UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

# CONTRIBUTIONS TO THE CONVERGENCE THEORY AND COMPUTATIONAL IMPLEMENTATION OF INTERIOR OPTIMIZATION METHODS FOR CONVEX PROBLEMS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MODELACIÓN MATEMÁTICA

NATALIA SOLEDAD KAREN RUIZ GARRIDO

PROFESOR GUÍA:
FELIPE ÁLVAREZ DAZIANO

MIEMBROS DE LA COMISIÓN:
LUIS BRICEÑO ARIAS
ALEJANDRO JOFRÉ CÁCERES
JUAN PEYPOUQUET URBANEJA
HÉCTOR RAMÍREZ CABRERA

SANTIAGO DE CHILE
2016

# Resumen

En esta tesis doctoral se estudian algoritmos para resolver problemas de optimización convexa con estructura separable, y problemas de equilibrio económico de Walras. Así, la tesis se divide en dos partes. La primera parte corresponde al estudio teórico y numérico de un método de direcciones alternantes de multiplicadores el cual usa un término proximal interior. La segunda parte está dedicada al estudio numérico de algoritmos de segundo orden para resolver problemas de maximización de utilidades que aparecen en problemas de equilibrio en economía.

En el primer capítulo se hace una revisión de algunos métodos de descomposición basados en el Lagrangeano aumentado. Luego, se hace una revisión de la definición y propiedades de las distancias proximales generalizadas.

En el segundo capítulo, se prueba la convergencia global del método propuesto bajo supuestos estándares. Este método es llamado Método de Direcciones Alternantes con Regularización Proximal Interior (RIPADM).

En el tercer capítulo, se establece la convergencia global de una variante del método RIPADM la cual añade un factor de relajación a la regla de actualización del multiplicador de Lagrange.

En el cuarto capítulo, se implementa computacionalmente en Matlab, el método RIPADM, el ADM original y el método proximal de multiplicadores (PMM) para resolver el problema LASSO restringido y un problema de máquinas de soporte vectorial. En la implementación computacional del método RIPADM se usa la distancia proximal Log-quad, y la distancia Kullback-Leibler.

En el quinto capítulo, se describe el modelo de intercambio puro de Arrow-Debreau, y se hace una revisión de un método recientemente propuesto para resolver problemas de equilibrio económico de Walras.

En el sexto capítulo, se implementa computacionalmente el método punto-interior primal-dual (PDIPM) y el método gradiente proyectado con aceleración (AGPM) para resolver problemas de maximización de utilidades que aparecen en problemas de equilibrio en economía.

# Abstract

This thesis deals with algorithms for solving convex optimization problems with separable structure, and Walras economic equilibrium problems. So the thesis is divided into two parts. The first part corresponds to a theoretical and numerical study of an alternating direction method of multipliers which uses an inner proximal term. The second one is focused on numerical study of algorithms for solving utility maximization problems that arise in equilibrium problems in economic.

The first chapter includes a review of some decomposition methods based on the augmented Lagrangian. Then, the definition and properties of generalized proximal distances are given.

In the second chapter, the global convergence of the proposed method is proved. This method is called Alternating Direction Method with Interior Proximal Regularization (RIPADM).

The third chapter contains the proof of the global convergence of a variant of the RIPADM method which adds a relaxation factor in the update rule of Lagrange multiplier.

The fourth chapter corresponds to the computational implementation in Matlab of the RIPADM method, the original ADM and proximal method of multipliers (PMM), to solve the LASSO problem and a support vector machine problem. In the computational implementation of the RIPADM method it is used the Log-quad distance, and also is used the Kullback-Leibler distance which is a Bregman distance.

The fifth chapter includes a description of the pure exchange model of Arrow-Debreau, and a review of a recently proposed method to solve Walras economic equilibrium problems.

The sixth chapter corresponds to the computational implementation of primal-dual interior-point method (PDIPM) and the projected gradient method with acceleration (AGPM) for solving utility maximization problems that appear in Walras economic equilibrium problems.

*A mis padres con cariño*
*Isabel y Hugo*

# Agradecimientos

Agradezco primero a Dios por poner en mi camino oportunidades, personas y recursos para poder realizar este doctorado. Agradezco a los profesores Juan Dávila e Isabel Flores por orientarme y apoyarme en el proceso de ingreso al doctorado. Agradezco de manera especial a mi profesor guía de tesis, Felipe Álvarez, por su apoyo y motivación en los momentos difíciles. Agradezco sinceramente a los profesores de la comisión, Juan Peypouquet, Héctor Ramírez, Alejandro Jofré y Luis Briceño por sus observaciones y correcciones de la tesis. Agradezco a CONICYT por darme la beca para poder realizar este doctorado. Agradezco sinceramente a todos los profesores del DIM y CMM que de una u otra forma me ayudaron en este proyecto, gracias por la oportunidad de asistir a escuelas y conferencias en el extranjero, sin duda grandes experiencias. Agradezco a todos los funcionarios del DIM que amablemente ayudan en todo. Agradezco de manera especial a Julio López quien me ayudo mucho en la etapa final de la tesis.

Agradezco de todo corazón a mis amigos del doctorado, Clarita, Luis, Miguel, Erwin y Héctor, por su cariño, honestidad y palabras de apoyo que se mantienen a pesar de las distancias. A todos mis compañeros del doctorado por compartir este camino. Agradezco a mis amigos, Anita, Luisa, Fabiola, Juan y Ruth por compartir tan lindos momentos y por trabajar juntos por los niños de la comuna de Lampa.

Agradezco en especial al Profesor Patricio Felmer quien me enseñó a pensar en las ideas de fondo y el por qué de las cosas, y por ser un ejemplo de persona que asume desafíos grandes. Muchas gracias a mis amigos y compañeros del proyecto en educación matemática ARPA, Yoya, Carlos, Lisa, Armando y Matías, son grandes personas.

Agradezco finalmente a toda mi familia, a aquellos que hoy están conmigo y quienes han partido (mis abuelos). Agradezco muy en especial a mis padres Hugo e Isabel, hermanas Claudia y Alejandra, a mis cuñados Osvaldo y Paolo, y a mis sobrinas Antonia y Julieta (ya te conoceremos Juli!). Gracias por el amor incondicional, y por brindarme toda la fuerza, motivación y tantas cosas más. Y para terminar agradezco nuevamente a Dios, quien estuvo conmigo en el inicio y está conmigo ahora en el término de este etapa.

# Contents

# Introduction

This thesis contains a study of some algorithms and computational techniques for solving certain convex optimization and economic equilibrium problems. With each of these classes of problems, both in optimization and economy, it is possible to associate a bivariate function, or briefly bifunction, which allows to characterize the solutions and provides a general framework to develop numerical methods. In fact, a Lagrangian function can be associated with a given convex optimization problem, whose saddle points are primal-dual solutions of the problem. On the other hand, a Walrasian function can be associated with a Walras economic equilibrium problem whose maxinf-points provide the equilibrium prices. These bifunctions can be used to build approaches that consist of steps or phases where an augmented bifunction is minimized (primal step) and, then this augmented bifunction is maximized (dual step). Such generic schemes are the base on which some numerical approximation algorithms can be developed and implemented computationally.

To get effective as well as efficient methods, it is key to exploit the underlying structure of the specific problems one is dealing with, but also to consider the properties of the data with a special focus on function regularity (differentiability). Indeed, some problems requiere to develop non-smooth techniques as the so-called proximal methods, while for other problems the more suitable choice is a variant of a second-order method as Newton's. In any case, the best computational performance that one can expect depends on the class of problems the corresponding algorithms have been developed for.

This thesis is divided into two parts. The first part is devoted to a theoretical and numerical study of a special class of algorithms for solving non-smooth convex optimization problems with separable structure. Specifically, in this part a variant of the alternating direction method of multipliers (ADM) is proposed, which adds (generalized) proximal regularization terms to an augmented Lagrangian function in order to deal with the primal subproblems of the ADM. The proposed algorithm is called regularized interior proximal alternating direction method (RIPADM). The analysis is focused on general conditions ensuring the global convergence of the method for non-smooth objective functions. Some accelerating techniques are also discussed, as the introduction of a relaxation parameter. This is supplemented with a complete computational implementation and numerical performance study, with some applications in statistics and machine learning that can be formulated as (non-smooth) convex optimization problems.

The second part of this thesis is devoted to a numerical study of algorithms and techniques for solving a particular class of subproblems that appear in a pure exchange model or, equivalently, a Walras barter problem. Each agent involved in the economy seeks to maximize its profits, which leads to solving a concave constrained optimization problem which is called the agent's problem. In this part, we conduct a study of different computational algorithms that solve the problem of utility of the agents who participate in a pure exchange economy, with the aim of contributing to the computational implementation of efficient algorithms for finding max-inf points of a Walrasian function, and thus to find the equilibrium price of the corresponding economy. We explore some second- order alternatives, all of them effective to solve the problem.

More precisely, the first part of the thesis consists of four chapters whose common thread is the RIPADM method which combines the ADM method with generalized proximal distances. Because of the latter, the first part of the thesis begins with a review of some decomposition methods and proximal distances. After this, the convergence analysis of RIPADM method is presented which is to demonstrate that primal-dual sequence generated by this method converges to a saddle point of the Lagrangian function associated with the problem under study. Subsequently, a relaxation factor is added to the RIPADM method and convergence of this relaxed method is demonstrated. To end the first, three problems in statistical and machine learning are presented, and computational results of some methods based on the augmented Lagrangian are compared.

In Chapter 1, two important subjects for the development of our study is presented: decomposition methods for convex optimization problems with separable structure, and generalized proximal distances. In the first issue, we explain the origin of the alternating direction method, and show some variants of this method. Moreover, we show a decomposition method proposed by Chen and Teboulle which uses a separation strategy of the augmented Lagrangian that is different from the strategy of ADM, and we refer to some recent studies based on this proposal. The second issue deals with generalized proximal distances which play the role of interior penalty functions.

In Chapter 2, the RIPADM method is presented in detail. This method solves a separable convex minimization problem where the feasible set is composed of linear constraints and one of the primal variables belonging to a closed nonempty convex set. The proposed decomposition method takes care of the last constraint adding a non-quadratic proximal term to the augmented Lagrangian of the problem under study. The RIPADM method works with different proximal distances, such as Bregman distances, homogeneous second order kernels and double regularization. Furthermore, in this chapter we prove the global convergence of the proposed method under standard assumptions.

In Chapter 3, the RIPADM method is modified slightly modify by adding a relaxation factor in the update rule of Lagrange multiplier. The relaxation factor plays a key role because numerical experiments show that this factor can accelerate the convergence of ADM methods. So, in this chapter we prove the global convergence of RIPADM when the relaxation factor takes values within a restricted range.

In Chapter 4, three new applications arising in statistics and machine learning are studied. The first application is the constrained LASSO problem, the second one is the constrained LASSO with a cost function and the third ones is the twin support vector machine problem (TWSVM). This chapter consists of four sections. The firt section is the introduction of the chapter. In the second section, each application is solved via three methods which are based on augmented Lagrangian: the RIPADM method, the original ADM and the proximal method of multipliers (PMM). Note that the ADM and RIPADM methods are decomposition methods, but PMM it is not. In the third section, the constrained LASSO and TWSVM problems are considered. Here, RIPADM and ADM methods with relaxation factor are applied to these two problems. In the second and third section of this chapter, the RIPADM method uses a Log-quad proximal distance. However, in the fourth section, the RIPADM method using a Bregman distance is analyzed, and the algorithm associated with this method is applied to the three problems discussed in this chapter.

The second part of the thesis consists of two chapters.

Chapter 5 begins with an introduction where some work on existence and calculation of economic equilibrium prices are mentioned. Continues with a description of the pure exchange model, ending with a review of a recently proposed method to solve equilibrium problems which is based on an augmented Walrasian technique and a lopsided convergence approximation procedure.

Chapter 6 begins with a description of two methods to solve the agent's problem: prima-dual interior-point method and projected gradient method with acceleration. Note that the primal-dual interior-point method and the projected gradient method with acceleration are methods exist in the literature. To conclude this chapter, the two methods mentioned above are applied to two utility functions frequently used in economics: Constant Elasticity of Subtitution (CES) function and Cobb-Douglas function.

# PART I

# Chapter 1

# Preliminaries on proximal algorithms and generalized distances

## 1.1 Introduction

Consider the equality-constrained convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper closed convex function, and $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are problem data. The augmented Lagrangian for (1.1) is

$$L_\lambda(x, y) = f(x) + \langle y, Ax - b\rangle + (\lambda/2)\|Ax - b\|^2, \tag{1.2}$$

where $\lambda > 0$ is given and it is called the *penalty parameter*.

An important method to solve the problem (1.1) is the the *method of multipliers* or *augmented Lagrangian method*. Augmented Lagrangians and the method of multipliers for constrained optimization were first proposed in the late 1960s by Hestenes [53, 84] and Powell

[42]. In [68], Rockafellar showed that the method of multipliers is obtained by applying the proximal point algorithm to the dual of (1.1).

Given $y^k \in \mathbb{R}^m$ and $\lambda > 0$, the method of multipliers consists of the following steps:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\ L_\lambda(x, y^k) \\ y^{k+1} &= y^k + \lambda(Ax^{k+1} - b). \end{aligned} \tag{1.3}$$

Below we give a general description that justifies the steps of the method of multipliers. The augmented Lagrangian (1.2) can be viewed as the (unaugmented) Lagrangian associated with the problem

$$\begin{aligned} \text{minimize} \quad & f(x) + (\lambda/2)\|Ax - b\|^2 \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{1.4}$$

This problem is clearly equivalent to the original problem (1.1), since the term added to the function $f$ is zero for any feasible point $x \in \mathbb{R}^n$ . The dual function for (1.4) is

$$g_\lambda(y) = \inf_x L_\lambda(x, y) = -h^*(-A^T y) - b^T y,$$

where

$$h(x) = f(x) + \frac{\lambda}{2}\|Ax - b\|^2,$$

and $y \in \mathbb{R}^m$ is the dual variable or Lagrange multiplier. Then the dual problem of (1.4) is

$$\text{maximize}_y \quad g_\lambda(y). \tag{1.5}$$

Suppose that $g_\lambda$ is a differentiable function, we solve the dual problem (1.5) using the gradient method, that is, given $\lambda > 0$ and $y^k \in \mathbb{R}^m$ the next iteration is obtained by the formula:

$$y^{k+1} = y^k + \lambda \nabla g_\lambda(y^k).$$

Assuming that $f$ is closed, the gradient $\nabla g_\lambda(y)$ can be evaluated as follows:

$$\nabla g_\lambda(y) = Ax^+ - b,$$

where

$$x^+ = \nabla h^*(-A^T y) = \operatorname*{argmin}_x L_\lambda(x, y).$$

This gave rise to the method of multipliers (1.3). One benefit of including the penalty term $(\lambda/2)\|Ax - b\|^2$ in the problem (1.4) is that the method of multipliers converges without assumptions like strict convexity or finiteness of $f$ (see [13]).

Another method to solve the problem (1.1) is the *proximal method of multipliers* (PMM) which is similar to the method multipliers (1.3) but it adds a quadratic proximal term to the minimization problem for $x$, that is, given $(x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\lambda > 0$, the next iteration $(x^{k+1}, y^{k+1})$ is generated by the steps:

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x \; L_\lambda(x, y^k) + \frac{1}{2\lambda}\|x - x^k\|^2 \\ y^{k+1} &= y^k + \lambda(Ax^{k+1} - b). \end{aligned} \tag{1.6}$$

The proximal method of multipliers is globally convergent under convexity of the objective function and existence of saddle points (see [68]). Moreover, Rockafellar [68] proved that the proximal method of multipliers is obtained by applying the proximal point algorithm to the primal-dual formulation of (1.1).

On the other hand, the primal variable in the problem (1.1) may be subject to other constraints such as belonging to the nonnegative orthant in $\mathbb{R}^n$. This results in a convex optimization problem of the form:

$$\min_x \; \left\{ f(x) \mid \; x \in \overline{C} \right\}, \tag{1.7}$$

where $\overline{C}$ denotes the closure of $C \subset \mathbb{R}^n$, an open nonempty convex set, and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper closed convex function.

There are several works in which methods to solve the problem (1.7) are proposed. These methods combine the *proximal point algorithm* [69] and non-quadratic proximal distances. For example, proximal methods using Bregman distances (see, e.g.,[20], [32]) or based on second order homogeneous kernels (see, e.g., [10], [9]). In [8], Auslender and Teboulle proposed a general scheme called interior proximal algorithm (IPA) which unifies several interior proximal methods, and they proved the convergence of this method.

Some problems that appear in several applied areas can be posed in the framework of convex optimization but are more structured than (1.1) and (1.7). In particular, we are interested in the case that the objective function is separable with respect to a splitting of the primal variable into subvectors, that is, the variable of the above problems, called $x$, has been split into two parts, called $x$ and $z$, with the objective function separable across this splitting.

In this thesis, we are interested in developing methods of decomposition, alternating and proximal, to solve a convex optimization problem with the following generic structure:

$$(P) \qquad \min_{x,z} \left\{ f(x) + g(z) \mid Ax + Bz = b,\ x \in \overline{C} \right\},$$

with variables $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, where $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$ and $b \in \mathbb{R}^p$ are problem data. Here $\overline{C}$ denotes the closure of $C \subset \mathbb{R}^n$, an open nonempty convex set. We will assume that $f$ and $g$ are proper closed convex functions. For solving the separable convex problem $(P)$, we will propose decomposition methods based on the method of multipliers, and the interior proximal algorithm (approaches described above).

To achieve the aim previously indicated, it is required to introduce some basic terminology, give some definitions and recall certain fundamental properties, which will be presented in the following sections. In Section 1.2, we will review the *alternating direction method of multipliers* and other decomposition methods that solve convex optimization problems with separable structure. In Section 1.3, we will give the definition of generalized proximal distances, recall some properties of these functions and give some examples. In addition, we will give the definition of induced proximal distances which is key to analyze the convergence of proximal methods.

## 1.2 Decomposition methods

### 1.2.1 Alternating direction method of multipliers

Consider the following convex minimization problem with separable structure:

$$\begin{array}{rl} \text{minimize} & f(x) + g(z) \\ \text{subject to} & Ax + Bz = b \end{array} \tag{1.8}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions, $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ are real matrices, and $b \in \mathbb{R}^p$ is a given vector. The classical Lagrangian function of the problem (1.8) is:

$$L(x, z, y) = f(x) + g(z) + \langle y, Ax + Bz - b\rangle, \tag{1.9}$$

and then its augmented Lagrangian function is:

$$L_\lambda(x, z, y) = L(x, z, y) + (\lambda/2)\|Ax + Bz - b\|^2. \tag{1.10}$$

The method of multipliers (1.3) applied to the problem (1.8) has the form

$$(x^{k+1}, z^{k+1}) = \operatorname*{argmin}_{x,z} L_\lambda(x, z, y^k) \tag{1.11}$$

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{1.12}$$

In the step (1.11), the augmented Lagrangian $L_\lambda$ is minimized jointly with respect to the two primal variables $x$ and $z$. We can realize that these variables are coupled in the quadratic term $\|Ax + Bz - b\|^2$ of the augmented Lagrangian $L_\lambda$ which prevents that the minimization subproblem (1.11) is divided into two separate problems. This means that basic method of multipliers cannot be used for decomposition. We will see how to address this issue next.

The alternating direction method of multipliers (ADM) consists in alternating the minimization of the augmented Lagrangian $L_\lambda$ (1.10) with respect to $x$ and $z$, and then updating the multiplier $y$, which accounts for the term *alternating direction*. That is, given a couple $(z^k, y^k) \in \mathbb{R}^m \times \mathbb{R}^p$ and a positive scalar $\lambda > 0$, the ADM consits of the following iterations:

$$\begin{array}{rcl} x^{k+1} & = & \underset{x}{\operatorname{argmin}}\, L_\lambda(x, z^k, y^k), \\ z^{k+1} & = & \underset{z}{\operatorname{argmin}}\, L_\lambda(x^{k+1}, z, y^k), \\ y^{k+1} & = & y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b), \end{array} \tag{1.13}$$

An important feature of ADM is that functions $f$ and $g$ are treated individually; thus the decomposed subproblems in (1.13) might be significantly easier than the original problem (1.8). Recently, the ADM has received wide attention from a broad spectrum of areas because of its easy implementation and impressive efficiency. We refer to [16, 48] for excellent review papers for the history and applications of ADM.

Assuming that the Lagrangian function $L$ (1.9) has a saddle point $(x^*, z^*, y^*)$, the sequence $\{(x^k, z^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ generated by the ADM satisfies the following convergence properties:

- *Residual convergence.* $Ax^k + Bz^k - b \to 0$ as $k \to \infty$, i.e., the iterates approach feasibility.

- *Objective convergence.* $f(x^k) + g(z^k) \to f(x^*) + g(z^*)$ as $k \to \infty$, i.e., the objective function of the iterates approaches the optimal value.

- *Dual variable convergence.* $y^k \to y^*$ as $k \to \infty$, where $y^*$ is a dual optimal point.

Note that $x^k$ and $z^k$ need not converge to optimal values, although such results can be shown under additional assumptions on the data.

The alternating direction method was originally proposed in the mid-1970s by Glowinski and Marroco [49], and Gabay and Mercier [46]. Gabay [43] showed that the alternating direction method of multipliers (ADM) is a special case of the Douglas-Rachford splitting method for finding a zero of the sum of two monotone operators. Later, Eckstein and Bertsekas [35] showed that the Douglas-Rachford splitting method is a special case of the proximal point algorithm (PPA). Therefore, applications of Douglas-Rachford splitting, such as the ADM for convex programming decomposition, are also special cases of the proximal point algorithm.

**Proximal Alternating Direction Method of Multipliers**

The proximal alternating direction method of multipliers is a decomposition method which consists in regularizing the ADM subproblems with proximal terms. The current literature is dominated by the utilization of quadratic proximal regularization to ensure a more stable and attractive numerical performance (see, e.g.,[51], [34]). For solving the separable convex optimization problem (1.8), the alternating direction method of multipliers with quadratic proximal regularization needs to solve the following subproblems to generate a new iteration:

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x L_\lambda(x, z^k, y^k) + \frac{1}{2\lambda}\|x - x^k\|^2, \\ z^{k+1} &= \operatorname*{argmin}_z L_\lambda(x^{k+1}, z, y^k) + \frac{1}{2\lambda}\|z - z^k\|^2, \\ y^{k+1} &= y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \end{aligned}$$

On the other hand, Yuan and Li [83] applied the logarithmic-quadratic proximal (LQP) terms to regularize the ADM subproblems, so they developed an LQP-based decomposition method for solving a class of variational inequalities and proved the global convergence of the method under standard assumptions. In [65], Li *et al.* study the inexact version of the ADM with the logarithmic-quadratic proximal regularization, and they prove the global convergence and establish worst-case convergence rates for the derived inexact algorithms.

### 1.2.2 Predictor corrector proximal multiplier method and some its variants

In [23], Chen and Teboulle proposed a decomposition method for solving convex programming problems with separable structure which is called *predictor corrector proximal multiplier* method (PCPM). They studied the problem (1.8) with $B = -I$ and $b = 0$ ($I$ is the $m \times m$ identity matrix), that is,

$$\begin{aligned} \text{minimize} \quad & f(x) + g(z) \\ \text{subject to} \quad & Ax = z \end{aligned} \tag{1.14}$$

Given a starting point $(x^0, y^0, z^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, the PCPM method applied to the problem (1.14) consists of the following steps:

**Step 1**. Compute

$$p^{k+1} = y^k + \lambda_k (Ax^k - z^k).$$

**Step 2**. Solve

$$\begin{aligned} x^{k+1} &= \operatorname*{Argmin}_{x \in \mathbb{R}^n} \; f(x) + \langle p^{k+1}, Ax \rangle + \frac{1}{2\lambda_k} \|x - x^k\|^2, \\ z^{k+1} &= \operatorname*{Argmin}_{z \in \mathbb{R}^m} \; g(z) - \langle p^{k+1}, z \rangle + \frac{1}{2\lambda_k} \|z - z^k\|^2. \end{aligned}$$

**Step 3**. Compute

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} - z^{k+1}).$$

Here, $\{\lambda_k\}$ is a sequence of positive scalars.

ADM and PCPM are different decomposition methods. On one hand, the ADM method is based on method of multipliers, and the PCPM in the proximal method of multipliers. On the other hand, the ADM achieves separation of augmented Lagrangian of the problem under study via alternating the primal variables, whereas the PCPM method linearizes the coupling quadratic term of the augmented Lagrangian at the current $k$-iterate.

The augmented Lagrangian associated with the problem (1.14) is:

$$L_\lambda(x, z, y) = f(x) + g(z) + \langle y, Ax - z\rangle + \frac{\lambda}{2}\|Ax - z\|^2. \tag{1.15}$$

The proximal method of multipliers (1.6) for solving the problem (1.14) reads as

$$(x^{k+1}, z^{k+1}) = \operatorname*{argmin}_{x,z} \; L_{\lambda_k}(x, z, y^k) + \frac{1}{2\lambda_k}(\|x - x^k\|^2 + \|z - z^k\|^2), \tag{1.16}$$

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} - z^{k+1}). \tag{1.17}$$

A main disadvantage of this method when used in the context of separable problems like (1.14) is the presence of the coupling quadratic term $\|Ax - z\|^2$ in the augmented Lagrangian $L_{\lambda_k}(x, z, y^k)$, which destroys the given separability and thus prevents to minimize separately the augmented Lagrangian $L_{\lambda_k}(x, z, y^k)$ in $x$ and $z$. To overcome this difficulty, Chen and Teboulle proposed a linearization of the coupling quadratic term in $L_{\lambda_k}(x, z, y^k)$ at the iterate $(x^k, z^k)$.

Set $u(x, z) = \frac{\lambda_k}{2}\|Ax - z\|^2$. Chen and Teboulle applied to (1.16) the following linearization strategy:

$$(x^{k+1}, z^{k+1}) \approx \underset{x,z}{\operatorname{argmin}}\ f(x)+g(z)+\langle y^k, Ax-z\rangle+\nabla u(x^k, z^k)^T(x,z)+\frac{1}{2\lambda_k}(\|x-x^k\|^2+\|z-z^k\|^2), \tag{1.18}$$

where

$$\nabla u(x^k, z^k)^T = (\lambda_k A^T(Ax^k - z^k), -\lambda_k(Ax^k - z^k)).$$

Rearranging the equation (1.18), we get:

$$(x^{k+1}, z^{k+1}) \approx \underset{x,z}{\operatorname{argmin}}\ f(x)+g(z)+\langle y^k+\lambda_k(Ax^k-z^k), Ax-z\rangle+\frac{1}{2\lambda_k}(\|x-x^k\|^2+\|z-z^k\|^2), \tag{1.19}$$

Defining

$$p^{k+1} = y^k + \lambda_k(Ax^k - z^k),$$

we can see that (1.19) is equivalent to the step 2 of the PCPM method. Moreover, Chen and Teboulle prove that the PCPM method is globally convergent and at a linear rate. The assumptions used by them to obtain the linear rate were suggested by the work of Rochafellar [68] to derive the rate of convergence of the proximal method of multipliers.

