequality follows from $(A_0)$ and the supergradient inequality. The result follows. □

**Lemma 3.2.7.** *If the sequence $\{z_k\}$ converges weakly to $0$, then $\{q_k\}$ converges to $\bar{q}$, the primal sequence $\{x_k\}$ is bounded, and all its weak accumulation points are primal solutions.*

*Proof.* Take an upper bound $r$ of $\{r_k\}$. It follows from Lemma 3.2.6(a) that

$$f(x_k, z_k) - \langle y_0, z_k\rangle \leq q_k + r_k \leq \bar{q} + r \quad \text{for all } k.$$

Rearranging this inequality and using Cauchy-Schwarz inequality, we obtain

$$f(x_k, z_k) \leq \|y_0\|\|z_k\| + \bar{q} + r \leq \tilde{b} := \|y_0\|b + \bar{q} + r \text{ for all } k, \tag{3.21}$$

where $b$ is an upper bound for $\{\|z_k\|\}$. Now by the level compactness assumption on $f$, there exists a weak open neighborhood $V \subset H$ of 0 and a weakly compact set $B \subset X$ such that

$$L_{V,f}(\tilde{b}) = \{x : f(x,z) \leq \tilde{b}\} \subset B, \text{ for all } z \in V.$$

Since $\{z_k\}$ is weakly convergent to 0, $z_k \in V$ for $k$ sufficiently large. Hence $x_k \in L_{V,f}(\tilde{b})$ for $k$ sufficiently large, by (3.21). Therefore $\{x_k\}$ is bounded. Take a weak accumulation point $\bar{x}$ of $\{x_k\}$. Thus there exists a subsequence $\{x_{k_j}\}$ which converges weakly to $\bar{x}$. In particular $\{(x_{k_j}, z_{k_j})\}$ converges weakly to $(\bar{x}, 0)$. Since $f(\cdot,\cdot)$ is w-lsc, we obtain

$$\varphi(\bar{x}) = f(\bar{x}, 0) \leq \liminf_j (f(x_{k_j}, z_{k_j}) - \langle y_0, z_{k_j}\rangle) \leq \liminf_j (q_{k_j} + r_{k_j}) \leq \bar{q}, \tag{3.22}$$

where the second inequality follows from Lemma 3.2.6 (a), and the third follows from the fact that $\{r_k\}$ converges to 0. Since $\bar{q} = M_p$ by Theorem 3.1.5, we obtain from (3.22) that $\varphi(\bar{x}) = M_p$, and then $\bar{x}$ is a primal solution. In particular, all inequalities in (3.22) are equalities. Since $\{r_k\}$ converges to 0, we obtain that $\liminf_j q_{k_j} = \bar{q}$. Since $\{q_k\}$ is increasing by Theorem 3.2.3, we conclude that $\{q_k\}$ converges to $\bar{q}$. The proof is complete. □

In order to obtain our results we consider the following assumption on the error sequence.

($\mathbf{A_1}$): There exists $R > 0$ such that $r_k \leq \bar{q} - q_k + R\sigma(z_k)$ for all $k$.

**Remark 3.2.8.** We mention that the verification of condition $(A_1)$ is not immediate, since in general at iteration $k$ we ignore the values of both $\bar{q}$ and $q_k$. One alternative is to think of an "a posteriori" verification of $(A_1)$, meaning that we check the boundedness of the sequence $\{\frac{r_k}{\sigma(z_k)}\}$. Indeed, note that when this sequence is bounded, then $(A_1)$ holds, because $\frac{\bar{q}-q_k}{\sigma(z_k)} \geq 0$. The situation improves considerably when we know the optimal dual value $\bar{q}$, a situation which occurs in some "real life" problems. In this case, we can verify condition $(A_1)$ along the iterations of the algorithm. In order to do this, we observe that the value $L_k := f(x_k, z_k) - \langle y_k, z_k \rangle + c_k \sigma(z_k)$ is computable, and satisfies $\hat{r}_k := L_k - q_k \leq r_k$ for each $k$. The (unknown) value $\hat{r}_k$ is in fact the actual error in the $k$th-iteration, while the (known) value $r_k$ can be seen as an estimate of $\hat{r}_k$. We assert that we can take $\hat{r}_k$ instead of $r_k$ in condition $(A_1)$. Observe that, since $L_k = q_k + \hat{r}_k$, we should verify the following condition:

$(\hat{\mathbf{A}}_\mathbf{1})$: There exists $R > 0$ such that $L_k \leq \bar{q} + R\sigma(z_k)$ for all $k$.

This condition is checkable when we know the optimal dual value $\bar{q} = M_p$. Another possibility is to think of IMSg as "measuring" at each iteration the boundedness of $\frac{r_k}{\sigma(z_k)}$, meaning that we observe whether the condition $\frac{r_k}{\sigma(z_k)} \leq R$ is satisfied, where $R > 0$ is an "a priori" given parameter. For those values of $k$ such that this inequality does not hold, one should consider an exact step ($r_k = 0$). Another option consists of applying the inexact algorithm IMSg just for a finite number of iterations and then switch to the exact version of IMSg, i.e., with $r_k = 0$. We mention that the exact version of IMSg is well defined, see Proposition 3.2.4.

**Lemma 3.2.9.** *If $\{c_k\}$ is bounded then $\{y_k\}$ is also bounded. If the dual solution set is nonempty then the converse of the previous statement holds.*

*Proof.* The first statement follows directly from (3.8) and (3.9). For proving the last statement, we rewrite the supergradient inequality as follows:

$$c_k \leq \bar{c} + \frac{q_k - \bar{q} + r_k - \langle \bar{y} - y_k, z_k \rangle}{\sigma(z_k)},$$

where $(\overline{y}, \overline{c}) \in D_*$. Using $(A_0), (A_1)$ and Cauchy-Schwarz inequality we obtain

$$c_k \leq \bar{c} + R + \|\bar{y} - y_k\| \tag{3.23}$$

where the constant $R$ is given by $(A_1)$. The last statement now follows from (3.23). □

Next we propose and analyze two algorithms related to IMSg. We remark that the difference between them lies in the stepsize selection rule.

### 3.2.1 Algorithm 1

Take two parameters $\beta > \eta > 0$. In Step-2 of the $k$-th iteration of IMSg, take $\eta_k := \min\{\eta, \|z_k\|\}$ and $\beta_k := \max\{\beta, \sigma(z_k)\}$, and choose a stepsize $s_k \in [\eta_k, \beta_k]$, for all $k$. We denote this algorithm by IMSg-1.

**Remark 3.2.10.** By definition of $\eta_k, \beta_k$ and $(A_0)$ we have $\eta_k \le \|z_k\| \le \sigma(z_k) \le \beta_k$. In particular we see that $\|z_k\|$ and $\sigma(z_k)$ are simple choices for the stepsize $s_k$. Observe that since $[\eta, \beta] \subset [\eta_k, \beta_k]$ for all $k$, we can choose any stepsize $s \in [\eta, \beta]$. In particular, a constant stepsize for all iterations is admissible.

The next theorem establishes some basic convergence properties of the dual sequence generated by IMSg-1.

**Theorem 3.2.11.** *Assume that* IMSg-1 *generates an infinite dual sequence* $\{(y_k, c_k)\}$. *If the dual optimal set is nonempty then* $\{(y_k, c_k)\}$ *is bounded and all its weak accumulation points are dual solutions; if the dual optimal set is empty then* $\{(y_k, c_k)\}$ *is unbounded.*

*Proof.* First, we prove that $\{(y_k, c_k)\}$ is bounded when $D_* \neq \emptyset$. Observe that $s_k \le \beta_k \le \max\{\beta, b\}$, where $b$ is an upper bound for $\sigma(z_k)$ (see Lemma 3.2.5). Thus $s_k\sigma(z_k) \le \hat{b} := \max\{b\beta, b^2\}$ for all $k$. Let $R$ be as in $(A_1)$, and take $(\bar{y}, \bar{c}) \in D_*$. If we show that $\{c_k\}$ is bounded, then $\{(y_k, c_k)\}$ will be bounded, by Lemma 3.2.9. Suppose by contradiction that $\{c_k\}$ is not bounded. Thus there exists $k_0$ such that $c_k \ge M := \dfrac{\hat{b}}{2} + R + \bar{c}$ for all $k \ge k_0$.

Using these estimates in (3.19), we obtain

$$\|y_{k+1} - \bar{y}\|^2 \leq \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left(\frac{\hat{b}}{2} + R + \bar{c} - c_k\right) \leq \|y_{k_0} - \bar{y}\|^2,$$

for all $k \geq k_0$. It follows that $\{y_k\}$ is bounded. This entails a contradiction, in view of Lemma 3.2.9. Therefore the dual sequence is bounded.

Let us prove now that all weak accumulation points of $\{(y_k, c_k)\}$ are dual solutions. In particular this also proves the last statement of the theorem by contradiction. Since $\{(y_k, c_k)\}$ is bounded, we know that $\sum_k s_k\sigma(z_k) < \infty$, by Proposition 3.2.1. In particular $\{s_k\sigma(z_k)\}$ converges to 0. On the other hand, using $(A_0)$ and the fact that $s_k \geq \min\{\eta, \|z_k\|\}$ we obtain

$$s_k\sigma(z_k) \geq \min\{\eta\|z_k\|, \|z_k\|^2\} > 0.$$

Since $\eta > 0$, we conclude that $\{\|z_k\|\}$ converge to 0. In particular $\{z_k\}$ converges weakly to 0. Now Lemma 3.2.7 ensures that $\{q_k\}$ converges to $\bar{q}$. Take a weak accumulation point $(\bar{y}, \bar{c})$ of $\{(y_k, c_k)\}$, so that there exists a subsequence $\{(y_{k_j}, c_{k_j})\}$ weakly convergent to $(\bar{y}, \bar{c})$. Since the dual function is w-usc (see Proposition 3.1.4), we have

$$\bar{q} \geq q(\bar{y}, \bar{c}) \geq \limsup_j q(y_{kj}, c_{kj}) = \lim_j q_{kj} = \bar{q}.$$

Hence $q(\bar{y}, \bar{c}) = \bar{q}$ and we conclude that $(\bar{y}, \bar{c})$ is a dual optimal solution. In particular, boundedness of the dual sequence implies that the dual solution set $D_*$ is nonempty, which establishes the theorem. □

Theorem 3.2.11 establishes dual convergence results of IMSg-1. The next theorem establishes primal convergence results.

**Theorem 3.2.12.** *Consider the primal sequence $\{x_k\}$ generated by* IMSg-1. *Suppose that there exists $\bar{\alpha} > 0$ such that $\bar{\alpha} \leq \alpha_k$ for all $k$. Then $\{q_k\}$ converges to $\bar{q}$, the primal sequence $\{x_k\}$ is bounded and all its weak accumulation points are primal solutions.*

*Proof.* Take the dual sequence $\{(y_k, c_k)\}$ generated by IMSg-1. If $\{(y_k, c_k)\}$ is bounded, then we can use the same argument as in the second part of the proof of Theorem 3.2.11 for ensuring that $\{z_k\}$ converges weakly to 0. Thus, in the case that $\{(y_k, c_k)\}$ is bounded the result follows from Lemma 3.2.7. Hence we just need to consider the case in which the dual sequence is unbounded. In this case we get from Lemma 3.2.6(a)

$$\sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j) \sigma(z_k) \le q_k - q_0 + r_k. \tag{3.24}$$

On the other hand, $\{r_k\}$ is bounded and $q_k - q_0 \le \bar{q} - q_0$ for all $k$. Thus there exists $\hat{M} > 0$ such that $q_k - q_0 + r_k \le \hat{M}$ for all $k$. Using this estimate in (3.24), together with the fact that $\alpha_k \ge \bar{\alpha}$ for all $k$, we obtain

$$\bar{\alpha} \left( \sum_{j=0}^{k-1} s_j \sigma(z_j) \right) \sigma(z_k) \le \hat{M} \tag{3.25}$$

for all $k \ge 1$. Since the dual sequence is unbounded, Proposition 3.2.1(ii) implies that $\sum_{j=0}^{\infty} s_j \sigma(z_j) = \infty$. Using this fact in (3.25), since $\bar{\alpha} > 0$, it follows that $\{\sigma(z_k)\}$ converges to zero. By assumption $(A_0)$ we get that $\{\|z_k\|\}$ converges to 0, and thus $\{z_k\}$ converges weakly to 0. The result now follows from Lemma 3.2.7. □

In order to establish convergence of the whole dual sequence, we need some preliminary material on Fejér convergence.

**Definition 3.2.1.** *Let $H$ be a Hilbert space and $V$ a nonempty subset of $H$. A sequence $\{z_k\} \subset H$ is said to be* quasi-Fejér *convergent to $V$ if and only if for all $\bar{z} \in V$ there exists some sequence $\{\mu_k\} \subset \mathbb{R}_+$ such that $\sum_k \mu_k < \infty$ and*

$$\|z_{k+1} - \bar{z}\|^2 \le \|z_k - \bar{z}\|^2 + \mu_k.$$

**Lemma 3.2.13.** *Consider a Hilbert space $H$ and a sequence $\{\xi_k\} \subset H$. If $\{\xi_k\}$ is* quasi-Fejér *convergent to some set $V \neq \emptyset$, then* *a) The sequence $\{\xi_k\}$ is bounded;*

*b) $\{\|\xi_k - \bar{v}\|\}$ is convergent for all $\bar{v} \in V$;*

*c) if all the weak accumulation points of $\{\xi_k\}$ are in $V$, then $\{\xi_k\}$ is weakly convergent to some $\bar{v} \in V$.*

*Proof.* See for example [1, Proposition 1]. □

Next we establish quasi-Fejér convergence of the dual sequence generated by IMSg-1 to an appropriate subset of the dual solution.

**Proposition 3.2.14.** *Consider the dual sequence $\{(y_k, c_k)\}$ generated by* IMSg-1*. If $D_*$ is nonempty, then $\{(y_k, c_k)\}$ is quasi-Fejér convergent to the set $V_* = \{(y, c) \in D_* : c \geq c_k \ \forall k\}$.*

*Proof.* Since $D_*$ is nonempty, it follows from Theorem 3.2.11 and Proposition 3.1.4 (ii) that there exists some $(\bar{y}, \bar{c}) \in D_*$ such that $\bar{c} \geq c_k$ for all $k$, i.e., $V_*$ is nonempty. Take any $(\bar{y}, \bar{c}) \in V_*$. Consider $d_k := \|(\bar{y}, \bar{c}) - (y_k, c_k)\|$. Using the updating formula for the dual sequence, we have, for all $k$,

$$\begin{aligned} d_{k+1}^2 &= \left\|(\bar{y}, \bar{c}) - \big(y_k - s_k z_k, c_k + (1+\alpha_k) s_k \sigma(z_k)\big)\right\|^2 \\ &= d_k^2 + s_k^2 \|z_k\|^2 + (1+\alpha_k)^2 s_k^2 \sigma(z_k)^2 + \\ &\quad 2 s_k \big[\langle \bar{y} - y_k, z_k \rangle - (1+\alpha_k)\sigma(z_k)(\bar{c} - c_k)\big] \\ &\leq d_k^2 + s_k^2 \sigma(z_k)^2 + (1+\alpha)^2 s_k^2 \sigma(z_k)^2 + \\ &\quad 2 s_k \big[\langle \bar{y} - y_k, z_k \rangle - \sigma(z_k)(\bar{c} - c_k)\big], \end{aligned}$$

where the inequality follows from $(A_0)$ and the fact that $\alpha > \alpha_k > 0$. Now, using the supergradient inequality, we obtain

$$d_{k+1}^2 \leq d_k^2 + \big(1 + (1+\alpha)^2\big) s_k^2 \sigma(z_k)^2 + 2 s_k (q_k - \bar{q} + r_k). \tag{3.26}$$

By $(A_1)$ we get $R > 0$ such that $q_k - \bar{q} + r_k \leq R\sigma(z_k)$. Using this estimate in (3.26) and considering $\hat{\alpha} := 1 + (1+\alpha)^2$, we have

$$d_{k+1}^2 \leq d_k^2 + \hat{\alpha} s_k^2 \sigma(z_k)^2 + 2 R s_k \sigma(z_k). \tag{3.27}$$

On the other hand, Theorem 3.2.11 ensures boundedness of the dual sequence. Hence we have $\sum_k s_k\sigma(z_k) < \infty$, by Proposition 3.2.1, which in turn implies that $\sum_k s_k^2\sigma(z_k)^2 < \infty$. Consider $\mu_k := \hat{\alpha}s_k^2\sigma(z_k)^2 + 2Rs_k\sigma(z_k)$ for all $k$. We see that $\sum_k \mu_k < \infty$ and by (3.27) we obtain

$$d_{k+1}^2 \leq d_k^2 + \mu_k$$

for all $k$. The result follows from Definition 3.2.1. □

Now we establish weak convergence of the whole dual sequence generated by IMSg-1 to a dual solution.

**Theorem 3.2.15.** *If the dual solution set is nonempty, then the dual sequence generated by* IMSg-1 *is weakly convergent to some dual solution.*

*Proof.* By Theorem 3.2.11, the dual sequence $\{(y_k, c_k)\}$ is bounded and all its weak accumulation points belong to $V_* = \{(y_*, c_*) \in D_* : c_* \geq c_k \ \forall k\}$ (observe that $\{c_k\}$ is increasing). By Proposition 3.2.14 this sequence is quasi-Fejér convergent to $V_*$. By Lemma 3.2.13(c), the sequence is weakly convergent to some $(\bar{y}, \bar{c}) \in V_* \subset D_*$. □

The argument is a standard one for proving weak convergence of the whole sequence generated by a subgradient method. Since in our approach the parameter $c$ is taken as variable, we can obtain more than just weak convergence. We obtain indeed that the whole dual sequence converges strongly when dual solutions exist.

**Theorem 3.2.16.** *If the dual optimal solution set is nonempty then the dual sequence generated by IMSg converges strongly to a dual solution.*

*Proof.* For proving this result, observe that for any $j > k > 0$ it holds:

$$\|y_j - y_k\| = \|\sum_{l=k}^{j-1} s_l z_l\| \leq \sum_{l=k}^{j-1} s_l\|z_l\| \leq \sum_{l=k}^{j-1} s_l\sigma(z_l). \tag{3.28}$$

By Theorem 3.2.11 and Proposition 3.2.1, if the dual optimal solution set is nonempty, then $\sum_{l=1}^{\infty} s_l \sigma(z_l)$ converges. In particular, from the estimates in (3.28) we conclude that $\{y_k\}$ is a Cauchy sequence and therefore strongly convergent, because $\{y_k\} \subset H$ and $H$ is a Hilbert space. Since, $\{c_k\}$ is monotonically increasing and, under nonemptyness of the dual solution set, $\{c_k\}$ is also bounded, it is convergent. We conclude that $\{(y_k, c_k)\}$ converges strongly to a dual solution, because we know that it converges strongly and each of its weak accumulation points is a dual solution, by Theorem (3.2.11). $\square$

### 3.2.2 Algorithm 2

In this section we propose a stepsize rule which ensures that IMSg converges in a finite number of steps.

Take $\beta > 0$ and a sequence $\{\theta_k\} \subset \mathbb{R}_+$ such that $\sum_j \theta_j = \infty$, and $\theta_k \leq \beta$ for all $k$. In Step-2 of the $k$-th iteration of IMSg, consider $\eta_k := \dfrac{\theta_k}{\sigma(z_k)}$ and $\beta_k := \dfrac{\beta}{\sigma(z_k)}$, and choose a stepsize $s_k \in [\eta_k, \beta_k]$, for all $k$. IMSg with this stepsize rule is denoted by IMSg-2.

**Theorem 3.2.17.** *a) Suppose that the dual solution set $D_*$ is nonempty. Let $\{(x_k, z_k)\}$ and $\{(y_k, c_k)\}$ be the primal and dual sequences generated by* IMSg-2. *Then there exists $\bar{k}$ such that* IMSg-2 *stops at iteration $\bar{k}$. As a consequence $x_{\bar{k}}$ and $(y_{\bar{k}}, c_{\bar{k}})$ are $\epsilon_*$-optimal primal and $\epsilon_*$-optimal dual solutions respectively.*

*b) Suppose that* IMSg-2 *generates infinite primal and dual sequences $\{(x_k, z_k)\}$ and $\{(y_k, c_k)\}$, (in this case $D_*$ is empty by (a)). Assume that $\alpha_k \geq \bar{\alpha} > 0$ for all $k$. Then $\{(y_k, c_k)\}$ is unbounded, $\{\|z_k\|\}$ converges to 0, and $\{q_k\}$ converges to the optimal value $\bar{q}$. The primal sequence $\{x_k\}$ is bounded and all its weak accumulation points are primal solutions.*

*Proof.* $a)$ Taking an upper bound $\hat{b}$ for $\{s_k \sigma(z_k)\}$ and repeating the first part of the proof of Theorem 3.2.11, it follows that $\{(y_k, c_k)\}$ is bounded. In particular, we obtain $\sum_j s_j\sigma(z_j) < \infty$, in view of Proposition 3.2.1. Observe now that the criterion $r_k < \varepsilon_*$ in Step-1 of IMSg-2 is satisfied after a finite number of iterations, because $\{r_k\}$ converges to 0. Suppose by contradiction that the stopping criterion of IMSg-2 is not satisfied. Therefore $z_k \neq 0$ for all $k$. By the stepsize selection rule of IMSg-2, it follows that $s_k\sigma(z_k) \geq \theta_k$ for all $k$. This is a contradiction, because $\infty > \sum_j s_j\sigma(z_j) \geq \sum_j \theta_j = \infty$. Therefore IMSg-2 stops at iteration some $\bar{k}$, and by Theorem 3.2.1 we conclude that $x_{\bar{k}}$ is an $\epsilon_*$-optimal primal solution and $\{(y_{\bar{k}}, c_{\bar{k}})\}$ is an $\epsilon_*$-optimal dual solution.

For proving $(b)$ we observe that since IMSg-2 generates infinite primal and dual sequences, it follows that $z_k \neq 0$ for all $k$. By the stepsize selection rule of IMSg-2 we have $s_k\sigma(z_k) \geq \theta_k$ for all $k$. In particular $\sum_k s_k\sigma(z_k) = \infty$, which is equivalent to unboundedness of $\{(y_k, c_k)\}$, by Proposition 3.2.1. Now the result follows by using the same argument as in the proof of the second part of Theorem 3.2.12. $\square$

The next proposition establishes that if Assumption $(A_0)$ does not hold, then the conclusion of Theorem 3.2.3 may fail.

**Proposition 3.2.18.** *Let $H$ be a Hilbert space. Suppose that there exists some $0 \neq \bar{u} \in H$ such that $\sigma(\bar{u}) = \gamma_1\|\bar{u}\|$ with $\gamma_1 < 1$. Moreover suppose that $\sigma(-\bar{u}) = \gamma_2\|\bar{u}\|$, with $\gamma_1\gamma_2 < 1$. In this situation, the conclusion of* Theorem 3.2.3 *may fail.*

*Proof.* Observe that we only need to find a problem such that $q_1 < q_0 < M_d$. First, consider a w-lsc function $g : H \to \mathbb{R}$, such that

$$\bar{u}, -\bar{u} \in \operatorname{Argmin}(g + \sigma) \quad \text{and} \quad \min(g + \sigma) = 0 < g(0),$$

(for example $g(x) = -\sigma(x)$ if $x \in \{\bar{u}, -\bar{u}\}$, $g(x) = 1$ otherwise). Let $K \subset H$ be a weakly compact set such that $\{\bar{u}, -\bar{u}, 0\} \subset K$. Consider the following primal problem

$$\min \;\; f(x) := g(x) + \langle \bar{u}, x\rangle \quad \text{s.t.} \;\; x \in K, \;\; h(x) := x = 0. \tag{3.29}$$

Let $C = \{x \in K : h(x) = 0\}$ and $\delta_C(x) = 0$ if $x \in C$, $\delta_C(x) = \infty$ otherwise (observe that $C = \{0\}$, because $0 \in K$ and $h(x) = x$). For each $x \in H$, let $\bar{\phi}(x) := f(x) + \delta_C(x)$. Therefore the problem (3.29) is equivalent to

$$\min \ \bar{\phi}(x) \ \text{ s.t. } x \in H,$$

Consider now a dualizing parameterization function given by

$$\bar{f}(x,u) = \begin{cases} f(x) & \text{if } x \in K \text{ and } x = u, \\ \infty & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \ell(x,y,c) &= \inf_u\{\bar{f}(x,u) - \langle y,u\rangle + c\sigma(u)\} \\ &= \begin{cases} f(x) - \langle y,x\rangle + c\sigma(x) & \text{if } x \in K \\ \infty & \text{otherwise,} \end{cases} \end{aligned}$$

and therefore

$$\begin{aligned} q(y,c) &= \inf_{x\in H} \ell(x,y,c) = \inf_{x\in K}\{f(x) - \langle y,x\rangle + c\sigma(x)\} \\ &= \inf_{x\in K}\{g(x) - \langle y - \bar{u},x\rangle + c\sigma(x)\}. \end{aligned}$$

Since $K$ is weakly compact and $\bar{f}$ is w-lsc, the hypotheses of Theorem 3.1.5 are satisfied, and therefore

$$\sup_{(y,c)\in H\times\mathbb{R}_+} q(y,c) = \inf_{x\in H} \bar{\phi}(x) =: v$$

(observe that $v = g(0)$). We can easily verify that

$$q(\overline{u},1) = \inf_{x\in K}\{g(x) + \sigma(x)\} = g(\overline{u}) + \sigma(\overline{u}) = 0.$$

Observe also that $0 = g(\overline{u}) + \sigma(\overline{u}) = \bar{f}(\overline{u},\overline{u}) - \langle\overline{u},\overline{u}\rangle + \sigma(\overline{u})$. Therefore, $(\overline{u},\overline{u}) \in A(\overline{u},1) := A_0(\overline{u},1)$, (see definition of $A_r(y,c)$ in (3.5)). Consider $y_0 = \overline{u}$, $c_0 = 1$ and $r_0 = 0$ in Step-0 of IMSg. If we take $(\overline{u},\overline{u}) \in A(y_0,c_0)$ as the solution of the subproblem (see IMSg) then

$$y_1 = y_0 - s_0\bar{u}, \text{ and } c_1 = c_0 + (1+\alpha_0)s_0\sigma(\bar{u}),$$

where $s_0 > 0$ is an initial stepsize and $\alpha_0$ satisfies $\alpha_0 < \dfrac{1}{\gamma_1\gamma_2} - 1$. Hence

$$\begin{aligned}
\ell(-\bar{u}, y_1, c_1) &= f(-\bar{u}) + \langle y_1, \bar{u}\rangle + c_1\sigma(-\bar{u}) \\
&= f(-\bar{u}) + \langle y_0 - s_0\bar{u}, \bar{u}\rangle + [c_0 + (1+\alpha_0)s_0\sigma(\bar{u})]\,\sigma(-\bar{u}) \\
&= f(-\bar{u}) + \|\bar{u}\|^2 + \sigma(-\bar{u}) + s_0\left[(1+\alpha_0)\sigma(\bar{u})\sigma(-\bar{u}) - \|\bar{u}\|^2\right]
\end{aligned}$$

where we use $y_0 = \overline{u}$ and $c_0 = 1$. Observing that $g(-\overline{u}) + \sigma(-\overline{u}) = 0$ and using the definition of $f$ we obtain

$$\begin{aligned}
\ell(-\bar{u}, y_1, c_1) &= s_0\left[(1+\alpha_0)\sigma(\bar{u})\sigma(-\bar{u}) - \|\bar{u}\|^2\right] \\
&= s_0\left[(1+\alpha_0)\gamma_1\gamma_2 - 1\right]\|\bar{u}\|^2 < 0,
\end{aligned}$$

where we use the assumptions on $\sigma(\overline{u}), \sigma(-\overline{u})$ and $\alpha_0$. Now by definition of $q(y_1, c_1)$ we get

$$q(y_1, c_1) \le \ell(-\bar{u}, y_1, c_1) < 0 = q(y_0, c_0) < g(0) = v,$$

that is, $q(y_1, c_1) < q(y_0, c_0) < v = M_d$. The proof is complete. $\square$

**Remark 3.2.19.** The assumption $\gamma_1\gamma_2 < 1$ used in Proposition 3.2.18 is satisfied when $\sigma$ is even (i.e., when $\sigma(z) = \sigma(-z)$ for all $z$) and there exists some $\bar{u}$ such that $\sigma(\bar{u}) < \|\bar{u}\|$. An example of such a function is $\sigma(z) = \|z\|^t$ for $t > 1$. Another choice of $\sigma$ for which Theorem 3.2.3 may be false is when $\sigma(\cdot) \ge \gamma\|\cdot\|$ with $0 < \gamma < 1$. In the latter case, the following simple modification of the algorithm ensures the increasing property of the dual values. Consider a sequence $\{t_k\}$ such that $t_k \ge \frac{1}{\gamma^2}$ for all $k$, and update the parameter $c_k$ as follows,

$$c_{k+1} := c_k + (\alpha_k + t_k)s_k\sigma(z_k).$$

The proof of the increasing property of the dual values is similar to the one given in Theorem 3.2.3 and it is omitted. We claim also that if we consider $\{t_k\} \subset [\frac{1}{\gamma^2}, d_1]$ for some $d_1 > \frac{1}{\gamma^2} > 0$, then Theorems 3.2.11, 3.2.12, 3.2.15 and Theorem 3.2.17 remain valid for this modification with essentially the same proofs.

### 3.2.3 IMSg with sharp Lagrangian

Modified subgradient algorithms (MSg) with sharp Lagrangian were proposed and analyzed in [12, 16, 23, 14] and Chapter 2, in finite dimensional spaces. In the present chapter we studied an inexact version of the MSg proposed in [16] (studied here in Chapter 2) and we extended the convergence results to augmented Lagrangians more general than the sharp Lagrangian. In [14] the authors proposed an inexact version of MSg proposed in [12]. In this section we compare our algorithm with these previous versions of MSg, giving special attention to the search direction. We also compare our assumption $(A_1)$ with the assumption on the error sequence $\{r_k\}$ used in [14]. For this purpose, consider Example 3.1.1, for which we have

$$\begin{aligned} A_r(y,c) &= \{(x,z) \in X \times H : f(x,z) - \langle y,z\rangle + c\|z\| \leq q(y,c) + r\} \\ &= \{(x,h(x)) \in K \times H : \psi(x) - \langle y,h(x)\rangle + c\|h(x)\| \leq q(y,c) + r\} \\ &= \{(x,h(x)) : x \in \Gamma_r(y,c)\}, \end{aligned}$$

where

$$\Gamma_r(y,c) := \{x \in K : L(x,y,c) := \psi(x) - \langle y,h(x)\rangle + c\|h(x)\| \leq q(y,c) + r\},$$

which is precisely the set $X_r(y,c)$ considered in [14, Equation (6)] in a finite dimensional setting. Moreover, the set $A_r(y,c)$ is completely determined if we know $\Gamma_r(y,c)$. Therefore, at iteration $k$ we update the search direction as $(-z_k, \sigma(z_k)) = (-h(x_k), \|h(x_k)\|)$ with $x_k \in \Gamma_{r_k}(y_k,c_k)$, which is the same search direction considered in [14]. In particular, when $r_k = 0$ for all $k$, we obtain the MSg studied in Chapter 2 and [16], which is the exact version of IMSg with sharp Lagrangian. Thus we have also extended to reflexive Banach spaces the MSg algorithm studied in Chapter 2. We also consider more general augmented Lagrangian functions than the sharp Lagrangian considered in [12, 14, 16, 23].

Next, we compare our assumption on the error sequence $\{r_k\}$ (condition $(A_1)$) with the assumption considered in [14]. First, we look again at Example

3.1.1 in a finite dimensional setting, with $\sigma(\cdot) = \|\cdot\|$, and $z_k = h(x_k)$ . It is easy to see that assumptions $(a_1)$ and $(a_2)$ presented in Section 1.6 of this thesis (considered in [14, Section 4]) are equivalent to the following assumption: there exist $\eta > 0$ and $M > 0$ such that, for all $k$,

$$\eta \frac{(\bar{q} - q_k + r_k)}{\sigma(z_k)} \leq s_k \sigma(z_k) \leq M.$$

In particular, from these inequalities we obtain $r_k < \frac{M}{\eta}\sigma(z_k)$ for all $k$. We remark that our assumption $(A_1)$ on the error sequence $\{r_k\}$ is an improvement over this last estimate.