Other decomposition methods are based on the PCPM method but its subproblems are regularized with generalized proximal terms (see sections 1.2 and 1.3). For example, Kyono and Fukushima [62] proposed a decomposition method which blends the PCPM method and the proximal point algorithm using Bregman distances. Other example, in [6] Auslender and Teboulle proposed a decomposition method for solving variational inequalities and convex optimization problems. The scheme combines the logarithmic-quadratic proximal theory, introduced by them in [10], with the Chen-Teboulle method, and so the method is called *entropic proximal decomposition method* (EPDM). The following convex optimization problem was considered in [6]:

$$\min\ \{f(x) + g(z) \mid Ax + Bz = b,\ x \in \mathbb{R}^n_+,\ z \in \mathbb{R}^m\},$$

and, in this special case, the EPDM algorithm reduces to:

**EPDM Algorithm**. Start with an initial arbitrary triple $(x^0, z^0, y^0) \in \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^p$, and given a positive parameter $\lambda_k$, the EPDM generates the sequence $\{(x^k, z^k, y^k)\} \subset \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^p$ by the following steps:

**Step 1**. Compute

$$p^{k+1} = y^k + (2\theta)^{-1}\lambda_k(Ax^k + Bz^k - b).$$

**Step 2**. Solve

$$\begin{aligned} x^{k+1} &= \operatorname*{Argmin}_{x \geq 0}\ f(x) + \langle p^{k+1}, Ax\rangle + \lambda_k^{-1} d_\varphi(x, x^k), \\ z^{k+1} &= \operatorname*{Argmin}_{z \in \mathbb{R}^m}\ g(z) + \langle p^{k+1}, Bz\rangle + \theta\lambda_k^{-1}\|z - z^k\|^2. \end{aligned}$$

**Step 3**. Compute

$$y^{k+1} = y^k + (2\theta)^{-1}\lambda_k(Ax^{k+1} + Bz^{k+1} - b).$$

Here, the function $d_\varphi$ is the logarithmic-quadratic proximal distance (1.38). Under the only assumption that the primal-dual problem have nonempty solution set, global convergence of the primal-dual sequences produced by the EPDM method is established.

It is important to note that the linearization strategy of the augmented Lagrangian seeks to obtain decomposition methods with simplified subproblems, that is, subproblems that have closed-form solutions or are easier to solve. The combination of decomposition methods based on the augmented Lagrangian method with linearization has resulted in some work such as the split inexact Uzawa method [38], some methods for image reconstruction problems [85] and for Dantzing Selector model [81].

*Remark* 1.1. Another interesting algorithm for convex problems was proposed by Chambolle and Pock [21]. They studied a first-order primal-dual algorithm for convex optimization problems with known saddle-point structure and proved convergence to a saddle-point in finite dimensions. They further showed accelerations of the proposed algorithm to yield optimal rates on easier problems.

## 1.3 Generalized proximal distances

### 1.3.1 Definition and main properties

Remember the convex optimization problem (1.7) (see [17]) given in the introduction of this chapter:

$$\min_{x} \left\{ f(x) \mid x \in \overline{C} \right\},$$

where $\overline{C}$ denotes the closure of $C \subset \mathbb{R}^n$, an open nonempty convex set.

**Definition 1.2.** [8, Definition 2.1] A function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called a proximal distance with respect to an open nonempty convex set $C \subset \mathbb{R}^n$ if for each $v \in C$ we have the following properties:

**($d_1$)** $d(\cdot, v)$ is proper, closed, convex and continuously differentiable on $C$;

**($d_2$)** $\operatorname{dom} d(\cdot, v) \subset \overline{C}$ and $\operatorname{dom} \partial_1 d(\cdot, v) = C$, where $\partial_1 d(\cdot, v)$ denotes the subgradient map of the function $d(\cdot, v)$ with respect to the first variable;

**($d_3$)** $d(\cdot, v)$ is level bounded on $\mathbb{R}^n$, i.e., $\lim_{\|u\| \to \infty} d(u, v) = +\infty$;

**($d_4$)** $d(v, v) = 0$.

We denote by $\mathcal{D}(C)$ the family of functions $d$ satisfying Definition 1.2.

An proximal iterative scheme given by Auslender and Teboulle [8] to solve the problem (1.7) is as follows:

$$x^{k+1} \in \underset{x\in\overline{C}}{\operatorname{argmin}}\, \lambda_k f(x) + d(x, x^k), \tag{1.20}$$

where $d$ is a proximal distance with respect to $C$, and $\lambda_k > 0$ is a positive scalar.

**Proposition 1.3.** *[8, Proposition 2.1] Let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, closed and convex function. Suppose that $F_* = \inf_{u\in\overline{C}} F(u) > -\infty$ and $\operatorname{dom} F \cap C \neq \emptyset$. Let $d \in \mathcal{D}(C)$, and for all $v \in C$ consider the optimization problem*

$$(P(v)) \qquad F_*(v) = \inf_{u\in\overline{C}} F(u) + d(u, v).$$

*Then the optimal set $S(v)$ of $P(v)$ is nonempty and compact, and for each $\epsilon \geq 0$ there exist $u(v) \in C$, $g \in \partial_\epsilon F(u(v))$ such that*

$$g + \nabla_1 d(u(v), v) = 0,$$

*where $\partial_\epsilon F(u(v))$ denotes the $\epsilon$-subdifferential of $F$ at $u(v)$. For such a $u(v) \in C$ we have*

$$F(u(v)) + d(u(v), v) \leq F_*(v) + \epsilon.$$

Thanks to the above proposition, the following algorithm proposed in [8] is well defined.

**Interior Proximal Algorithm (IPA)**

Given $d \in \mathcal{D}(C)$, start with a point $x^0 \in C$, and for $k = 1, 2, ...$ with $\lambda_k > 0$, $\epsilon_k \geq 0$, generate a sequence

$$\{x^k\} \in C \text{ with } g^k \in \partial_{\epsilon_k} f(x^k) \tag{1.21}$$

such that

$$\lambda_k g^k + \nabla_1 d(x^k, x^{k-1}) = 0. \tag{1.22}$$

### 1.3.2 Compatible proximal pairs and convergence of proximal methods

We associate with $d \in \mathcal{D}(C)$ a corresponding induced proximal distance $H$ satisfying some desirable properties needed to analyze proximal-type methods. Next, we present a variant of the definition of induced proximal distance introduced by Auslender and Teboulle [8, Definition 2.2].

**Definition 1.4.** Given an open nonempty convex set $C \subset \mathbb{R}^n$ and $d \in \mathcal{D}(C)$, we say that $(d, H)$ is a compatible proximal pair associated with $\overline{C}$ if $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ is a finite valued function on $\overline{C} \times C$ and for each $a, b \in C$ satisfies:

(i) $H(a, a) = 0$;

(ii) $\langle c - b, \nabla_1 d(b, a)\rangle \leq H(c, a) - H(c, b) - \kappa H(b, a)$, $\forall c \in \overline{C}$, and some fixed $\kappa > 0$;

(iii) For all $c \in \overline{C}$, $H(c, \cdot)$ is level bounded on $C$.

We write $(d, H) \in \mathcal{F}(\overline{C})$ when a triple $[\overline{C}, d, H]$ satisfies the premises of Definition 1.4. We also denote $(d, H) \in \mathcal{F}(C)$ if there exists $H$ which is finite valued on $C \times C$ satisfying $(i)$ and $(ii)$ with $\kappa = 0$, for each $c \in C$.

The motivation behind the construction of compatible proximal pair $(d, H)$ is not as misterious as it might look at first sight. Indeed, notice that the classical proximal algorithm (PA), which corresponds to the special case $C = \overline{C} = \mathbb{R}^n$, $d = 2^{-1}\|x - y\|^2$ and the induced proximal distance $H$ being exactly $d$, clearly satisfies (ii) of the Definition 1.4, thanks to the well-known identity:

$$\|z - x\|^2 = \|z - y\|^2 + \|y - x\|^2 + 2\langle z - y, y - x\rangle.$$

The requested properties for the function $H$ emerge naturally from the analysis of the classical PA and later extended for various specific classes of IPA.

Next, we present the theorem of convergence in limit points of the produced sequence by IPA given in [8], and we include the proof for completeness. To derive the global convergence of the sequence to an optimal solution of the problem (1.7), Auslender and Teboulle needed additional assumptions on the induced proximal distance $H$, akin to the properties of norms. These additional properties and the global convergence theorem will be presented later.

**Lemma 1.5.** *[67] Let $\{\nu_k\}$, $\{\gamma_k\}$ and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying $\nu_{k+1} \leq (1+\gamma_k)\nu_k + \beta_k$ and such that $\sum_{k=1}^{\infty} \beta_k < \infty$, $\sum_{k=1}^{\infty} \gamma_k < \infty$. Then, the sequence $\{\nu_k\}$ converges.*

**Lemma 1.6.** *[67] Let $\{\lambda_k\}$ be a sequence of positive numbers, $\{a_k\}$ a sequence of real numbers, and $b_n = \sigma_n^{-1} \sum_{k=1}^{n} \lambda_k a_k$, where $\sum_{k=1}^{n} \lambda_k$. If $\sigma_n \to \infty$,*

**(i)** $\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup a_n$,

**(ii)** $\lim b_n = a$ *whenever* $\lim a_n = a$.

**Theorem 1.7.** *[8, Theorem 2.1] Let $(d, H) \in \mathcal{F}(C)$ and let $\{x^k\}$ the sequence generated by IPA. Set $\sigma_n = \sum_{k=1}^{n} \lambda_k$. Then the following hold:*

**(i)** $f(x^n) - f(x) \leq \sigma_n^{-1} H(x, x^0) + \sigma_n^{-1} \sum_{k=1}^{n} \sigma_k \epsilon_k, \ \ \forall x \in C.$

**(ii)** *If* $\lim \sigma_n = +\infty$ *and* $\epsilon_k \to 0$ *then* $\liminf f(x^n) = f_*$, *and the sequence* $\{f(x^k)\}$ *converges to* $f_*$ *whenever* $\sum_{k=1}^{\infty} \epsilon_k < \infty$.

**(iii)** *Furthermore, suppose the optimal set $X_*$ of the problem (1.7) is nonempty, and consider the following cases:*

**(a)** *$X_*$ is bounded,*

**(b)** $\sum_{k=1}^{\infty} \lambda_k \epsilon_k < \infty$ *and* $(d, H) \in \mathcal{F}(\overline{C})$.

*Then, under either (a) or (b), the sequence $\{x^k\}$ is bounded with all its limit points in $X_*$.*

*Proof.* **(i)** From (1.22), since $g^k \in \partial_{\epsilon_k} f(x^k)$ we have:

$$\lambda_k(f(x^k) - f(x)) \leq \langle x - x^k, \nabla_1 d(x^k, x^{k-1})\rangle + \lambda_k \epsilon_k, \ \ \forall x \in C. \tag{1.23}$$

Using (ii) of the definition 1.4 at the points $c = x$, $a = x^{k-1}$, $b = x^k$ and with $\kappa = 0$, the above inequality implies that

$$\lambda_k(f(x^k) - f(x)) \leq H(x, x^{k-1}) - H(x, x^k) + \lambda_k \epsilon_k, \ \ \forall x \in C. \tag{1.24}$$

Summing over $k = 1, ..., n$ we obtain:

$$-\sigma_n f(x) + \sum_{k=1}^{n} \lambda_k f(x^k) \leq H(x, x^0) - H(x, x^n) + \sum_{k=1}^{n} \lambda_k \epsilon_k. \tag{1.25}$$

Now setting $x = x^{k-1}$ in (1.24), we obtain

$$f(x^k) - f(x^{k-1}) \leq \epsilon_k. \tag{1.26}$$

Multiplying the latter inequality by $\sigma_{k-1}$ (with $\sigma_0 \equiv 0$) and summing over $k = 1, ..., n$, we obtain, after some algebra,

$$\sigma_n f(x^n) - \sum_{k=1}^{n} \lambda_k f(x^k) \leq \sum_{k=1}^{n} \sigma_{k-1} \epsilon_k. \tag{1.27}$$

Adding this inequality to (1.25) and recalling that $\lambda_k + \sigma_{k-1} = \sigma_k$, it follows that

$$f(x^n) - f(x) \leq \sigma_n^{-1}[H(x, x^0) - H(x, x^n)] + \sigma_n^{-1} \sum_{k=1}^{n} \sigma_k \epsilon_k \ \ \forall x \in C, \tag{1.28}$$

proving (i), since $H(\cdot, \cdot) \geq 0$.

**(ii)** If $\sigma_n \to +\infty$ and $\epsilon_k \to 0$, then dividing (1.25) by $\sigma_n$ and invoking Lemma 1.6 (i), we obtain from (1.25) that

$$\liminf f(x^n) \leq \inf\{f(x)|x \in C\},$$

which together with

$$\inf\{f(x)|x \in C\} \leq f(x^n)$$

implies that

$$\liminf f(x^n) = \inf\{f(x)|x \in C\} = f_*.$$

From (1.26) we have

$$0 \leq f(x^k) - f_* \leq f(x^{k-1}) - f_* + \epsilon_k.$$

Then using Lemma 1.5 it follows that the sequence $\{f(x^k)\}$ converges to $f_*$ whenever $\sum_{k=1}^{\infty} \epsilon_k < \infty$.

**(iii)** Case (a): If $X_*$ is bounded, then $f$ is level bounded over $\overline{C}$, and since the sequence $\{f(x^k)\}$ converges to $f_*$, it follows that the sequence $\{x^k\}$ is bounded. Since $f$ is closed, passing to the limit, and recalling that $\{x^k\} \subset C$, it follows that each limit point is an optimal solution.

Case (b): Here, we suppose that $\sum_{k=1}^{\infty} \lambda_k \epsilon_k < \infty$ and that $(d, H) \in \mathcal{F}(\overline{C})$. Then (1.24) holds for each $x \in \overline{C}$, and in particular for $x \in X_*$, so that

$$H(x, x^k) \leq H(x, x^{k-1}) + \lambda_k \epsilon_k \;\; \forall x \in X_*. \tag{1.29}$$

Summing over $k = 1, ..., n$, we obtain

$$H(x, x^n) \leq H(x, x^0) + \sum_{k=1}^{\infty} \lambda_k \epsilon_k.$$

But, since in this case $H(x, \cdot)$ is level bounded, the last inequality implies that the sequence $\{x^k\}$ is bounded, and thus as in Case (a) it follows that all its limit points are in $X_*$.

□

To establish the *global convergence* of the sequence $\{x^k\}$ to an optimal solution of problem (1.7), we need to make further assumptions on the induced proximal distance $H$, mimicking the behavior of norms.

Let $(d, H) \in \mathcal{F}_+(\overline{C}) \subset \mathcal{F}(\overline{C})$ be such that the function $H$ satisfies the following two additional properties:

$(a_1)$ $\forall y \in \overline{C}$ and $\forall \{y_k\} \subset C$ bounded with $\lim H(y, y_k) = 0$, we have $\lim y_k = y$;

$(a_2)$ $\forall y \in \overline{C}$ and $\forall \{y_k\} \subset C$ converging to $y$, we have $\lim H(y, y_k) = 0$.

With these aditional hypotheses on $H$ we inmediately obtain the global convergence of the interior proximal algorithm IPA.

**Theorem 1.8.** *[8, Theorem 2.2] Let $(d, H) \in \mathcal{F}_+(\overline{C})$ and let $\{x^k\}$ be the sequence generated by IPA. Suppose that the optimal set $X_*$ of (1.4) is nonempty, $\sigma_n = \sum_{k=1}^n \lambda_k \to \infty$, $\sum_{k=1}^\infty \lambda_k \epsilon_k < \infty$, and $\sum_{k=1}^\infty \epsilon_k < \infty$. Then the sequence $\{x^k\}$ converges to an optimal solution of (1.4).*

*Proof.* Let $x \in X_*$. Then, since $(d, H) \in \mathcal{F}_+(\overline{C})$, from (1.29) with $\sum_{k=1}^n \lambda_k \epsilon_k < +\infty$ and Lemma 1.5 we obtain that the sequence $\{H(x, x^k)\}$ converges to some $a(x) \in \mathbb{R}$ $\forall x \in X_*$. Let $x_\infty$ be the limit of a subsequence $\{x^{k_l}\}$. Obviously, from Theorem 1.7, $x_\infty \in X_*$. Then by assumption $(a_2)$ $\lim H(x_\infty, x^{k_l}) = 0$, so that $\lim H(x_\infty, x^k) = 0$, and by assumption $(a_1)$ it follows that the sequence $\{x^k\}$ converges to $x_\infty$. □

### 1.3.3 Examples of compatible proximal pairs

In the literature one can find various proximal distances which are obtained as compatible proximal pairs $(d, H)$. For example, Bregman distances, second order homogeneous proximal distances and double regularizations.

**Bregman distances**

A classical example of proximal distance with respect to $C$ is the Bregman distance $D_h$ : $\mathbb{R}^n \times \mathbb{R}^n \to (-\infty, +\infty]$ defined by:

$$D_h(u,v) = h(u) - h(v) - \langle u - v, \nabla h(v)\rangle, \ \ \forall (u,v) \in \overline{C} \times C,$$

where $h$ is a strictly convex and continuous function on $\overline{C}$, it is continuously differentiable on $C$ and $\operatorname{dom} \nabla h = C$. The function $h$ is called a Bregman function with zone $C$. Every Bregman distance $D_h$ satisfies the following property, called *three point identity* (see [22]):

$$D_h(c,a) = D_h(c,b) - D_h(b,a) + \langle \nabla h(b) - \nabla h(a), c - b\rangle, \ \forall a, b \in C, \ \forall c \in \overline{C}. \tag{1.30}$$

If $(d, H) = (D_h, D_h)$, the three point identity says that this pair satisfies (ii) of the Definition 1.4. The properties (i) and (iii) of the Definition 1.4 come from the definition of Bregman distance. Therefore, $(d, H) = (D_h, D_h) \in \mathcal{F}(\overline{C})$. Some standard examples of Bregman distances are:

1. For $C = \mathbb{R}^n$ and $h(u) = \frac{1}{2}\|u\|^2$ one obtains $D_h(u,v) = \frac{1}{2}\|u - v\|^2$.

2. Let $C = \mathbb{R}^n_{++} = \{u \in \mathbb{R}^n | \ u_i > 0, \ i = 1, ..., n\}$. For $h(u) = \sum_{i=1}^n \ u_i \ \ln(u_i)$ extended with continuity to the boundary of $\mathbb{R}^n_{++}$ with the convention that $0 \ln 0 = 0$, for all $(u,v) \in \mathbb{R}^n_+ \times \mathbb{R}^n_{++}$ we obtain:

$$D_h(u,v) = \sum_{i=1}^n \ u_i \ \ln\left(\frac{u_i}{v_i}\right) + v_i - u_i, \tag{1.31}$$

   which is the Kullback-Leibler distance.

3. Let $C = \mathbb{R}^n_{++}$. For $\alpha \in (0,1)$, consider the family of functions $h_\alpha(u) = \sum_{i=1}^n \ \frac{\alpha u_i - u_i^\alpha}{1-\alpha}$. Then

$$D_{h_\alpha}(u,v) = \sum_{i=1}^n \ v_i^{\alpha-1}\left(v_i + \frac{\alpha}{1-\alpha}u_i\right) - \frac{1}{1-\alpha}u_i^\alpha.$$

An interesting particular case is for $\alpha = 1/2$, which gives $h(u) = \sum_{i=1}^{n} u_i - 2\sqrt{u_i}$ and the distance:

$$D_h(u,v) = \sum_{i=1}^{n} \frac{(\sqrt{u_i} - \sqrt{v_i})^2}{\sqrt{v_i}}, \quad \forall (u,v) \in \mathbb{R}_+^n \times \mathbb{R}_{++}^n.$$

*Remark* 1.9. We denote by $S^n$ the linear space of symmetric real matrices equipped with the trace inner product $\langle x, y\rangle = \operatorname{tr}(xy)$ and $\|x\| = \sqrt{\operatorname{tr}(x^2)}$, for all $x, y \in S^n$. The cone of $n \times n$ symetric positive semidefinite (positive definite) matrices is denoted by $S_+^n$ ($S_{++}^n$). Let $C = S_{++}^n$ and $\bar{C} = S_+^n$. Let $h_1 : S_+^n \to, h_1(x) = \operatorname{tr}(x \log x)$ and $h_2 : S_{++}^n \to, h_2(x) = -\operatorname{tr}(\log x) = -\log\det(x)$ ($\det(x)$ is the determinant of the matrix $x$). For any $y \in S_{++}^n$, let

$$\begin{aligned} d_1(x,y) &= \operatorname{tr}(x\log x - x\log y + y - x) \text{ with } \operatorname{dom} d_1(\cdot, y) = S_+^n, && (1.32)\\ d_2(x,y) &= \operatorname{tr}(-\log x + \log y + xy^{-1}) - n && (1.33)\\ &= -\log\det(xy^{-1}) + \operatorname{tr}(xy^{-1}) - n \text{ with } \operatorname{dom} d_2(\cdot, y) = S_{++}^n. && (1.34) \end{aligned}$$

The proximal distances $d_1$ and $d_2$ are Bregman type corresponding to $h_1$ and $h_2$, respectively, and were proposed by Doljansky and Teboulle in [30], who derived convergence results for the associated IPA. From the results of [30] it is easy to see that $d_i \in \mathcal{D}(C)$, $i = 1, 2$, and with $H(x,y) = d_i(x,y)$ it follows that $(d_1, H) \in \mathcal{F}(S_+^n)$ and $(d_2, H) \in \mathcal{F}(S_{++}^n)$ so that the convergence results of [30] are recovered through Theorem 1.7. However, as noticed in a counterexample [30, Example 4.1], property $(a_2)$ does not hold even for $d_1$ and therefore $(d_i, H) \notin \mathcal{F}_+(\bar{C})$, $i = 1, 2$. Consequently, Theorem 1.8 does not apply, i.e., global convergence to an optimal solution cannot be guaranted.

### Second order homogeneous proximal distances

Other proximal distances are the second order homogeneous proximal distances. Let $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a closed, convex, proper function such that $\operatorname{dom}\varphi \subseteq \mathbb{R}_+$ and $\operatorname{dom}\partial\varphi = \mathbb{R}_{++}$. We suppose that $\varphi$ is strictly convex on its domain, and $\varphi$ is $\mathcal{C}^2$ on $\mathbb{R}_{++}$ with $\varphi(1) = \varphi'(1) = 0$. We denote by $\Phi$ the class of such kernels. The class $\Phi_2$ (subclass of $\Phi$) consists of kernels

$p \in \Phi$ which satisfy

$$p''(1)\left(1-\frac{1}{t}\right) \leq p'(t) \leq p''(1)(t-1), \forall t > 0. \tag{1.35}$$

Given $\varphi \in \Phi$, the $\varphi$-divergence proximal distance is defined by

$$d_\varphi(u,v) = \sum_{j=1}^{n} v_j^2 \, \varphi\left(\frac{u_j}{v_j}\right), \tag{1.36}$$

and it follows that $d_\varphi \in \mathcal{D}(C)$ with $C = \mathbb{R}^n_{++}$ since $\varphi$ is coercive. The second order homogeneous proximal distances are defined by (1.36) where

$$\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2, \tag{1.37}$$

with $\nu \geq \mu p''(1) > 0$ and $p \in \Phi_2$. Therefore the second order homogeneous proximal distances have the form:

$$d_\varphi(u,v) = \mu d_p(u,v) + \frac{\nu}{2}\|u-v\|^2,$$

and due to (1.37) and (1.35) it has been showed the following key inequality [9, Lemma 3.4 and formula 3.1 there]

$$\langle c-b, \nabla_1 d_\varphi(b,a)\rangle \leq \left(\frac{\nu+\mu p''(1)}{2}\right)(\|c-a\|^2 - \|c-b\|^2) - \left(\frac{\nu-\mu p''(1)}{2}\right)\|b-a\|^2,$$

for all $a, b \in \mathbb{R}^n_{++}$ and $c \in \mathbb{R}^n_+$. Therefore with $H(u,v) = \left(\frac{\nu+\mu p''(1)}{2}\right)\|u-v\|^2$ it follows that $(d_\varphi, H)$ is a compatible proximal pair associated with $\overline{C} = \mathbb{R}^n_+$, that is, $(d_\varphi, H) \in \mathcal{F}(\mathbb{R}^n_+)$.

In particular, $p(t) = -\log(t) + t - 1$ gives the so-called Logarithmic-Quadratic Proximal distance (see [10]) and then

$$d_\varphi(u,v) = \sum_{i=1}^{n} \mu\left(v_i^2 \log\left(\frac{v_i}{u_i}\right) + u_i v_i - v_i^2\right) + \frac{\nu}{2}(u_i - v_i)^2, \forall u, v \in \mathbb{R}_{++}. \tag{1.38}$$

### Double regularizations

Let $B$ be a box set of the form

$$B = \{x \in \mathbb{R}^n | \ a_i \leq x_i \leq b_i, \ i = 1, ..., n\},$$

where $a_i \in [-\infty, +\infty)$ and $b \in (\infty, +\infty]$, $a \leq b$. A *double regularitazion* for the box $B$ is a separable function $\widetilde{D} : B \times \mathrm{int}B \to \mathbb{R}$ of the form:

$$\widetilde{D}(u, v) = \sum_{i=1}^{n} \widetilde{d}_i(u_i, v_i) = \sum_{i=1}^{n} d_i(u_i, v_i) + \frac{\nu}{2}(u_i - v_i)^2, \tag{1.39}$$

where $\nu \geq 1$ is a scalar and the individual $\widetilde{d}_i$, $i = 1, ..., n$, conform to the following assumptions [71, Assumptions 2.1.1-2.1.3]:

- For all $v_i \in (a_i, b_i)$, $\widetilde{d}_i(\cdot, v_i)$ is closed and strictly convex, with its minimum at $v_i$. Moreover, $\mathrm{int\,dom}\, \widetilde{d}_i(\cdot, v_i) = (a_i, b_i)$.
- $\widetilde{d}_i$ is differentiable with respect to its first argument on $(a_i, b_i) \times (a_i, b_i)$, and this partial derivative is continuous at all points of the form $(u_i, u_i) \in (a_i, b_i) \times (a_i, b_i)$.
- For all $v_i \in (a_i, b_i)$, $\widetilde{d}_i(\cdot, v_i)$ is essentially smooth.

Assume that the first term of $\widetilde{D}$, that is, $\sum_{i=1}^{n} d_i(u_i, v_i)$ is coercive, meaning that its gradient with respect to $u$ becomes infinite as $u$ approaches the boundary of $B$. We can see that $\widetilde{D}$ is a proximal distance with respect to $\mathrm{int}\, B$. Moreover, we assume that the individual terms $d_i$ satisfy the following assumption given in [71]:

For $i = 1, \ldots, n$, let $d_i : \mathbb{R} \times (a_i, b_i) \to (-\infty, \infty]$ and $u_i, v_i \in (a_i, b_i)$. Then,

- If $a_i$ and $b_i$ are both finite,

$$\frac{(u_i - v_i)(v_i - a_i)}{u_i - a_i} \leq d_i'(u_i, v_i) \leq \frac{(u_i - v_i)(b_i - v_i)}{b_i - u_i}.$$

- Otherwise, we take the respective limits as $a_i \to -\infty$ or $b_i \to \infty$ in the above relation.

– If only $a_i$ is finite:

$$\frac{(u_i - v_i)(v_i - a_i)}{u_i - a_i} \leq d_i'(u_i, v_i) \leq u_i - v_i.$$

– If only $b_i$ is finite:

$$u_i - b_i \leq d_i'(u_i, v_i) \leq \frac{(u_i - v_i)(b_i - v_i)}{b_i - u_i}.$$

– $(a_i, b_i) = \mathbb{R}$

$$d_i'(u_i, v_i) = u_i - v_i.$$

This assumption on $d_i$ generalizes the assumption (1.35) on $p \in \Phi_2$.