# Chapter 4

# Exact penalty properties

In this chapter we shall consider the same primal-dual scheme as the one in Chapter 3. We analyze some properties related to the dual problem and relate differentiability of the dual function at a dual solution with an exact penalty property. We show how this result can be used for obtaining primal convergence of a modified subgradient algorithm. The outline of this chapter is as follows. In Section 4.1 we recall the primal-dual scheme and state some basic results. In Section 4.2 we define a penalty mapping and show some examples illustrating these results. We also show that exactness of this penalty mapping is equivalent, under mild assumptions, to differentiability of the dual function at a dual solution. In Section 4.2.1 we study properties of this penalty mapping. In Section 4.3 we relate these results to primal convergence of a modified subgradient algorithm. We apply these results to the constrained optimization problem in Section 4.4.

## 4.1 Preliminary

The primal-dual scheme of this chapter is the same as the one in Chapter 3. $X$ is a reflexive Banach space, and $H$ a Hilbert space. We denote by $\|\cdot\|$ the norm in both $X$ and $H$. The inner product of $z, y \in H$ is denoted by $\langle y, z\rangle$.

The primal optimization problem is

$$\text{minimize } \varphi(x) \text{ subject to } x \text{ in } X, \tag{4.1}$$

where the function $\varphi : X \to \mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ is a proper and weakly lower semicontinuous (w-lsc) function. Let $f : X \times H \to \bar{\mathbb{R}}$ be a dualizing parameterization function for (4.1).

The *augmented Lagrangian* function $\ell : X \times H \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ corresponding to the dualizing parameterization function $f$ and the augmenting function $\sigma$ is given by

$$\ell(x, y, r) := \inf_{z \in H} \{f(x, z) - \langle z, y\rangle + r\sigma(z)\}. \tag{4.2}$$

The dual function $q : H \times \mathbb{R}_+ \to \mathbb{R}_{-\infty}$ is $q(y, r) := \inf_{x \in X} \ell(x, y, r)$. The dual problem is stated as

$$\text{maximize } q(y, r) \text{ subject to } (y, r) \in H \times \mathbb{R}_+. \tag{4.3}$$

We denote by $M_p := \inf_{x \in X} \varphi(x)$ and by $M_d := \sup_{(y,r) \in H \times \mathbb{R}_+} q(y, r)$ the optimal values of the primal and dual problem, respectively. The primal and dual solution sets are denoted by $P_*$ and $D_*$, respectively.

We will only consider dualizing parameterization functions which are proper, weakly level-compact (see Definition 3.1.2) and w-lsc. In fact, we assume that the hypotheses of Theorem 3.1.5 are satisfied. In particular there is no duality gap between the primal and dual problems.

Fix $(y, c) \in H \times \mathbb{R}_+$ and consider the function $\Phi_{(y,c)} : X \times H \to \bar{\mathbb{R}}$ defined by

$$\Phi_{(y,c)}(x, z) := f(x, z) - \langle y, z\rangle + c\sigma(z), \tag{4.4}$$

and let $A(y, c) := \operatorname{Argmin}_{(x,z)} \Phi_{(y,c)}(x, z) \subset X \times H$. We also consider the projection of $A(y, c)$ in $X$, denoted by $P(y, c)$. More precisely,

$$P(y, c) := \{x \in X : (x, z) \in A(y, c), \text{ for some } z \in H\}.$$

Analogously, we consider the projection of $A(y, c)$ in $H$, denoted by $P_H(y, c)$:

$$P_H(y, c) := \{z \in H : (x, z) \in A(y, c), \text{ for some } x \in X\}.$$

Note that $q(y,c) = \Phi_{(y,c)}(x,z)$ whenever $(x,z) \in A(y,c)$.

**Proposition 4.1.1.** *The following statements hold:*

*a)* $(y,c) \in D_*$ *if and only if* $0 \in P_H(y,c)$. *If* $(y,c) \in D_*$ *then* $P_* \subset P(y,c)$;

*b) if* $P_H(y,c) = \{0\}$ *then* $P(y,c) = P_*$;

*c) if* $(y,c) \in D_*$ *then* $P_H(y,t) = \{0\}$ *for all* $t > c$.

*Proof.* For proving $(a)$ take an arbitrary $x^* \in P_*$. Hence

$$M_p = \varphi(x^*) = f(x^*,0) - \langle y,0\rangle + c\sigma(0) = \Phi_{(y,c)}(x^*,0). \tag{4.5}$$

Since $M_p = M_d$ (Theorem 3.1.5) we conclude from (4.5) that

$$q(y,c) = M_d \Longleftrightarrow q(y,c) = \Phi_{(y,c)}(x^*,0) \Longleftrightarrow 0 \in P_H(y,c).$$

In particular, when $(y,c) \in D_*$, it follows that $(x^*,0) \in A(y,c)$, which implies $x^* \in P(y,c)$. Since $x^* \in P_*$ is arbitrary we obtain that $P_* \subset P(y,c)$.

For proving $(b)$ let $P_H(y,c) = \{0\}$. Then for any $x \in P(y,c)$ it follows that $(x,0) \in A(y,c)$. Therefore

$$q(y,c) = \Phi_{(y,c)}(x,0) = f(x,0) - \langle y,0\rangle + c\sigma(0) = \varphi(x) \geq M_p \geq q(y,c),$$

which proves that $x \in P_*$ and $(y,c) \in D_*$. Since $x \in P(y,c)$ is arbitrary we obtain that $P(y,c) \subset P_*$. By $(a)$ we conclude that $P(y,c) = P_*$.

For proving $(c)$ let $t > c$. By Proposition 3.1.4 we obtain that $(y,t) \in D_*$, and by $(a)$ we have $0 \in P_H(y,t)$. Take an arbitrary $z \in P_H(y,t)$, we want to prove that $z = 0$. We know that there exists $x \in X$ such that $(x,z) \in A(y,t)$ and then

$$\Phi_{(y,t)}(x,z) = q(y,t) = M_d.$$

On the other hand, a simple manipulation shows that

$$\Phi_{(y,t)}(x,z) = \Phi_{(y,c)}(x,z) + (t-c)\sigma(z) \geq q(y,c) + (t-c)\sigma(z).$$

Therefore $\sigma(z) = 0$, because $t > c$, $q(y,c) = M_d$ and $\sigma$ is nonnegative. As a consequence, $z = 0$ and then $P_H(y,t) = \{0\}$. $\square$

The next example shows that the inclusion $P_* \subsetneq P(y,c)$ can be strict.

**Example 4.1.2.** Consider the problem

$$\min_{x\in\mathbb{R}} \varphi(x) := x + \delta_{[0,1]}(x)$$

with optimal value 0 and $P^* = \{0\}$. Take the parameterization function given by

$$f(x,z) = |x-z| + \delta_{[0,1]}(x) + \delta_{[0,1]}(z)$$

and the augmenting function $\sigma(z) = |z|$. It follows that

$$q(u,c) = \min_{x,z\in[0,1]} \{|x-z| + (c-u)z\}.$$

It is easy to see that $D^* = \{(u,c) \in \mathbb{R}\times\mathbb{R}_+ \; : \; c \geq u\}$ and that $P(t,t) = [0,1] \supsetneq P^*$ for any $t \in [0,1]$.

The preceding discussion and example motivate us to study under which conditions we obtain $P(y,c) = P_*$, or the stronger condition $P_H(y,c) = \{0\}$. As we will see, the condition $P_H(y,c) = \{0\}$ is equivalent to the differentiability of $q$ in $(y,c)$.

## 4.2 Exact Penalty Map

Consider a penalty map $E : H \to \mathbb{R}_+$ given by

$$E(y) = \inf\{c \geq 0 : (y,c) \in D_*\}. \tag{4.6}$$

The infimum above is assumed to be $+\infty$ when the argument of the infimum in (4.6) is empty. The weakly upper semicontinuity of the dual function $q$ implies that $(y, E(y)) \in D_*$ when $y \in \operatorname{dom} E$. Therefore by Proposition 4.1.1 we see that if $(y,c) \in D_*$ then $P_* \subsetneq P(y,c)$ can hold only when $c = E(y)$.

**Remark 4.2.1.** From (4.6) we see that the penalty map $E$ is nonnegative. However, since the dual function can be extended to $H\times\mathbb{R}$, it is still possible to consider negative values of $E$, see Example 4.2.2 for which $E(0) = -1$.

**Definition 4.2.1.** The penalty map $E$ is said to be *exact* at $y \in \operatorname{dom} E$ if and only if $P(y, E(y)) = P_*$. The penalty map $E$ is said *strongly exact* at $y \in \operatorname{dom} E$ if and only if $P_H(y, E(y)) = \{0\}$.

It follows from Proposition 4.1.1 $(b)$ that if $E$ is strongly exact at $y$ then $E$ is exact at $y$. We give next an example in which $E$ is exact but not strongly exact.

**Example 4.2.2.** Consider the following primal problem

$$\text{minimize } \phi(x) := \begin{cases} ln(x+1), & \text{if } \ x \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

We have that $P_* = \{0\}$. Take $\sigma(z) = |z|$ and a parameterization function $f$ given by

$$f(x,z) := \begin{cases} ln(x+1) + z, & \text{if } \ x \geq 0, z \in [0,1] \\ +\infty, & \text{otherwise.} \end{cases}$$

It is straightforward to show that

$$q(y,c) = \begin{cases} c - y + 1, & \text{if } c < y - 1, \\ 0, & \text{if } c \geq y - 1. \end{cases}$$

Also $E(y) = y - 1$ and $A(y, E(y)) = \{(0,t) : t \in [0,1]\}$. In particular, $P(y, E(y)) = P_*$ and $P_H(y, E(y)) = [0,1]$. Therefore $E$ is exact but not strongly exact. We mention that $q$ is not differentiable at $(y, E(y))$.

The next example illustrates a situation in which the penalty map $E$ is strongly exact.

**Example 4.2.3.** Consider the following primal problem

$$\text{minimize } \phi(x) := \begin{cases} ln(x+1) + 1, & \text{if } \ x \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Let $\sigma(z) = |z|$ and a parameterization function $f$ given by

$$f(x,z) := \begin{cases} ln(x+1) + \exp(-z), & \text{if } \ x,z \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

In this case we have

$$q(y,c) = \begin{cases} -\infty, & \text{if } c < y, \\ 0, & \text{if } c = y, \\ (c-y)(1-\ln(c-y)), & \text{if } y < c \leq y+1, \\ 1, & \text{if } c > y+1. \end{cases}$$

Also $E(y) = y+1$ and $A(y, E(y)) = \{(0,0)\}$. In particular, $P(y, E(y)) = P_*$ and $P_H(y, E(y)) = \{0\}$. Therefore $E$ is strongly exact. We observe that in this example $q$ is differentiable at $(y, E(y))$.

We mention that the existence of $y$ supporting an exact penalty representation is equivalent to $(y,c) \in D_*$ for some $c > 0$. Moreover the value $E(y)$ is the threshold for $y$, that is, the infimum of $c$ such that $(y,c) \in D_*$ and $\operatorname*{Argmin}_x \varphi(x) = \operatorname*{Argmin}_x l(x,y,r)$. We also mention that Definition 1.4.2 does not require nonemptiness of $A(y,c)$, while Definition 4.2.1 requires nonemptiness of $A(y,c)$. In fact the only situation in which $A(y,c)$ is empty, under the compactness assumption on the sub-levels of $f$ (Definition 3.1.2), is when $q(y, c-\varepsilon) = -\infty$ for all $\varepsilon > 0$. It is obvious that $A(y,c) = \emptyset$ when $q(y,c) = -\infty$. Also, $q(y,c) = -\infty$ implies that $q(y, c-\varepsilon) = -\infty$ for all $\varepsilon > 0$, by monotonicity of $q(y,\cdot)$. Example 4.2.3 illustrates a situation in which $q(y,y) > -\infty$, $A(y,y) = \emptyset$ for any $y \in \mathbb{R}$, and $q(y, c-\varepsilon) = -\infty$ for all $\varepsilon > 0$. As a consequence of the next lemma we have $A(y,c) \neq \emptyset$ when $(y,c) \in \operatorname{int}(\operatorname{dom} q)$. We use the notation $B(y,\varepsilon) := \{z \in H : \|z-y\| < \varepsilon\}$. Recall that $\Phi_{(u,t)}(x,z) := f(x,z) - \langle u, z\rangle + t\sigma(z)$. The next lemma is similar to Lemma 3.2.2. We assume from now on in this section that $\sigma(\cdot) \geq \|\cdot\|$.

**Lemma 4.2.4.** *Let $(y, c_y) \in H \times \mathbb{R}_{++}$ be such that $(y, c_y)$ belongs to* $\operatorname{int}(\operatorname{dom} q)$*. Then for each $b \geq M_p$ there exist a weak compact set $B \subset X \times H$ and $\varepsilon > 0$ such that*

$$L_b(w, c) := \{(x, z) \in X \times H : \Phi_{(w,c)}(x, z) \leq b\} \subset B$$

*for all $(w, c) \in B(y, \varepsilon) \times (c_y - \varepsilon, c_y + \varepsilon)$. As a consequence, for each $(w, c) \in B(y, \varepsilon) \times (c_y - \varepsilon, c_y + \varepsilon)$ there exists some $(\tilde{x}, \tilde{z}) \in A(w, c)$, that is,*

$$q(w, c) = f(\tilde{x}, \tilde{z}) - \langle \tilde{z}, w \rangle + c\sigma(\tilde{z}).$$

*Proof.* Since $(y, c_y)$ belongs to $\operatorname{int}(\operatorname{dom} q)$, there exists $0 < r < c_y$ such that $B(y, r) \times [c_y - r, c_y + r] \subset \operatorname{dom} q$. Take $\varepsilon := \frac{r}{3}$. Let us show that there exists a weak compact set $B \subset X \times H$ such that $L_b(u, t) \subset B$ for all $(u, t) \in B(y, \varepsilon) \times [c_y - \varepsilon, c_y + \varepsilon]$. Suppose by contradiction that this is not true. Therefore there exist a sequence $\{(y_k, c_k)\} \subset B(y, \varepsilon) \times [c_y - \varepsilon, c_y + \varepsilon]$ and a sequence $\{(x_k, z_k)\}$ satisfying $\lim_k \|(x_k, z_k)\| = \infty$ and $(x_k, z_k) \in L_b(y_k, c_k)$ for all $k \in \mathbb{N}$. Let $d := c_y - r > 0$, and observe that $(y, d) \in \operatorname{dom} q$. It follows that

$$\Phi_{(y_k,c_k)}(x_k, z_k) = f(x_k, z_k) - \langle y_k, z_k \rangle + c_k\sigma(z_k) \leq b \text{ for each } k, \tag{4.7}$$

which implies

$$\begin{aligned} b \geq \Phi_{(y_k,c_k)}(x_k, z_k) &= \Phi_{(y,d)}(x_k, z_k) + \langle y - y_k, z_k \rangle + (c_k - d)\sigma(z_k) \\ &\geq q(y, d) - \|y_k - y\|\|z_k\| + 2\varepsilon\sigma(z_k) \\ &\geq q(y, d) - \varepsilon\|z_k\| + 2\varepsilon\|z_k\| \\ &= q(y, d) + \varepsilon\|z_k\| \end{aligned} \tag{4.8}$$

where in the second inequality we use Cauchy-Schwarz inequality and the fact that $c_k \geq c_y - \varepsilon = d + 2\varepsilon$, and in the last inequality we use the fact that $\sigma(\cdot) \geq \|\cdot\|$ and $y_k \in B(y, \varepsilon)$. Let $a := \frac{b - q(y,d)}{\varepsilon}$. Therefore we obtain from (4.8) that $\{z_k\} \subset B(0, a)$. Thus there exists a subsequence $\{z_{k_j}\}$ weakly convergent to some $z$. Take $\alpha := b + a(\|y\| + \varepsilon)$. By the compactness assumption on the sublevel of $f$ (see Definition 3.1.2), we have that there exist a weak neighborhood $W$ of $z$ and a weak compact set $B$ such that $L_{f,W}(\alpha) := \{x : f(x,u) \leq \alpha\} \subset B$ for all $u \in W$. In particular, since $\{z_{k_j}\}$ is weakly convergent to $z$, we obtain that $z_{k_j} \in W$ for all $j$ sufficiently large. Observing that $\sigma \geq 0$ and $\{y_k\} \subset B(y, \varepsilon)$, we obtain from (4.7) that $f(x_{k_j}, z_{k_j}) \leq \alpha$ for all $j$. In particular $x_{k_j} \in B$ for all $j$ sufficiently large. Therefore $\{x_{k_j}\}$ is bounded. Hence $\{(x_{k_j}, z_{k_j})\}$ is bounded, which is a contradiction with the fact that $\lim_k \|(x_k, z_k)\| = \infty$. Since each function composing $\Phi_{(u,t)}(\cdot,\cdot)$ is w-lsc, the last statement follows by the first part already proved and the fact that every w-lsc function assumes its minimum on a weak compact set. $\square$

The next theorem shows the equivalence between the differentiability of the dual function $q$ at $(y, E(y))$ and the strong exactness property of $E$.

**Theorem 4.2.5.** *Let $y \in H$ be such that $E(y) < +\infty$ and $(y, E(y)) \in$ $\operatorname{int}(\operatorname{dom} q)$. Then the dual function $q$ is differentiable at $(y, E(y))$ iff $E$ is strongly exact at $y$.*

*Proof.* Suppose that $q$ is differentiable at $(y, E(y))$. We need to prove that $P_H(y, E(y)) = \{0\}$. We get from (4.5) that $0 \in P_H(y, E(y))$, because $(y, E(y)) \in D_*$. Let us show that 0 is the only element in $P_H(y, E(y))$. In order to show this, take an arbitrary $z \in P_H(y, E(y))$. Thus, there exists $x \in X$ such that $(x, z) \in A(y, E(y))$. In particular $(-z, \sigma(z)) \in \partial q(y, E(y))$ by Proposition 3.2.1. On the other hand, $(0, 0) \in \partial q(y, E(y))$ because $(y, E(y))$ maximizes the concave function $q$. Since $q$ is differentiable at $(y, E(y))$, we conclude that $\partial q(y, E(y)) = \{(-z, \sigma(z))\} = \{(0, 0)\} \subset H \times \mathbb{R}$. Therefore $z = 0$ and then $P_H(y, E(y)) = \{0\}$, which proves the "only if" statement of the theorem. In order to prove the "if" statement, suppose that $E$ is strongly exact at $y$. By Lemma 4.2.4 we have that there exists a weak compact set $B$

such that

$$\emptyset \neq \{(x,z) : \Phi_{(w,c)}(x,z) \leq M_p\} \subset B,$$

for all $(w,c) \in B(y,\varepsilon) \times (E(y)-\varepsilon, E(y)+\varepsilon)$ and some $0 < \varepsilon < E(y)$. It follows that

$$q(w,c) = \min\{\Phi_{(w,c)}(x,z) : (x,z) \in B\}.$$

Note that for each $(x,z) \in B$ the function $\Phi_{(w,c)}(x,z)$ (as function of $(w,c)$) is an affine function and its derivative at $(w,c)$ is $(-z,\sigma(z))$. Therefore, we obtain from Theorem 1.7.1 that the superdifferential of $q$ at $(y,E(y))$ is given by

$$\partial q(y,E(y)) = \overline{\text{co}}^w\{(-z,\sigma(z)) : (x,z) \in A(y,E(y)) \text{ for some } x \in X\}.$$

Since $\{0\} = P_H(y,E(y)) = \{z : (x,z) \in A(y,E(y)) \text{ for some } x \in X\}$, we obtain that $\partial q(y,E(y)) = \overline{\text{co}}^w\{(0,0) \in H \times \mathbb{R}\} = \{(0,0)\}$. Therefore $q$ is differentiable at $(y,E(y))$, and we conclude the proof. $\square$

The next proposition considers a general dual sequence converging to a dual solution. As we have seen in previous chapters, subgradient type methods are suitable methods for generating such a sequence.

**Proposition 4.2.6.** *Consider a dual sequence $\{(y_k,c_k)\}$ converging strongly to $(\bar{y},\bar{c}) \in D_*$. Let $\{(x_k,z_k)\}$ be such that $(x_k,z_k) \in A(y_k,c_k)$, and suppose that $\{z_k\}$ is weakly convergent. If $P(\bar{y},\bar{c}) = P_*$ then each (if any) weak accumulation point of $\{x_k\}$ is a primal solution.*

*Proof.* let $z$ be the weak limit point of $\{z_k\}$ and $\bar{x}$ the weak limit point of some subsequence $\{x_{k_j}\}$. Under the assumption of the proposition, it follows that $\langle y_k, z_k\rangle$ converges to $\langle \bar{y}, z\rangle$. Therefore, using the weak lower semicontinuity of the functions involved in the augmented Lagrangian we obtain

$$\begin{aligned} M_d = q(\bar{y},\bar{c}) &\leq f(\bar{x},z) - \langle \bar{y}, z\rangle + \bar{c}\sigma(z) \\ &\leq \liminf_j f(x_{k_j},z_{k_j}) - \langle y_{k_j}, z_{k_j}\rangle + c_{k_j}\sigma(z_{k_j}) \\ &= \liminf_j q(y_{k_j},c_{k_j}) \leq M_d. \end{aligned}$$

In particular we have $(\bar{x}, z) \in A(\bar{y}, \bar{c})$ and then $\bar{x} \in P(\bar{y}, \bar{c})$. Since $P_* = P(\bar{y}, \bar{c})$, we conclude that $\bar{x}$ is a primal solution. □

**Remark 4.2.7.** Proposition 4.2.6 assumes that $\{y_k\}$ is strongly convergent. If we have that $\{y_k\}$ converges weakly and $\{z_k\}$ converges strongly, the result is still true with a similar proof. In fact, we only use the strong convergence to ensure that $\langle y_k, z_k\rangle$ converges to some $\langle y, z\rangle$. We also recall that the modified subgradient algorithm considered in Chapter 3 generates a dual sequence $\{(y_k, c_k)\}$ which is strongly convergent to a dual solution.

### 4.2.1 Properties of the penalty map

We will see that some properties of the function $\Phi_{(y,c)}(x, z)$ (defined in (4.4)) imply some interesting properties on $E$ and the dual solution set. The following estimate will be useful in the sequel. For each $y, w \in H$ and $c \in \mathbb{R}_+$,

$$\Phi_{(y,c)}(x, z) \le \Phi_{(w,c+\|w-y\|)}(x, z), \quad \text{for all} \quad (x, z) \in X \times H. \tag{4.9}$$

Indeed,

$$\begin{aligned} \Phi_{(y,c)}(x, z) &= f(x, z) - \langle y, z\rangle + c\sigma(z) \\ &= f(x, z) - \langle w, z\rangle + (c + \|w - y\|)\sigma(z) \\ &+ \langle w - y, z\rangle - \|w - y\|\sigma(z), \end{aligned}$$

now using the definition of $\Phi$ and Cauchy-Schwarz inequality we obtain

$$\Phi_{(y,c)}(x, z) \le \Phi_{(w,c+\|w-y\|)}(x, z) + \|w - y\|(\|z\| - \sigma(z)),$$

and the result follows because $\sigma(\cdot) \ge \|\cdot\|$.

Next we consider some results related to the dual solution set, extending Proposition 2.1.1 to the general setting of this chapter.

**Proposition 4.2.8.** *Take $(y, c^*) \in D_*$. Then $(w, c^* + \|w - y\|) \in D_*$ for all $w \in H$.*

*Proof.* Using that $(y, c^*) \in D_*$ and inequality (4.9) we obtain

$$\begin{aligned} \bar{q} = q(y, c^*) &= \inf_{(x,z)} \Phi_{(y,c^*)}(x, z) \\ &\leq \inf_{(x,z)} \Phi_{(w,c^*+\|w-y\|)}(x, z) \\ &= q(w, c^* + \|w - y\|) \leq \bar{q}. \end{aligned}$$

Therefore all the inequalities are equalities and $(w, c^* + \|w - y\|) \in D_*$, completing the proof. $\square$

**Corollary 4.2.9.** *Take* $(y, c^*) \in D_*$, *and* $\rho > 0$. *Then*

$$\Delta_\rho := \{(w, c) \in H \times \mathbb{R}_+ : c \geq c^* + \rho, \text{ and } \|w - y\| \leq \rho\} \subset D_*.$$

*Proof.* The result follows directly from Propositions 4.2.8 and 3.1.4 (ii). $\square$

Consider the mapping $T : H \times H \to \mathbb{R} \cup \{+\infty\}$ defined by

$$T(y, w) = E(y) + \|w - y\|.$$

We mention that the dual optimal solution set $D_*$ is convex, because the dual function $q$ is concave. In particular, $E$ is a convex function, because it is a lower envelope of a convex set. It follows that $T$ is a convex function and $T(y, y) = E(y)$ for all $y \in H$. Next we relate this function with the results proved above and then we prove that $E$ is a Lipschitz continuous mapping.

**Corollary 4.2.10.** *If* $D_*$ *is nonempty then* $E(w) < \infty$ *for all* $w \in H$. *Also* $(0, t) \in D_*$ *for all* $t \geq T(w, 0)$.

*Proof.* Consider $(y, c^*) \in D_*$ and fix $w \in H$. It follows from Proposition 4.2.8 that $(w, c^* + \|w - y\|)$ is a dual solution. In particular, $E(w) \leq c^* + \|w - y\| < \infty$, by definition of $E$. The last statement follows from Corollary 4.2.9, by taking $\rho := \|w\|, c^* := E(w)$ and using the definition of $T(w, 0)$. $\square$

**Corollary 4.2.11.** *If* $E(y) < \infty$ *for some* $y \in H$ *then* $(w, T(y, w)) \in D_*$ *for all* $w \in H$. *As a consequence,* $E(w) \leq T(y, w)$ *for all* $y, w \in H$ *and* $E$ *is a Lipschitz continuous mapping with Lipschitz constant* 1, *that is, a nonexpansive mapping.*

*Proof.* Take $y \in H$ such that $E(y) < \infty$. It follows from upper semicontinuity of $q$ and definition of $E$ that $(y, E(y)) \in D_*$. Take an arbitrary $w \in H$. Using the definition of $T(y, w)$ and taking $c^* = E(y)$ we obtain from Proposition 4.2.8 that $(w, T(y, w)) \in D_*$. Therefore, by definition of $E$ we obtain that $E(w) \leq T(y, w)$. In particular $E(w) < \infty$ for all $w \in H$ and the inequality $E(w) \leq T(y, w)$ yields $E(w) - E(y) \leq \|y - w\|$. Since $E(w) < \infty$ for all $w \in H$, the last conclusion is true for arbitrary $y, w \in H$. Therefore $|E(y) - E(w)| \leq \|y - w\|$ for any $y, w \in H$, concluding the corollary. $\square$

**Remark 4.2.12.** It follows from Corollary 4.2.11 that either $\operatorname{dom} E = \emptyset$ or $\operatorname{dom} E = H$.

## 4.3 Exact penalty map and modified subgradient algorithm

In this section we show how the results above relate to the convergence properties of MSg (exact version of IMSg). Next we recall MSg in the framework of this chapter.

**Step** 0. Choose $(y_0, c_0) \in H \times \mathbb{R}_+$ such that $q(y_0, c_0) > -\infty$, and parameters $\{\alpha_k\} \subset [\alpha, \hat{\alpha}]$ for some $\hat{\alpha} > \alpha > 0$. Let $k := 0$.

**Step** 1. (Subproblem and stopping criterion)

$a)$ Find $(x_k, z_k) \in A(y_k, c_k)$,

$b)$ if $z_k = 0$ then stop,

$c)$ if $z_k \neq 0$ go to Step 2.

**Step** 2. (Selection of the stepsize and updating the variables)

Choose a stepsize $s_k$ and update the dual variables,

$y_{k+1} := y_k - s_k z_k$,

$c_{k+1} := c_k + (\alpha_k + 1) s_k \sigma(z_k)$,

$k := k + 1$, go to Step 1.

We recall for further use the following estimates for the dual sequence generated by MSg:

$$\|y_{k+j} - y_k\| \leq \sum_{l=k}^{k+j-1} \|y_{l+1} - y_l\| = \sum_{l=k}^{k+j-1} s_l \|z_l\| \leq \sum_{l=k}^{k+j-1} s_l \sigma(z_l).$$

$$c_{k+j} - c_k = \sum_{l=k}^{k+j-1} c_{l+1} - c_l = \sum_{l=k}^{k+j-1} (\alpha_l + 1) s_l \sigma(z_l).$$

As a consequence of theses estimates we obtain for all $k, j \in \mathbb{N}$,

$$c_{k+j} - c_k \geq (1 + \alpha)\|y_{k+j} - y_k\|. \tag{4.10}$$

The next proposition relates exactness property of the penalty map $E$ and convergence of MSg. We mention that no specific stepsize selection rule is assumed.

**Proposition 4.3.1.** *Let* $\{(x_k, z_k)\}$ *and* $\{(y_k, c_k)\}$ *be generated by* MSg. *Suppose that* $\{(y_k, c_k)\}$ *converges to some* $(\bar{y}, \bar{c}) \in D_*$*, and that there exists a subsequence* $\{\sigma(z_{k_j})\}$ *converging to zero. Then the penalty map* $E$ *is strongly exact at* $\bar{y}$*.*

*Proof.* We need to prove that $P_H(\bar{y}, \bar{c}) = \{0\}$. It follows from Proposition 4.1.1(a) that $0 \in P_H(\bar{y}, \bar{c})$. Take an arbitrary $\bar{z} \in P_H(\bar{y}, \bar{c})$ and suppose for the sake of contradiction that $\bar{z} \neq 0$. In particular, $\sigma(\bar{z}) > 0$ and since $\{\sigma(z_{k_j})\}$ converges to 0, we can consider $j_0$ such that

$$\sigma(z_{k_j}) \leq \frac{\alpha\sigma(\bar{z})}{2(2 + \alpha)} < \sigma(\bar{z}) \text{ for all } j \geq j_0, \tag{4.11}$$

where $\alpha$ is given in Step 0 of MSg. It follows from Proposition 3.2.1(i) that $(-\bar{z}, \sigma(\bar{z})) \in \partial q(\bar{y}, \bar{c})$ and $(-z_k, \sigma(z_k)) \in \partial q(y_k, c_k)$. Therefore, we obtain by antimonotonicity of $\partial q$ that

$$\langle(-\bar{z}, \sigma(\bar{z})) - (-z_{k_j}, \sigma(z_{k_j})), (\bar{y}, \bar{c}) - (y_{k_j}, c_{k_j})\rangle \leq 0,$$

which is equivalent to

$$\langle z_{k_j} - \bar{z}, \bar{y} - y_{k_j}\rangle + (\sigma(\bar{z}) - \sigma(z_{k_j}))(\bar{c} - c_{k_j}) \leq 0.$$

Using Cauchy Schwarz inequality we obtain

$$-\|\bar{z}-z_{k_j}\|\|\bar{y}-y_{k_j}\|+(\sigma(\bar{z})-\sigma(z_{k_j}))(\bar{c}-c_{k_j})\leq 0.$$

The estimate (4.10) implies that $\bar{c}-c_{k_j}\geq(1+\alpha)\|\bar{y}-y_{k_j}\|$. Therefore

$$-\|\bar{z}-z_{k_j}\|\|\bar{y}-y_{k_j}\|+(\sigma(\bar{z})-\sigma(z_{k_j}))(1+\alpha)\|\bar{y}-y_{k_j}\|\leq 0.$$

Using triangle inequality and the fact that $\sigma(\cdot)\geq\|\cdot\|$ we obtain

$$\left[-\sigma(\bar{z})-\sigma(z_{k_j})+(1+\alpha)(\sigma(\bar{z})-\sigma(z_{k_j}))\right]\|\bar{y}-y_{k_j}\|\leq 0$$

and then

$$\left[\alpha\sigma(\bar{z})-(2+\alpha)\sigma(z_{k_j})\right]\|\bar{y}-y_{k_j}\|\leq 0. \tag{4.12}$$

In particular, using (4.11) in (4.12), we obtain for $j\geq j_0$ that

$$0<\frac{\alpha}{2}\sigma(\bar{z})\|\bar{y}-y_{k_j}\|\leq 0,$$

which is a contradiction. The proof is complete. □

The following result fully characterizes convergence of MSg in terms of the map $E$.