**Lemma 1.10.** *[37, Lemma 1] Suposse $\widetilde{D}$ is a double regularization for the box $B$ with regularization factor $\nu \geq 1$. Under the last assumption we have*

$$\langle c - b, \nabla_1 \widetilde{D}(b, a) \rangle \leq \frac{\nu + 1}{2} \left( \|c - a\|^2 - \|c - b\|^2 \right) - \frac{\nu - 1}{2} \|a - b\|^2, \tag{1.40}$$

*for any $c \in B$ and $a, b \in \operatorname{int} B$.*

We see that if $\nu > 1$ and $H(u, v) = \frac{\nu+1}{2}\|u - v\|^2$ it follows that $(\widetilde{D}, H) \in \mathcal{F}(B)$. For examples of double regularization see [71].

# Chapter 2

# Alternanting direction method with interior proximal regularization

## 2.1 Introduction

In this chapter, we develop a special alternating direction method (ADM) for solving iteratively a convex optimization problem with separable structure as follows:

$$(P) \qquad \min_{x,z} \left\{ f(x) + g(z) \mid Ax + Bz = b,\ x \in \overline{C} \right\},$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions, $A$ and $B$ are $p \times n$ and $p \times m$ real matrices, respectively, and $b \in \mathbb{R}^p$ is a given vector. Here $\overline{C}$ denotes the closure of $C \subset \mathbb{R}^n$, an open nonempty convex set; a typical example being the strictly positive orthant $C = \mathbb{R}^n_{++}$ so that $x \in \overline{C} = \mathbb{R}^n_+$ corresponds to $x \geq 0$.

Under the following natural condition:

$$\operatorname{dom} f \cap C \neq \emptyset,$$

we are interested in designing a method which exploits the separable structure of $(P)$. With such a motivation, we propose the following iterative algorithm for $(P)$: given $(x^k, z^k, y^k) \in C \times \mathbb{R}^m \times \mathbb{R}^p$, compute consecutively

$$
\begin{aligned}
x^{k+1} &\approx \operatorname*{argmin}_{x \in \overline{C}} \; f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k), && (2.1)\\
z^{k+1} &\approx \operatorname*{argmin}_{z \in \mathbb{R}^m} \; g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, && (2.2)\\
y^{k+1} &= y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). && (2.3)
\end{aligned}
$$

Here $\lambda > 0$ is a positive scalar and $d(\cdot, x^k)$ is a given *proximal distance* with respect to $C$ (see chapter 1, definition 1.2). The latter enforces that $x^{k+1} \in C$, that is, the $x$-iterate lies in the interior of the additional constraints set $\overline{C}$. Two popular choices for $d$ include either a Bregman distance (see, e.g., [19], [72]) or a proximal distance based on second order homogeneous kernels (see, e.g., [10], [9]). Another possible choice is the double regularization technique introduced by [71], which extends the notion of second order homogeneous kernels. Any of these choices will be covered by our analysis.

In order to explain the ideas behind the iterative scheme given by (2.1)-(2.3), let us introduce the restricted Lagrangian function $\ell : \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ given by

$$\ell(x, z, y) = f(x) + g(z) + \langle y, Ax + Bz - b\rangle, \tag{2.4}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{R}^p$. Throughout this paper, we assume the following: There exists a saddle point $(x^*, z^*, y^*) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, that is,

$$\ell(x^*, z^*, y) \leq \ell(x^*, z^*, y^*) \leq \ell(x, z, y^*), \;\; \forall (x, z, y) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p. \tag{2.5}$$

In this setting, if we introduce the augmented Lagrangian $\ell_\lambda : \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ by

$$\ell_\lambda(x, z, y) = \ell(x, z, y) + \frac{\lambda}{2}\|Ax + Bz - b\|^2, \tag{2.6}$$

then, from a given $(x^k, z^k, y^k) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, the classical ADM produces iterates via the following consecutive steps:

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_{x\in\overline{C}} \ell_\lambda(x, z^k, y^k) = \operatorname*{argmin}_{x\in\overline{C}}\ f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2, && (2.7)\\ z^{k+1} &= \operatorname*{argmin}_{z\in\mathbb{R}^m} \ell_\lambda(x^{k+1}, z, y^k) = \operatorname*{argmin}_{z\in\mathbb{R}^m}\ g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2, && (2.8)\\ y^{k+1} &= \operatorname*{argmax}_{y\in\mathbb{R}^p}\ \ell_\lambda(x^{k+1}, z^{k+1}, y) - \frac{1}{2\lambda}\|y - y^k\|^2 = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). && (2.9) \end{aligned}$$

ADM was originally proposed by [46], and since then, it has received intensive attention for others researchers; see, e.g., [76]. This method is closely related to the Douglas-Rachford operator splitting algorithm which solves monotone inclusion problems; see, e.g., [31], [25], [74], [15] (see [24, 77, 18, 44] for others important operator splitting algorithms).

On the other hand, the conceptual primal iterations given by (2.1) and (2.2) can be equivalently written as follows:

$$\begin{aligned} x^{k+1} &\approx \operatorname*{argmin}_{x\in C}\ \ell_\lambda(x, z^k, y^k) + \frac{1}{2\lambda} d(x, x^k),\\ z^{k+1} &\approx \operatorname*{argmin}_{z\in\mathbb{R}^m}\ \ell_\lambda(x^{k+1}, z, y^k) + \frac{1}{2\lambda}\|z - z^k\|^2. \end{aligned}$$

In this alternate proximal procedure, minimization steps are performed consecutively on the primal variables, first with respect to $x$ by using a *regularized interior proximal* term so that all the iteration points lie in the interior of the feasible set (see, e.g., [8]), and then with respect to $z$ using a standard quadratic proximal metric. Therefore, we call the iterative scheme given by (2.1), (2.2) and (2.3) the *regularized interior proximal alternating direction method* (RIPADM).

The main theoretical result in this chapter is a general convergence result for a special approximate version of RIPADM (*cf.* Theorem 2.2), bridging two different areas of proximal algorithm theory: ADM on the one hand, and generalized nonquadratic proximal distance on the other. To the best of our knowledge, the only partial positive result in this direction is the recent work by [83] (see [65] for a version with inexact computations), where these authors investigate a particular PADM for variational inequalities under non-negativity constraints in which the regularizing proximal term in the ADM subproblems is induced by the logarithmic-quadratic proximal (LQP) kernel, previously developed in [10]. In the case of optimization problems, the LQP approach can be viewed as a particular instance of our general setting.

Along a complementary line of research, an entropic proximal decomposition method (EPDM) has been introduced by [6] for solving general variational inequality problems with particular separable structure. The EPDM combines the LQP theory for non-negativity constraints with a predictor-corrector proximal multiplier method developed previously by [23]. EPDM differs from RIPADM as the predicted multiplier estimate is used to compute proximal steps *in parallel* for both primal variables, thanks to the fact that the decoupling is obtained by a sort of linearization of the squared Euclidean norm in the augmented Lagrangian at the current iterate, following the strategy proposed by [73]. Thus, EPDM is a parallel method in the primal variables, which is different from the ADM approach (the latter is closer to a Gauss-Seidel scheme).

This chapter is organized as follows. In section 2, we give an approximate version of the RIPADM scheme (2.1)-(2.2), in which the stopping criteria for the inner inexact computations are fairly practical to check. In section 3, we prove our main full convergence result for the inexact version of RIPADM.

## 2.2 The inexact RIPADM algorithm

Consider problem $(P)$ and assume that $\operatorname{dom} f \cap C \neq \emptyset$. Let $d \in \mathcal{D}(C)$ and $\lambda$ be a positive scalar. Starting from a point $(x^0, z^0, y^0) \in C \times \mathbb{R}^m \times \mathbb{R}^p$, we generate the sequence $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$, and sequences of errors $\{a^k\}$ and $\{b^k\}$ via the following steps:

**Step 1.** Find $(x^{k+1}, a^{k+1}) \in C \times \mathbb{R}^n$ with $r^{k+1} \in \partial f(x^{k+1})$ solving:

$$a^{k+1} = r^{k+1} + A^t[y^k + \lambda(Ax^{k+1} + Bz^k - b)] + \frac{1}{2\lambda}\nabla_1 d(x^{k+1}, x^k), \tag{2.10}$$

where the error $a^{k+1}$ satisfies the conditions

$$\|a^{k+1}\| \le \epsilon_{k+1}, \quad \|a^{k+1}\|\|x^{k+1}\| \le \eta_{k+1}. \tag{2.11}$$

**Step 2.** Find $(z^{k+1}, b^{k+1}) \in \mathbb{R}^m \times \mathbb{R}^m$ with $\overline{r}^{k+1} \in \partial g(z^{k+1})$ solving:

$$b^{k+1} = \overline{r}^{k+1} + B^t[y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b)] + \frac{1}{\lambda}(z^{k+1} - z^k), \tag{2.12}$$

where the error $b^{k+1}$ satisfies the conditions

$$\|b^{k+1}\| \le \bar{\epsilon}_{k+1}, \quad \|b^{k+1}\|\|z^{k+1}\| \le \overline{\eta}_{k+1}. \tag{2.13}$$

**Step 3.** Compute

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{2.14}$$

The positive scalars $\epsilon_k, \eta_k, \bar{\epsilon}_k, \overline{\eta}_k > 0$ satisfy the following conditions:

$$\sum_{k=1}^{\infty} \epsilon_k < \infty, \quad \sum_{k=1}^{\infty} \eta_k < \infty, \quad \sum_{k=1}^{\infty} \bar{\epsilon}_k < \infty, \quad \text{and} \quad \sum_{k=1}^{\infty} \overline{\eta}_k < \infty. \tag{2.15}$$

*Remark* 2.1. In the exact case where $a_k = 0 \in \mathbb{R}^n$ and $b_k = 0 \in \mathbb{R}^m$ for all $k$, steps 1 and 2 in the previous scheme amount to find exact solutions to the minimization problems in (2.1) and (2.2), respectively. Due to Proposition 1.3, such an exact RIPADM is well defined.

## 2.3 Global convergence of the inexact RIPADM

Let us return to problem $(P)$. Consider the Lagrangian function $\ell$ defined by (2.4). We suppose:

**Assumption A**

$(A_1)$ $\operatorname{dom} f \cap C \neq \emptyset$.

$(A_2)$ There exists a saddle point $(x^*, z^*, y^*) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ (see (2.5)).

**Assumption B**

There exists a compatible proximal pair $(d, H)$ associated with $\overline{C}$ which verifies:

$(B_1)$ If $\{u^k\} \subset \overline{C}$ and $\{v^k\} \subset C$ are sequences such that

$$\lim_k H(u^k, v^k) = 0,$$

and one of the sequences ($\{u^k\}$ or $\{v^k\}$) converges, then the other one also converges to the same limit.

$(B_2)$ The function $H : \overline{C} \times C \to \mathbb{R}$ extends by continuity to $\overline{C} \times \overline{C}$.

$(B_3)$ $\forall u \in \overline{C}$ and $\forall \{u^k\} \subset C$ converging $u$, we have $\lim_{k\to\infty} H(u, u^k) = 0$.

If $d = \widetilde{D}$ is a double regularization, it is possible to take $H = \frac{\nu+1}{2}\|u - v\|^2$ with $\nu > 1$ (see lemma 1.10), then the pair $(d, H)$ satisfies the assumption B. On the other hand, [72] proved that the Bregman distances satisfy the condition $(B_1)$ (see Theorem 2.4 of [72]). Moreover, the Bregman distances satisfy the condition $(B_3)$ but not necessarily the condition $(B_2)$.

We will prove some results related to a primal-dual sequence generated by the RIPADM method:

1. Under Assumption A, the primal-dual sequence $(x^k, z^k, y^k)$ is bounded.

2. Additionally, if the proximal pair $(d, H)$ verify $(B_1)$, all limit points of $\{(x^k, z^k)\}$ are solutions of the problem (P), for example this is the case of Bregman distances.

3. Moreover, if the proximal pair $(d, H)$ satisfy the Assumption B (for example double regularization), the sequence globally converges to a saddle point of the Lagrangian function $\ell$ of the problem (P).

Below we state our principal result:

**Theorem 2.2.** *Let $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$ be a primal-dual sequence generated by the RIPADM method (2.10)-(2.14).*

**(i)** *Under Assumptions A and $(B_1)$, all limit points of the primal sequence $\{(x^k, z^k)\}$ are solutions of the problem $(P)$.*

**(ii)** *If moreover the Assumptions $(B_2)$ and $(B_3)$ are satisfied, the primal-dual sequence $\{(x^k, z^k, y^k)\}$ globally converges to a saddle point $(x^\infty, z^\infty, y^\infty)$ of the Lagrangian $\ell$ of the problem $(P)$.*

We have divided the proof of Theorem 2.2 into a sequence of lemmas and propositions.

We begin stating two lemmas, the first one is a classical result on scalar sequences, and the second one gives us two saddle point type inequalities. These inequalities are used in the proof of the Proposition 2.5 below which asserts that the sequence $\{(x^k, z^k, y^k)\}$ generated by RIPADM is bounded if at least one saddle point $(x^*, z^*, y^*)$ of $\ell$ exists.

Finally, under Assumptions A and B, we prove that there exists a unique limit point $(x^\infty, z^\infty, y^\infty)$ of the primal-dual sequence $\{(x^k, z^k, y^k)\}$ and so we get the global convergence of our algorithm.

**Lemma 2.3.** *[67, Section 2.2] Suppose $\{w_k\}$, $\{\beta_k\} \subset \mathbb{R}$ are sequences such that $\{w_k\}$ is bounded below, $\sum_{k=0}^{\infty} \beta_k$ exists and is finite, and the recursion $w_{k+1} \leq w_k + \beta_k$ holds for all $k$. Then $\{w_k\}$ is convergent.*

**Lemma 2.4.** *Let $(d, H) \in \mathcal{F}(\overline{C})$. Consider the sequence $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$, and errors sequences $\{a^k\} \subset \mathbb{R}^n$ and $\{b^k\} \subset \mathbb{R}^m$ being generated by the RIPADM algorithm given by (2.10)-(2.14). Under Assumption $(A_1)$, for all $x \in \mathrm{dom}(f)$, $z \in \mathrm{dom}(g)$, $y \in \mathbb{R}^p$, we have:*

$$
\begin{aligned}
\ell(x^{k+1}, z^{k+1}, y^k) - \ell(x, z, y^k) \quad \leq \quad & \frac{\lambda}{2}(\|Ax + Bz^k - b\|^2 - \|Ax^{k+1} + Bz^k - b\|^2 \\
& -\|Ax^{k+1} - Ax\|^2) + \frac{\lambda}{2}(\|Ax^{k+1} + Bz - b\|^2 \\
& -\|Ax^{k+1} + Bz^{k+1} - b\|^2 - \|Bz^{k+1} - Bz\|^2) \\
& +\frac{1}{2\lambda}\left(H(x, x^k) - H(x, x^{k+1}) - \kappa H(x^{k+1}, x^k)\right) \\
& +\frac{1}{2\lambda}\left(\|z - z^k\|^2 - \|z - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2\right) \\
& +\langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle.
\end{aligned} \tag{2.16}
$$

*and*

$$
\ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) \quad = \quad \frac{1}{2\lambda}\left(\|y - y^k\|^2 - \|y - y^{k+1}\|^2 + \|y^{k+1} - y^k\|^2\right) \tag{2.17}
$$

*Proof. Lemma 2.4.* From steps (2.10) and (2.12) of our algorithm, since $r^{k+1} \in \partial f(x^{k+1})$ and $\overline{r}^{k+1} \in \partial g(z^{k+1})$ we have:

$$
f(x^{k+1}) + \left\langle \frac{-1}{2\lambda}\nabla_1 d(x^{k+1}, x^k) - A^t[y^k + \lambda(Ax^{k+1} + Bz^k - b)] + a^{k+1}, x - x^{k+1} \right\rangle \quad \leq \quad f(x), \tag{2.18}
$$

and

$$
g(z^{k+1}) + \left\langle \frac{-1}{\lambda}(z^{k+1} - z^k) - B^t[y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b)] + b^{k+1}, z - z^{k+1} \right\rangle \quad \leq \quad g(z), \tag{2.19}
$$

for all $x \in \mathrm{dom}(f)$ and $z \in \mathrm{dom}(g)$.

Adding and rearranging these inequalities, we obtain

$$\begin{aligned} f(x^{k+1}) + g(z^{k+1}) + \langle y^k, Ax^{k+1} + Bz^{k+1}\rangle & \\ - \big(f(x) + g(z) + \langle y^k, Ax + Bz\rangle\big) \;\; & \leq \;\; \lambda\langle Ax^{k+1} + Bz^k - b, Ax - Ax^{k+1}\rangle \\ & \quad + \lambda\langle Ax^{k+1} + Bz^{k+1} - b, Bz - Bz^{k+1}\rangle \\ & \quad + \frac{1}{2\lambda}\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \\ & \quad + \frac{1}{\lambda}\langle z^{k+1} - z^k, z - z^{k+1}\rangle \\ & \quad + \langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle. \end{aligned}$$

Considering the definition of the Lagrangian function $\ell$ and using the relation $\|m-n\|^2 = \|m\|^2 + \|n\|^2 - 2\langle m, n\rangle$ in the last inequality, we obtain

$$\begin{aligned} \ell(x^{k+1}, z^{k+1}, y^k) - \ell(x, z, y^k) \;\; & \leq \;\; \frac{\lambda}{2}(\|Ax + Bz^k - b\|^2 - \|Ax^{k+1} + Bz^k - b\|^2 \\ & \quad - \|Ax^{k+1} - Ax\|^2) + \frac{\lambda}{2}(\|Ax^{k+1} + Bz - b\|^2 \\ & \quad - \|Ax^{k+1} + Bz^{k+1} - b\|^2 - \|Bz^{k+1} - Bz\|^2) \\ & \quad + \frac{1}{2\lambda}\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \\ & \quad + \frac{1}{2\lambda}\left(\|z - z^k\|^2 - \|z^{k+1} - z^k\|^2 - \|z - z^{k+1}\|^2\right) \\ & \quad + \langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle. \end{aligned} \tag{2.20}$$

Using (ii) of Definition 1.4 with $a = x^k$, $b = x^{k+1}$ and $c = x$, we have

$$\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \leq H(x, x^k) - H(x, x^{k+1}) - \kappa H(x^{k+1}, x^k). \tag{2.21}$$

From (2.20) and (2.21), we obtain the inequality (2.16).

Finally we prove the equality (2.17). We have

$$\ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) = \langle y - y^k, Ax^{k+1} + Bz^{k+1} - b\rangle, \tag{2.22}$$

but from (2.14) we have $Ax^{k+1} + Bz^{k+1} - b = \frac{1}{\lambda}(y^{k+1} - y^k)$, then

$$\begin{aligned} \ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) &= \frac{1}{\lambda}\langle y - y^k, y^{k+1} - y^k\rangle \\ &= \frac{1}{2\lambda}\left(\|y - y^k\|^2 - \|y - y^{k+1}\|^2 + \|y^{k+1} - y^k\|^2\right). \end{aligned}$$

□

**Proposition 2.5.** *Suppose that the hypothesis of Lemma 2.4 holds. Then:*

**(i)** *For any saddle point $(\overline{x}, \overline{z}, \overline{y})$ of $\ell$, the sequence $\{E_k(\overline{x}, \overline{z}, \overline{y})\}$ is convergent where*

$$E_k(\overline{x}, \overline{z}, \overline{y}) = \frac{1}{2\lambda}\left(H(\overline{x}, x^k) + \|z^k - \overline{z}\|^2 + \|y^k - \overline{y}\|^2\right) + \frac{\lambda}{2}\|B(z^k - \overline{z})\|^2. \tag{2.23}$$

**(ii)** *If at least one saddle point $(\overline{x}, \overline{z}, \overline{y})$ of $\ell$ exists then the sequence $\left\{(x^k, z^k, y^k)\right\}$ is bounded in $C \times \mathbb{R}^m \times \mathbb{R}^p$, the quantities $H(x^{k+1}, x^k)$, $\|z^{k+1} - z^k\|^2$ and $\|Ax^{k+1} + Bz^k - b\|^2$ are summable, hence they vanish as $k$ goes to $+\infty$.*

*Proof. Proposition 2.5.* Let $(\overline{x}, \overline{z}, \overline{y}) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ be a saddle point of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, that is,

$$\ell(\overline{x}, \overline{z}, y) \leq \ell(\overline{x}, \overline{z}, \overline{y}) \leq \ell(x, z, \overline{y}), \forall x \in \overline{C},\ \forall z \in \mathbb{R}^m,\ \forall y \in \mathbb{R}^p.$$

Then we have:

$$\ell(\overline{x}, \overline{z}, y^k) - \ell(x^{k+1}, z^{k+1}, \overline{y}) \leq 0. \tag{2.24}$$

Adding the inequality (2.16) with $x = \overline{x}$ and $z = \overline{z}$, and (2.24) we get

$$\begin{aligned}
\ell(x^{k+1}, z^{k+1}, y^k) - \ell(x^{k+1}, z^{k+1}, \overline{y}) \quad &\leq \frac{\lambda}{2}(\|A\overline{x} + Bz^k - b\|^2 - \|Ax^{k+1} + Bz^k - b\|^2) \\
&+\frac{\lambda}{2}(-\|Ax^{k+1} - A\overline{x}\|^2 + \|Ax^{k+1} + B\overline{z} - b\|^2) \\
&+\frac{\lambda}{2}(-\|Ax^{k+1} + Bz^{k+1} - b\|^2 - \|Bz^{k+1} - B\overline{z}\|^2) \\
&+\frac{1}{2\lambda}\left(H(\overline{x}, x^k) - H(\overline{x}, x^{k+1}) - \kappa H(x^{k+1}, x^k)\right) \\
&+\frac{1}{2\lambda}(\|\overline{z} - z^k\|^2 - \|z^{k+1} - z^k\|^2 - \|\overline{z} - z^{k+1}\|^2) \\
&+\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle.
\end{aligned} \tag{2.25}$$

From the equality (2.17) of the Lemma 2.4 with $y = \overline{y}$ we have

$$\ell(x^{k+1}, z^{k+1}, y^k) - \ell(x^{k+1}, z^{k+1}, \overline{y}) \;=\; \frac{1}{2\lambda}\left(\|\overline{y} - y^{k+1}\|^2 - \|\overline{y} - y^k\|^2 - \|y^{k+1} - y^k\|^2\right). \tag{2.26}$$

Replacing the left side of (2.25) with (2.26), and since $A\overline{x} + B\overline{z} = b$ we obtain:

$$\begin{aligned}
\frac{1}{2\lambda}(\|\overline{y} - y^{k+1}\|^2 - \|\overline{y} - y^k\|^2 & \\
-\|y^{k+1} - y^k\|^2) \;\leq\; & \frac{\lambda}{2}\left(\|Bz^k - B\overline{z}\|^2 - \|Ax^{k+1} + Bz^k - b\|^2\right) \\
&+\frac{\lambda}{2}\left(-\|Ax^{k+1} + Bz^{k+1} - b\|^2 - \|Bz^{k+1} - B\overline{z}\|^2\right) \\
&+\frac{1}{2\lambda}\left(H(\overline{x}, x^k) - H(\overline{x}, x^{k+1}) - \kappa H(x^{k+1}, x^k)\right) \\
&+\frac{1}{2\lambda}\left(\|\overline{z} - z^k\|^2 - \|z^{k+1} - z^k\|^2 - \|\overline{z} - z^{k+1}\|^2\right) \\
&+\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle,
\end{aligned}$$

and using the equation (2.14), i.e., $y^{k+1} - y^k = \lambda(Ax^{k+1} + Bz^{k+1} - b)$, results:

$$\begin{aligned} \frac{1}{2\lambda}(\|\overline{y} - y^{k+1}\|^2 - \|\overline{y} - y^k\|^2) \;\; \leq \;\; & \frac{\lambda}{2}\left(\|Bz^k - B\overline{z}\|^2 - \|Ax^{k+1} + Bz^k - b\|^2\right) \\ & -\frac{\lambda}{2}\|Bz^{k+1} - B\overline{z}\|^2 \\ & +\frac{1}{2\lambda}\left(H(\overline{x}, x^k) - H(\overline{x}, x^{k+1}) - \kappa H(x^{k+1}, x^k)\right) \\ & +\frac{1}{2\lambda}\left(\|\overline{z} - z^k\|^2 - \|z^{k+1} - z^k\|^2 - \|\overline{z} - z^{k+1}\|^2\right) \\ & +\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle, \end{aligned} \tag{2.27}$$

Set

$$E_k(\overline{x}, \overline{z}, \overline{y}) = \frac{1}{2\lambda}\left(H(\overline{x}, x^k) + \|z^k - \overline{z}\|^2 + \|y^k - \overline{y}\|^2\right) + \frac{\lambda}{2}\|Bz^k - B\overline{z}\|^2.$$

Rearranging the last inequality we obtain:

$$\begin{aligned} E_{k+1}(\overline{x}, \overline{z}, \overline{y}) + \frac{\kappa}{2\lambda}H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 & \\ +\frac{\lambda}{2}\|Ax^{k+1} + Bz^k - b\|^2 \;\; \leq \;\; & E_k(\overline{x}, \overline{z}, \overline{y}) + \langle a^{k+1}, x^{k+1} - \overline{x}\rangle \\ & +\langle b^{k+1}, z^{k+1} - \overline{z}\rangle. \end{aligned} \tag{2.28}$$

The sum $\sum_{k=0}^{\infty} \langle a^{k+1}, x^{k+1} - \overline{x}\rangle$ exists and is finite. In effect, we have that $\langle a^{k+1}, x^{k+1} - \overline{x}\rangle = \langle a^{k+1}, x^{k+1}\rangle - \langle a^{k+1}, \overline{x}\rangle$ and

$$0 \leq \sum_{k=0}^{\infty} |\langle a^{k+1}, x^{k+1}\rangle| \leq \sum_{k=0}^{\infty} \|a^{k+1}\|\|x^{k+1}\| = \sum_{k=1}^{\infty} \|a^k\|\|x^k\| \leq \sum_{k=1}^{\infty} \eta_k,$$

$$0 \leq \sum_{k=0}^{\infty} |\langle a^{k+1}, \overline{x}\rangle| \leq \sum_{k=0}^{\infty} \|a^{k+1}\|\|\overline{x}\| = \|\overline{x}\| \sum_{k=1}^{\infty} \|a^k\| \leq \|\overline{x}\| \sum_{k=1}^{\infty} \epsilon_k. \tag{2.29}$$

Because of $\sum_{k=1}^{\infty} \epsilon_k < \infty$ and $\sum_{k=1}^{\infty} \eta_k < \infty$, we get $\sum_{k=0}^{\infty} |\langle a^{k+1}, x^{k+1}\rangle|$ and $\sum_{k=0}^{\infty} |\langle a^{k+1}, \overline{x}\rangle|$ are finite. Then $\sum_{k=0}^{\infty} \langle a^{k+1}, x^{k+1}\rangle$ and $\sum_{k=0}^{\infty} \langle a^{k+1}, \overline{x}\rangle$ exist and are finite. In similar way, we can prove that $\sum_{k=0}^{\infty} \langle b^{k+1}, z^{k+1} - \overline{z}\rangle$ is convergent.