**Corollary 4.3.2.** *Let $\{(x_k,z_k)\}$ and $\{(y_k,c_k)\}$ be bounded sequences generated by* MSg. *Suppose that $\{(y_k,c_k)\}$ converges to $(\bar{y},\bar{c})\in D_*\cap\operatorname{int}(\operatorname{dom}q)$. The following statements are equivalent:*

*a) there exists a subsequence of $\{\sigma(z_k)\}$ converging to 0;*

*b) the sequence $\{\sigma(z_k)\}$ converges to 0;*

*c) the dual function $q$ is differentiable at $(\bar{y},\bar{c})$;*

*d) the penalty map $E$ is strongly exact at $\bar{y}$.*

*Moreover, under any of these statements as assumption, the sequence $\{z_k\}$ converges strongly to 0 and every accumulation point of $\{x_k\}$ is a primal solution.*

*Proof.* By Theorem 4.2.5 we have the equivalence between (c) and (d). It is obvious that (b) implies (a). Proposition 4.3.1 shows that (a) implies (d). Therefore for proving the corollary we just need to show that (d) implies (b). An argument similar to the one used in the proof of Proposition 4.2.6 shows that any weak accumulation point of $\{z_k\}$ belongs to $P_H(\bar{y}, \bar{c})$. Since we are assuming (d), we have $P_H(\bar{y}, \bar{c}) = \{0\}$ and then we obtain that $\{z_k\}$ converges weakly to 0. We observe that $\sigma$ is just w-lsc (our prototypical of augmenting function is $\sigma(\cdot) = \|\cdot\|$, which is w-lsc but not weakly continuous). Therefore we cannot conclude yet that (b) holds. Consider the following sequences

$$m_k := f(x_k, z_k) - \langle y_k, z_k\rangle \ \text{ and } \ q_k := m_k + c_k\sigma(z_k), \quad \text{for all } \ k.$$

Since $\{y_k\}$ is strongly convergent to $\bar{y}$ and $\{z_k\}$ is weakly convergent to 0, we have that $\{\langle y_k, z_k\rangle\}$ converges to 0. We know that $\{q_k\}$ converges to $M_d$. Let us show now that $\{m_k\}$ also converges to $M_d$.

Fix a subsequence $\{m_{k_j}\}$ convergent to $m := \liminf m_k$. Take a subsequence $\{x_{k_{j_n}}\}$ weakly convergent to some $\hat{x}$. In particular $\{m_{k_{j_n}}\}$ converges to $m$. From weak lower semicontinuity of $f$ and the fact that $f(\cdot, 0) = \varphi(\cdot)$, we have

$$\begin{aligned} M_p \le \varphi(\hat{x}) &\le \liminf_n f(x_{k_{j_n}}, z_{k_{j_n}}) - \langle y_{k_{j_n}}, z_{k_{j_n}}\rangle = \lim_n m_{k_{j_n}} \\ &= \liminf_k m_k \le \limsup_k m_k \\ &\le \limsup_k q_k = M_p. \end{aligned}$$

Therefore $\{m_k\}$ converges to $m = M_p$. Since $\{q_k\}$ also converges to $M_p$ we obtain that

$$0 \le \sigma(z_k) = \frac{q_k - m_k}{c_k} \xrightarrow{k} \frac{M_p - M_p}{\bar{c}} = 0.$$

That is, $\{\sigma(z_k)\}$ converges to 0, and then (b) holds. We also obtain that $\{z_k\}$ converges strongly to 0, because we are assuming that $\sigma(\cdot) \ge \|\cdot\|$. Taking an arbitrary subsequence $\{x_{k_l}\}$ weakly convergent to some $\bar{x}$ we obtain

$$\varphi(\bar{x}) \le \liminf_l f(x_{k_l}, z_{k_l}) - \langle y_{k_l}, z_{k_l}\rangle = \lim_l m_{k_l} = M_p,$$

and then $\bar{x}$ is a primal solution, proving that every weak accumulation point of $\{x_k\}$ is a primal solution. The proof is complete. □

**Remark 4.3.3.** From the proof of Corollary 4.3.2 we see that if $\{z_k\}$ is weakly convergent then it is strongly convergent.

**Lemma 4.3.4.** *Consider a sequence $\{(u_k, t_k)\}$ converging to some $(\bar{u}, \bar{t}) \in D_*$, with $\bar{t} > 0$. Assume that $P_H(u_k, t_k) \neq \{0\}$ for all $k$. Then $\bar{t} = E(\bar{u})$.*

*Proof.* Take $d := \bar{t} - E(\bar{u})$. Since $(\bar{u}, \bar{t}) \in D_*$, $d \geq 0$ by definition of $E$. In order to prove the lemma we need to show that $d = 0$. Suppose for the sake of contradiction that $d > 0$. By Proposition 4.2.8 we know that $(u, t) \in D_*$ for each $u \in H$ and $t \geq E(\bar{u}) + \|u - \bar{u}\|$. Take $\bar{k}$ such that $\|u_k - \bar{u}\| \leq \frac{d}{3}$ and $t_k \geq \bar{t} - \frac{d}{3}$ for all $k \geq \bar{k}$. It follows that

$$t_k \geq \bar{t} - \frac{d}{3} = E(\bar{u}) + \frac{2d}{3} \geq E(\bar{u}) + \|u_k - \bar{u}\| + \frac{d}{3}. \tag{4.13}$$

Therefore, Proposition 4.2.8 implies that $(u_k, t_k) \in D_*$. It follows from (4.13) and Proposition 4.1.1 that $P_H(u_k, t_k) = \{0\}$ for all $k \geq \bar{k}$, contradicting the hypothesis. Thus $d = 0$ and the result follows. □

Example 2.4.1 of Chapter 2 and [12, Example 1] show that MSg with a Polyak stepsize can fail to obtain a primal solution via accumulation points of the primal sequence. In [12] an auxiliary sequence $\{\tilde{x}_k\}$ such that all its accumulation points are primal solutions was considered for an equality constrained problem in finite dimensional spaces, see Theorem 1.6.6 of this thesis. In order to obtain this sequence one needs to solve two subproblems. Next we recall how this auxiliary sequence is obtained and we show that it has finite convergence, that is, there exists $\bar{k} > 0$ such that $\tilde{x}_{\bar{k}} \in P_*$. We consider the auxiliary sequence in the general framework of augmented Lagrangian considered here. When the primal problem is an equality constrained problem and we use a canonical parameterization function (Example 3.1.1), this sequence is the one considered in [12].

Consider $\{(y_k, c_k)\}$ generated by MSg, and fix $\beta > 0$. Let $\{\tilde{x}_k\}$ be an auxiliary primal sequence satisfying $\tilde{x}_k \in P(y_k, c_k + \beta)$ for all $k$.

**Proposition 4.3.5.** *Suppose that $D_*$ is nonempty, and assume that MSg generates an infinite dual sequence $\{(y_k, c_k)\}$ convergent to a dual solution $(\bar{y}, \bar{c})$. Fix $\beta > 0$ and consider the auxiliary sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k \in P(y_k, c_k + \beta)$ for all $k$. Then there exists $\bar{k}$ such that $P_H(y_{\bar{k}}, c_{\bar{k}} + \beta) = \{0\}$. In particular $\tilde{x}_{\bar{k}}$ is a primal solution.*

*Proof.* If the conclusion is not true, we have $P_H(y_k, c_k + \beta) \neq \{0\}$ for each $k$. Use now Lemma 4.3.4 for the choice $(u_k, t_k) = (y_k, c_k + \beta)$, which converges to $(\bar{y}, \bar{c} + \beta) \in D_*$, to obtain

$$E(\bar{y}) = \bar{c} + \beta > \bar{c}. \tag{4.14}$$

On the other hand, since $(\bar{y}, \bar{c}) \in D_*$ we have $E(\bar{y}) \leq \bar{c}$ by definition of $E$, which is a contradiction with (4.14). The last statement follows from Proposition 4.1.1. □

## 4.4 Exact penalty map and equality constrained problems

Consider the equality constrained optimization problem $(P)$ of Chapter 2. Let us denote by

$$\Gamma(y, c) := \{x \in K : L(x, y, c) := \psi(x) - \langle y, h(x)\rangle + c\|h(x)\| = q(y, c)\}.$$

In this context, the sets $A(y, c)$ and $P(y, c)$ defined in Section 4.1 become $A(y, c) = \{(x, h(x)) : x \in \Gamma(y, c)\}$ and $P(y, c) = \Gamma(y, c)$. In particular, $P_H(y, c) = \{0\} \Leftrightarrow h(x) = 0$ for every $x \in \Gamma(y, c)$. Therefore we have that $P_H(y, c) = \{0\} \Leftrightarrow \Gamma(y, c) = P_*$. In particular, in the context of equality constrained optimization problem, the dual penalty map $E$ is exact if and only if it is strongly exact. As a consequence of these facts we obtain the following theorem. We remark that we do not assume any specific rule for the stepsize in MSg.

**Theorem 4.4.1.** *Consider the sequences* $\{(y_k, c_k)\}$ *and* $\{x_k\}$ *generated by* MSg *applied to the dual problem* $(D)$ *of Chapter 2. Suppose that* $\{(y_k, c_k)\}$ *converges to* $(\bar{y}, \bar{c}) \in D_*$. *Then the following statements are equivalent:*

*a) the dual function* $q$ *is differentiable at* $(\bar{y}, \bar{c})$*;*

*b) the sequence* $\{h(x_k)\}$ *converges to* $0$*;*

*c) the dual penalty map* $E$ *is exact at* $\bar{y}$*;*

*d) all weak accumulation points of* $\{x_k\}$ *are primal solutions.*

*Proof.* Observing that $\sigma(z) = \|z\|$ for all $z \in \mathbb{R}^m$ and $z_k = h(x_k)$ for all $k$, the statements $(a), (b)$ and $(c)$ are equivalent by Corollary 4.3.2. The equivalence between $(b)$ and $(d)$ follows easily from the continuity of $h$. $\square$

# Chapter 5

# General augmented Lagrangian

In this chapter, we consider a primal problem of minimizing an extended real-valued function in a Hausdorff topological space. A main tool in our analysis is abstract convexity, which recently became a natural language to investigate duality-schemes via augmented Lagrangian type functions, see for example Burachik and Rubinov [17], Nedić et al. [41], Penot and Rubinov [44], Rubinov and Yang [56], and Rubinov et al. [58]. With abstract convexity tools, we propose and analyze a duality scheme induced by a general augmented Lagrangian function. We consider a valley at zero type property on the coupling (augmenting) function, which generalizes the valley at zero type property proposed in the related literature (e.g., [17] and references therein, [65, Section 3.1]), see Sections 1.5 and 5.4. We show that our duality scheme has a zero duality gap property. A sub-optimal path related to the dual problem is considered, and we prove that all its cluster points are primal solutions. A criterion for exact penalization was presented in Rockafellar and Wets [54, Theorem 11.61]. We also extend this criterion to our general setting. Since no linearity on the augmented Lagrangian is assumed, this allows us to consider our primal-dual scheme in Hausdorff topological spaces. It is also worthwhile to note that the general augmented Lagrangian considered here, for which the valley at zero type property is assumed directly at the coupling function $\rho$ (see Section 5.1), has not been considered in the literature even in finite dimensional spaces.

The outline of this chapter is as follows. Section 5.1 contains basic definitions and assumptions. Also, our primal-dual scheme is stated. In Section 5.2 we show that our duality scheme provides strong duality, and a criterion to exact penalty representation is presented. In Section 5.3 we study the convergence properties of a sub-optimal path related to our dual problem. In Section 5.4 we present some examples and compare our setting with the ones considered in [17] and [65, Section 3.1].

## 5.1 Statement of the problem and basic assumptions

Let $Y$ be an arbitrary (nonempty) set. Let $X$ and $Z$ be Hausdorff topological spaces. We consider the optimization problem

$$\text{minimize} \quad \varphi(x) \quad \text{subject to } x \text{ in } X, \tag{5.1}$$

where the function $\varphi : X \to \mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ is a proper extended real-valued function. We fix a base point in $Z$ and denote it by 0. We use a duality parameterization for $\varphi$, which is a function $f : X \times Z \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ satisfying $f(x, 0) = \varphi(x)$ for all $x \in X$. We also consider a perturbation function $\beta : Z \to \bar{\mathbb{R}}$, related to this duality parameterization, given by

$$\beta(z) := \inf_{x \in X} f(x, z).$$

The properness of $\varphi$ implies that $\beta(0) < +\infty$.

In what follows, we consider a coupling function $\rho : Z \times Y \times \mathbb{R}_+ \to \mathbb{R}$ that satisfies the following basic assumptions:

$C_1$) For any $(y, r) \in Y \times \mathbb{R}_+$ the function $\rho(\cdot, y, r)$ is upper semicontinuous at 0, and $\rho(0, y, r) = 0$.

$C_2$) For every neighborhood $V \subset Z$ of 0, and for every $(y, \overline{r}) \in Y \times \mathbb{R}_+$, it holds that

i) $A^V_{y,\overline{r}}(r) := \inf\limits_{z \in V^c} \{\rho(z, y, \overline{r}) - \rho(z, y, r)\} > 0, \;\text{ for all }\; r > \overline{r};$

ii) $\lim\limits_{r \to \infty} A^V_{y,\overline{r}}(r) = \infty.$

**Remark 5.1.1.** Condition $C_1$ is a basic assumption considered in the related literature. Condition $C_2$ is a valley at zero type property, which generalizes similar properties for augmenting functions recently introduced in the literature. Item (i) in condition $C_2$ ensures that the function $\rho(z, y, \cdot)$ is strictly decreasing for any fixed $(y, z) \in Y \times Z \setminus \{0\}$. In particular, the function $A^V_{y,\overline{r}} : (\overline{r}, \infty) \to \mathbb{R}_+$ is nondecreasing, ensuring the existence of $\lim\limits_{r \to \infty} A^V_{y,\overline{r}}(r)$. In Section 5.4 we compare condition $C_2$ with related assumptions in the literature.

The *augmented Lagrangian function* $\ell : X \times Y \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ induced by the coupling function $\rho$ is defined by

$$\ell(x, y, r) := \inf_{z \in Z} \{f(x, z) - \rho(z, y, r)\}. \tag{5.2}$$

The dual function $q : Y \times \mathbb{R}_+ \to \bar{R}$ is defined as $q(y, r) := \inf_{x \in X} \ell(x, y, r)$ and therefore the dual problem is stated as

$$\text{maximize } \;q(y, r)\; \text{ subject to } (y, r) \text{ in } Y \times \mathbb{R}_+. \tag{5.3}$$

It is clear that $q(y, r) = \inf_{z \in Z}\{\beta(z) - \rho(z, y, r)\}$, where $\beta$ is the perturbation function. We denote by $M_p := \inf_{x \in X} \varphi(x)$ the optimal value of the primal problem, and by $M_d := \sup_{(y,r) \in Y \times \mathbb{R}_+} q(y, r)$ the optimal value of the dual problem.

Since $f$ is a parameterization function, condition $C_1$ easily implies the weak duality property for our scheme, that is, $M_d \leq M_p$. In this section we show that our duality scheme enjoys a strong duality property, that is to say, the zero duality gap property holds ($M_p = M_d$).

Next, for the convenience of the reader, we recall some of the definitions related to abstract convexity presented in Chapter 1. Let $\Omega = Y \times \mathbb{R}_+$ and $g : Z \to \bar{\mathbb{R}}$. Given $\varepsilon \geq 0$, an element $(y, r)$ is an $\varepsilon-$*abstract subgradient of* $g$ in $\overline{z}$ (with respect to $\rho$) iff

$$g(z) - \rho(z, y, r) \geq g(\overline{z}) - \rho(\overline{z}, y, r) - \varepsilon \text{ for all } z \in Z. \tag{5.4}$$

The set of all $\varepsilon-$subgradients of $g$ at $\bar{z}$ is called $\varepsilon-$subdifferential of $g$ at $\bar{z}$, and it is denoted by $\partial_{\rho,\varepsilon} g(\bar{z})$.

**Remark 5.1.2.** It follows from $C_1$ and the definition of $\partial_{\rho,\varepsilon} g(0)$, that if $(y, r_0) \in \partial_{\rho,\varepsilon} g(0)$ then $(y, r) \in \partial_{\rho,\varepsilon} g(0)$ for all $r \geq r_0$, using the fact that $\rho(z, y, \cdot)$ is decreasing and $\rho(0, y, r) = 0$.

We recall that $\beta(0) = M_p$ and $\beta^{\rho\rho}(0) = M_d$. In particular weak and strong duality are rewritten, respectively, as

$$\beta^{\rho\rho}(0) \leq \beta(0) \quad \text{and} \quad \beta^{\rho\rho}(0) = \beta(0).$$

In this context strong duality is related to abstract convexity at 0 of the perturbation function $\beta$ with respect to the family elementary functions

$$H_\rho := \{\rho(\cdot, y, r) + c : (y, r, c) \in Y \times \mathbb{R}_+ \times \mathbb{R}\}.$$

Recall that the support of $\beta$ with respect to $H_\rho$ is the set

$$\mathrm{Supp}(\beta, H_\rho) := \{h \in H_\rho : h \leq \beta\}.$$

We mention that Proposition 1.5.2 relates the sets $\partial_\rho \beta(0)$, $\mathrm{Supp}(\beta, H_\rho)$, and $\mathrm{dom}\, \beta^\rho$ to the dual function $q$.

## 5.2 Strong duality and exact penalty representation

The next theorem ensures that, under mild assumptions, for every $\varepsilon > 0$ the $\varepsilon$-abstract subgradient of $\beta$ at 0 is nonempty. As a consequence of this fact, we establish strong duality. In Example 5.3.5 we consider a constrained optimization problem for which the hypothesis of the next theorem, regarding lower semicontinuity of $\beta$ at 0, is satisfied. Proposition 5.3.4 presents some conditions on the parameterization function that guarantee the lower semicontinuity of $\beta$ at 0.

**Theorem 5.2.1.** *Assume that $C_1$ and $C_2$ hold, that $\beta$ is lower semicontinuous* (lsc) *at* $0$*, and that there exists $(\overline{y}, \overline{r}) \in \operatorname{dom} \beta^\rho$. Then $\partial_{\rho,\varepsilon}\beta(0) \neq \emptyset$ for all $\varepsilon > 0$. Moreover, for any $\varepsilon > 0$ there exists $r_0$ such that $(\overline{y}, r) \in \partial_{\rho,\varepsilon}\beta(0)$ for all $r \geq r_0$.*

*Proof.* The properness of the primal function $\varphi$ implies that $\beta(0) < \infty$. We have that $\beta(0) > -\infty$ by weak duality and the assumption that $\operatorname{dom} \beta^\rho$ is nonempty. Therefore, $\beta(0) \in \mathbb{R}$. Observe that we just need to prove the last statement of the theorem. In order to arrive at a contradiction, suppose that there exists $\bar{\varepsilon} > 0$ such that for any $k > 0$ there exists $r_k \geq k, z_k \in Z$ satisfying:

$$\beta(z_k) - \rho(z_k, \overline{y}, r_k) < \beta(0) - \bar{\varepsilon}. \tag{5.5}$$

Suppose that $\{z_k\}_{k\in\mathbb{N}}$ converges to 0. Thus

$$\beta(0) - \bar{\varepsilon} > \beta(z_k) - \rho(z_k, \overline{y}, r_k) > \beta(z_k) - \rho(z_k, \overline{y}, \overline{r})$$

for all $k \geq k_0 > \overline{r}$. Hence, using $C_1$ and the lower semicontinuity of $\beta$ at 0, we have

$$\beta(0) - \bar{\varepsilon} \geq \liminf_{k\to\infty}\{\beta(z_k) - \rho(z_k, \overline{y}, \overline{r})\} \geq \beta(0) - \rho(0, \overline{y}, \overline{r}) = \beta(0),$$

which is a contradiction. Therefore $\{z_k\}_{k\in\mathbb{N}}$ does not converge to 0, which implies that there exists some open neighborhood $V \subset Z$ of 0, and a subsequence $\{z_{k_j}\}_{j\in\mathbb{N}} \subset V^c$. Now, using (5.5) and the fact that there exists $\overline{c}$ such that $\rho(\cdot,\overline{y},\overline{r})+\overline{c}\in \mathrm{Supp}(\beta,H_\rho)$ (see Proposition 1.5.2), we have

$$\begin{aligned}\beta(0)-\bar{\varepsilon} &> \beta(z_{k_j})-\rho(z_{k_j},\overline{y},r_{k_j})\\ &= \beta(z_{k_j})-\rho(z_{k_j},\overline{y},\overline{r})+\rho(z_{k_j},\overline{y},\overline{r})-\rho(z_{k_j},\overline{y},r_{k_j})\\ &\geq \overline{c}+\inf_{z\in V^c}\{\rho(z,\overline{y},\overline{r})-\rho(z,\overline{y},r_{k_j})\}.\end{aligned}$$

Henceforth,

$$A^V_{\overline{y},\overline{r}}(r_{k_j}) := \inf_{z\in V^c}\{\rho(z,\overline{y},\overline{r})-\rho(z,\overline{y},r_{k_j})\}\leq \beta(0)-\bar{\varepsilon}-\overline{c},$$

which contradicts $C_2$, because $\lim_j r_{k_j}=\infty$. The result follows. $\square$

The next corollary, which extends [17, Proposition 4.2], shows that in order to check if the abstract subgradient of $\beta$ at 0 is nonempty we just need to verify that there exists an element $(y,r)\in Y\times R_+$ satisfying the abstract subgradient inequality (5.4) in a neighborhood of 0. As we will see in Theorem 5.2.5, under mild assumptions, this fact is equivalent to the existence of an exact penalty representation.

**Corollary 5.2.2.** *Suppose that the assumptions of Theorem 5.2.1 hold. Suppose also that there exists an open neighborhood $V\subset Z$ of 0 such that $\beta(z)-\rho(z,\overline{y},\overline{r})\geq\beta(0)$ for all $z\in V$, with $(\overline{y},\overline{r})\in\mathrm{dom}\,\beta^\rho$. Then there exists $r_0$ such that $\beta(z)-\rho(z,\overline{y},r)\geq\beta(0)$ for all $z\in Z$ and $r\geq r_0$, i.e., $(\overline{y},r)\in\partial_\rho\beta(0)$ for all $r\geq r_0$.*

*Proof.* Take $V$ as in the assumption. Consider $z\in V^c$ and $\varepsilon>0$. By Theorem 5.2.1 there exists $r_\varepsilon>0$ such that $(\overline{y},r_\varepsilon)\in\partial_{\rho,\varepsilon}\beta(0)$. Thus

$$\begin{aligned}\beta(z) &\geq \beta(0)+\rho(z,\overline{y},r_\varepsilon)-\varepsilon\\ &= \beta(0)+\rho(z,\overline{y},r)+\rho(z,\overline{y},r_\varepsilon)-\rho(z,\overline{y},r)-\varepsilon \qquad (5.6)\\ &\geq \beta(0)+\rho(z,\overline{y},r)+\inf_{u\in V^c}\{\rho(u,\overline{y},r_\varepsilon)-\rho(u,\overline{y},r)\}-\varepsilon.\end{aligned}$$

By $C_2$ there exists $r_1>r_\varepsilon$ such that $\inf_{u\in V^c}\{\rho(u,\overline{y},r_\varepsilon)-\rho(u,\overline{y},r)\}>\varepsilon$ for all $r\geq r_1$. Using this estimate in (5.6) we obtain $\beta(z)\geq\beta(0)+\rho(z,\overline{y},r)$ for all $z \in V^c$ and $r \geq r_1$. Since, by assumption, $\beta(z) \geq \beta(0) + \rho(z, \overline{y}, \overline{r})$ for all $z \in V$, the result follows by taking $r_0 = \max\{\overline{r}, r_1\}$ and observing that $\rho(z, y, \cdot)$ is nonincreasing in $\mathbb{R}_+$ for each $(z, y) \in Z \times Y$. $\square$

**Theorem 5.2.3.** *Under the assumptions of Theorem 5.2.1, zero duality gap property holds for the primal-dual pair of problems* (5.1)-(5.3).

*Proof.* Take $\varepsilon > 0$. By Theorem 5.2.1, there exists $(\overline{y}, r_\varepsilon) \in \partial_{\rho,\varepsilon}\beta(0)$. Hence we have

$$\begin{aligned} \beta^{\rho\rho}(0) &= \sup_{(y,r)}\{\rho(0, y, r) - \beta^\rho(y, r)\} = \sup_{(y,r)} -\beta^\rho(y, r) \\ &\geq -\beta^\rho(\overline{y}, r_\varepsilon) = \inf_z\{\beta(z) - \rho(z, \overline{y}, r_\varepsilon)\} \geq \beta(0) - \varepsilon, \end{aligned}$$

using $C_1$ in the second equality and the fact that $(\overline{y}, r_\varepsilon) \in \partial_{\rho,\varepsilon}\beta(0)$ in the last inequality. It follows that $\beta^{\rho\rho}(0) \geq \beta(0) - \varepsilon$. Since $\varepsilon$ is arbitrary, we have that $\beta^{\rho\rho}(0) \geq \beta(0)$, and the reverse inequality is the weak duality property. We conclude that $\beta^{\rho\rho}(0) = \beta(0)$, i.e., zero duality gap property holds, completing the proof. $\square$

**Remark 5.2.4.** Corollary 5.2.1 and Theorem 5.2.3 generalize [17, Propositions 4.2 and 4.1], respectively. Observe also that we just use the lower semicontinuity of $\beta$ at 0, while in [17] $\beta$ is assumed to be lsc in all the space.

Exact penalty representation for augmented Lagrangian function was defined and studied by Rockafellar and Wets in [54, Chapter 11]. A criterion for such a representation was presented in [54, Theorem 11.61]; see Theorem 1.4.3. This criterion has been studied for more general augmented Lagrangians, for instance, [17] and [26]. In the next theorem we extend this criterion to our more general setting. We recall now the definition of exact penalty representation.

**Definition 5.2.1.** *Consider the primal and dual problems* (5.1)-(5.3). *An element $\overline{y} \in Y$ is said to* support an exact penalty representation for problem (5.1) *if there exists $r_0 \in \mathbb{R}_+$ such that for any $r > r_0$,*

$E_1$) $\beta(0) = q(\overline{y}, r)$;

$E_2$) $\operatorname*{Argmin}_x \varphi(x) = \operatorname*{Argmin}_x l(x, \overline{y}, r)$.

**Theorem 5.2.5.** *Assume that*

*a) the primal optimal solution set is nonempty, and the parameterization function $f$ satisfies: $f(x, \cdot)$ is* lsc *at* $0$ *for every $x \in X$;*

*b) the perturbation function $\beta$ is lsc at* $0$*;*

*c) conditions $C_1$ and $C_2$ are satisfied;*

*d) there exists $(\overline{y}, \overline{r}) \in \operatorname{dom} \beta^\rho$.*

*Then the following assertions are equivalent:*

*i) There exist an open neighborhood $V \subset Z$ of* $0$ *and $r_0 > 0$ such that*

$$\beta(z) \geq \beta(0) + \rho(z, \overline{y}, r_0) \ \ \textit{for all} \ \ z \in V;$$

*ii) $\overline{y}$ supports an exact penalty representation for problem* (5.1).

*Proof.* First, we prove that $(ii) \Rightarrow (i)$. By $E_1$ there exists $r_0 > 0$ such that $\forall r \geq r_0$

$$\beta(0) = q(\overline{y}, r) = \inf_{z \in Z} \{\beta(z) - \rho(z, \overline{y}, r)\}.$$

In particular, for any open neighborhood $V \subset Z$ of 0, we have

$$\beta(0) \leq \beta(z) - \rho(z, \overline{y}, r) \quad \text{for all} \quad z \in V \ \text{ and } \ r \geq r_0,$$

which proves $(i)$.

Let us now prove that $(i) \Rightarrow (ii)$. By condition $(i)$ and Corollary 5.2.2 we obtain that there exists $r_1$ such that $(\overline{y}, r) \in \partial_\rho \beta(0)$, for all $r \geq r_1$. From Proposition 1.5.2 we conclude that $E_1$ holds for all $r \geq r_1$. Take $r \geq r_1$. We prove now that $E_2$ holds.

$\subset$) Take $x^* \in \operatorname{Argmin}_x \varphi(x)$. Then

$$\begin{aligned} l(x^*, \overline{y}, r) &= \inf_z \{f(x^*, z) - \rho(z, \overline{y}, r)\} \le f(x^*, 0) - \rho(0, \overline{y}, r) = \varphi(x^*) \\ &= \beta(0) = q(\overline{y}, r) = \inf_x l(x, \overline{y}, r), \end{aligned}$$

where the second equality follows from $C_1$ and the fact that $f(x, 0) = \varphi(x)$ for all $x \in X$, and the fourth equality follows from $E_1$ (already proved). From these estimates we obtain that $x^* \in \operatorname{Argmin}_x l(x, \overline{y}, \overline{r})$. Since $x^*$ is arbitrary, we conclude that the announced inclusion holds.

$\supset$) Consider $r > r_1$ and take $x_r \in \operatorname{Argmin}_x l(x, \overline{y}, r)$. We know that $E_1$ holds, and therefore

$$\begin{aligned} \beta(0) &= q(\overline{y}, r) = \inf_x l(x, \overline{y}, r) = l(x_r, \overline{y}, r) \\ &= \inf_z \{f(x_r, z) - \rho(z, \overline{y}, r)\} \\ &= \lim_{k\to\infty} \{f(x_r, z_k) - \rho(z_k, \overline{y}, r)\} \end{aligned} \tag{5.7}$$

for some minimizing sequence $\{z_k\}_{k\in\mathbb{N}}$. We analyze two possible cases:

1) the sequence $\{z_k\}_k$ converges to 0;

2) the sequence $\{z_k\}_k$ does not converge to 0.

In the first case we get from (5.7) that

$$\begin{aligned} \beta(0) &= \lim_{k\to\infty} \{f(x_r, z_k) - \rho(z_k, \overline{y}, r)\} \\ &= \liminf_{k\to\infty} \{f(x_r, z_k) - \rho(z_k, \overline{y}, r)\} \\ &\ge f(x_r, 0) - \rho(0, \overline{y}, r) = f(x_r, 0) = \varphi(x_r) \end{aligned}$$

where the inequality follows from $(a)$ and $C_1$, included in $(b)$, and the third equality also follows from $C_1$. We conclude that in this case $x_r \in \operatorname{Argmin}_x \varphi(x)$. Since $x_r$ is arbitrary, the proof will be complete if we prove that case (2) cannot occur. Suppose by contradiction that case (2) holds. Thus there exist an open neighborhood $V \subset Z$ of 0 and a subsequence $z_{k_j} := z_j$, such that $z_j \in V^c$ for all $j \in \mathbb{N}$. Then,

$$\begin{aligned}
f(x_r, z_j) - \rho(z_j, \overline{y}, r) &= f(x_r, z_j) - \rho(z_j, \overline{y}, r_1) + \rho(z_j, \overline{y}, r_1) - \rho(z_j, \overline{y}, r) \\
&\geq \inf_z \{f(x_r, z) - \rho(z, \overline{y}, r_1)\} + \\
&\quad \inf_{z \in V^c} \{\rho(z, \overline{y}, r_1) - \rho(z, \overline{y}, r)\} \\
&\geq q(\overline{y}, r_1) + \inf_{z \in V^c} \{\rho(z, \overline{y}, r_1) - \rho(z, \overline{y}, r)\} \\
&= \beta(0) + \inf_{z \in V^c} \{\rho(z, \overline{y}, r_1) - \rho(z, \overline{y}, r)\}.
\end{aligned}$$

Taking limits with $j \to \infty$ in the inequalities above and using (5.7), we obtain that

$$\beta(0) \geq \beta(0) + \inf_{z \in V^c} \{\rho(z, \overline{y}, r_1) - \rho(z, \overline{y}, r)\}. \tag{5.8}$$

Since $\beta(0) \in \mathbb{R}$ and $r > r_1$, (5.8) contradicts condition $C_2(i)$, included in $(b)$. We conclude that case (2) cannot occur, which completes the proof. $\square$

**Remark 5.2.6.** We mention that assumption $(a)$ in Theorem 1.4.3 does not imply assumption $(b)$. Indeed, consider a primal problem $\min_{x \in \mathbb{R}} \varphi(x)$, where $\varphi(x) = x^2$. Let a continuous parameterization function $f(x, z) = x^2 + zx^3$. It follows that assumption $(a)$ in Theorem 1.4.3 holds, but assumption $(b)$ does not hold, because $\beta(0) = 0$ and $\beta(z) = -\infty$ for all $z \neq 0$. Proposition 5.3.4 presents some assumptions under which $\beta$ is lsc at 0.

## 5.3 Sub-optimal path

In general, getting an exact optimal solution of an optimization problem is very hard or even impossible, but when the optimal value is finite, approximate solutions always exist and in general they are easier to find than exact solutions. In [66], the authors defined a sub-optimal path related to the dual problem and established some convergence results in finite dimensional spaces. A sub-optimal path is presented in Definition 1.5.3. In this section we recall a sub-optimal path related to our duality scheme and analyze its convergence properties. This result is related to [17, Theorem 6.1], where the authors consider an optimal path in the sense that all the subproblems are solved exactly. Also, as we will see in Section 5.4, our duality scheme includes the one considered in [17].