Now, we apply Lemma 2.3 to (2.28) with $w_k = E_k(\overline{x}, \overline{y}, \overline{z})$ and $\beta_k = \langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle$ establishing that $\{E_k(\overline{x}, \overline{y}, \overline{z})\}$ is convergent (with that we have proved (i)) and then $\{E_k(\overline{x}, \overline{y}, \overline{z})\}$ is bounded.

Moreover, from Definition 1.4 (iii) we have that $H(\overline{x}, \cdot)$ is level bounded on $C$ and therefore $\{(x^k, z^k, y^k)\}$ is bounded.

From (2.28) we have that the squares of the quantities $\|z^{k+1} - z^k\|$ and $\|Ax^{k+1} + Bz^k - b\|$, and $H(x^{k+1}, x^k)$ are summable, hence vanish as $k$ goes to $+\infty$. The proof is complete. $\square$

Now, we are ready to give the proof of our main result.

*Proof. Theorem 2.2.*

**(i)** By Assumption $(A_2)$ we know that there exists a saddle point $(x^*, z^*, y^*)$ of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^n$.

From Proposition 2.5 (ii) it follows that $\{(x^k, z^k, y^k)\}$ is bounded and then this sequence has a subsequence $\{(x^{k_j}, z^{k_j}, y^{k_j})\}$ which converges to some $(x^\infty, z^\infty, y^\infty) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ (as $x^{k_j} \in C$ in the limit we get $x^\infty \in \overline{C}$).

We will demostrate that $(x^\infty, z^\infty)$ is a feasible point for $(P)$ and then we will prove that $f(x^\infty) + g(z^\infty) \leq f(x^*) + g(z^*)$.

From Proposition 2.5 (ii) we have that the quantities $H(x^{k+1}, x^k)$ and $\|Ax^{k+1} + Bz^k - b\|$ vanish as $k$ goes to $+\infty$. Then $H(x^{k_j+1}, x^{k_j}) \to 0$, and due to $(B_1)$ we have $x^{k_j+1} \to x^\infty$.

The quantity $\|Ax^{k_j} + Bz^{k_j} - b\|$ converge to zero because $\|x^{k_j+1} - x^{k_j}\| \to 0$ and $\|Ax^{k_j+1} + Bz^{k_j} - b\| \to 0$. Then, $0 = \lim_{j \to +\infty} \|Ax^{k_j} + Bz^{k_j} - b\| = \|Ax^\infty + Bz^\infty - b\|$ and so $Ax^\infty + Bz^\infty = b$. Therefore, $(x^\infty, z^\infty)$ is a feasible point for $(P)$.

Considering the inequality (2.16) of the Lemma 3 on the sequence $\{(x^{k_j}, z^{k_j}, y^{k_j})\}$, and with $x = x^*$ and $z = z^*$ we have:

$$
\begin{aligned}
\ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x^*, z^*, y^{k_j}) \;\leq\;& \frac{\lambda}{2}(\|Ax^* + Bz^{k_j} - b\|^2 - \|Ax^{k_j+1} + Bz^{k_j} - b\|^2 \\
& -\|Ax^{k_j+1} - Ax^*\|^2) + \frac{\lambda}{2}(\|Ax^{k_j+1} + Bz^* - b\|^2 \\
& -\|Ax^{k_j+1} + Bz^{k_j+1} - b\|^2 - \|Bz^{k_j+1} - Bz^*\|^2) \\
& +\frac{1}{2\lambda}\left(H(x^*, x^{k_j}) - H(x^*, x^{k_j+1}) - \kappa H(x^{k_j+1}, x^{k_j})\right) \\
& +\frac{1}{2\lambda}\left(\|z^* - z^{k_j}\|^2 - \|z^* - z^{k_j+1}\|^2 - \|z^{k_j+1} - z^{k_j}\|^2\right) \\
& +\langle a^{k_j+1}, x^{k_j+1} - x^*\rangle + \langle b^{k_j+1}, z^{k_j+1} - z^*\rangle.
\end{aligned}
$$

From the definition of $E_k(x^*, z^*, y^*)$ (cf. (2.23)) and since $Ax^* + Bz^* = b$, the last inequality is equivalent to:

$$
\begin{aligned}
\ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x^*, z^*, y^{k_j}) \;\leq\;& E_{k_j}(x^*, z^*, y^*) - E_{k_j+1}(x^*, z^*, y^*) \\
& -\frac{1}{2\lambda}\left(\|y^{k_j} - y^*\|^2 - \|y^{k_j+1} - y^*\|^2\right) \\
& -\frac{\lambda}{2}\|Ax^{k_j+1} + Bz^{k_j} - b\|^2 \\
& +\frac{\lambda}{2}\|Ax^{k_j+1} + Bz^{k_j+1} - b\|^2 \\
& -\frac{\kappa}{2\lambda}H(x^{k_j+1}, x^{k_j}) - \frac{1}{2\lambda}\|z^{k_j+1} - z^{k_j}\|^2 \\
& +\langle a^{k_j+1}, x^{k_j+1} - x^*\rangle + \langle b^{k_j+1}, z^{k_j+1} - z^*\rangle.
\end{aligned}
\tag{2.30}
$$

The right side of inequality (2.30) goes to zero as $j$ tends to $+\infty$. Indeed, from the Proposition 2 (ii) we have that $\|z^{k+1} - z^k\| \to 0$, and $z^{k_j} \to z^\infty$, then $z^{k_j+1} \to z^\infty$ as $j$ goes to $+\infty$.

Moreover, since $y^{k_j+1} = y^{k_j} + \lambda(Ax^{k_j+1} + Bz^{k_j+1} - b)$, $y^{k_j} \to y^\infty$ and $Ax^\infty + Bz^\infty = b$, we have $y^{k_j+1} \to y^\infty$.

Therefore, the subsequence $\{(x^{k_j+1}, z^{k_j+1}, y^{k_j+1})\}$ converges to $(x^\infty, z^\infty, y^\infty)$ as $j$ goes to $+\infty$.

On the other hand, from the Proposition 2(ii) we have $E_{k_j}(x^*, z^*, y^*) - E_{k_j+1}(x^*, z^*, y^*) \to 0$.

Also, the quantities $\|Ax^{k_j+1} + Bz^{k_j} - b\|^2$ and $\|Ax^{k_j+1} + Bz^{k_j+1} - b\|^2$ converge to $\|Ax^\infty + Bz^\infty - b\|^2$ and since $Ax^\infty + Bz^\infty - b = 0$, they vanish as $j$ goes to $+\infty$. Moreover, by the Cauchy-Schawrz we have:

$$|\langle a^{k_j+1}, x^{k_j+1} - x^*\rangle| \leq \|a^{k_j+1}\|\|x^{k_j+1} - x^*\|.$$

From (2.11), (2.13) and (2.15) we have that $\|a^{k+1}\| \to 0$. Then

$$|\langle a^{k_j+1}, x^{k_j+1} - x^*\rangle| \to 0,$$

as $j$ goes to $+\infty$. In the similar way, we can prove that: $|\langle b^{k_j+1}, z^{k_j+1} - z^*\rangle| \to 0$.

Taking limit as $j$ goes to $+\infty$ in the inequality (2.30) we have that the right side goes to zero, and due to $f$ and $g$ are closed functions we obtain:

$$f(x^\infty) + g(z^\infty) \leq f(x^*) + g(z^*).$$

**(ii)** We will demonstrate that $(x^\infty, z^\infty, y^\infty)$ is a saddle point of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, that is,

$$\ell(x^\infty, z^\infty, y) \leq \ell(x^\infty, z^\infty, y^\infty) \leq \ell(x, z, y^\infty), \forall x \in \overline{C},\ \forall z \in \mathbb{R}^m,\ \forall y \in \mathbb{R}^p. \tag{2.31}$$

From part (i), we know $(x^\infty, z^\infty)$ is a feasible point of $(P)$, then

$$\ell(x^\infty, z^\infty, y) = \ell(x^\infty, z^\infty, y^\infty),\ \forall y \in \mathbb{R}^p. \tag{2.32}$$

Now we'll prove the rigth inequality of (2.31). Consider the inequality (2.16) in Lemma 2.4 where $(x^k, z^k, y^k)$ and $(x^{k+1}, z^{k+1})$ are replaced by $(x^{k_j}, z^{k_j}, y^{k_j})$ and $(x^{k_j+1}, z^{k_j+1})$

respectively. Moreover, applying the Cauchy-Schwarz inequality we get

$$\begin{aligned}\ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x, z, y^{k_j}) \;\leq\; & \frac{\lambda}{2}(\|Ax + Bz^{k_j} - b\|^2 - \|Ax^{k_j+1} + Bz^{k_j} - b\|^2 \\ & -\|Ax^{k_j+1} - Ax\|^2) + \frac{\lambda}{2}(\|Ax^{k_j+1} + Bz - b\|^2 \\ & -\|Ax^{k_j+1} + Bz^{k_j+1} - b\|^2 - \|Bz^{k_j+1} - Bz\|^2) \\ & +\frac{1}{2\lambda}\left(H(x, x^{k_j}) - H(x, x^{k_j+1}) - \kappa H(x^{k_j+1}, x^{k_j})\right) \\ & +\frac{1}{2\lambda}\left(\|z - z^{k_j}\|^2 - \|z^{k_j+1} - z^{k_j}\|^2 - \|z - z^{k_j+1}\|^2\right) \\ & +\|a^{k_j+1}\|\|x^{k_j+1} - x\| + \|b^{k_j+1}\|\|z^{k_j+1} - z\|. \quad (2.33)\end{aligned}$$

Taking $\liminf$ as $j \to \infty$ on both sides of (2.33), and since $Ax^\infty + Bz^\infty - b = 0$ we obtain

$$\liminf \left\{\ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x, z, y^{k_j})\right\} \;\leq\; \liminf \left\{\frac{1}{2\lambda}\left(H(x, x^{k_j}) - H(x, x^{k_j+1})\right)\right\}.$$

From the assumption $(B_2)$ we have that $\lim H(x, x^{k_j}) = H(x, x^\infty) = \lim H(x, x^{k_j+1})$, and since $f$ and $g$ are closed functions we obtain:

$$\ell(x^\infty, z^\infty, y^\infty) \leq \ell(x, z, y^\infty), \;\; \forall x \in \mathrm{dom}(f), \; \forall z \in \mathrm{dom}(g). \tag{2.34}$$

From (2.32) and (2.34) we have $(x^\infty, z^\infty, y^\infty)$ is a saddle point of $\ell$.

Now we prove that the sequence $\{(x^k, z^k, y^k)\}$ globally converges to the saddle point $(x^\infty, z^\infty, y^\infty)$. From Proposition 2.5.(i) the following limit exists:

$$E = \lim_{k\to\infty} E_k(x^\infty, z^\infty, y^\infty)$$

. But for the subsequence $\{k_j\}$ we have that $(x^{k_j}, z^{k_j}, y^{k_j}) \to (x^\infty, z^\infty, y^\infty)$ and by Assumption $(B_3)$ we have $\lim_{k\to\infty} H(x^\infty, x^{k_j}) = 0$. Hence $\lim_{j\to\infty} E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$. Then $E = \lim_{j\to\infty} E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$, and so $H(x^\infty, x^k) \to 0$. Moreover, due to $(B_1)$

we obtain $x^k \to x^\infty$, consequently, we get $(x^k, z^k, y^k) \to (x^\infty, z^\infty, y^\infty)$, and the proof is complete. $\square$

# Chapter 3

# Relaxation factor

## 3.1 Introduction

In this chapter, we consider the convex optimization problem with separable structure $(P)$ presented in the chapter 2.

In the literature on ADMs, the update ruler for the Lagrange multiplier (2.9) has been generalized to:

$$y^{k+1} = y^k + \rho\lambda(Ax^{k+1} + Bz^{k+1} - b), \tag{3.1}$$

where $\rho$ is a relaxation factor and $\lambda > 0$ is a given penalty parameter for the linearly constrained equation $Ax + Bz - b = 0$. Most ADM papers use $\rho = 1$, however numerical experiments show that the relaxation factor $\rho$ can accelerate the convergence of the ADM methods.

In [47], Glowinski (1984) proposed the original ADM with relaxation factor, and he proved that if the relaxation factor belongs to the region $\left(0, \frac{\sqrt{5}+1}{2}\right)$ then this more general ADM converges. The original ADM with relaxation factor $\rho$ applied to the problem $(P)$ consists of the following steps:

$$x^{k+1} = \operatorname*{argmin}_{x\in\overline{C}} \; f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2, \tag{3.2}$$

$$z^{k+1} = \operatorname*{argmin}_{z\in\mathbb{R}^m} \; g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2, \tag{3.3}$$

$$y^{k+1} = y^k + \rho\lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{3.4}$$

In [82], Xu (2007) studied the convergence of the alternating direction method with quadratic proximal regularization (QPADM) when a relaxation factor is incorporated into the QPADM in the same way as the original ADM, and the restriction region for the relaxation factor is the same as the original ADM.

Inspired by the studies mentioned above, we propose a more general version of the RIPADM method for solving the problem $(P)$ which we call *RIPADM method with relaxation factor* (or *relaxed RIPADM*), and consists of the following iterations:

$$x^{k+1} \approx \operatorname*{argmin}_{x\in\overline{C}} \; f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2 + \frac{1}{2\lambda}d(x, x^k), \tag{3.5}$$

$$z^{k+1} \approx \operatorname*{argmin}_{z\in\mathbb{R}^m} \; g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \tag{3.6}$$

$$y^{k+1} = y^k + \rho\lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{3.7}$$

We demonstrate that the proposed method converges to a saddle point of the Lagrangian of the problem $(P)$ under standard assumptions.

## 3.2 Convergence results

We remember the Assumptions A and B of the Chapter 2.

#### Assumption A

$(A_1)$ dom $f \cap C \neq \emptyset$.

$(A_2)$ There exists a saddle point $(x^*, z^*, y^*) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ (see (2.5)).

**Assumption B**

There exists a compatible proximal pair $(d, H)$ associated with $\overline{C}$ which verifies:

$(B_1)$ If $\{u^k\} \subset \overline{C}$ and $\{v^k\} \subset C$ are sequences such that

$$\lim_k H(u^k, v^k) = 0,$$

and one of the sequences ($\{u^k\}$ or $\{v^k\}$) converges, then the other one also converges to the same limit.

$(B_2)$ The function $H : \overline{C} \times C \to \mathbb{R}$ extends by continuity to $\overline{C} \times \overline{C}$.

$(B_3)$ $\forall u \in \overline{C}$ and $\forall \{u^k\} \subset C$ converging $u$, we have $\lim_{k\to\infty} H(u, u^k) = 0$.

**Theorem 3.1.** *Let us consider the inexact RIPADM method with relaxation factor given by (2.10)-(2.13) and (3.1). Under Assumptions A and B, if*

$$\rho \in \left(0, \frac{\sqrt{5}+1}{2}\right)$$

*and moreover*

$$\sum_{k=1}^{\infty} |\langle b^{k+1} - b^k, z^{k+1} - z^k \rangle| < \infty$$

*then each sequence $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$ generated by the proposed algorithm converges to a saddle point $(x^\infty, z^\infty, y^\infty)$ of the Lagrangian $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$.*

*Remark* 3.2. It may appear somewhat unnatural that the summability condition $\sum_{k=1}^{\infty} |\langle b^{k+1} - b^k, z^{k+1} - z^k \rangle| < \infty$ involves not only the error sequence $\{b^k\}$, but also the iterates $\{z^k\}$, which are unknown a priori. Still, this condition might not be too difficult to enforce in practice: for example, if one shows that $\{z^k\}$ is bounded then this condition is a direct consequence of $\|b^{k+1}\| \leq \bar{\epsilon}_{k+1}$.

We begin the proof of Theorem 3.1 with the Lemma 3.3 which give us two saddle point type inequalities. Then, inspired by Xu [82, Theorem 3.1] we state the Proposition 3.4 that gives a fundamental inequality from which we can deduce the boundedness of the sequence $\{(x^k, z^k, y^k)\}$ generated by the relaxed RIPADM. Then we end this chapter with the proof of Theorem 3.1.

**Lemma 3.3.** *Let $(d, H) \in \mathcal{F}(\overline{C})$. Consider the sequence $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$, and errors sequences $\{a^k\} \subset \mathbb{R}^n$ and $\{b^k\} \subset \mathbb{R}^m$ being generated by the RIPADM algorithm with relaxation factor given by (2.10)-(2.13) and (3.1). Under Assumption $(A_1)$, for all $x \in \mathrm{dom}(f)$, $z \in \mathrm{dom}(g)$, $y \in \mathbb{R}^p$, the following are satisfied:*

$$\begin{aligned} \ell(x^{k+1}, z^{k+1}, y^k) - \ell(x, z, y^k) \quad \leq \quad & \lambda\langle Ax^{k+1} + Bz^{k+1} - b, Ax + Bz - b\rangle \\ & +\lambda\langle Bz^k - Bz^{k+1}, Ax - Ax^{k+1}\rangle \\ & -\lambda\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ & +\frac{1}{2\lambda}\left(H(x, x^k) - H(x, x^{k+1}) - \alpha H(x^{k+1}, x^k)\right) \\ & +\frac{1}{2\lambda}\left(\|z - z^k\|^2 - \|z - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2\right) \\ & +\langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle. \end{aligned} \tag{3.8}$$

*and*

$$\ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) \quad = \quad \frac{1}{2\rho\lambda}\left(\|y - y^k\|^2 - \|y - y^{k+1}\|^2 + \|y^{k+1} - y^k\|^2\right) \tag{3.9}$$

*Proof. Lemma 3.3.* Consider the inequalities (2.18), (2.19) and (2.21) of the Proof of the Lemma 2.4. We can rewrite (2.18) as:

$$\begin{aligned} f(x) \quad \geq \quad & f(x^{k+1}) + \left\langle \frac{-1}{2\lambda}\nabla_1 d(x^{k+1}, x^k), x - x^{k+1} \right\rangle \\ & + \left\langle -A^t[y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b) + \lambda(Bz^k - Bz^{k+1})], x - x^{k+1} \right\rangle + \langle a^{k+1}, x - x^{k+1}\rangle. \end{aligned} \tag{3.10}$$

Adding (3.10) and (2.19), and rearranging these inequalities we obtain:

$$
\begin{aligned}
f(x^{k+1}) + g(z^{k+1}) + \langle y^k, Ax^{k+1} + Bz^{k+1}\rangle & \\
- \big(f(x) + g(z) + \langle y^k, Ax + Bz\rangle\big) \;\; \leq \;\; & \lambda\langle Ax^{k+1} + Bz^{k+1} - b, Ax + Bz\rangle \\
& -\lambda\langle Ax^{k+1} + Bz^{k+1} - b, Ax^{k+1} + Bz^{k+1}\rangle \\
& +\lambda\langle Bz^k - Bz^{k+1}, Ax - Ax^{k+1}\rangle \\
& +\frac{1}{2\lambda}\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \\
& +\frac{1}{\lambda}\langle z^{k+1} - z^k, z - z^{k+1}\rangle \\
& +\langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle,
\end{aligned}
$$

that is,

$$
\begin{aligned}
f(x^{k+1}) + g(z^{k+1}) + \langle y^k, Ax^{k+1} + Bz^{k+1}\rangle & \\
- \big(f(x) + g(z) + \langle y^k, Ax + Bz\rangle\big) \;\; \leq \;\; & \lambda\langle Ax^{k+1} + Bz^{k+1} - b, Ax + Bz - b\rangle \\
& -\lambda||Ax^{k+1} + Bz^{k+1} - b||^2 \\
& +\lambda\langle Bz^k - Bz^{k+1}, Ax - Ax^{k+1}\rangle \\
& +\frac{1}{2\lambda}\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1}\rangle \\
& +\frac{1}{\lambda}\langle z^{k+1} - z^k, z - z^{k+1}\rangle \\
& +\langle a^{k+1}, x^{k+1} - x\rangle + \langle b^{k+1}, z^{k+1} - z\rangle.
\end{aligned}
$$

Considering the definition of the Lagrangian function $\ell$, using the inequality (2.21) and the relation $\|m - n\|^2 = \|m\|^2 + \|n\|^2 - 2\langle m, n\rangle$ in the last inequality, we obtain (3.8).

For the equality (3.9) we can proceed analogously to the Proof of the Lemma 2.4 for the equality (2.17).

□

The next result is inspired from *Theorem 3.1* by Xu [82].

**Proposition 3.4.** *Under the same hypotheses of the Lemma 3.3, for any saddle point* $(\overline{x}, \overline{z}, \overline{y})$ *of* $\ell$ *on* $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$*, the sequence* $\{(x^k, z^k, y^k)\}$ *satisfies the following inequality:*

$$\begin{aligned} E_k(\overline{x}, \overline{z}, \overline{y}) & \\ +\langle b^{k+1} - b^k, z^{k+1} - z^k\rangle & \\ +\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle \quad \geq \quad & E_{k+1}(\overline{x}, \overline{z}, \overline{y}) \\ & +\frac{\alpha}{2\lambda} H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 \\ & +\frac{\lambda}{2}\left(1 - \frac{(1-\rho)^2}{\tau^2}\right)\|Bz^k - Bz^{k+1}\|^2 \\ & +\lambda\left(1 - \frac{\rho}{2} - \frac{\tau^2}{2}\right)\|Ax^{k+1} + Bz^{k+1} - b\|^2, \end{aligned} \tag{3.11}$$

*where*

$$\begin{aligned} E_k(\overline{x}, \overline{z}, \overline{y}) \quad = \quad & \frac{1}{2\lambda} H(\overline{x}, x^k) + \frac{1}{2\lambda}\|z^k - \overline{z}\|^2 + \frac{1}{2\lambda\rho}\|y^k - \overline{y}\|^2 + \frac{\lambda}{2}\|Bz^k - B\overline{z}\|^2 \\ & +\frac{\lambda\tau^2}{2}\|Ax^k + Bz^k - b\|^2 + \frac{1}{2\lambda}\|z^k - z^{k-1}\|^2. \end{aligned}$$

*With some adaptive values of the parameter* $\tau$*, the factors* $\left(1 - \frac{(1-\rho)^2}{\tau^2}\right)$ *and* $\left(1 - \frac{\rho}{2} - \frac{\tau^2}{2}\right)$ *are non-negative when* $\rho \in \left(0, \frac{\sqrt{5}+1}{2}\right)$*.*

*Proof. Proposition 3.4.* Following the same steps of the proof of the Proposition 2.5, but considering the inequality (3.8) and the equality (3.9) instead of (2.16) and (2.17), we obtain the following inequality:

$$\begin{aligned} \frac{1}{2\lambda\rho}(\|\overline{y} - y^{k+1}\|^2 - \|\overline{y} - y^k\|^2) \quad \leq \quad & \lambda\langle A\overline{x} - Ax^{k+1}, Bz^k - Bz^{k+1}\rangle \\ & +\lambda\left(\frac{\rho}{2} - 1\right)\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ & +\frac{1}{2\lambda}\left(H(\overline{x}, x^k) - H(\overline{x}, x^{k+1}) - \alpha H(x^{k+1}, x^k)\right) \\ & +\frac{1}{2\lambda}\left(\|\overline{z} - z^k\|^2 - \|\overline{z} - z^{k+1}\|^2 - \|z^{k+1} - z^k\|^2\right) \\ & +\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle. \end{aligned}$$

Rearranging the last inequality and using that

$$\begin{aligned}\langle A\overline{x} - Ax^{k+1}, Bz^{k+1} - Bz^k\rangle &= \langle Ax^{k+1} + Bz^{k+1} - b, Bz^k - Bz^{k+1}\rangle \\ &+\frac{1}{2}\left[\|Bz^{k+1} - B\overline{z}\|^2 - \|Bz^k - B\overline{z}\|^2\right] \\ &+\frac{1}{2}\|Bz^k - Bz^{k+1}\|^2,\end{aligned}$$

we get:

$$\begin{aligned}&\frac{1}{2\lambda}H(\overline{x}, x^k) + \frac{1}{2\lambda}\|\overline{z} - z^k\|^2 \\ &+\frac{1}{2\lambda\rho}\|\overline{y} - y^k\|^2 + \frac{\lambda}{2}\|Bz^k - B\overline{z}\|^2 \\ &+\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle \geq \frac{1}{2\lambda}H(\overline{x}, x^{k+1}) + \frac{1}{2\lambda}\|\overline{z} - z^{k+1}\|^2 \\ &+\frac{1}{2\lambda\rho}\|\overline{y} - y^{k+1}\|^2 + \frac{\lambda}{2}\|Bz^{k+1} - B\overline{z}\|^2 \\ &+\lambda\left(1 - \frac{\rho}{2}\right)\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ &+\lambda\langle Ax^{k+1} + Bz^{k+1} - b, Bz^k - Bz^{k+1}\rangle \\ &\frac{\alpha}{2\lambda}H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 + \frac{\lambda}{2}\|Bz^k - Bz^{k+1}\|^2\end{aligned}$$

Set

$$\widetilde{E}_k(\overline{x}, \overline{z}, \overline{y}) = \frac{1}{2\lambda}H(\overline{x}, x^k) + \frac{1}{2\lambda}\|z^k - \overline{z}\|^2 + \frac{1}{2\lambda\rho}\|y^k - \overline{y}\|^2 + \frac{\lambda}{2}\|Bz^k - B\overline{z}\|^2.$$

Then

$$\begin{aligned}&\widetilde{E}_k(\overline{x}, \overline{z}, \overline{y}) \\ &+\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle \geq \widetilde{E}_{k+1}(\overline{x}, \overline{z}, \overline{y}) \\ &+\frac{\alpha}{2\lambda}H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 \\ &+\frac{\lambda}{2}\|Bz^k - Bz^{k+1}\|^2 \\ &+\lambda(1 - \frac{\rho}{2})\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ &+\lambda\langle Ax^{k+1} + Bz^{k+1} - b, Bz^k - Bz^{k+1}\rangle.\end{aligned} \tag{3.12}$$

Let us treat the last term on the right-hand side of (3.12). From Step 2 formula (2.12) we have

$$\begin{aligned} g(z^k) &\geq g(z^{k+1}) + \left\langle -B^t[y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b)] + b^{k+1}, z^k - z^{k+1} \right\rangle \\ &\quad + \left\langle \frac{-1}{\lambda}(z^{k+1} - z^k), z^k - z^{k+1} \right\rangle. \end{aligned}$$

Due to the same reason in the $(k-1)$th iteration, it follows that

$$\begin{aligned} g(z^{k+1}) &\geq g(z^k) + \left\langle -B^t[y^{k-1} + \lambda(Ax^k + Bz^k - b)] + b^k, z^{k+1} - z^k \right\rangle \\ &\quad + \left\langle \frac{-1}{\lambda}(z^k - z^{k-1}), z^{k+1} - z^k \right\rangle. \end{aligned}$$

Adding and rearranging the above two inequalities, and since $y^k - y^{k-1} = \lambda\rho(Ax^k + Bz^k - b)$ we obtain

$$\begin{aligned} \lambda\langle Ax^{k+1} + Bz^{k+1} - b, Bz^k - Bz^{k+1}\rangle &\geq \lambda(1-\rho)\langle Ax^k + Bz^k - b, Bz^k - Bz^{k+1}\rangle \\ &\quad + \frac{1}{\lambda}\|z^k - z^{k+1}\|^2 + \frac{1}{\lambda}\langle z^k - z^{k-1}, z^k - z^{k+1}\rangle \\ &\quad - \langle b^{k+1} - b^k, z^{k+1} - z^k\rangle. \end{aligned} \tag{3.13}$$

Using the relation $\|m+n\|^2 = \|m\|^2 + 2\langle m, n\rangle + \|n\|^2$ we have:

$$\|z^{k+1} - z^{k-1}\|^2 = \|z^{k+1} - z^k\|^2 + 2\langle z^{k+1} - z^k, z^k - z^{k-1}\rangle + \|z^k - z^{k-1}\|^2,$$

that is,

$$-\langle z^{k+1} - z^k, z^k - z^{k-1}\rangle = \frac{1}{2}\|z^{k+1} - z^k\|^2 + \frac{1}{2}\|z^k - z^{k-1}\|^2 - \frac{1}{2}\|z^{k+1} - z^{k-1}\|^2.$$

Then

$$\begin{aligned} \|z^k - z^{k+1}\|^2 + \langle z^k - z^{k-1}, z^k - z^{k+1}\rangle &= \|z^k - z^{k+1}\|^2 + \frac{1}{2}\|z^{k+1} - z^k\|^2 \\ &\quad + \frac{1}{2}\|z^k - z^{k-1}\|^2 - \frac{1}{2}\|z^{k+1} - z^{k-1}\|^2. \end{aligned} \tag{3.14}$$

Using the identity of the parallelogram $2\|m\| + 2\|n\| = \|m+n\|^2 + \|m-n\|^2$ we have:

$$\|(z^k - z^{k+1}) + (z^k - z^{k-1})\|^2 = 2\|z^k - z^{k+1}\|^2 + 2\|z^k - z^{k-1}\|^2 - \|z^{k-1} - z^{k+1}\|^2,$$

Multiplying 1/2 the last equality and rearranging this we have

$$\|z^k - z^{k+1}\|^2 + \frac{1}{2}\|z^k - z^{k-1}\|^2 - \frac{1}{2}\|z^{k+1} - z^{k-1}\|^2 = \frac{1}{2}\|2z^k - z^{k+1} - z^{k-1}\|^2 - \frac{1}{2}\|z^k - z^{k-1}\|^2. \quad (3.15)$$

From (3.14) and (3.15), we obtain:

$$\begin{aligned}\|z^k - z^{k+1}\|^2 + \langle z^k - z^{k+1}, z^k - z^{k-1}\rangle &= \frac{1}{2}\|2z^k - z^{k+1} - z^{k-1}\|^2 \\ &\quad + \frac{1}{2}\|z^k - z^{k+1}\|^2 - \frac{1}{2}\|z^{k-1} - z^k\|^2.\end{aligned} \quad (3.16)$$

From (3.13) and (3.16) we have

$$\begin{aligned}\lambda\langle Ax^{k+1} + Bz^{k+1} - b, Bz^k - Bz^{k+1}\rangle &\geq \lambda(1-\rho)\langle Ax^k + Bz^k - b, Bz^k - Bz^{k+1}\rangle \\ &\quad + \frac{1}{2\lambda}\|z^k - z^{k+1}\|^2 - \frac{1}{2\lambda}\|z^{k-1} - z^k\|^2 \\ &\quad - \langle b^{k+1} - b^k, z^{k+1} - z^k\rangle.\end{aligned}$$

Due to the last inequality and (3.12) we get

$$\begin{aligned}&\widetilde{E}_k(\overline{x}, \overline{z}, \overline{y}) + \frac{1}{2\lambda}\|z^{k-1} - z^k\|^2 \\ &\quad + \langle b^{k+1} - b^k, z^{k+1} - z^k\rangle \\ &+ \langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle \geq \widetilde{E}_{k+1}(\overline{x}, \overline{z}, \overline{y}) + \frac{1}{2\lambda}\|z^k - z^{k+1}\|^2 \\ &\quad + \frac{\alpha}{2\lambda}H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 \\ &\quad + \frac{\lambda}{2}\|Bz^k - Bz^{k+1}\|^2 \\ &\quad + \lambda\left(1 - \frac{\rho}{2}\right)\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ &\quad + \lambda(1-\rho)\langle Ax^k + Bz^k - b, Bz^k - Bz^{k+1}\rangle.\end{aligned} \quad (3.17)$$

Let us bound the last term on the right-hand of (3.17). Choosing $\tau > 0$ and using the identity $\|m+n\|^2 = \|m\|^2 + 2\langle m,n\rangle + \|n\|^2$ with $m = \tau(Ax^k + Bz^k - b)$ and $n = \frac{1-\rho}{\tau}(Bz^k - Bz^{k+1})$, we have

$$\begin{aligned}\|\tau(Ax^k + Bz^k - b) + \frac{1-\rho}{\tau}(Bz^k - Bz^{k+1})\|^2 \;&=\; \tau^2\|Ax^k + Bz^k - b\|^2 \\ &\quad +2(1-\rho)\langle Ax^k + Bz^k - b, Bz^k - Bz^{k+1}\rangle \\ &\quad +\frac{(1-\rho)^2}{\tau^2}\|Bz^k - Bz^{k+1}\|^2.\end{aligned}$$

Then

$$\begin{aligned}\lambda(1-\rho)\langle Ax^k + Bz^k - b, Bz^k - Bz^{k+1}\rangle \;&\geq\; \frac{-\lambda\tau^2}{2}\|Ax^k + Bz^k - b\|^2 \\ &\quad -\lambda\frac{(1-\rho)^2}{2\tau^2}\|Bz^k - Bz^{k+1}\|^2.\end{aligned} \tag{3.18}$$

From (3.17) and (3.18), we obtain the inequality:

$$\begin{aligned}&\widetilde{E}_k(\overline{x},\overline{z},\overline{y}) + \frac{1}{2\lambda}\|z^{k-1} - z^k\|^2 \\ &\quad +\langle b^{k+1} - b^k, z^{k+1} - z^k\rangle \\ &+\langle a^{k+1}, x^{k+1} - \overline{x}\rangle + \langle b^{k+1}, z^{k+1} - \overline{z}\rangle \;\geq\; \widetilde{E}_{k+1}(\overline{x},\overline{z},\overline{y}) + \frac{1}{2\lambda}\|z^k - z^{k+1}\|^2 \\ &\qquad +\frac{\alpha}{2\lambda}H(x^{k+1}, x^k) + \frac{1}{2\lambda}\|z^{k+1} - z^k\|^2 \\ &\qquad +\frac{\lambda}{2}\left(1 - \frac{(1-\rho)^2}{\tau^2}\right)\|Bz^k - Bz^{k+1}\|^2 \\ &\qquad +\lambda\left(1 - \frac{\rho}{2}\right)\|Ax^{k+1} + Bz^{k+1} - b\|^2 \\ &\qquad -\frac{\lambda\tau^2}{2}\|Ax^k + Bz^k - b\|^2.\end{aligned}$$

Set $E_k(\overline{x},\overline{z},\overline{y}) = \widetilde{E}_k(\overline{x},\overline{z},\overline{y}) + \frac{1}{2\lambda}\|z^k - z^{k-1}\|^2 + \frac{\lambda\tau^2}{2}\|Ax^k + Bz^k - b\|^2$. Then the last inequality is equivalent to (3.11).

Through simply inspection, we have that if $\tau = \frac{\sqrt{5}-1}{2}$, $\rho \in \left[\frac{3-\sqrt{5}}{2}, \frac{\sqrt{5}+1}{2}\right)$ or $\tau = 1$,

$\rho \in (0,1)$, then $\left(1 - \frac{(1-\gamma)^2}{\tau^2}\right) \geq 0$ and $(2 - \rho - \tau^2) > 0$. $\square$

**Corollary 3.5.** *Assume that* $\rho \in \left(0, \frac{\sqrt{5}+1}{2}\right)$ *and* $\sum_{k=1}^{\infty} |\langle b^{k+1} - b^k, z^{k+1} - z^k \rangle| < \infty$.

**(i)** *For any saddle point* $(\overline{x}, \overline{z}, \overline{y})$ *of* $\ell$*, the sequence* $\{E_k(\overline{x}, \overline{z}, \overline{y})\}$ *is convergent.*

**(ii)** *If at least one saddle point* $(\overline{x}, \overline{z}, \overline{y})$ *of* $\ell$ *exists then the sequence* $\left\{(x^k, z^k, y^k)\right\}$ *is bounded in* $C \times \mathbb{R}^m \times \mathbb{R}^p$*, the quantities* $H(x^{k+1}, x^k)$*,* $\|z^{k+1} - z^k\|^2$ *and* $\|Ax^{k+1} + Bz^{k+1} - b\|^2$ *are summable, hence they vanish as* $k$ *goes to* $+\infty$.

*Proof.* For (i).If $\rho \in \left(0, \frac{\sqrt{5}+1}{2}\right)$, from the inequiality (3.11) of the Proposition 2.4 we have that

$$E_{k+1}(\overline{x}, \overline{z}, \overline{y}) \leq E_k(\overline{x}, \overline{z}, \overline{y}) + \beta_k,$$

for all $k \geq 0$, where $\beta_k = \langle b^{k+1} - b^k, z^{k+1} - z^k \rangle + \langle a^{k+1}, x^{k+1} - \overline{x} \rangle + \langle b^{k+1}, z^{k+1} - \overline{z} \rangle$. Moreover, $\sum_{k=0}^{\infty} \beta_k$ is convergent because it is the sum of three convergent series. Therefore the sequence $\{E_k(\overline{x}, \overline{z}, \overline{y})\}$ is convergent due to Lemma 2.3.

For (ii). From (i) we know that $\{E_k(\overline{x}, \overline{z}, \overline{y})\}$ is convergent. On one hand, this implies that the sequence $\{E_k(\overline{x}, \overline{z}, \overline{y})\}$ is bounded and so the sequence $\{(x^k, z^k, y^k)\}$ is bounded too. On the other hand, considering the inequality (3.11) we obtain that the squares of the quantities $\|z^{k+1} - z^k\|$ and $\|Ax^{k+1} + Bz^{k+1} - b\|$, and $H(x^{k+1}, x^k)$ are summable, hence vanish as $k$ goes to $+\infty$. $\square$

Now, we are ready to give the proof of our main result of this chapter.

*Proof. Theorem 3.1.* By Assumption $(A_2)$ we know that there exists a saddle point $(x^*, z^*, y^*)$ of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^n$. From Corollary 3.5 (ii) it follows that $\{(x^k, z^k, y^k)\}$ is bounded and then this sequence has a subsequence $\{(x^{k_j}, z^{k_j}, y^{k_j})\}$ which converges to some $(x^\infty, z^\infty, y^\infty) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$.

We will prove that $(x^\infty, z^\infty, y^\infty)$ is a saddle point of $\ell$ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$. We have that $\|Ax^k + Bz^k - b\| \to 0$ because $\|Ax^{k+1} + Bz^{k+1} - b\| \to 0$ (Corollary 3.5 ), then $\|Ax^{k_j} +$

$Bz^{k_j} - b\| \to 0$. Also, $\|Ax^{k_j} + Bz^{k_j} - b\| \to \|Ax^\infty + Bz^\infty - b\|$. Thus, $\|Ax^\infty + Bz^\infty - b\| = 0$. On the other hand, we have that $\ell(x^\infty, z^\infty, y) - \ell(x^\infty, z^\infty, y^\infty) = \langle y - y^\infty, Ax^\infty + Bz^\infty - b\rangle$ for all $y \in \mathbb{R}^p$. Therefore,

$$\ell(x^\infty, z^\infty, y) = \ell(x^\infty, z^\infty, y^\infty),\ \forall y \in \mathbb{R}^p. \tag{3.19}$$

In addition, we have that

$$\ell(x^\infty, z^\infty, y^\infty) \leq \ell(x, z, y^\infty),\ \ \forall x \in \mathrm{dom}(f),\ \forall z \in \mathrm{dom}(g), \tag{3.20}$$

This is obtained taking $\liminf$ over the appropriate subsequence on both sides of the inequality (2.16). From (3.19) and (3.20) we conclude that $(x^\infty, z^\infty, y^\infty)$ is a saddle point of $\ell$.

From Corollary 3.5 (i) the following limit exists: $L = \lim_{k\to\infty}\ E_k(x^\infty, z^\infty, y^\infty)$. But for the subsequence $\{k_j\}$ we have that $(x^{k_j}, z^{k_j}, y^{k_j}) \to (x^\infty, z^\infty, y^\infty)$ and by Assumption $(B_3)$ we have $\lim_{k\to\infty} H(x^\infty, x^{k_j}) = 0$. Hence $\lim_{j\to\infty}\ E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$. Then $L = \lim_{j\to\infty}\ E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$, and so by Assumption $(B_1)$ we obtain $x^k \to x^\infty$ , consequently, we get $(x^k, z^k, y^k) \to (x^\infty, z^\infty, y^\infty)$, and the proof is complete. $\square$

# Chapter 4

# Numerical Experiences

## 4.1 Introduction

In this chapter, we study three types of problems in the form $(P)$ with $C = \mathbb{R}^n_{++}$, that is, of the form:

$$(P) \qquad \min_{x,z} \{f(x) + g(z) \mid Ax + Bz = b,\ x \in \mathbb{R}^n_+\}.$$

The first problem is the constrained LASSO problem, the second one is a modification of the constrained LASSO problem, and the third one is a learning machine problem.

In the section 4.2, we solve each of these applications using three different methods: our approach RIPADM which was introduced in chapter 2 (see (2.1)-(2.3)), the proximal method of multipliers (PMM), and the classical ADM (see (2.7)-(2.9)). Recall that the PMM (proposed by Rockafellar [68]) is obtained by applying the proximal algorithm to the primal-dual system of the problem under study. The PMM applied to the problem $(P)$ generates a sequence $\{(x^k, y^k, z^k)\}$ via the following iterates

$$\begin{aligned} (x^{k+1}, z^{k+1}) &= \arg\min_{x \geq 0,\, z} \ell_\lambda(x, z, y^k) + \frac{1}{2\lambda}(\|x - x^k\|^2 + \|z - z^k\|^2) &\quad (4.1)\\ y^{k+1} &= y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b), &\quad (4.2) \end{aligned}$$

where $\ell_\lambda$ denotes the augmented Lagrangian of the problem (P) (cf. (2.6)). These three methods are based on the augmented Lagrangian, but methods ADM and RIPADM are decomposition methods which separate the primal variables $x$ and $z$ via alternating. The difference between the ADM and RIPADM is that in the latter an interior proximal distance is used to treat positivity constraint $x \geq 0$. In this section, we use the Log-quad distance (1.38) (with $\mu = 1$ and $\nu = 2$) to regularize the $x$-subproblem of the RIPADM method.

For each of the applications: contrained LASSO problem, constrained LASSO problem with a cost function, and norm-mixed twin support vector machine (TWSVM) problem, we present the computational results obtained of the three methods mentioned above, and give some conclusions from these results.

In the section 4.3, we present the computational results obtained from the application of RIPADM method with relaxation factor (3.5)-(3.7), which was introduced in the chapter 3, to the constrained LASSO problem and norm-mixed TWSVM problem.

The relaxation factor $\rho$ can assume different values in the region of convergence $\left(0, \frac{\sqrt{5}+1}{2}\right)$, specifically, we study the relaxed RIPADM method with $\rho = 1.62, 1.6$ and $0.7$. Note that in section 4.2, we consider $\rho = 1$. Interesting results are obtained from these numerical experiments. Moreover, we show the numerical results obtained of applying the relaxed ADM (3.2)- (3.4) with $\rho = 1.62, 1.6$ and $0.7$ to the two problems under study in this section. Then we can compare the computational results of relaxed RIPADM and relaxed ADM for each of the problems.

In the section 4.4, we propose an algorithm which is a variant of the RIPADM method, and solves the problem $(P)$ with $f$ being a quadratic function. This variant uses the Bregman kernel $h(x) = \sum_{i=1}^{n} x_i \log(x_i) - x_i$. Due to the work of Eckstein [32] on proximal algorithms using Bregman functions, solve the x-subproblem of the RIPADM method is equivalent to solving a problem without restrictions and then calculate $x^{k+1}$ using a formula. This gives rise to a variant of RIPADM which we call *ADM-type algorithm with Bregman regularization*. We give the computational results obtained of applying the proposed algorithm to the three applications.

The numerical experiments were done on a laptop with Intel Core i3 processor, 1.80 GHz, 4 GB RAM. The operating system of the laptop is Windows 8, and the program used is MATLAB version 8.2.0.701 (R2013b).

In our experiments, the following stopping rule was taken: $|val_k - val^*| < 10^{-5}$, where $val_k$ denotes the value of the objective function at the iteration $k$ of the appropriate algorithm, and $val^*$ denotes the optimal value obtained by CVX solver, which is available in http://cvxr.com/cvx. This solver also was used for solving the inner problems that appear in the three methods: RIPADM, PMM and ADM.

## 4.2 Applications

### 4.2.1 Constrained LASSO problem

Consider the standard linear regression model:

$$d = Dz + \epsilon,$$

where $D \in \mathbb{R}^{r \times m}$ is a matrix of predictors, $d \in \mathbb{R}^r$ is a response vector, $z \in \mathbb{R}^m$ is a vector of regression coefficients, and $\epsilon \in \mathbb{R}^r$ is a vector of random noises. In high-dimensional setting where the number of responses is much smaller than the number of regression coefficients, $r \ll m$, the traditional least-squares method does not perform well. To overcome this difficulty, certain sparsity conditions are assumed on the vector of regression coefficients, that is, one consider the following problem

$$\min_z \left\{ \frac{1}{2}\|Dz - d\|_2^2 + \gamma \|z\|_1 \right\},$$

where $\gamma > 0$ is a running parameter, and $\|\cdot\|_1$ denotes the 1-norm in $\mathbb{R}^m$. This problem is known as *LASSO problem* (see [75]) and it can be interpreted as finding a sparse solution to a least squares or linear regression problem, that is, the LASSO problem is $L_1$-regularized linear regression ([**?**]).

Inspired by significant applications such as portfolio selection ([**?**]) and monotone regression ([55]), the following problem was proposed recently in [55]:

$$\min_z \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 \mid Bz \leq b \right\}, \tag{4.3}$$

where $D \in \mathbb{R}^{r\times m}$, $d \in \mathbb{R}^r$, $B \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, and $\gamma > 0$ are problem data. This problem is known as *constrained LASSO problem.*

By introducing a slack variable $x \in \mathbb{R}^n$, we can rewrite the problem (4.3) like $(P)$ with $f(x) = 0$, $g(z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1$, and $A = I$. That is, the problem (4.3) is equivalent to:

$$\min_z \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 \mid x + Bz = b,\ \ x \geq 0 \right\}. \tag{4.4}$$

Thus, the RIPADM, PMM, and ADM are applicable.

**RIPADM:** This iterative scheme consists of the following steps:

$$x^{k+1} \approx \operatorname*{argmin}_x \ \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda}d(x, x^k), \tag{4.5}$$

$$z^{k+1} \approx \operatorname*{argmin}_z \ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \tag{4.6}$$

$$y^{k+1} = y^k + \lambda(x^{k+1} + Bz^{k+1} - b). \tag{4.7}$$

*Remark* 4.1. If we replace the distance $d$ by the Log-quad distance (1.38), the $x$-problem

(4.5) is rewrite as:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}} \ \langle y^k, x\rangle + \frac{\lambda}{2}\|x+q^k\|^2 \\ &+\frac{1}{2\lambda}\sum_{i=1}^{n}\left[\mu\left((x_i^k)^2\log\left(\frac{x_i^k}{x_i}\right)+x_ix_i^k-(x_i^k)^2\right)+\frac{\nu}{2}(x_i-x_i^k)^2\right], \end{aligned}$$

where $q^k = Bz^k - b$. The optimality condition of this problem give us the following equations:

$$y_i^k + \lambda(x_i^{k+1}+q_i^k) + \frac{1}{2\lambda}\left(-\mu\frac{(x_i^k)^2}{x_i^{k+1}}+\mu x_i^k+\nu x_i^{k+1}-\nu x_i^k\right)=0, \quad i=1,\ldots,n.$$

Then, from this, we can obtain the following closed-form solution

$$x_i^{k+1} = \frac{-\tilde{b}_i+\sqrt{\tilde{b}_i^2-4ac_i}}{2a}, \quad i=1,\ldots,n, \tag{4.8}$$

where $a=\lambda+\frac{\nu}{2\lambda}$, $\tilde{b}_i = y_i^k+\lambda q_i^k+((\mu-\nu)/2\lambda)x_i^k$, and $c_i=(-\mu/2\lambda)(x_i^k)^2$.

**PMM:** Here, we compute the sequences $\{(x^k,z^k,y^k)\}$ via (4.1)-(4.2) with augmented Lagrangian

$$\ell_\lambda(x,y,z)=\frac{1}{2}\|Dz-d\|^2+\gamma\|z\|_1+\langle y, x+Bz-b\rangle+\frac{\lambda}{2}\|x+Bz-b\|^2. \tag{4.9}$$

**ADM:** This method computes the sequences $\{(x^k,z^k,y^k)\}$ via (2.7)-(2.9) with augmented Lagrangian given by (4.9).

In Table 4.1 we present the numerical results obtained when we apply the RIPADM, PMM, and ADM methods for solving the problem (4.3). For each algorithm, we take the same starting point $(x^0,z^0,y^0)=(\mathbf{1},\mathbf{1},\mathbf{3})\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^n$, $\gamma=1$, and set $m=n$. The problem data were randomly generated from the uniform distribution and with a single stream fixed, that is, `s=RandStream('mt19937ar','Seed',1)`, $D$=`rand(s,r,m)`, $B$=`rand(s,n,m)`, $d$=`rand(s,r,1)`, and $b$=`rand(s,n,1)`.

| $r$ | $n$ | | RIPADM | PMM | ADM |
|---|---|---|---|---|---|
| 10 | 30 | CPU time(s) | **192.29** | 291.33 | 225.80 |
| | | Objective Value | 1.309512 | 1.309514 | 1.309513 |
| | | Iterations | 231 | 299 | **178** |
| 30 | 50 | CPU time(s) | **79.21** | 113.56 | 149.75 |
| | | Objective Value | 3.343769 | 3.343757 | 3.343770 |
| | | Iterations | 93 | **88** | 90 |
| 50 | 100 | CPU time(s) | 140.17 | **90.47** | 162.56 |
| | | Objective Value | 4.103247 | 4.103251 | 4.103239 |
| | | Iterations | 123 | **51** | 72 |
| 70 | 200 | CPU time(s) | **337.70** | 388.60 | 648.58 |
| | | Objective Value | 6.354824 | 6.354806 | 6.354807 |
| | | Iterations | 158 | **102** | 128 |
| 100 | 300 | CPU time(s) | 765.18 | **725.76** | 1081.92 |
| | | Objective Value | 7.855478 | 7.855494 | 7.855478 |
| | | Iterations | 151 | **73** | 100 |

Table 4.1: Numerical results for the constrained LASSO problem.

Table 4.1 shows the average of runtime in seconds (run five times), the objective values and the number of iterations of the RIPADM, PMM, and ADM methods apply to Problem (4.3) for different values of $r$ and $n$. The best results in terms of CPU time and number of iterations is highlighted in bold type. From Table 4.1, we observe that the PMM method uses fewer iterations (four out of five experiments) in comparison with the other two methods, however our approach uses less CPU time in three out of five experiments.

### 4.2.2 Constrained LASSO problem with a cost function

In this section, we consider a cost function associated to the problem (4.3), specifically, we consider the following problem:

$$\min_{x,z} \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \frac{\beta}{2}\|x\|^2 \mid x + Bz = b,\ x \geq 0 \right\}, \tag{4.10}$$

with $\beta > 0$. Clearly, this problem has the form of $(P)$ with $f(x) = \frac{\beta}{2}\|x\|^2$, $g(z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1$, $A = I$ and $\overline{C} = \mathbb{R}^n_+$. Thus the RIPADM, PMM, and ADM are again applicable.

**RIPADM:** This iterative scheme consists of the following steps:

$$
\begin{aligned}
x^{k+1} &\approx \operatorname*{argmin}_{x} \ \frac{\beta}{2}\|x\|^2 + \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k),\\
z^{k+1} &\approx \operatorname*{argmin}_{z} \ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2,\\
y^{k+1} &= y^k + \lambda(x^{k+1} + Bz^{k+1} - b).
\end{aligned}
$$

*Remark* 4.2. In a way similar to Remark 4.1, if we replace the distance $d$ by the Log-quad distance (1.38), we can obtain the following closed-form solution for the $x$-subproblem

$$
x_i^{k+1} = \frac{-\tilde{b}_i + \sqrt{\tilde{b}_i^2 - 4ac_i}}{2a}, \quad i = 1, \ldots, n, \tag{4.11}
$$

where $a = \beta + \lambda + \frac{\nu}{2\lambda}$, $\tilde{b}_i = y_i^k + \lambda q_i^k + ((\mu - \nu)/2\lambda)x_i^k$, and $c_i = (-\mu/2\lambda)(x_i^k)^2$.

**PMM:** Here, we compute the sequences $\{(x^k, z^k, y^k)\}$ via (4.1)-(4.2) with augmented Lagrangian

$$
\ell_\lambda(x, y, z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \frac{\beta}{2}\|x\|^2 + \langle y, x + Bz - b\rangle + \frac{\lambda}{2}\|x + Bz - b\|^2. \tag{4.12}
$$

**ADM:** This method computes the sequences $\{(x^k, z^k, y^k)\}$ via (2.7)-(2.9) with augmented Lagrangian given by (4.12).

In Table 4.2 we summarize the numerical results obtained when we apply the RIPADM, PMM, and ADM methods for solving the problem (4.10). For each method, we take the same starting point $(x^0, z^0, y^0) = (\mathbf{1}, \mathbf{1}, \mathbf{3}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, $\gamma = \beta = 1$, and set $m = n$. The problem data were randomly generated in similar way to the above case.

It can be seen in Table 4.2 that the RIPADM method is much faster than the other two methods (see highlighted in bold type) for the different values of $r$ and $n$. For instance, for

| $r$ | $n$ | | RIPADM | PMM | ADM |
|---|---|---|---|---|---|
| 10 | 30 | CPU time(s) | **232.71** | 355.69 | 479.47 |
| | | Objective Value | 3.715823 | 3.715823 | 3.715823 |
| | | Iterations | **324** | 326 | 315 |
| 30 | 50 | CPU time(s) | **41.10** | 133.51 | 111.32 |
| | | Objective Value | 6.855128 | 6.855116 | 6.855122 |
| | | Iterations | **51** | 104 | 67 |
| 50 | 100 | CPU time(s) | **140.34** | 249.80 | 286.16 |
| | | Objective Value | 10.501275 | 10.501275 | 10.501275 |
| | | Iterations | **131** | 141 | 143 |
| 70 | 200 | CPU time(s) | **220.39** | 427.66 | 505.65 |
| | | Objective Value | 14.609375 | 14.609376 | 14.609376 |
| | | Iterations | **100** | 112 | 114 |
| 100 | 300 | CPU time(s) | **658.07** | 1152.91 | 1560.74 |
| | | Objective Value | 23.198968 | 23.198968 | 23.198968 |
| | | Iterations | **137** | 139 | 146 |

Table 4.2: Numerical results for the constrained LASSO problem with cost function.

$r = 70$, $n = 200$ the PMM method needs almost twice the CPU time of the RIPADM one, while the ADM scheme requires more than twice of CPU time of our approach.

### 4.2.3 Twin support vector machine classifier

Support vector machines (SVM) is a new machine learning method which is developed on the basic of statistical learning theory and structural risk minimization [26, 78]. One of the most popular SVM in classification is the “maximum margin" one that attempts to find the optimal separating hyperplane maximizing the margin between two disjoint half planes (associated to positive and negative samples). The resulting optimization task involves the minimization of a convex quadratic function subject to linear inequality constraints.