Recall that the calculation of the dual function leads to the following problem:

$$\inf\{f(x,z) - \rho(z,y,r) : (x,z) \in X \times Z\}. \tag{5.9}$$

**Definition 5.3.1.** *Let $I \subset \mathbb{R}_+$ be unbounded above, and for each $r \in I$ take $\varepsilon_r \geq 0$. The set $\{(x_r,z_r)\}_{r\in I} \subset X \times Z$ is called a* sub-optimal path *of problem* (5.9) *if*

$$f(x_r,z_r) - \rho(z_r,y,r) \leq q(y,r) + \varepsilon_r \tag{5.10}$$

*for all $r \in I$. When $(x_r,z_r)$ satisfies* (5.10) *with $\varepsilon_r = 0$ for all $r \in I$, the set $\{(x_r,z_r)\}_{r\in I}$ is called an* optimal path.

In the next theorem we analyze limit points of a sub-optimal path, where $\{\varepsilon_r\}_{r\in I}$ satisfies $\lim\limits_{r\in I, r\to\infty} \varepsilon_r = 0$.

**Theorem 5.3.1.** *Assume that*

*a) there exists $(\overline{y},\overline{r}) \in \operatorname{dom}\beta^\rho$;*

*b) the parameterization function $f$ is* lsc *at $(x,0)$ for each $x \in X$, and there exist an open neighborhood $W \subset Z$ of $0$, a real number $\alpha > \beta(0)$, and a compact subset $B \subset X$ such that*

$$L_{f,W}(\alpha) := \{x \in X : f(x,z) \leq \alpha\} \subset B, \quad \text{for all } z \in W.$$

*Then*

*i) there exists a sub-optimal path $\{(x_r,z_r)\}_{r\geq\overline{r}}$.*

*ii) Take a set $I \subset \mathbb{R}_+$ unbounded above and consider a sub-optimal path $\{(x_r, z_r)\}_{r\in I}$ satisfying $\lim\limits_{r\in I, r\to\infty} \varepsilon_r = 0$. Then $\{z_r\}_{r\in I}$ converges to 0, and the set of cluster points of $\{x_r\}_{r\in I}$ is a nonempty set contained in the primal optimal solution set.*

*Proof.* Since $\rho(z, y, \cdot)$ is a nonincreasing function, we have that $q(\overline{y}, \cdot)$ is nondecreasing. Thus, if $r \geq \overline{r}$ then $q(\overline{y}, r) > -\infty$, by item $(a)$ and Proposition 1.5.2. Thus the existence of a sub-optimal path is trivially ensured, which proves $(i)$.

For proving $(ii)$, let $\{(x_r, z_r)\}_{r\in I}$ be a sub-optimal path. Assume that $\lim\limits_{r\in I, r\to\infty} \varepsilon_r = 0$. Suppose by contradiction that $\{z_r\}_{r\in I}$ does not converge to 0 when $r \to \infty$. Thus there exist an open neighborhood $V \subset Z$ of 0 and $J \subset I$, unbounded above, such that $\{z_r\}_{r\in J} \subset V^c$, (for instance, we can take $J_k := I \cap [k, \infty)$, for $k \in \mathbb{N}$, and hence there exists $r_k \in J_k$ such that $z_{r_k} \in V^c$; then $J = \{r_k\}_k$ is unbounded above). Therefore we have

$$\begin{aligned}
\beta(0) + \varepsilon_r &\geq q(\overline{y}, r) + \varepsilon_r \geq f(x_r, z_r) - \rho(z_r, \overline{y}, r) \\
&= f(x_r, z_r) - \rho(z_r, \overline{y}, \overline{r}) + \rho(z_r, \overline{y}, \overline{r}) - \rho(z_r, \overline{y}, r) \\
&\geq q(\overline{y}, \overline{r}) + \inf_{z\in V^c} \{\rho(z, \overline{y}, \overline{r}) - \rho(z, \overline{y}, r)\}.
\end{aligned}$$

Since $\lim\limits_{r\in I, r\to\infty} \varepsilon_r = 0$, we conclude that there exists $r_0 \in J$ such that for all $r \geq r_0, r \in J$, we have

$$\beta(0) + 1 - q(\overline{y}, \overline{r}) \geq \inf_{z\in V^c} \{\rho(z, \overline{y}, \overline{r}) - \rho(z, \overline{y}, r)\},$$

which contradicts $C_2$, because $J$ is unbounded above and $\beta(0), q(\overline{y}, \overline{r}) \in \mathbb{R}$. It follows that $\{z_r\}_{r\in I}$ converges to 0. Consider an open neighborhood $W \subset Z$ of 0 and $\alpha > \beta(0)$ as in assumption $(b)$. Since $\{z_r\}_{r\in I}$ converges to 0, there exists $r_0 \in I$ such that $\{z_r\}_{r\geq r_0, r\in I} \subset W$. Take $t := \alpha - \beta(0) > 0$. The function $\rho(\cdot, \overline{y}, \overline{r})$ is upper semicontinuous at 0 by condition $C_1$. Thus there exists some $r_1 \geq \max\{r_0, \overline{r}\}$ such that $\rho(z_r, \overline{y}, \overline{r}) \leq \frac{t}{2}$ and $\varepsilon_r \leq \frac{t}{2}$ for all

$r \geq r_1, r \in I$. Therefore, for all $r \geq r_1, r \in I$,

$$\begin{aligned} \beta(0) + \tfrac{t}{2} &\geq q(\overline{y}, r) + \varepsilon_r \geq f(x_r, z_r) - \rho(z_r, \overline{y}, r) \\ &\geq f(x_r, z_r) - \rho(z_r, \overline{y}, \overline{r}) \geq f(x_r, z_r) - \tfrac{t}{2}. \end{aligned}$$

Hence

$$f(x_r, z_r) \leq \beta(0) + t = \alpha, \quad \text{for all } \; r \geq r_1, r \in I,$$

that is to say $\{x_r\}_{r \geq r_1, r \in I} \subset L_{f,W}(\alpha) \subset B$, where $B$ is a compact set and the last inclusion follows by Assumption $(b)$. In particular, since $\{z_r\}_{r \in I}$ converges to 0, the set of cluster points of the sub-optimal path $\{(x_r, z_r) : r \in I\}$ is nonempty. Moreover every cluster point has the form $(x^*, 0)$. Let us prove that $x^*$ is a primal optimal solution, where $x^*$ is an arbitrary cluster point of $\{x_r\}_{r \in I}$. Take a subnet $\{x_{r_j}\}_{j \in J}$ converging to $x^*$, and $j_0 \in J$ satisfying $r_j \geq \overline{r}$ for all $j \geq j_0, j \in J$. Observe that $\{z_{r_j}\}_{j \in J}$ converges to 0. Thus

$$\begin{aligned} \beta(0) + \varepsilon_{r_j} &\geq q(\overline{y}, r_j) + \varepsilon_{r_j} \\ &\geq f(x_{r_j}, z_{r_j}) - \rho(z_{r_j}, \overline{y}, r_j) \\ &\geq f(x_{r_j}, z_{r_j}) - \rho(z_{r_j}, \overline{y}, \overline{r}) \end{aligned}$$

for all $j \geq j_0, j \in J$. If we take the $\liminf_{j \in J}$ in these inequalities, we obtain

$$\beta(0) \geq f(x^*, 0) - \rho(0, \overline{y}, \overline{r}) = f(x^*, 0) = \varphi(x^*),$$

using conditions $(b)$ and $C_1$. Thus $x^*$ is a primal solution. The theorem is proved. □

**Remark 5.3.2.** In connection with the compactness assumption of Theorem 5.3.1, we mention that when $X$ is an infinite dimensional reflexive Banach space with the weak topology (which is not metrizable), Banach-Alaoglu's Theorem implies that a set is weakly compact if and only if it is bounded and weakly closed. In particular, closed balls (in the strong topology) are weakly compact in such spaces. Thus, a parameterization function $f$ such that some sub-level set of $f(\cdot, z)$ is weakly closed and uniformly bounded when $z$ runs over a neighborhood of 0, provides an example for which assumption (b) of Theorem 5.3.1 holds. This situation is indeed a prototypical and nontrivial case to which Theorem 5.3.1 applies. We remind also that sub-level sets of convex and lsc functions are always weakly closed, so that in the convex case it suffices to check the uniform boundedness of the sub-level sets of $f(\cdot, z)$.

**Remark 5.3.3.** Theorem 5.3.1 is related to [17, Theorem 6.1], where the authors considered an optimal path (in a reflexive Banach space) instead of a sub-optimal path, and the compactness assumption on the sub-level sets of $f(\cdot, z)$ is assumed locally at all $z$, instead of just at $z = 0$, as assumed here. Also, in [17, Theorem 6.1] it is assumed that the compactness property holds for all sub-level sets of $f(\cdot, z)$, while Theorem 5.3.1 assumes compactness of just one of them, corresponding to $\alpha > \beta(0)$. Since we are not assuming convexity of $f(\cdot, z)$, compactness of just one sub-level set of $f(\cdot, z)$ is not equivalent to compactness of all of them.

**Proposition 5.3.4.** *Let $f : X \times Z \to \mathbb{R}$ be* lsc *at $(x, 0)$ for each $x \in X$. Take $\beta(z) := \inf_x f(x, z)$. Suppose that $\beta(0) > -\infty$ and that there exist an open neighborhood $W \subset Z$ of $0$, $\alpha \geq \beta(0)$ and a compact subset $B \subset X$, such that*

$$L_{f,W}(\alpha) := \{x \in X : f(x, z) \leq \alpha\} \subset B, \quad \textit{for all } z \in W.$$

*Then the perturbation function $\beta$ is* lsc *at* 0.

*Proof.* Let $J$ be the set of all neighborhoods of 0. We know that $J$ is a directed set with the partial order $V_1 \geq V_2$ iff $V_1 \subset V_2$. Suppose by contradiction that $\beta$ is not lsc at 0. Then there exists $\varepsilon > 0$ such that

$$\sup_{V \in J} \inf_{v \in V} \beta(v) < \beta(0) - \varepsilon.$$

Thus, $\inf\limits_{v \in V} \beta(v) < \beta(0) - \varepsilon$ for all $V \in J$. In particular for each $V \in J$ there exists $z_V \in V$ such that $\beta(z_V) < \beta(0) - \varepsilon$, which in turn implies that for each $V \in J$ there exists $x_V \in X$ satisfying

$$f(x_V, z_V) < \beta(0) - \varepsilon. \tag{5.11}$$

By construction the net $\{z_V\}_{V\in J}$ converges to 0. Taking $W$ and $\alpha$ as in the assumption. It follows from (5.11) that $I := \{V \in J : V \geq W\}$ is a terminal subset of $J$ such that $\{z_V\}_{V\in I} \subset W$ and $\{x_V\}_{V\in I} \subset L_{f,W}(\alpha) \subset B$, where $B$ is the compact set given by hypothesis. Hence there exists a subnet $\{\eta_s\}_{s\in S}$ of $\{x_V\}_{V\in I}$ convergent to some $\overline{x}$. This means that $\eta_s = x_{g(s)}$, where $S$ is a directed set and $g : S \to I$ is a function such that for every $U \in I$ there exists an $s_U \in S$ satisfying $g(s) \geq U$ for all $s \geq s_U$, $s \in S$. In particular, the set $\{t_s\}_{s\in S}$, where $t_s := z_{g(s)}$ for all $s \in S$, is a subnet of $\{z_V\}_{V\in I}$ converging to 0, and $f(\eta_s, t_s) < \beta(0) - \varepsilon$ for all $s \in S$, by (5.11). Therefore, using the lower semicontinuity of $f$ in $(\overline{x}, 0)$ we obtain

$$\beta(0) \leq f(\overline{x}, 0) \leq \liminf_{s\in S} f(\eta_s, t_s) \leq \beta(0) - \varepsilon,$$

entailing a contradiction. $\square$

**Example 5.3.5.** Consider the following constrained optimization problem

$$\text{minimize} \quad h(x) \quad \text{subject to } x \text{ in } C, \tag{5.12}$$

where $h : X \to \mathbb{R}$ is a lsc function such that $L_\alpha := \{x \in X : h(x) \leq \alpha\}$ is compact for some $\alpha > \inf_{x\in C} h(x)$, and $C$ is a closed subset of $X$. Take a mapping $D : Z \rightrightarrows X$ such that $D(0) = C$ and suppose that $D$ has a closed graph, that is, the set $\{(z, u) : u \in D(z), z \in Z\}$ is closed (in the case that $C := \{x : g_j(x) \leq 0,\ j = 1, ..., m\}$, where $g_j : X \to \mathbb{R}$ is lsc for $j = 1, ..., m$, a canonical such mapping is $D(z) = \{x : g_j(x) \leq z_j,\ j = 1, ..., m\}$). A canonical dualizing parameterization function for problem (5.12) is $f(x, z) = h(x) + \delta_{D(z)}(x)$, where $\delta_V(x) = 0$ if $x \in V$ and $\delta_V(x) = \infty$ otherwise. It is not difficult to see that $f$ satisfies the assumptions of Proposition 5.3.4. Thus the perturbation function $\beta(z) = \inf_{x\in D(z)} h(x)$ is lsc at 0.

Next we show some examples of general augmented Lagrangians and compare our setting with the ones considered in [17] and [65, Section 3.1].

## 5.4 Augmented Lagrangian type functions

Consider a coupling function $p : Z \times Y \times \mathbb{R}_{++} \to \mathbb{R}$ such that $p(z, y, \cdot)$ is differentiable. We state next, in our general setting, the valley at zero property given in [65, Section 3.1] and presented in Section 1.5.1 of this thesis.

$A_1$) There exists $\alpha \in [0, 1)$ such that, for every open neighborhood $V \subset Z$ of 0, and $y \in Y$,

$$M_{V,\varepsilon} := \inf_{u \in V^c, \tau \geq \varepsilon} \tau^\alpha p'_r(u, y, \tau) > 0$$

for all $\varepsilon > 0$.

**Remark 5.4.1.** In [65, Section 3.1] the authors also assume that $p'_r(0, y, r) = 0$, while here we do not assume this condition. Regarding our condition $C_1$, it is a standard assumption, used in both [17] and [65]. Therefore, is enough for us to study the relationship between our condition $C_2$ and related assumptions in the aforementioned papers.

We mention that [65] uses as coupling function $\rho := -p$ in the construction of the Lagrangian function.

**Proposition 5.4.2.** *Take a function $p$ satisfying ($A_1$). Then the function $\rho := -p$ satisfies condition $C_2$.*

*Proof.* Fix an open neighborhood $V \subset Z$ of 0, $y \in Y$, and $\hat{r} > 0$. For every $z \in V^c$ and $r > \hat{r}$ there exists $\theta_r \in (\hat{r}, r)$ such that

$$p(z, y, r) - p(z, y, \hat{r}) = p'_r(z, y, \theta_r)(r - \hat{r}) \geq (r^{1-\alpha} - \hat{r}^{1-\alpha})\theta_r^\alpha p'_r(z, y, \theta_r), \quad (5.13)$$

where the inequality follows from the following estimates:

$$r > \theta_r \Rightarrow r = r^{1-\alpha} r^\alpha \geq r^{1-\alpha}\theta_r^\alpha,$$

where $\alpha \in [0, 1)$ is given by ($A_1$); analogously we have $\hat{r} = \hat{r}^{1-\alpha}\hat{r}^\alpha \leq \hat{r}^{1-\alpha}\theta_r^\alpha$. Thus we get

$$r - \hat{r} \geq r^{1-\alpha}\theta_r^\alpha - \hat{r}^{1-\alpha}\theta_r^\alpha = \theta_r^\alpha(r^{1-\alpha} - \hat{r}^{1-\alpha}).$$

Take $0 < \varepsilon < \hat{r}$. From (5.13) we obtain

$$\begin{aligned} p(z,y,r) - p(z,y,\hat{r}) &\geq (r^{1-\alpha} - \hat{r}^{1-\alpha}) \inf_{\tau \geq \varepsilon} \tau^{\alpha} p'_r(z,y,\tau) \\ &\geq (r^{1-\alpha} - \hat{r}^{1-\alpha}) \inf_{u \in V^c, \tau \geq \varepsilon} \tau^{\alpha} p'_r(u,y,\tau) \\ &= (r^{1-\alpha} - \hat{r}^{1-\alpha}) M_{V,\varepsilon} \end{aligned}$$

for all $z \in V^c$. Therefore

$$\inf_{z \in V^c} \{p(z,y,r) - p(z,y,\hat{r})\} \geq (r^{1-\alpha} - \hat{r}^{1-\alpha}) M_{V,\varepsilon}.$$

It is easy to see that $C_2$ follows from the last estimate above and $(A_1)$, observing that $\rho = -p$ and $\alpha \in [0,1)$. $\square$

The above result shows that our setting is more general than the one considered in [65]. In order to show that our setting is more general than the one considered in [17], we recall next their main assumptions. We mention that these assumptions are stated in Chapter 1, we recall them here for the convenience of the reader.