In order to reduce the computational cost of SVM, [59] proposed a nonparallel hyperplane classifier, called Twin support vector machines (TWSVM) for binary classification. TWSVMs construct two nonparallel hyperplanes such that each one is closer to one of the two training dataset and is as far as possible from the other. The above hyperplanes are obtained by solving two small quadratic programming problems.

On the other hand, an important task in classification is to identify a subset of features which contribute most to classification. The benefit of feature selection is crucial for achieving good classification accuracy in the presence of redundant features. To select important groups of features automatically and simultaneously, [86] proposed to consider the $F_\infty$-norm SVM.

Motivated from the works on TWSVM, and $F_\infty$-norm SVMs, for linearly separable case, we propose to consider the norm-mixed TWSVM given by the following problems

$$\begin{aligned} \min_{w_1,t_1} \quad & \|D_1 w_1 + \mathbf{e}_1 t_1\|_\infty + \frac{c_1}{2}(\|w_1\|^2 + t_1^2) \\ \text{s.t.} \quad & -(D_2 w_1 + \mathbf{e}_2 t_1) \geq \mathbf{e}_2, \end{aligned} \tag{4.13}$$

and

$$\begin{aligned} \min_{w_2,t_2} \quad & \|D_2 w_2 + \mathbf{e}_2 t_2\|_\infty + \frac{c_2}{2}(\|w_2\|^2 + t_2^2) \\ \text{s.t.} \quad & (D_1 w_2 + \mathbf{e}_1 t_2) \geq \mathbf{e}_1, \end{aligned} \tag{4.14}$$

where $\|\cdot\|_\infty$ denotes the infinity-norm in Euclidean space, $D_1 \in \mathbb{R}^{m_1 \times n}$ and $D_2 \in \mathbb{R}^{m_2 \times n}$ are matrices containing the training dataset of positive and negative class, respectively, $c_1$, and $c_2$ are positive parameters, and $\mathbf{e}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{e}_2 \in \mathbb{R}^{m_2}$ are vectors of ones.

We will focus on solving the problem (4.13). Set $z = [w_1^\top, t_1]^\top \in \mathbb{R}^{n+1}$. Then the problem (4.13) can be written in the form:

$$\min_z \left\{ \|[D_1 \ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 | \ -[D_2 \ \mathbf{e}_2]z \geq \mathbf{e}_2 \right\}. \tag{4.15}$$

By introducing a slack variable $x \in \mathbb{R}^{m_2}$ to the inequality constraints of the above problem, we can reformulate it in the form of $(P)$ with $f(x) = 0$, $g(z) = \|[D_1 \ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2$, $A = I$,

$B = [D_2\ \mathbf{e}_2] \in \mathbb{R}^{m_2 \times n+1}$ and $b = -\mathbf{e}_2$. That is, the problem (4.15) is equivalent to:

$$\min_z \left\{ \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 |\ \ x + Bz = -\mathbf{e}_2,\ x \geq 0 \right\}. \tag{4.16}$$

Then, the RIPADM, PMM, and ADM are applicable.

**RIPADM:** This iterative scheme consists of the following steps:

$$\begin{aligned} x^{k+1} &\approx \underset{x}{\operatorname{argmin}}\ \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k), \\ z^{k+1} &\approx \underset{z}{\operatorname{argmin}}\ \ \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \\ y^{k+1} &= y^k + \lambda(x^{k+1} + Bz^{k+1} - b). \end{aligned} \tag{4.17}$$

**PMM:** Here, the sequences $\{(x^k, z^k, y^k)\}$ are compute via (4.1)-(4.2) with

$$\ell_\lambda(x, y, z) = \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 + \langle y, x + Bz - b\rangle + \frac{\lambda}{2}\|x + Bz - b\|^2. \tag{4.18}$$

**ADM:** This method computes the sequences $\{(x^k, z^k, y^k)\}$ via (2.7)-(2.9) with augmented Lagrangian given by (4.18).

In order to solve numerically the norm-mixed TWSVM problem (4.15) with the RIPADM, PMM, and ADM methods, we use four real data sets taken from the UCI Repository ([2]): Liver Disorders (BUPA), Australian Credit (AUS), Wisconsin Breast Cancer (WBC), and Pima Indians Diabetes (DIA). A brief information regarding each of the data sets is given below:

- **BUPA**: It contains $m = 345$ samples of patients, divided into $m_1 = 145$ and $m_2 = 200$, with $n = 6$ attributes.

- **AUS**: It concerns credit card applications and contains $m = 690$ samples, divided into $m_1 = 145$ and $m_2 = 200$, with $n = 14$ features.

- **WBC**: It contains $m = 569$ observations of tissue samples ($m_1 = 212$ diagnosed as malignant and $m_2 = 357$ as benign tumors) described by $n = 30$ features.
- **DIA**: It contains $m = 768$ samples of patients, divided into $m_1 = 268$ (tested positive) and $m_2 = 500$ (tested negative), with $n = 8$ attributes.

For each scheme and data set, we take the same starting point $(x^0, z^0, y^0) = \frac{1}{10}(\mathbf{1}, \mathbf{0}, \mathbf{0}) \in \mathbb{R}^{m_2} \times \mathbb{R}^{n+1} \times \mathbb{R}^{m_1}$ and $c_1 = 1$. In Table 4.3, we present the results that have been obtained in the numerical tests. The best method in terms of CPU time and number of iterations is highlighted in bold type.

| Data set | | RIPADM | PMM | ADM |
|---|---|---|---|---|
| BUPA | CPU time(s) | 922.73 | 1342.85 | **684.79** |
| | Objective Value | 1.350705 | 1.350714 | 1.350709 |
| | Iterations | 306 | **192** | **192** |
| AUS | CPU time(s) | **508.39** | 1815.14 | 724.12 |
| | Objective Value | 1.260135 | 1.260127 | 1.260134 |
| | Iterations | 101 | 148 | **99** |
| WBC | CPU time(s) | **5434.05** | 18213.59 | 7826.76 |
| | Objective Value | 1.496976 | 1.496976 | 1.496976 |
| | Iterations | 1407 | 1469 | **1406** |
| DIA | CPU time(s) | **67.08** | 318.25 | 82.04 |
| | Objective Value | 1.500000 | 1.500002 | 1.499990 |
| | Iterations | 13 | 24 | **9** |

Table 4.3: Numerical results for the norm-mixed TWSVM problem.

From this Table we observe that the best results (in number of iterations) were achieved using the ADM scheme in all data sets, however, the RIPADM one has better CPU time in three data sets (AUS, WBC and DIA).

## 4.3 Relaxated Decompositon Methods

In this section, we apply the RIPADM method with relaxation factor given by (3.5)-(3.7) to solves two examples of the previous section: constrained LASSO problem and norm-mixed

TWSVM problem. We want to show that for the relaxed RIPADM method, the general case with $\rho \in \left(1, \frac{1+\sqrt{5}}{2}\right)$ could lead to better numerical result than the special case with $\rho = 1$.

### 4.3.1 Constrained LASSO problem

The Constrained LASSO problem (4.3) can be reformulated introducing a slack variable as:

$$\min_{z} \left\{ \frac{1}{2} \|Dz - d\|^2 + \gamma \|z\|_1 \mid x + Bz = b, \;\; x \geq 0 \right\},$$

where $D \in \mathbb{R}^{r \times m}$, $d \in \mathbb{R}^r$, $B \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, and $\gamma > 0$ are problem data, and $m = n$. The problem data are randomly generated in the same way as in the subsection 4.2.1 and we take $\gamma = 1$. Here we only consider three different dimensions of the problem data $(r, n)$.

In the relaxed RIPADM method, the primal variable $x^{k+1}$ is computed by the closed formula (4.8), the primal variable $z^{k+1}$ is obtained by solving the problem (4.6) via CVX solver and the dual multiplier is calculated by the update ruler (3.7).

In Table 4.4, we list the averaged computing time in seconds, the objective function values and iterations number for relaxed RIPADM with $\rho = 1.62$, 1.6, 0.7 and 1. In this table, the computational results show that the RIPADM method with $\rho = 1.62$ is faster than with $\rho = 1$.

In Table 4.5, we show the computational results obtained from the application of the relaxed ADM method (see (3.2)-(3.4)) to the contrained LASSO problem. The relaxation factor $\rho$ takes the values 1.62, 1.6, 0.7 and 1. The ADM subproblems are solved by CVX solver. The results of this table show that the ADM with relaxation factor $\rho = 1.6$ is faster than with other values. Moreover, the relaxed RIPADM method with $\rho = 1$ or $\rho = 1.62$ is faster than the relaxed ADM method.

| r | n | | RIPADM | RIPADM | RIPADM | RIPADM |
|---|---|---|---|---|---|---|
| | | | $\rho = 1.62$ | $\rho = 1.6$ | $\rho = 0.7$ | $\rho = 1$ |
| 70 | 200 | CPU time(s) | **265.39** | 295.09 | 289.35 | 337.70 |
| | | Objective Value | 6.354818 | 6.354821 | 6.354808 | 6.354824 |
| | | Iterations | 135 | 148 | 142 | 158 |
| 100 | 300 | CPU time(s) | **545.97** | 774.39 | 792.64 | 765.18 |
| | | Objective Value | 7.855478 | 7.855489 | 7.855478 | 7.855478 |
| | | Iterations | 139 | 157 | 161 | 151 |
| 150 | 400 | CPU time(s) | 1594.23 | **1534.48** | 1674 | 2213.95 |
| | | Objective Value | 10.084388 | 10.084380 | 10.084391 | 10.084378 |
| | | Iterations | 145 | 144 | 158 | 203 |

Table 4.4: Numerical results of the relaxed RIPADM applied to the constrained LASSO problem.

| | | | ADM | ADM | ADM | ADM |
|---|---|---|---|---|---|---|
| | | | $\rho = 1.62$ | $\rho = 1.6$ | $\rho = 0.7$ | $\rho = 1$ |
| 70 | 200 | CPU time(s) | 444.22 | **291.88** | 882.46 | 648.58 |
| | | Objective Value | 6.354824 | 6.354808 | 6.354805 | 6.354808 |
| | | Iterations | 125 | 66 | 196 | 128 |
| 100 | 300 | CPU time(s) | 1322.09 | 1162.04 | 1220.48 | **1081.92** |
| | | Objective Value | 7.855489 | 7.855492 | 7.855491 | 7.855478 |
| | | Iterations | 126 | 108 | 117 | 100 |
| 150 | 400 | CPU time(s) | 2358.68 | **2166.24** | 2541.93 | 2758.52 |
| | | Objective Value | 10.084378 | 10.084397 | 10.084395 | 10.084378 |
| | | Iterations | 93 | 87 | 108 | 111 |

Table 4.5: Numerical results of the relaxed ADM applied to the constrained LASSO problem.

### 4.3.2 Twin support vector machine classiffier

The norm-mixed Twin Support Vector Machine problem (4.15) can be reformulated introducing a slack variable as:

$$\min_z \left\{ \|[D_1 \ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 |\ x + [D_2 \ \mathbf{e}_2]z = -\mathbf{e}_2 \right\}.$$

where $D_1 \in \mathbb{R}^{m_1 \times n}$ and $D_2 \in \mathbb{R}^{m_2 \times n}$ are real matrices, $c_1 = 1 \in \mathbb{R}$, and $\mathbf{e}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{e}_2 \in \mathbb{R}^{m_2}$ are vectors of ones. The problem data are real data sets taken from the UCI repository in the same way as in the subsection 4.2.3.

In the relaxed RIPADM method, the primal variable $x^{k+1}$ is computed by the closed formula (4.8), the z-problem of (4.17) is solved by CVX solver and the dual multiplier is calculated by the update ruler (3.7).

In Table 4.6, we list the averaged computing time in seconds, the objective function value and iterations number of relaxed RIPADM with $\rho = 1.62$, 1.6, 0.7 and 1. In this table, the computational results show that the RIPADM method with relaxation factor $\rho = 1.62$ is faster than with $\rho = 1$ in most cases.

In Table 4.7, we show the computational results obtained from the application of the relaxed ADM method (see (3.2)-(3.4)) to the norm-mixed TWSVM problem. The relaxation factor $\rho$ takes the values 1.62, 1.6, 0.7 and 1. The ADM subproblems are solved by CVX solver. For data sets BUPA and WBC, the results of this table show that the ADM with relaxation factor $\rho = 1.6$ is faster than with other values.

If we compare Tables 4.6 and 4.7, we can see that the CPU times of relaxed RIPADM method are lower than those of relaxed ADM method for data sets AUS, WBC and DIA.

| | | RIPADM | RIPADM | RIPADM | RIPADM |
|---|---|---|---|---|---|
| | | $\rho = 1.62$ | $\rho = 1.6$ | $\rho = 0.7$ | $\rho = 1$ |
| BUPA | CPU time(s) | **846.42** | 888.94 | 925.93 | 922.73 |
| | Objective Value | 1.350717 | 1.350717 | 1.350716 | 1.350705 |
| | Iterations | 290 | 293 | 317 | 306 |
| AUS | CPU time(s) | **495.20** | 824.43 | 765.65 | 508.39 |
| | Objective Value | 1.260125 | 1.260130 | 1.260133 | 1.260135 |
| | Iterations | 96 | 148 | 158 | 101 |
| WBC | CPU time(s) | **3369.13** | 4102.60 | 9169.87 | 5434.05 |
| | Objective Value | 1.496976 | 1.496976 | 1.496976 | 1.496976 |
| | Iterations | 869 | 880 | 2009 | 1407 |
| DIA | CPU time(s) | 201.41 | 91.76 | 80.50 | **67.08** |
| | Objective Value | 1.500005 | 1.500010 | 1.500010 | 1.500000 |
| | Iterations | 18 | 16 | 14 | 13 |

Table 4.6: Numerical results of relaxed RIPADM for the norm-mixed TWSVM problem.

| | | ADM | ADM | ADM | ADM |
|---|---|---|---|---|---|
| | | $\rho = 1.62$ | $\rho = 1.6$ | $\rho = 0.7$ | $\rho = 1$ |
| BUPA | CPU time(s) | 450.37 | **284.96** | 957.37 | 684.79 |
| | Objective Value | 1.350715 | 1.350718 | 1.350704 | 1.350709 |
| | Iterations | 114 | 69 | 245 | 192 |
| AUS | CPU time(s) | **673.28** | 1168.85 | 1553.64 | 724.12 |
| | Objective Value | 1.260128 | 1.260139 | 1.260126 | 1.260134 |
| | Iterations | 94 | 151 | 219 | 99 |
| WBC | CPU time(s) | 6511.15 | **5037.53** | 15938.98 | 7826.76 |
| | Objective Value | 1.496976 | 1.496976 | 1.496976 | 1.496976 |
| | Iterations | 868 | 879 | 2000 | 1406 |
| DIA | CPU time(s) | 119.24 | 119.93 | 97.29 | **82.04** |
| | Objective Value | 1.500007 | 1.500004 | 1.499990 | 1.499990 |
| | Iterations | 16 | 16 | 13 | 9 |

Table 4.7: Numerical results of relaxed ADM for the norm-mixed TWSVM problem.

## 4.4 RIPADM with Bregman Regularization

### 4.4.1 Prelimaries

Consider the following convex minimization problem:

$$u^+ = \underset{u \in \overline{C}}{\operatorname{argmin}}\ F(u) + \frac{1}{2\lambda} D_h(u, u^0),$$

where $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a closed proper convex function, and $h$ is a Bregman function with zone $C \subset \mathbb{R}^n$ ($C$ is an open nonempty convex set). As a byproduct of the analysis given in [32, Theorem 6], one gets the following dual characterization of $u^+$:

**Lemma 4.3.** *Suppose that $F(u) = (\phi^*)(-u) + \langle c, u\rangle$, where $\phi^*$ is the conjugate of a closed proper convex function $\phi$ and $c \in \mathbb{R}^n$ is a given vector.*

*If the pair $(w^+, u^+)$ satisfies*

$$\begin{aligned} w^+ &\in \underset{q \in \mathbb{R}^n}{\operatorname{argmin}}\ \phi(w) + \frac{1}{2\lambda} h^*(\nabla h(u^0) + 2\lambda(w - c)), && (4.19)\\ u^+ &= \nabla h^*(\nabla h(u^0) + 2\lambda(w^+ - c)). && (4.20)\end{aligned}$$

*Then*

$$u^+ = \underset{u \in \overline{C}}{\operatorname{argmin}}\ F(u) + \frac{1}{2\lambda} D_h(u, u^0). \tag{4.21}$$

*Proof.* From (4.19) we have that:

$$0 \in \partial \left[\phi(w) + \frac{1}{2\lambda} h^*(\nabla h(u^0) + 2\lambda(w - c))\right]_{q=q^+}.$$

Note that $h^*$ is differentiable by the strict convexity of $h$. Since $\operatorname{dom} \nabla(h^*) = \operatorname{im}(\nabla h) = \mathbb{R}^n$, $h^*$ is defined and differentiable everywhere, whence:

$$\frac{d}{dq}\left[h^*(\nabla h(u^0) + 2\lambda(w - c))\right] = 2\lambda \nabla h^*(\nabla h(u^0) + 2\lambda(w - c)).$$

Because $h^*(\nabla h(u^0) + 2\lambda(\cdot - c))$ is defined and differentiable for all $w \in \mathbb{R}^n$, [70, Theorem 23.8] gives that the definition of $w^+$ is equivalent to:

$$\begin{aligned}
&0 \in \partial\phi(w^+) + \nabla h^*(\nabla h(u^0) + 2\lambda(w^+ - c)) \\
&\Leftrightarrow 0 \in \partial\phi(w^+) + u^+ \\
&\Leftrightarrow -u^+ \in \partial\phi(w^+) \\
&\Leftrightarrow w^+ \in \partial\phi^*(-u^+)
\end{aligned}$$

Then $-w^+ \in -\partial\phi^*(-u^+)$. From [70, Theorem 23.9] we have

$$-w^+ \in \partial_u[\phi^*(-u)]_{u=u^+}. \tag{4.22}$$

Since $F(u) = \phi^*(-u) + \langle c, u\rangle$, [70, Theorems 23.8 and 23.9] give

$$\partial_u[\phi^*(-u)]_{u=u^+} + c \subseteq \partial F(u^+). \tag{4.23}$$

Thus, $\nabla h$ and $\nabla h^*$ being functional inverses, and by (4.20)

$$\begin{aligned}
\frac{1}{2\lambda}(\nabla h(u^0) - \nabla h(u^+)) &= \frac{1}{2\lambda}(\nabla h(u^0) - \nabla h(u^0) - 2\lambda(w^+ - c)) \\
&= -w^+ + c.
\end{aligned}$$

By (4.22) we have $-w^+ + c \in \partial_u[\phi^*(-u)]_{u=u^+} + c$, and then by (4.23) we get

$$-w^+ + c \in \partial F(u^+).$$

Therefore

$$\frac{1}{2\alpha}(\nabla h(u^0) - \nabla h(u^+)) \in \partial F(u^+),$$

which is equivalent to (4.21). $\square$

The interest of this result is that $u^+$ is obtained explicitly in terms of $w^+$, which in turn is obtained by solving an auxiliary penalized problem with penalty term given by the conjugate of the Bregman kernel.

### 4.4.2 Algorithm

We write the problem (P) under study:

$$(P) \qquad \min_{x,z} \{f(x) + g(z) \mid Ax + Bz = b,\ x \in \mathbb{R}^n_+\}.$$

Recall that the RIPADM method consists of the following steps:

$$\begin{aligned}
x^{k+1} &\approx \operatorname*{argmin}_{x \in \mathbb{R}^n_+} \ \ell_\lambda(x, z^k, y^k) + \frac{1}{2\lambda} d(x, x^k), \\
z^{k+1} &\approx \operatorname*{argmin}_{z \in \mathbb{R}^m} \ \ell_\lambda(x^{k+1}, z, y^k) + \frac{1}{2\lambda} \|z - z^k\|^2, \\
y^{k+1} &= y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b).
\end{aligned}$$

Suppose that $f$ is a quadratic function and $d$ is a Bregman distance. Then the Lemma 4.3 can be applied to the step 1 of the RIPADM method which produces a new strategy to obtain $x^{k+1}$. With this in mind, one may suggest a variant of our method.

Suppose that the function $f$ is a quadratic function of the form:

$$f(x) = \frac{1}{2}\langle Mx, x\rangle + \langle n, x\rangle. \tag{4.24}$$

Then

$$\ell_\lambda(x, z^k, y^k) = \frac{1}{2}\langle Qx, x\rangle + \langle c^k, x\rangle + \delta_k, \tag{4.25}$$

with

$$
\begin{aligned}
Q &= M + \lambda A^T A, && (4.26) \\
c^k &= n + A^T(y^k + \lambda(Bz^k - b)), && (4.27) \\
\delta_k &= \frac{\lambda}{2}\|Bz^k - b\|^2.
\end{aligned}
$$

Suppose that the matrix $A$ is full rank. Then

$$
\ell_\lambda(x, z^k, y^k) = \phi^*(-x) + \langle c^k, x\rangle + \delta_k
$$

with $\phi(x) = \frac{1}{2}\langle Q^{-1}x, x\rangle$.

Moreover, if we consider the Bregman kernel $h(x) = \sum_{i=1}^n x_i \log(x_i) - x_i$, then $h^*(x) = \sum_{i=1}^n \exp(x_i)$, and the Bregman distance $D_h$ is given by:

$$
D_h(x, x^k) = \sum_{i=1}^n x_i \log\left(\frac{x_i}{x_i^k}\right) - x_i + x_i^k, \forall x \in \mathbb{R}^n_{++}. \tag{4.28}
$$

Therefore, if $f$ is a quadratic function (4.24) and $d$ is the Bregman distance (4.28), the $x$-problem of the RIPADM method is equivalent to:

$$
x^{k+1} \approx \underset{x \in C}{\operatorname{argmin}} \ \frac{1}{2}\langle Qx, x\rangle + \langle c^k, x\rangle + \sum_{i=1}^n x_i \log\left(\frac{x_i}{x_i^k}\right) - x_i + x_i^k,
$$

where $Q$ and $c^k$ are given by (4.26) and (4.27), respectively.

Then the hypothesis of the Lemma 4.3 are satisfied with $F(x) = \ell_\lambda(x, z^k, y^k)$, and the Bregman kernel $h(x) = \sum_{i=1}^n x_i \log(x_i) - x_i$. Hence, the iteration $x^{k+1}$ is obtained by the formula:

$$
x_i^{k+1} = x_i^k \exp(2\lambda(q_i^k - c_i^k)), \ i = 1, 2, \ldots, n,
$$

where $q^k$ is the solution of the smooth unconstrained minimization problem:

$$q^k = \underset{q \in \mathbb{R}^n}{\operatorname{argmin}} \ \Phi_k(q),$$

with

$$\Phi_k(q) = \frac{1}{2}\langle Q^{-1}q, q\rangle + \frac{1}{2\lambda}\sum_{i=1}^{n} x_i^k \exp(2\lambda(q_i - c_i^k)).$$

Consider problem $(P)$ where $f$ is a quadratic function given by (4.24), and assume that $\operatorname{dom} f \cap C \neq \emptyset$. Let $d \in \mathcal{D}(C)$ and $\lambda$ be a positive scalar. Starting from a point $(x^0, z^0, y^0) \in C \times \mathbb{R}^m \times \mathbb{R}^p$, we generate the sequence $\{(x^k, z^k, y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$ via the following steps:

**Step 1.** Compute:

$$Q = M + \lambda A^T A \quad \text{and} \quad c^k = n + A^T(y^k + \lambda(Bz^k - b)). \tag{4.29}$$

Find the auxiliar variable $q^k$ solving:

$$q^k \approx \underset{q \in \mathbb{R}^n}{\operatorname{argmin}} \ \frac{1}{2}\langle Q^{-1}q, q\rangle + \frac{1}{2\lambda}\sum_{i=1}^{n} x_i^k \exp(2\lambda(q_i - c_i^k)). \tag{4.30}$$

Compute:

$$x_i^{k+1} = x_i^k \exp(2\lambda(q_i^k - c_i^k)), \ \forall i = 1.2, \ldots, n, \tag{4.31}$$

**Step 2.** Find $z^{k+1}$ solving:

$$z^{k+1} \approx \underset{z \in \mathbb{R}^m}{\operatorname{argmin}} \ g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2. \tag{4.32}$$

**Step 3.** Compute

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{4.33}$$

### 4.4.3 Applications

**Constrained LASSO problem**

Recall the constrined LASSO problem (4.4) again:

$$\min_{z} \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 \mid x + Bz = b,\ \ x \geq 0 \right\}.$$

This problem has the form of the problem (P) with $f(x) = 0$, $g(z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1$, $A = I$ and $\overline{C} = \mathbb{R}^n_+$. Then the RIPADM method with Bregman regularization (4.29)-(4.33) applied to this problem consists of the following steps:

**Step 1.** Compute:

$$Q = \lambda I \quad \text{and} \quad c^k = y^k + \lambda(Bz^k - b).$$

Find the auxiliar variable $q^k$ solving:

$$q^k \approx \operatorname*{argmin}_{q \in \mathbb{R}^n} \ \frac{1}{2\lambda}\|q\|^2 + \frac{1}{2\lambda}\sum_{i=1}^{n} x_i^k \exp(2\lambda(q_i - c_i^k)).$$

Compute:

$$x_i^{k+1} = x_i^k \exp(2\lambda(q_i^k - c_i^k)),\ \ \forall i = 1.2, \ldots, n,$$

**Step 2.** Find $z^{k+1}$ solving:

$$z^{k+1} \approx \operatorname*{argmin}_{z \in \mathbb{R}^m} \ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2.$$

**Step 3.** Compute

$$y^{k+1} = y^k + \lambda(x^{k+1} + Bz^{k+1} - b).$$

In Table 4.8 we present the computational results obtained when we apply the RIPADM method with Bregman regularization for solving the problem (4.4). We consider five schemes are characterized by different dimensions of the data $(r, n)$. For each scheme, we take the same starting point $(x^0, z^0, y^0) = (\mathbf{1}, \mathbf{1}, \mathbf{3}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, $\gamma = 1$, and set $m = n$. The problem data were randomly generated from the uniform distribution and with a single stream fixed, that is, `s=RandStream('mt19937ar','Seed',1)`, $D$=`rand(s,r,m)`, $B$=`rand(s,n,m)`, $d$=`rand(s,r,1)`, and $b$=`rand(s,n,1)`.

| | $r = 10$<br>$n = 30$ | $r = 30$<br>$n = 50$ | $r = 50$<br>$n = 100$ | $r = 70$<br>$n = 200$ | $r = 100$<br>$n = 300$ |
|---|---|---|---|---|---|
| CPU time(s) | 757.21 | 1014.12 | 6304.09 | $\infty$ | $\infty$ |
| Objective Value | 1.309526 | 3.343757 | 4.103253 | – | – |
| Iterations | 565 | 646 | 2121 | – | – |

Table 4.8: Numerical results for the constrained LASSO problem.

The Table 4.8 shows the solution times in seconds, each averaged over five instances with the same values of $r$ and $n$. If a method could not solve any of the five instances within the time limit (12 hours) for given $r$ and $n$, we denote this by an average solution time of $\infty$. Moreover, this Table shows the objective values and number of iterations for each scheme.

If we compare the Table 4.8 to Table 4.1, we can realize that both CPU times as the iterations of the Bregman method are much higher than RIPADM, PMM and ADM. For example, when $r = 50$ and $n = 100$ the CPU time of Bregman method is at least 38 times larger than the other methods.