Consider a function $s : \mathbb{R}^2 \to \mathbb{R}$ such that $s(0,0) = 0$ and for every $a \in \mathbb{R}$ and $b_1 \geq b_2$, it satisfies

$$s(a,b_1) - s(a,b_2) \geq \psi(b_1 - b_2), \tag{5.14}$$

where $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing function such that $\psi(0) = 0$ and $\psi$ is coercive, that is, $\lim_{t \to \infty} \psi(t) = \infty$.

Let $\{\nu_\eta\}_{\eta \in U_1}$ be a family of upper semicontinuous functions satisfying

$$\nu_\eta(0) = 0 \text{ for all } \eta \in U_1, \tag{5.15}$$

and $\{\sigma_\mu\}_{\mu \in U_2}$ be a family of augmenting functions which have a valley at zero property, that is, for every $\mu \in U_2$, $\sigma_\mu : Z \to \mathbb{R}_+$ is proper, lsc and satisfies

$$\sigma_\mu(0) = 0 \quad \text{and} \quad \inf_{z \in V^c} \sigma_\mu(z) > 0, \tag{5.16}$$

for every open neighborhood $V \subset Z$ of 0.

The coupling function considered in [17] is $\rho$, given by

$$\rho(z,(\eta,\mu),r) = s(\nu_\eta(z), -r\sigma_\mu(z))$$

where $s$, $\{\nu_\eta\}_{\eta\in U_1}$ and $\{\sigma_\mu\}_{\mu\in U_2}$ satisfy (5.14)-(5.16).

Since we are not supposing any structure on the set $Y$, we can consider $Y := U_1 \times U_2$. In the next proposition we show that our primal-dual scheme includes the one in [17].

**Proposition 5.4.3.** *Take $\rho(z,y,r) := s(\nu_\eta(z), -r\sigma_\mu(z))$, where $y = (\eta,\mu)$ and the functions $\{\nu_\eta\}_{\eta\in U_1}$ and $\{\sigma_\mu\}_{\mu\in U_2}$ satisfy* (5.14)-(5.16). *Then condition $C_2$ is satisfied.*

*Proof.* Fix an open neighborhood $V \subset Z$ of 0, $y \in Y$ and $\overline{r} > 0$. For all $r > \overline{r}$ and $z \in V^c$ we have

$$\begin{aligned} \rho(z,y,\overline{r}) - \rho(z,y,r) &= s(\nu_\eta(z), -\overline{r}\sigma_\mu(z)) - s(\nu_\eta(z), -r\sigma_\mu(z)) \\ &\geq \psi((r-\overline{r})\sigma_\mu(z)) \\ &\geq \psi((r-\overline{r})M_V), \end{aligned}$$

where the first inequality follows from the property of the function $s$, and the second inequality follows from the fact that $\psi$ is increasing and $M_V := \inf\limits_{u\in V^c} \sigma_\mu(u) > 0$. It follows that

$$\inf_{z\in V^c} \rho(z,y,\overline{r}) - \rho(z,y,r) \geq \psi((r-\overline{r})M_V).$$

Using this last estimate and the property of the function $\psi$, we conclude that $C_2$ is satisfied. □

**Remark 5.4.4.** The coercivity property $\lim_{t\to\infty}\psi(t) = \infty$ was not explicitly required in [17], but it was used in the proof of [17, Theorem 4.1], and this theorem is applied throughout the paper.

**Example 5.4.5.** Let $Y$ be a Banach space and $Z$ be a reflexive Banach space. Consider a coupling function $g : Y \times Z \to \mathbb{R}$ such that $g(y, \cdot)$ is weakly upper semicontinuous and $g(y, 0) = 0$ for each $y \in Y$. Let

$$\rho(z, y, r) := g(y, z) - r\sigma(z),$$

where $\sigma$ is an augmenting function with a valley at zero (i.e., $\sigma$ satisfies (5.16)). In this case, we recover the augmented Lagrangian studied by Zhou and Yang [71]:

$$\ell(x, y, r) = \inf_z \{\phi(x, z) - g(y, z) + r\sigma(z)\}.$$

**Example 5.4.6.** Let $Z$ be a Hilbert space. Consider a continuous and invertible map $A : Z \to Z$, and suppose that $Y = Z$. Let the coupling function $\rho$ be defined by $\rho(z, y, r) = \langle y, Az\rangle - r\sigma(Az)$, where $\sigma : Z \to \mathbb{R}$ is an augmenting function, i.e. a proper, lsc and convex function satisfying:

$$\sigma(0) = 0 \text{ and } \operatorname{Argmin} \sigma = \{0\}.$$

In this context our general augmented Lagrangian is the $A$-augmented Lagrangian proposed and studied by Yang and Zhang [68]:

$$\ell_A(x, y, r) = \inf_{z \in Z} \{\phi(x, z) - \langle y, Az\rangle + r\sigma(Az)\}.$$

The $A$-augmented Lagrangian was studied in finite dimensional space, and some additional conditions are imposed on the mapping $A$, see [68]. In particular, in the finite dimensional setting, when $A = I$, that is, $Az = z$ for all $z \in Z$, we recover the augmented Lagrangian proposed by Rockafellar and Wets in [54, Chapter 11], presented in Section 1.4.

# Bibliography


[1] Alber, Ya.I., Iusem, A.N. and Solodov, M.V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* **81** (1998), 23–35.

[2] Allen, E., Helgason, R., Kennington, J. and Shetty, B. A generalization of Polyak's convergence result for subgradient optimization. *Mathematical Programming* **37** (1987), 309–317.

[3] Andreani, R., Birgin, E.G., Martínez, J.M. and Schuverdt, M.L. On Augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization* **18** (2007), 1286–1309.

[4] Bazaraa, M.S. and Sherali, H.D. On the choice of stepsize in subgradient optimization. *European Journal of Operational Research* **7** (1981), 380–388.

[5] Bertsekas, D.P. and Mitter, S.K. A descent numerical method for optimization problems with nondifferentiable cost functionals. *SIAM Journal on Control and Optimization* **11** (1973), 637–652 .

[6] Bertsekas, D.P. On penalty and multiplier methods for constrained optimization problems. *SIAM Journal on Control and Optimization* **14** (1976), 216–235.

[7] Bertsekas, D.P. *Constrained optimization and Lagrangian multiplier methods.* Academic Press, New York (1982).

[8] Bertsekas, D.P. *Nonlinear programming.* Athena Scientific, Belmont, Massachusetts (1995).

[9] Birgin, E.G., Castillo, R. and Martínez, J.M. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. *Computational Optimization and Applications* **31** (2005), 31-56.

[10] Brännlund, U. *On relaxation methods for nonsmooth convex optimization.* Doctoral Thesis, Royal Institute of Technology, Stockholm, Sweden, (1993).

[11] Brännlund, U. A generalized subgradient method with relaxation step. *Mathematical Programming* **71** (1995), 207–219.

[12] Burachik, R.S., Gasimov, R.N., Ismayilova, N.A. and Kaya, C.Y. On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian. *Journal of Global Optimization* **34** (2006), 55–78.

[13] Burachik, R.S. and Kaya, C.Y. An update rule and a convergence result for a penalty function method. *Journal of Industrial and Management Optimization* **3** (2007), 381–398.

[14] Burachik, R.S., Kaya, C.Y. and Mammadov, M. An inexact modified subgradient algorithm for nonconvex optimization. *Computational Optimization and Applications*, (2008) DOI 10.1007/s10589-008-9168-7.

[15] Burachik, R.S. and Iusem, A.N. *Set-valued mappings and enlargements of monotone operators.* Springer, Berlin, (2008).

[16] Burachik, R.S., Iusem, A.N. and Melo, J.G. A primal dual modified subgradient algorithm with sharp Lagrangian. *Journal of Global Optimization*, (2009) DOI 10.1007/s10898-009-9429-8.

[17] Burachik, R.S. and Rubinov, A.M. Abstract convexity and augmented Lagrangians. *SIAM Journal on Optimization* **18** (2007), 413–436.

[18] Burachik, R.S. and Rubinov, A.M. On the absence of duality gap for Lagrange-type functions. *Journal of Industrial and Management Optimization* **1** (2005), 33–38.

[19] Buys, J.D. *Dual algorithms for constrained optimization problems.* Doctoral Thesis, University of Leiden, Leiden, The Netherlands (1972).

[20] Clarke, F.H. *Optimization and nonsmooth analysis.* John Wiley, New York, (1983).

[21] Dem'yanov, V.F. and Vasil'ev, L.V. *Nondifferentiable optimization.* Optimization Software, New York, (1985).

[22] Di Pillo, G. and Grippo, L. A new class of augmented Lagrangians in nonlinear programming. *SIAM Journal on Control and Optimization* **17** (1979), 618-628.

[23] Gasimov, R.N. Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. *Journal of Global Optimization* **24** (2002), 187–203.

[24] Gasimov, R.N. and Rubinov, A.M. On augmented Lagrangians for optimization problems with a single constraint. *Journal of Global Optimization* **28** (2004), 153–173.

[25] Hestenes, M.R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications* **4** (1969), 303–320.

[26] Huang, X.X. and Yang, X.Q. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research* **28** (2003), 533–552.

[27] Huang, X.X. and Yang, X.Q. Further study on augmented Lagrangian duality theory. *Journal of Global Optimization* **31** (2005), 193–210.

[28] Ito, K. and Kunisch, S. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces *Mathematical Programming.* **46** (1990), 341–360.

[29] Ito, K. and Kunisch, S. Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. *Nonlinear Analysis* **41** (2000),591–616.

[30] Iusem, A.N. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa* **8** (1999), 11–50.

[31] Kiwiel, K.C. A method of centers with approximate subgradient linearizations for nonsmooth convex optimization. *SIAM Journal on Optimization* **18** (2008), 1467-1489.

[32] Kim, S., Ahn, H. and Cho, S.-C. Variable target value subgradient method. *Mathematical Programming* **49** (1991), 359–369.

[33] Lemaréchal, C. and Correa, R. Convergence of some algorithms for convex minimization. *Mathematical Programming* **62** (1993), 261-275.

[34] Lemaréchal, C., Nemirovskii, A.S. and Nesterov, Yu.E. New variants of bundle methods. *Mathematical Programming* **69** (1995), 111-147.

[35] Mangasarian, O.L. Unconstrained Lagrangians in nonlinear programming. *SIAM Journal on Control and Optimization* **13** (1975), 772–791.

[36] Nedić, A. and Bertsekas, D.P. Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization* **12** (2001), 109–138.

[37] Nedić, A. and Ozdaglar, A. A geometric framework for nonconvex optimization duality using augmented Lagrangian functions. *Journal of Global Optimization* **40** (2008), 545–573.

[38] Nedić, A and Ozdaglar, A. Separation of nonconvex sets with general augmenting functions. *Mathematics of Operations Research* **33** (2008), 587–605.

[39] Nedić, A. and Ozdaglar, A. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization* **19** (2009), 1757–1780.

[40] Nedić, A. and Ozdaglar, A. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Application*, (2009), DOI 10.1007/s10957-009-9522-7.

[41] Nedić, A., Ozdaglar, A. and Rubinov, A.M. Abstract convexity for nonconvex optimization duality. *Optimization* **56** (2007), 655–674.

[42] Oettli, W. and Yang, X.Q. Modified Lagrangian and least root approaches for general nonlinear optimization problems. *Acta Mathematicae Applicatae Sinica, English Series* **18** (2002), 147–152.

[43] Penot, J.-P. Augmented Lagrangians, duality and growth conditions. *Journal of Nonlinear and Convex Analysis* **3** (2002), 283–302.

[44] Penot, J.-P. and Rubinov, A.M. Multipliers and general Lagrangians. *Optimization* **54** (2005), 443–467.

[45] Polyak, B.T. A general method of solving extremum problems. *Soviet Mathematics Doklady* **8** (1967), 593–597.

[46] Polyak, B.T. Minimization of unsmooth functionals. *Zurnal Vychislitelnoy Matematiki i Matematicheslcoy Fizilci* **9** (1969), 509–521.

[47] Powell, M.J.D. *Method for nonlinear constraints in minimization problems.* In Optimization. Academic Press, London (1969).

[48] Rockafellar, R.T. *Convex analysis*, Princeton University Press, Princeton, New Jersey, (1970).

[49] Rockafellar, R.T. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization* **12** (1974), 268-285.

[50] Rockafellar, R.T. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming* **5** (1973), 354–373.

[51] Rockafellar, R.T. The multiplier method of Hestenes and Powell applied to convex programming, *Journal of Optimization Theory and Applications*, **12** (1973), 555–562.

[52] Rockafellar, R.T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research* **1** (1976), 97–116.

[53] Rockafellar, R.T. Lagrange multipliers and optimality. *SIAM Review* **35** (1993), 183–238.

[54] Rockafellar, R.T. and Wets, R.J.-B. *Variational analysis.* Springer, Berlin, (1998).

[55] Rubinov, A.M. *Abstract convexity and global optimization.* Kluwer Academic Publishers, Dordrecht, The Netherlands, (2000).

[56] Rubinov, A.M. and Yang, X.Q. *Lagrange-type functions in constrained non-convex optimization.* Kluwer Academic Publishers, Dordrecht, The Netherlands, (2003).

[57] Rubinov, A.M., Glover, B.M. and Yang, X.Q. Modified Lagrangian and penalty functions in continuous optimization. *Optimization* **46** (1999), 327-351.

[58] Rubinov, A.M., Huang, X.X. and Yang, X.Q. The zero duality gap property and lower semicontinuity of the perturbation function. *Mathematics of Operations Research* **27** (2002), 775–791.

[59] Schirotzek, W. *Nonsmooth analysis.* Springer, Berlin, (2007).

[60] Shor, N.Z. *Minimization methods for nondifferentiable functions.* Springer, Berlin, (1985).

[61] Singer, I. *Abstract convex analysis.* John Wiley, New York, (1997).

[62] Volkwein, S. Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces. *SIAM Journal on Control and Optimization* **38** (2000), 767–785.

[63] Wang, C.Y. and Li, D. Unified theory of augmented Lagrangian methods for constrained global optimization. *Journal of Global Optimization*, (2008), DOI 10.1007/s10898-008-9347-1.

[64] Wang, C.Y., Yang, X.Q. and Yang, X.M. Nonlinear Lagrange duality theorems and penalty function methods in continuous optimization. *Journal of Global Optimization* **27** (2003), 473–484.

[65] Wang, C.Y., Yang, X.Q. and Yang, X.M. Unified nonlinear Lagrangian approach to duality and optimal paths. *Journal of Optimization Theory and Applications.* **135** (2007), 85–100.

[66] Wang, C.Y., Yang, X.Q. and Yang, X.M. Zero duality gap and convergence of sub-optimal paths for optimization problems via nonlinear augmented Lagrangian (to appear).

[67] Yang, X.Q. and Huang, X.X. A nonlinear Lagrangian approach to constrained optimization problems. *SIAM Journal on Optimization* **14** (2001), 1119–1144.

[68] Zhang, L. and Yang, X.Q. An augmented Lagrangian approach with a variable transformation in non linear programming. *Nonlinear Analysis: Theory, Methods and Applications* **69** (2008), 2095–2113.

[69] Zhou, Y.Y. and Yang, X.Q. Some results about duality and exact Penalization. *Journal of Global Optimization* **29** (2004), 497–509.

[70] Zhou, Y.Y. and Yang, X.Q. Augmented Lagrangian function, non-quadratic growth condition and exact penalization. *Operations Research Letters* **34** (2006), 127–134.

[71] Zhou, Y.Y. and Yang, X.Q. Duality and penalization in optimization via an augmented Lagrangian function with applications. *Journal of Optimization Theory and Applications* **140** (2009), 171–188.