**Contrained LASSO problem with a cost function**

We remember the constrained LASSO problem with a cost function (4.10) studied in subsection 4.2.2:

$$\min_{x,z} \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \frac{\beta}{2}\|x\|^2 \mid x + Bz = b,\ x \geq 0 \right\},$$

with $\beta > 0$. Clearly, this problem has the form of $(P)$ with $f(x) = \frac{\beta}{2}\|x\|^2$, $g(z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1$, $A = I$ and $\overline{C} = \mathbb{R}^n_+$. The problem data are randomly generated in the same way as in the subsection 4.2.2, and we take $\gamma = 1 = \beta$. Then RIPADM method with Bregman regularization (4.29)-(4.33) applied to this problem consists of the following steps:

**Step 1.** Compute:

$$Q = (\beta + \lambda)I \quad \text{and} \quad c^k = y^k + \lambda(Bz^k - b).$$

Find the auxiliar variable $q^k$ solving:

$$q^k \approx \underset{q\in\mathbb{R}^n}{\text{argmin}} \;\; \frac{1}{2(\beta+\lambda)}\|q\|^2 + \frac{1}{2\lambda}\sum_{i=1}^{n} x_i^k \exp(2\lambda(q_i - c_i^k)).$$

Compute:

$$x_i^{k+1} = x_i^k \exp(2\lambda(q_i^k - c_i^k)), \;\; \forall i = 1.2, \ldots, n,$$

**Step 2.** Find $z^{k+1}$ solving:

$$z^{k+1} \approx \underset{z\in\mathbb{R}^m}{\text{argmin}} \;\; \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2.$$

**Step 3.** Compute

$$y^{k+1} = y^k + \lambda(x^{k+1} + Bz^{k+1} - b).$$

In Table (4.9) we present the computational results obtained when we apply the RIPADM method with Bregman regularization for solving the problem (4.10). We consider five schemes that differ in the size of the problem data. For each scheme, we take the same starting point $(x^0, z^0, y^0) = (\mathbf{1}, \mathbf{1}, \mathbf{3}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, and set $m = n$.

| | $r = 10$<br>$n = 30$ | $r = 30$<br>$n = 50$ | $r = 50$<br>$n = 100$ | $r = 70$<br>$n = 200$ | $r = 100$<br>$n = 300$ |
|---|---|---|---|---|---|
| CPU time(s) | 392.71 | 895.10 | 5893.81 | $\infty$ | $\infty$ |
| Objective Value | 3.715827 | 6.855118 | 10.501293 | – | – |
| Iterations | 280 | 442 | 1693 | – | – |

Table 4.9: Computational results for the constrained LASSO problem with a cost function.

If we compare the Table 4.9 to Table 4.2, we can see that both CPU times as the iterations of the RIPADM method with Bregman regularization are much higher than RIPADM, PMM and ADM for all the schemes.

**Twin support vector machine classiffier**

Recall the norm-mixed Twin Support Vector Machine problem (4.16) again:

$$\min_z \left\{ \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 |\ x + Bz = -\mathbf{e}_2,\ x \geq 0 \right\}.$$

This problem has the form of the problem (P) with $f(x) = 0$, $g(z) = \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2$, $A = I$, $B = [D_2\ \mathbf{e}_2]$, and $\overline{C} = \mathbb{R}^n_+$. The problem data are real data sets taken from the UCI repository in the same way as in the subsection 4.2.3, and we take $c_1 = 1$. Then the RIPADM method with Bregman regularization (4.29)-(4.33) applied to this problem consists of the following steps:

**Step 1.** Compute:

$$Q = \lambda I \quad \text{and} \quad c^k = y^k + \lambda(Bz^k - b).$$

Find the auxiliar variable $q^k$ solving:

$$q^k \approx \operatorname*{argmin}_{q\in\mathbb{R}^n}\ \frac{1}{2\lambda}\|q\|^2 + \frac{1}{2\lambda}\sum_{i=1}^n x_i^k \exp(2\lambda(q_i - c_i^k)).$$

Compute:

$$x_i^{k+1} = x_i^k \exp(2\lambda(q_i^k - c_i^k)),\ \ \forall i = 1.2, \ldots, n,$$

**Step 2.** Find $z^{k+1}$ solving:

$$z^{k+1} \approx \operatorname*{argmin}_z\ \ \|[D_1\ \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2.$$

**Step 3.** Compute

$$y^{k+1} = y^k + \lambda(x^{k+1} + Bz^{k+1} - b).$$

In Table 4.10, we present the computational results obtained when we apply the RIPADM algorithm with Bregman regularization for solving the problem (4.16). We use four real data sets taken from the UCI Repository ([2]): Liver Disorders (BUPA), Australian Credit (AUS), Wisconsin Breast Cancer (WBC), and Pima Indians Diabetes (DIA) (see subsection 4.2.3 for more details). For each data set, we take the same starting point $(x^0, z^0, y^0) = \frac{1}{10}(\mathbf{1}, \mathbf{0}, \mathbf{0}) \in \mathbb{R}^{m_2} \times \mathbb{R}^{n+1} \times \mathbb{R}^{m_1}$ and $c_1 = 1$.

| | BUPA | AUS | WBC | DIA |
|---|---|---|---|---|
| CPU time(s) | 8010.57 | 1997.49 | 1511.35 | 9454.13 |
| Objective Value | 1.350715 | 1.260124 | 1.496984 | 1.500009 |
| Iterations | 978 | 163 | 131 | 769 |

Table 4.10: Computational results for the norm-mixed TWSVM problem.

The Table 4.10 shows the solution times in seconds, each averaged over five instances with the same data set. Moreover, this Table shows the objective values and number of iterations for each scheme.

If we compare the Table 4.10 to Table 4.3, we can see that both CPU times as the iterations of the algorithm with Bregman regularization are much higher than RIPADM, PMM and ADM for data sets BUPA, AUS and DIA. However, for the data set WBC the CPU time of algorithm with Bregman regularization is much less than the RIPADM method with Log-quad regularization and relaxation factor 1.62.

# PART II

# Chapter 5

# Preliminaries on economic equilibrium model

## 5.1 Introduction

It is important to have computational methods that provide approximate solutions to the economic problem of finding equilibrium prices in a market economics, and to have computational procedures for the study of the potential effects of variations in the parameters and specifications that determine an economic model, since such methods will allow us to perform quantitative analysis on policies to be implemented.

Generally speaking, most of the algorithms developed to find and analyze equilibria in economic models have been based on the available existence proofs. The equilibrium existence proofs have as a central argument the existence of fixed points for a function based on the conditions of supply and aggregate demand (see, e.g., [1], [28], [66]). Thus, standard algorithms for the approximate computation of economic equilibria are based on variations to the successive approximation of fixed points, by exploiting the special characteristics of this problem. These algorithms have good results in small and particular instances, but have important deficiencies and drawbacks in medium and large problems, and have unstable behaviors with respect to changes in the parameters of the problem.

More recently, algorithms have been developed to calculate equilibrium prices as max-inf points of underlying bifunction problems. This is due to new equilibrium existence proofs through the application of variational analysis (see, e.g., [5], [57]). Thus the focus has changed from the use of fixed-point algorithms to construct algorithms that can find the max-inf points of a bifunction that is characteristic of the economy in study, and which collects all parameters that determine it, which is known as the Walrasian function in this case.

## 5.2 Walrasian function

We consider a pure exchange model (economy without production) or equivalently a Walras barter problem, of Arrow-Debreau [28]. In this economy, there are $I$ agents (consumers) $i \in \mathcal{I} = \{1, ..., I\}$, and $L$ commodities $j \in \mathcal{L} = \{1, ..., L\}$. Each agent $i$ has an initial endowment $e_i \in \mathbb{R}^L$ (a bundle of goods to be traded), a survival set $X_i \subset \mathbb{R}^L_+$ and an utility function $u_i : X_i \to \mathbb{R}$. These endowments, survival sets and utilities are the primitives of the pure exchange model, so the economy is described by

$$\mathcal{E} = \{(u_i, X_i, e_i)|\ i \in \mathcal{I}\}.$$

Trading takes place at a per-unit market price $p_j$ for good $j$, $j = 1, ..., L$. The bundle of goods agent $i$ could acquire is limited by the budgetary contraint $p^T x \leq p^T e_i$. It is assumed that agents act as utility maximizers. Thus, given $p \in \mathbb{R}^L_+$, each agent $i \in \mathcal{I}$ will end up with its (consumption) demand:

$$c_i(p) = \operatorname*{argmax}_{x \in X_i} \left\{ u_i(x) \mid p^T x \leq p^T e_i \right\}. \tag{5.1}$$

which we assume to be well (uniquely) defined. Note that $c_i(p) = c_i(\alpha p)$ for any positive scalar $\alpha$, i.e., the agents demand functions are homogeneous of degree 0 with respect to prices. So, we may as well restrict the choice of $p$ to $\Delta = \{p \in \mathbb{R}^L_+ \mid \sum_{j=1}^L p_j = 1\}$.

**Definition 5.1.** The excess supply function of the economy $\mathcal{E}$ is defined as:

$$s(p) = \sum_{i=1}^{I} s_i(p), \tag{5.2}$$

where $s_i(p) = e_i - c_i(p)$.

Following Walras [80], a price vector $\bar{p} \in \Delta$ such that $s(\bar{p}) \geq 0$ is called an *equilibrium price*. On the other hand, Jofré and Wets [57] introduced the Walrasian function which allows to give a new characterization of equilibrium price.

**Definition 5.2.** For a pure exchange economy $\mathcal{E}$, the Walrasian function is defined as

$$W(p,q) = q^T s(p),$$

where $(p,q) \in \Delta \times \Delta$, and $s$ is the excess supply function defined by (5.2).

**Definition 5.3.** Given $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, and a bifunction $F : X \times Y \to \mathbb{R}$, a point $\bar{x}$ is a maxinf-point of $F$ if:

$$\bar{x} \in \operatorname*{argmaxinf}_{(x,y)\in X\times Y} F \;= \operatorname*{argmax}_{x\in X} \left( \inf_{y\in Y} F(x,y) \right).$$

**Lemma 5.4.** *[58, Proposition 2.4] Every maxinf-point $\bar{x} \in \Delta$ of the Walrasian function $W$ such that $W(\bar{p},\cdot) \geq 0$ on $\Delta$ is an equilibrium price.*

Jofré et al. [56] proposed a method to solve Walras equilibrium models that is based on finding a maxinf-point of the Walrasian function $W$ by an approximating squeme. The numerical procedure relies on an augmentation of this function. The proposed method generates a sequence of approximate maxinf-point $\{p^\nu\}$ of a sequence of augmented Walrasian function $\{W^\nu\}$, and, under some assumptions, the sequence $\{p^\nu\}$ converges to $\bar{p}$ a maxinf-point of $W$.

Given the augmenting function $\sigma$ and a scalar $r > 0$, *the augmented Walrasian*, by definition, is

$$\tilde{W}_r(p,q) = \sup_{u\in\mathbb{R}^n} \{q^T u - V(p,u) - r\sigma(u)\},$$

where

$$V(p,u) = \sup_{z\in\Delta} [u^T z - W(p,z)].$$

One can re-write the definition of $\tilde{W}_r$ as (see [56] for more details):

$$\tilde{W}_r(p,q) = \inf_z \{W(p,q-z) + r\sigma^*(r^{-1}z)\},$$

where $\sigma^*$ is the conjugate of $\sigma$. For example, one can consider $\sigma = \frac{1}{2}|\cdot|_2^2$, the augmented Walrasian takes the form:

$$\tilde{W}_r(p,q) = \min_{z\in\Delta} \left[ W(p,z) + \frac{1}{2r^\nu}|z-q|_2^2 \right].$$

The proposed method in [56] consists of the followings steps:

- At iteration $\nu+1$, given $p^\nu \in \Delta$ with $r = r_{\nu+1}$ ($\geq r_\nu$), the Phase I (or primal) consists in solving

$$q^{\nu+1} \in \operatorname*{argmin}_{q\in\Delta} \tilde{W}^{\nu+1}(p^\nu, q).$$

- How to carry out the next step will depend on the shape and the properties of the demand functions. For example, this turns out to be rather simple when the utility functions are the Cobb-Douglas type, defining the Phase II (or dual) as finding

$$p^{\nu+1} \in \operatorname*{argmax}_{p\in\Delta} \tilde{W}^{\nu+1}(p, q^{\nu+1}).$$

We can see that in the augmented Walrasian $\tilde{W}_r$ appears the Walrasian function $W$, and therefore the excess supply function $s$ (5.2). Then in the computational implementation of this method one needs to solve the utility problem for each agent participating in the economic.

# Chapter 6

# Computational experiences on utility maximization problems

## 6.1 Methods for solving the agent problem

In the previous chapter we saw that the proposed method in [56] for solving economic equilibrium problems requires solving utility maximization problems of the agents. In this chapter we study two methods to solve the agent problem: primal-dual interior-point method and gradient projection with acceleration. Note that these methods exist in the literature (see [41], bertsekasnonlinear), but we are interested in the computational implementation of them.

Below we give the assumptions for the agent problem, and then give a detailed description of the algorithms under study.

Each agent $i$ is associated with a survival set $X_i \subset \mathbb{R}^L_+$, an utility function $u_i : X_i \to \mathbb{R}$ and initial endowment $e_i \in \mathbb{R}^L_+$. The main assumptions are, for all agent $i \in I$,

- $X_i = \mathbb{R}^L_+$,
- $u_i$ is continuous,

- $u_i$ is strictly concave,
- $u_i$ is nondecresing in each argument, and
- $e_i \in \mathbb{R}^L_{++}$.

Each agent markets its products in order to maximize profits. Then, given a price $p \in \mathbb{R}^L_+$, for each agent $i$ we have to solve the following maximization problem:

$$\begin{aligned} \max \quad & u_i(x) \\ \text{subject to} \quad & p^T x \leq p^T e_i \\ & x \in \mathbb{R}^L_+ \end{aligned} \tag{6.1}$$

Due to the assumption that $u_i$ is nondecreasing in each argument, we have that the optimal solution satisfies the condition $p^T x = p^T e_i$. The problem (6.1) is equivalent to the following convex minimization problem:

$$\begin{aligned} -\min \quad & -u_i(x) \\ \text{subject to} \quad & p^T x = p^T e_i \\ & x \in \mathbb{R}^L_+ \end{aligned}$$

We rewrite the last problem as:

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned} \tag{6.2}$$

where $f = -u_i(x)$ is a convex function, and the vectors $p \in \mathbb{R}^L_+$ and $e \in \mathbb{R}^L_{++}$ are data problem. Below, we study two methods for solving the convex minimization problem (6.2).

### 6.1.1 Primal-dual interior-point method

When $f$ is $C^2$, the problem (6.2) can probably be solved easily with a primal-dual interior-point method (PDIPM). Moreover, when $f$ is separable, the interior-point solver can easily exploit this structure. We sketch the method below.

We introduce a dual slack variable $s \in \mathbb{R}^L$, along with the notation $\mathbf{1} = (1, 1, \dots, 1)^T \in \mathbb{R}^L$, and a vector-valued function $F : \mathbb{R}^L \times \mathbb{R} \times \mathbb{R}^L \to \mathbb{R}^L \times \mathbb{R} \times \mathbb{R}^L$ defined as follows:

$$F(x, \mu, s) := \begin{bmatrix} \nabla f(x) + \mu p - s \\ p^T(x - e) \\ XS\mathbf{1} \end{bmatrix}, \tag{6.3}$$

where

$$X = \mathrm{diag}(x_1, x_2, ..., x_L), \quad S = \mathrm{diag}(s_1, s_2, ..., s_L).$$

Using this notation, we can state the optimality conditions for the problem (6.2) as follows:

$$F(x, \mu, s) = 0, \;\; (x, s) \geq 0. \tag{6.4}$$

At each iterate of a basic primal-dual interior-point method, we calculate a modified Newton step for the nonlinear equations $F(x, \mu, s) = 0$, and take a step along this direction, choosing the steplength so as to continue to satisfy strictly the inequalities $(x, s) \geq 0$. Specifically, given the current iterate $(x, \mu, s)$ with $(x, s) > 0$, we obtain the new iterate $(x^+, \mu^+, s^+)$ by taking the step:

$$(x^+, \mu^+, s^+) = (x, \mu, s) + \alpha(\Delta x, \Delta\mu, \Delta s), \tag{6.5}$$

where $\alpha$ is chosen to ensure that $(x^+, s^+) > 0$. Moreover, the step $(\Delta x, \Delta\mu, \Delta s)$ is obtained by solving the system:

$$\begin{bmatrix} \nabla^2 f(x) & p & -I \\ p^T & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \mu \\ \Delta s \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \mu p - s \\ p^T(x-e) \\ XS\mathbf{1} - \gamma(x^T s/L)\mathbf{1} \end{bmatrix}, \tag{6.6}$$

where $\gamma \in (0,1)$ is the centering parameter, usually chosen adaptively according to a formula of Mehrotra. (See Wright [41] for details on the choice of this parameter in the case of linear programming algorithms, which generalizes in a straightforward way to the nonlinear case).

A common formula for $\alpha$ is

$$\alpha = .95 \times \min\left(1, \min_{\Delta x_i < 0} \frac{x_i}{\Delta x_i}, \min_{\Delta s_i < 0} \frac{s_i}{\Delta s_i}\right). \tag{6.7}$$

A second-order correction strategy may also be applied; again see Wright [41] for details.

### 6.1.2 Gradient projection with acceleration

We next describe a gradient projection approach for (6.2) with simple Newton acceleration (AGP). The key operation is projection onto the feasible set for that is

$$\{x \geq 0 \,:\, p^T x = b\}, \tag{6.8}$$

where $b := p^T e$. We actually work with a slightly more general version of this set, since for some utility functions we need to keep the values of $x$ away from zero. Hence, our generalized feasible set is

$$\Omega := \{x \geq \epsilon \mathbf{1} \,:\, p^T x = b\}, \tag{6.9}$$

where $\epsilon$ is a user-supplied nonnegative parameter and $\mathbf{1}$ is the vector whose elements are all 1. Of course, we must assume that this set is nonempty, that is, $\epsilon p^T \mathbf{1} \leq b$. Note that $\Omega$ is convex. In this notation, we can state (6.2) as follows:

$$\min_{x \in \Omega} f(x). \tag{6.10}$$

We denote by $P_\Omega$ the operation of projecting onto the set $\Omega$, that is,

$$P_\Omega(z) := \arg\min_t \frac{1}{2}\|t - z\|_2^2 \text{ s.to } p^T t = b,\ t \geq \epsilon \mathbf{1}. \tag{6.11}$$

### Gradient Projection

A standard gradient projection method for (6.10) (see for example Bertsekas [12]) can be stated as follows.

**Algorithm 1** Gradient Projection with Backtracking

Given $f : \mathbb{R}^L \to \mathbb{R}$, $p \in \mathbb{R}^L$ with $p \geq 0$, $b > 0$;
Choose $x^0 \in \Omega$, $\bar{\alpha}_0 > 0$, $c_1 \in (0,1)$, and $\tau > 1$;
**for** $k = 0, 1, 2, \ldots$ **do**
  Set $\alpha_k \leftarrow \bar{\alpha}_k$; {Assign initial guess for step length}
  **while** $f(P_\Omega(x^k - \alpha_k \nabla f(x^k))) > f(x^k) + c_1 \nabla f(x^k)^T [P_\Omega(x^k - \alpha_k \nabla f(x^k)) - x^k]$ **do**
    $\alpha_k \leftarrow \alpha_k / \tau$; {Decrease step length}
  **end while**
  $x^{k+1} \leftarrow P_\Omega(x^k - \alpha_k \nabla f(x^k))$; {Take the step}
  $\bar{\alpha}_{k+1} \leftarrow \tau \alpha_k$; {Assign starting value for next iteration}
**end for**

### Projection

The most complicated operation in this algorithm is computation of the projection $P_\Omega$ (6.11), but this can be performed efficiently, as pointed out by Dai and Fletcher [27]. We outline the basic approach here.

First note that the KKT conditions for the solution $t$ of (6.11) are as follows:

$$0 \leq t - z + \mu p \perp t - \epsilon \mathbf{1} \geq 0, \quad p^T t = b. \tag{6.12}$$

where $\mu$ is the Lagrange multiplier for the constraint $p^T t = b$. For a given $\mu$, the vector $t(\mu)$ that satisfies the conditions $0 \le t - z - \mu p \perp t \ge 0$ can be written in closed form, as follows:

$$t(\mu) = \max(\epsilon \mathbf{1}, z - \mu p). \tag{6.13}$$

We seek a value of $\mu$ for which $p^T t(\mu) = b$. This will yield a solution of the projection problem (6.11). Essentially this task reduces to solving the scalar but nonsmooth equation $\phi(\mu) = 0$, where

$$\phi(\mu) := p^T t(\mu) - b. \tag{6.14}$$

We note the following facts about $\phi(\mu)$, all proved easily.

- $\phi(\mu)$ is a decreasing function of $\mu$. Proof: Since $p \ge 0$, each $t_i(\mu)$ is a decreasing function of $\mu$. Thus $p^T t(\mu)$ is also decreasing in $\mu$, and the claim follows.

- $\phi(\mu)$ is piecewise linear, with breakpoints at $(z_i - \epsilon)/p_i$ for those indices $i$ with $p_i > 0$.

- $\lim_{\mu \to \infty} \phi(\mu) = -b + \epsilon p^T \mathbf{1}$. As proof we note that $\phi(\mu) = \sum_{i=1}^L p_i t_i(\mu) - b$. When $p_i > 0$, we have $t_i(\mu) = \epsilon$ for all $\mu$ sufficiently large, so that $p_i t_i(\mu) = \epsilon p_i$ for all such $i$. When $p_i = 0$, we have $p_i t_i(\mu) = 0$ for *all* values of $\mu$. We conclude that $\sum_{i=1}^L p_i t_i(\mu) = \epsilon p^T \mathbf{1}$ for all $\mu$ sufficiently large.

- $\lim_{\mu \to -\infty} \phi(\mu) = \infty$. As proof, we note that $p_i > 0$ for at least one index $i$. For this index, we have $t_i(\mu) \uparrow \infty$ as $\mu \downarrow -\infty$. Thus, for this $i$, we have $p^T t(\mu) \ge p_i t_i(\mu) \uparrow \infty$, proving the claim.

These observations suggest an efficient approach for finding the optimal $\mu$ and thus solving the projection subproblem (6.11). First, we calculate all the breakpoints and order them: an $O(L \log L)$ operation in general. We can then work through them in decreasing order, calculating the function value at each, and stopping when $\phi$ increases above zero. (The recalculation of $\phi$ at each breakpoint requires just a few operations.) The optimal $\mu$ can thus be identified by interpolating between the last two values of $\mu$.

The procedure for finding $\mu$ has been implemented in Matlab. We give a few additional details of the procedure here. First, we define the breakpoints as follows:

$$\mu_i' := \begin{cases} (z_i - \epsilon)/p_i & \text{if } p_i > 0 \\ -\infty & \text{otherwise.} \end{cases}$$

It makes use of the following form of $\phi(\mu)$:

$$\phi(\mu) = \sum_{\mu \le \mu_i'} p_i z_i + \sum_{\mu > \mu_i'} p_i \epsilon - \mu \sum_{\mu \le \mu_i'} p_i^2 - b,$$

We initialize partial sums $S_p = \sum_{i=1}^{L} p_i^2$ and $S_z = \sum_{i=1}^{L} p_i z_i$ and note that for all $\mu$ sufficiently negative, we have $\phi(\mu) = S_z - \mu S_p - b$. We sort the breakpoints $\mu_i'$, in increasing order and work through them one at a time, performing the following operations at each index.

- For the current values of $S_z$ and $S_p$, calculate $\mu = (S_z - b)/S_p$, and choose this value as the optimal $\mu$ if it is smaller than the next breakpoint.

- Otherwise, if $j$ is the index for the next breakpoint, decrement the partial sums as follows: $S_z \leftarrow S_z - p_j z_j + p_j \epsilon$, $S_p \leftarrow S_p - p_j^2$.

This procedure is guaranteed to terminate before the final breakpoint is reached, because $\phi$ takes on the nonpositive value $-b + \epsilon p^T \mathbf{1}$ at and above this breakpoint.

The cost of this procedure is $O(L \log L)$. Calculation of the breakpoints is $O(L)$, the operations at each iteration require $O(1)$ (and there are a maximum of $L$ iterations), and the sort requires $O(L \log L)$.

### Accelerated Gradient Projection

At each iteration $x^k$ of Algorithm 1, we denote the set of nonzero components by $\mathcal{A}_k$, that is,

$$\mathcal{A}_k := \{i \,:\, x_i^k > 0\}. \tag{6.15}$$

(We denote $\mathcal{A}_k^c := \{1, 2, \dots, L\} \setminus \mathcal{A}_k$.) We can accelerate the algorithm by calculating a reduced Newton step, in which the zero components of $x^k$ are fixed at zero. If this step yields a reduction in the objective $f$ and also maintains nonnegativity of the nonzero components of $x^k$, we accept it. Otherwise, we discard the step and move to the next iteration of Algorithm 1.

The reduced Newton step from a point $x \in \Omega$ whose nonzero set is $\mathcal{A} := \{i \,:\, x_i > 0\}$ is defined as follows:

$$\min_d \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d \;\text{ s.t. }\; d_i = 0 \text{ for } i \notin \mathcal{A}, \;\; p^T d = 0. \tag{6.16}$$

Denoting by $p_{\mathcal{A}}$ the subvector of $p$ that corresponds to the indices $i \in \mathcal{A}$ (similarly for $d_{\mathcal{A}}$ and $[\nabla f(x)]_{\mathcal{A}}$) and by $[\nabla^2 f(x)]_{\mathcal{A}\mathcal{A}}$ the submatrix of $\nabla^2 f(x)$ whose row and column indices are both in $\mathcal{A}$, we can obtain the solution of (6.16) by solving the following linear system:

$$\begin{bmatrix} [\nabla^2 f(x)]_{\mathcal{A}\mathcal{A}} & p_{\mathcal{A}}^T \\ p_{\mathcal{A}} & 0 \end{bmatrix} \begin{bmatrix} d_{\mathcal{A}} \\ \nu \end{bmatrix} = \begin{bmatrix} -[\nabla f(x)]_{\mathcal{A}} \\ 0 \end{bmatrix}. \tag{6.17}$$

We obtain the solution $d$ of (6.16) by filling out $d_{\mathcal{A}}$ with zeros.

To test whether $d$ is a suitable step, we test two criteria.

- We must have $x_i + d_i \geq 0$ for all $i \in \mathcal{A}$;
- We must have $f(x + d) \leq f(x) + c_1 \nabla f(x)^T d$, for some small positive constant $c_1$ (sufficient decrease condition).

If either of these conditions does not hold, we reject the step and return to the standard gradient projection iteration of Algorithm 1.

We need only calculate these reduced Newton steps when the active set $\mathcal{A}$ has "settled down," that is, did not change on the last iteration.

We now an describe enhanced version of Algorithm 1 that incorporates Newton acceleration.

**Algorithm 2** Gradient Projection with Backtracking and Acceleration

Given $f : \mathbb{R}^L \to \mathbb{R}$, $p \in \mathbb{R}^L$ with $p \geq 0$, $b > 0$;
Choose $x^0 \in \Omega$, $\bar{\alpha} > 0$, $c_1 \in (0, 1)$, and $\tau > 1$;
Set **accel** $\leftarrow$ false and $\mathcal{A}_0 \leftarrow \{i \,:\, x_i^0 > 0\}$;
**for** $k = 0, 1, 2, \ldots$ **do**
  **if accel** == true **then**
    Compute $d^k_{\mathcal{A}_k}$ and $\nu_k$ by solving (6.17) with $x = x^k$ and $\mathcal{A} = \mathcal{A}_k$, and set $d^k_{\mathcal{A}_k^c} \leftarrow 0$;
    **if** $x^k_{\mathcal{A}_k} + d^k_{\mathcal{A}_k} > 0$ and $f(x^k + d^k) \leq f(x^k) + c_1 \nabla f(x^k)^T d^k$ **then**
      Set $x^{k+1} \leftarrow x^k + d^k$, $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k$; {Successful reduced Newton step}
    **else**
      Set $x^{k+1} \leftarrow x^k$;
      **accel** $\leftarrow$ false; {Do a gradient projection step instead}
    **end if**
  **else**
    Set $\alpha_k \leftarrow \bar{\alpha}$; {Initial choice of $\alpha_k$ based on last projected gradient steplength}
    **while** $f(P_\Omega(x^k - \alpha_k \nabla f(x^k))) > f(x^k) + c_1 \nabla f(x^k)^T [P_\Omega(x^k - \alpha_k \nabla f(x^k)) - x^k]$ **do**
      $\alpha_k \leftarrow \alpha_k / \tau$; {Decrease step length}
    **end while**
    $x^{k+1} \leftarrow P_\Omega(x^k - \alpha_k \nabla f(x^k))$; {Take the step}
    $\mathcal{A}_{k+1} := \{i = 1, 2, \ldots, n \,:\, x_i^{k+1} > 0\}$;
    **if** $\mathcal{A}_{k+1} = \mathcal{A}_k$ **then**
      Set **accel** $\leftarrow$ true; {Try acceleration on next iteration}
    **else**
      Set **accel** $\leftarrow$ false;
    **end if**
    $\bar{\alpha} \leftarrow \tau \alpha_k$; {Assign starting value for next iteration}
  **end if**
**end for**

## 6.2 Numerical experiments

This section shows the numerical results of the PDIPM and AGP which were applied to the problem of agent using two types of utility functions: Constant elasticity of subtitution and Cobb-Douglas.

In the computational implementation of the proposed method in [56], the problem of agent is solved using the interior point method, Ipopt, implemented by [79].

The numerical experiments were done on a laptop with Intel Core i3 processor, 1.80 GHz, 4 GB RAM. The operating system of the laptop is Windows 8, and the program used is MATLAB version 8.2.0.701 (R2013b).

### 6.2.1 Constant elasticity of subtitution utility function

The Constant Elasticity of Subtitution (CES) utility function is defined as:

$$u(x) = \left( \sum_{j=1}^{L} (a_j)^{\frac{1}{b}} (x_j)^{\frac{b-1}{b}} \right)^{\frac{b}{b-1}},$$

where the paremeters $a_j$ and $b$ are positive scalars. We consider $b = \frac{1}{2}$, and $a_j = 1$ for all $j$.

The utility maximization problem has the form:

$$\begin{aligned} \max \quad & u(x) \\ \text{s. t.} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned}$$

For practical purposes, this problem can be stated as:

$$\begin{aligned} \max \quad & \ln(u(x)) \\ \text{s. t.} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned}$$

This problem is equivalent to:

$$\begin{aligned} -\min \quad & -\ln(u(x)) \\ \text{s. t.} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned} \tag{6.18}$$

In the particular case of the CES function, the above problem is:

$$\begin{aligned} -\min \quad & \frac{b}{1-b} \ln\left(\sum_{j=1}^{L} (a_j)^{\frac{1}{b}} (x_j)^{\frac{b-1}{b}}\right) \\ \text{s. t.} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned} \tag{6.19}$$

In this section we show the computational results obtained applying the primal-dual interior-point method (PDIPM) and the accelerated gradient projection method (AGPM) to the problem (6.19). The problem data were randomly generated from the uniform distribution and with a single stream fixed, that is, `s=RandStream('mt19937ar','Seed',1)`, $p$=`rand(s,L,1)` and $e$=`rand(s,L,1)`.

In primal-dual interior-point method, we take the starting point $(x^0, \mu^0, s^0) = (\mathbf{100}, 0, \mathbf{100})$ where $\mathbf{100} = (100, 100, ..., 100)^T$, $\gamma = 0.2$ and $\alpha$ is given by the formula (6.7). We define the primal residual $r_p = p^T(x - e)$ and the dual residual $r_d = \nabla f(x) + \mu p - s$. The stopping criteria of this method is given by:

$$\max\left(\|r_p\|,\ \|r_d\|,\ \frac{x^T s}{L}\right)\ \leq\ tol,$$

where $tol = 10^{-5}$.

In accelerated gradient projection method, we take the starting point $x^0 = P_\Omega(x)$ where $x = rand(s, n, 1)$, $\bar{\alpha} = 1$, $c_1 = 10^{-3}$ and $\tau = 0.5$. The stopping criteria of this method is given by:

$$f(x^{k-1}) - f(x^k)\ \leq\ tol,$$

where $tol = 10^{-5}$.

In Table 6.1 we present the numerical results obtained when we apply the PDIPM and AGPM methods for solving the CES minimization problem (6.19).

| $L$ | | PDIPM | AGPM |
|---|---|---|---|
| 10 | CPU time(s) | 0.03316 | 0.02984 |
| | Objective Value | 3.00148214 | 3.00148214 |
| | Iterations | 120 | 14 |
| 50 | CPU time(s) | 0.10359 | 0.04066 |
| | Objective Value | 4.79846905 | 4.79846904 |
| | Iterations | 127 | 14 |
| 75 | CPU time(s) | 0.18475 | 0.05654 |
| | Objective Value | 5.12140176 | 5.12140174 |
| | Iterations | 128 | 14 |
| 100 | CPU time(s) | 0.24772 | 0.08387 |
| | Objective Value | 5.32303457 | 5.32303457 |
| | Iterations | 117 | 16 |

Table 6.1: Numerical results for the CES utility problem.

Table 6.1 shows the average of runtime in seconds (run five times), the objective values and the number of iterations of the PDIPM and AGPM methods apply to Problem (6.19) for different values of $L$. From Table 6.1, we observe that the AGPM method uses fewer iterations and less CPU time than PDIPM method in all experiments.

### 6.2.2 Cobb-Douglas utility function

The Cobb-Douglas utility function is defined by:

$$u(x) = \prod_{j=1}^{L} x_j^{\alpha_j},$$

where $\sum_{j=1}^{L} \alpha_j = 1$, $\alpha_j > 0$ for all $j$. In this particular case, the utility maximization problem (6.18) take the form:

$$\begin{aligned} -\min \quad & -\sum_{j=1}^{L} \alpha_j \ln(x_j) \\ \text{s. t.} \quad & p^T x = p^T e \\ & x \geq 0 \end{aligned} \tag{6.20}$$

In this section we show the computational results obtained applying the primal-dual interior-point method (PDIPM) and the accelerated gradient projection method (AGPM) to the problem (6.20). The problem data were randomly generated from the uniform distribution and with a single stream fixed, that is, `s=RandStream('mt19937ar','Seed',1)`, $p$=`rand(s,L,1)` and $e$=`rand(s,L,1)`.

In Table 6.2 we present the numerical results obtained when we apply the PDIPM and AGPM methods for solving the Cobb Douglas utility maximization problem (6.20).

Table 6.2 shows the average of runtime in seconds (run five times), the objective values and the number of iterations of PDIPM and AGPM methods apply to Problem (**??**) for different values of $L$. From Table 6.2, we observe that the AGPM method uses fewer iterations in comparison with the PDIPM method, moreover AGPM uses less CPU time in three out of four experiments.

| $L$ | | PDIPM | AGPM |
|---|---|---|---|
| 10 | CPU time(s) | 0.02414 | 0.02740 |
| | Objective Value | 0.53427517 | 0.53427517 |
| | Iterations | 120 | 12 |
| 50 | CPU time(s) | 0.08959 | 0.02941 |
| | Objective Value | 0.72716729 | 0.72716729 |
| | Iterations | 127 | 11 |
| 75 | CPU time(s) | 0.14313 | 0.04110 |
| | Objective Value | 0.64766544 | 0.64766544 |
| | Iterations | 128 | 12 |
| 100 | CPU time(s) | 0.22231 | 0.05502 |
| | Objective Value | 0.53881219 | 0.53881225 |
| | Iterations | 117 | 14 |

Table 6.2: Numerical results for the Cobb Douglas utility problem.

## Acknowledgment

We want to thank Professor Stephen Wright for his important collaboration in computational implementation of primal-dual interior-point method and gradient projection with acceleration.

# Conclusions and future works

## Conclusions

In the first part of the thesis, it is demonstrated the global convergence of a special alternating direction method which uses a generalized proximal distance to penalize one of its subproblems, and also its iterations are calculated approximately. The proposed method is called RIPADM. Note that there are several works on convergence of proximal algorithms with nonquadratic penalty terms, but these generalizations do not appear to carry over in a straightforward manner from non-decomposition augmented Lagrangian methods to ADM-type methods. The conditions for the convergence of the RIPADM method are of two types, some assumptions are related to the separable convex optimization problem under study, and the other cases are related to generalized proximal distances. The assumptions for the problem under study are standar assumptions of ADM-type methods with nonquadratic proximal distances. On the other hand, the second order homogeneous proximal distances and double regularizations satisfy all the assumptions ($B_1$-$B_2$-$B_3$) associated with proximal distances that arise in the analysis of convergence of RIPADM method. Instead, Bregman distances do not necessarily satisfy the assumption $B_2$, for example the Kullback-Leibler distance (1.31) does not satisfy this condition. The assumption $B_2$ is key to prove that every limit point of the sequence generated by RIPADM method is a saddle point of the Lagrangian function associated with the problem under study, and then with the assumptions $B_1$ and $B_3$ the global convergence of the sequence is obtained. Therefore, when a proximal distance does not satisfy $B_2$, but if all the other assumptions, the Theorem 2.2 states that every limit point of the sequence generated via RIPADM is primal solution of the problem under study.

Furthermore, it is demonstrated the global convergence of the RIPADM method when a relaxation parameter is added to the update ruler for the Lagrange multiplier that in most literature is set to 1. In the computational implementation of the RIPADM method using the Log-quad distance, numerical results show that the relaxation factor equal to 1.62 tends to make the algorithm faster than when this factor is equal to 1.

In the numerical experiments of the first part of the thesis, two algorithms based on RIPADM method are proposed, one using the Log-quad distance and the other using the Kullback-Liebler Bregman distance. Both algorithms are applied to three problems that appear in statistics and machine learning. In the algorithm using Log-quad distance, the $x$-subproblem has a closed-form solution. Instead, the algorithm using Bregman distance gives an approximate solution for the $x$-subproblem. In both algorithms, the $z$-subproblem is solved using CVX solver. Almost all numerical results show that the algorithm using Log-quad distance have less CPU time and number of iterations than the algorithm using Bregman distance. There was only one instance where the algorithm using Bregman distance had less time and number of iterations, the twin support vector machine problem with WBC data set. Moreover, the ADM and PMM were implemented computationally. Note that the PMM is not a decomposition method but can be applied to convex optimization problems with or without separable structure. These methods were applied to the problems of statistics and machine learning seen in the thesis. The numerical results show that the RIPADM method had less CPU time (thirteen out of fourteen experiments) and number of iterations (nine out of fourteen experiments) in comparasion with the ADM. Moreover, the numerical results show that the RIPADM method had less CPU time (twelve out of fourteen experiments) and number of iterations (nine out of fourteen experiments) in comparasion with the PMM.

In the second part of the thesis, the numerical results of two algorithms to solve the agent's problem are shown. These methods are primal-dual interior-point method (PDIPM) and gradient projection with acceleration (AGPM), both methods use a modified newton step. Then, they can be applied to twice continuously differentiable functions as CES utility function but can not be applied to non-differentiable utility functions. The numerical results for the problem of agent with CES and Cobb-Douglas function show that the AGPM method has better results in time and number of iterations than the PDIPM method.

# Future works

We propose two decomposition methods based on the ADM that could be studied in the future, and are inspired by recent work on generalized alternating direction methods of multipliers (see [29], [40]). The first method uses a linearization strategy for both primal subproblems of ADM whose purpose is to obtain simplified subproblems, i.e., subproblems that are easier to solve or have closed-form solutions. The second method proposes the addition of a relaxation factor to accelerate convergence in a different way than proposed in Chapter 3 of this thesis.

## Proposal 1: Linearized RIPADM Method

Let us consider the following problem:

$$(P') \qquad \min_{x,z} \left\{ f(x) + g(z) \mid Ax + Bz = b,\ x \in \bar{C},\ z \in \bar{K} \right\},$$

where $C \subset \mathbb{R}^n$ and $K \subset \mathbb{R}^m$ are nonempty open convex sets, $\bar{C}$ and $\bar{K}$ denote the closure of $C$ and $K$, respectively, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions, and $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$, $b \in \mathbb{R}^p$.

In this section, we propose a method for solving the problem $(P')$. Given $(x^0, z^0, y^0)$, the proposed method consists of the following steps:

$$x^{k+1} \approx \underset{x\in\overline{C}}{\operatorname{argmin}} \ \ell_\lambda(x, z^k, y^k) + \frac{1}{2}(x-x^k)^T P(x-x^k) + \frac{1}{2\lambda} d_1(x, x^k), \tag{6.21}$$

$$z^{k+1} \approx \underset{z\in\mathbb{R}^m}{\operatorname{argmin}} \ \ell_\lambda(x^{k+1}, z, y^k) + \frac{1}{2}(z-z^k)^T Q(z-z^k) + \frac{1}{2\lambda} d_2(z, z^k), \tag{6.22}$$

$$y^{k+1} = y^k + \rho\lambda(Ax^{k+1} + Bz^{k+1} - b), \tag{6.23}$$

where $P \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{m\times m}$ are real matrix, and $\rho > 0$ is a relaxation factor. We call *Linearized RIPADM* to the proposed method given by (6.21)- (6.23). Compared to RIPADM method, linearized RIPADM adds $\frac{1}{2}(x-x^k)^T P(x-x^k)$ to the $x$-subproblem, $(z-z^k)^T Q(z-z^k)$ to the $z$-subproblem and $\rho$ to the update ruler for Lagrange multiplier. The term $(z-z^k)^T Q(z-z^k)$ makes the z-subproblem much easier to solve or have closed-form solution, and the relaxation factor $\rho$ accelerates the convergence of the method. The same is true for $x$.

Like a motivation, we analize the constrained Lasso problem. We recall this problem:

$$\min_z \ \left\{ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 \mid x + Bz = b, \ \ x \geq 0 \right\},$$

where $D \in \mathbb{R}^{r\times m}$, $d \in \mathbb{R}^r$, $B \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, and $\gamma > 0$ are problem data, and $m = n$.

The RIPADM method applied to constrained Lasso consists of the following steps:

$$x^{k+1} \approx \underset{x}{\operatorname{argmin}} \ \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k),$$

$$z^{k+1} \approx \underset{z}{\operatorname{argmin}} \ \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2,$$

$$y^{k+1} = y^k + \lambda(x^{k+1} + Bz^{k+1} - b).$$

We remember that when the proximal distance $d(\cdot, x^k)$ is the Log-quad distance, we obtain a closed-form solution for the x-subproblem given by (4.11). However, in the z-subproblem, we can not to obtain a closed-form solution. So to overcome this difficulty, we can linearize the smooth terms of the augmented Lagrangean $\ell_\lambda(x^{k+1}, z, y^k)$ resulting:

$$z^{k+1} \approx \underset{z}{\operatorname{argmin}} \;\; \gamma\|z\|_1 + \langle v^k, z - z^k\rangle + \frac{1}{2\lambda}\|z - z^k\|^2, \tag{6.24}$$

where

$$v^k := D^T(Dz^k - d) + B^T y^k + \lambda B^T(x^{k+1} + Bz^k - b), \tag{6.25}$$

and this vector is the gradient of the terms $\frac{1}{2}\|Dz - d\|^2 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2$ at $z = z^k$.

The subproblem (6.24) is equivalent to (subject to a constant difference in the objective)

$$z^{k+1} \approx \underset{z}{\operatorname{argmin}} \;\; \gamma\|z\|_1 + \frac{1}{2\lambda}\|z - (z^k - \lambda v^k)\|^2, \tag{6.26}$$

whose closed-form solution is given by:

$$z^{k+1} = \operatorname{shrinkage}(z^k - \lambda v^k, \lambda). \tag{6.27}$$

## Proposal 2: Generalized RIPADM Method

Let us consider the following problem:

$$(P') \qquad \min_{x,z} \left\{ f(x) + g(z) \mid Ax + Bz = b,\ x \in \bar{C},\ z \in \bar{K} \right\},$$

where $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$, $b \in \mathbb{R}^p$, $C \subset \mathbb{R}^n$ and $K \subset \mathbb{R}^m$ are nonempty open convex sets, $\bar{C}$ and $\bar{K}$ denote the closure of $C$ and $K$, respectively, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions.

In [45] the ADM was explained as an application of the well-known Douglas-Rachford splitting method (DRSM); and in [35] , the DRSM was further explained as an application of the proximal point algorithm (PPA). Therefore, it was suggested in [35] to apply the acceleration scheme for the PPA to accelerate the original ADM, and this idea gave rise to a variation on the ADM called *generalized alternating direction method of multipliers* (GADM). The GADM applied to the problem $(P')$ consists of the following steps:

$$x^{k+1} \approx \operatorname*{Argmin}_{x \in \bar{C}} \; f(x) + \langle y^k, Ax \rangle + \frac{\lambda}{2} \|Ax + Bz^k - b\|^2, \tag{6.28}$$

$$z^{k+1} \approx \operatorname*{Argmin}_{z \in \bar{K}} \; g(z) + \langle y^k, Bz \rangle + \frac{\lambda}{2} \|\beta Ax^{k+1} + (1-\beta)(b - Bz^k) + Bz - b\|^2, \tag{6.29}$$

$$y^{k+1} = y^k + \lambda(\beta Ax^{k+1} + (1-\beta)(b - Bz^k) + Bz^{k+1} - b), \tag{6.30}$$

where the parameter $\beta \in (0, 2)$ is a relaxation factor; when $\beta > 1$, this technique is called over-relaxation, and when $\beta < 1$, it is called under-relaxation. The difference between ADM and GADM is that the quantity $Ax^{k+1}$ is replaced with $\beta Ax^{k+1} + (1-\beta)(b - Bz^k)$ in the z-problem (2.8) and the update ruler (2.9) for $y$ of the ADM. We can see that the generalized ADM reduces to the original ADM when $\beta = 1$. Preserving the main advantage of the original ADM in treating the objective functions $f$ and $g$ individually, the GADM enjoys the same easiness in implementation while experiments in [33, 36] suggest that over-relaxation with $\beta \in [1.5, 1.8]$ can numerically accelerate the ADM.

Recently, Min Li et. al. [64] proposed a exact method which combines the GADM with the logarithmic-quadratic proximal regularization for solving a variational inequality with separable structure. They proved the global convergence and established the worst case convergence rate measured by the iteration complexity in both the ergodic and nonergodic senses for the resulting algorithm.

We propose a method for solving $(P')$ which is based on GADM but we regularize the $x$-problem (6.28) and $z$-problem (6.29) with generalized proximal distances, that is, the proposed method consists of the following steps:

$$x^{k+1} \approx \operatorname*{Argmin}_{x \in X} \; f(x) + \langle y^k, Ax \rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2 + d_1(x, x^k), \tag{6.31}$$

$$z^{k+1} \approx \operatorname*{Argmin}_{z \in Z} \; g(z) + \langle y^k, Bz \rangle + \frac{\lambda}{2}\|\beta Ax^{k+1} + (1-\beta)(b - Bz^k) + Bz - b\|^2 + d_2(z, z^k), \tag{6.32}$$

$$y^{k+1} = y^k + \lambda(\beta Ax^{k+1} + (1-\beta)(b - Bz^k) + Bz^{k+1} - b), \tag{6.33}$$

where $d_1(\cdot, x^k)$ and $d_2(\cdot, z^k)$ are proximal distances with respect to $\bar{C}$ and $\bar{K}$, respectively. This approach covers the method given in [64] because the GADM subproblems are regularized by generalized proximal distances.

# Bibliography


[1] K.J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pages 265–290, 1954.

[2] A. Asuncion and D. Newman. Uci machine learning repository, 2007.

[3] H. Attouch, A. Cabot, P. Frankel, and J. Peypouquet. Alternating proximal algorithms for linearly constrained variational inequalities: Application to domain decomposition for PDE's . *Nonlinear Analysis: Theory, Methods and Applications* , 74(18):7455 – 7473, 2011.

[4] H. Attouch and M. Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces . *Pacific Journal of Optimization* , 5:17 – 37, 2009.

[5] J-P Aubin and I. Ekeland. *Applied nonlinear analysis*. Courier Dover Publications, 2006.

[6] A. Auslender and M. Teboulle. Entropic proximal decomposition methods for convex programs and variational inequalities. *Mathematical Programming*, 91(1):33–47, 2001.

[7] A. Auslender and M. Teboulle. Interior gradient and epsilon-subgradient descent methods for constrained convex minimization. *Mathematics of Operations Research*, 29(1):1–26, 2004.

[8] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.

[9] A. Auslender, M. Teboulle, and S. Ben-Tiba. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research*, 24(3):645–668, 1999.

[10] A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications*, 12:31–40, 1999.

[11] H. Bauschke, J. Borwein, and P. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[12] D. Bertsekas. *Nonlinear programming*. 1999.

[13] D. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[14] A. Bnouhachem and Q. Hasan. A descent lqp alternating direction method for solving variational inequality problems with separable structure. *Applied Mathematics and Computation*, 246(0):519 – 532, 2014.

[15] R. Bot and C. Hendrich. A douglas–rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. *SIAM Journal on Optimization*, 23(4):2541–2565, 2013.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[17] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[18] L. Briceño-Arias and P. Combettes. A monotone+ skew splitting model for composite monotone inclusions in duality. *SIAM Journal on Optimization*, 21(4):1230–1250, 2011.

[19] R. Burachik and A. Iusem. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM Journal on Optimization*, 8(1):197–216, 1998.

[20] Y. Censor and S. A. Zenios. Proximal minimization algorithm with d-functions. *Journal of Optimization Theory and Applications*, 73(3):451–464, 1992.

[21] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[22] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[23] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64(1-3):81–101, 1994.

[24] P. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004.

[25] P. Combettes and J.-C. Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):564–574, 2007.

[26] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[27] Y-H Dai and R. Fletcher. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Mathematical Programming*, 106(3):403–421, 2006.

[28] G. Debreu. *Theory of value: An axiomatic analysis of economic equilibrium*. Number 17. Yale University Press, 1987.

[29] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66(3):889–916, 2016.

[30] M. Doljansky and M. Teboulle. An interior proximal algorithm and the exponential multiplier method for semidefinite programming. *SIAM Journal on Optimization*, 9(1):1–13, 1998.

[31] J. Douglas and H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, pages 421–439, 1956.

[32] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.

[33] J. Eckstein. Parallel alternating direction multiplier decomposition of convex programs. *Journal of Optimization Theory and Applications*, 80(1):39–62, 1994.

[34] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optimization methods and software*, 4:75–83, 1994.

[35] J. Eckstein and D.P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.

[36] J. Eckstein and M. C. Ferris. Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS Journal on Computing*, 10(2):218–235, 1998.

[37] J. Eckstein and P. Silva. Proximal methods for nonlinear programming: double regularization and inexact subproblems. *Computational Optimization and Applications*, 46(2):279–304, 2010.

[38] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[39] J. Fan, J. Zhang, and K. Yu. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606, 2012.

[40] E.X. Fang, B. He, H. Liu, and X. Yuan. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Mathematical Programming Computation*, 7(2):149–187, 2015.

[41] M. Ferris, O. Mangasarian, and S. Wright. *Linear programming with MATLAB*, volume 7. SIAM, 2007.

[42] R. Fletcher. *Optimization*. Academic Press, 1969.

[43] M. Fortin and R. Glowinski. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Studies in Mathematics and its Applications. North-Holland, 1983.

[44] P. Frankel and J. Peypouquet. Lagrangian-penalization algorithm for constrained optimization and variational inequalities. *Set-Valued and Variational Analysis*, 20(2):169–185, 2012.

[45] D. Gabay. Applications of the method of multipliers to variational inequalities. *Augmented Lagrangian Methods: Applications to the Solution of Boundary-valued Problems*, pages 299–331, 1983.

[46] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40, 1976.

[47] R. Glowinski. *Numerical Methods for Non-Linear Variational Problems*. Scientific Computation. Springer, 2008.

[48] R. Glowinski. On alternating direction methods of multipliers: a historical perspective. In *Modeling, Simulation and Optimization for Science and Technology*, pages 59–82. Springer, 2014.

[49] R. Glowinski and A. Marroco. Sue l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de Dirichlet non lineares. *Revue Francaise d'Automatique, Informatique, et Recherche Opérationelle*, 9:41–76, 1975.

[50] R. Glowinski and P.L. Tallec. *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1989.

[51] B.S. He, L.Z. Liao, D.R. Han, and H. Yang. A new inexact alternating directions method for monotone variational inequalities. *Mathematical Programming*, 92(1):103–118, 2002.

[52] B.S. He, S.L. Wang, and H. Yang. A modified variable-penalty alternating directions method for monotone variational inequalities. *Journal of Computational Mathematics*, 21(4):495 – 504, 2003.

[53] M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

[54] A.N. Iusem. Some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications*, 85(3):593–612, 1995.

[55] G.M. James, C. Paulson, and P. Rusmevichientong. The constrained lasso. Technical report, Citeseer, 2012.

[56] A. Jofré, J. Deride, and R. J-B Wets. Solving deterministic ans stochastic equilibrium problems via augmented walrasian. *Computational Economics (submitted)*, 2016.

[57] A. Jofré and R. J-B Wets. Continuity properties of walras equilibrium points. *Annals of Operations Research*, 114(1-4):229–243, 2002.

[58] A. Jofré and R. J-B Wets. Variational convergence of bifunctions: motivating applications. *SIAM Journal on Optimization*, 24(4):1952–1979, 2014.

[59] R. Khemchandani, S. Chandra, et al. Twin support vector machines for pattern classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(5):905–910, 2007.

[60] K. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization*, 35(4):1142–1168, 1997.

[61] S. Kontogiorgis and R. Meyer. A variable-penalty alternating directions method for convex optimization. *Mathematical Programming*, 83(1-3):29–53, 1998.

[62] M. Kyono and M. Fukushima. Nonlinear proximal decomposition method for convex programming. *Journal of optimization theory and applications*, 106(2):357–372, 2000.

[63] M. Li. A hybrid lqp-based method for structured variational inequalities. *International Journal of Computer Mathematics*, 89(10):1412–1425, 2012.

[64] M. Li, X. Li, and X. Yuan. Convergence analysis of the generalized alternating direction method of multipliers with logarithmic–quadratic proximal regularization. *Journal of Optimization Theory and Applications*, 164(1):218–233, 2015.

[65] M. Li, L.Z. Liao, and X. Yuan. Inexact alternating direction methods of multipliers with logarithmic–quadratic proximal regularization. *Journal of Optimization Theory and Applications*, 159(2):412–436, 2013.

[66] L. W. McKenzie. On the existence of general equilibrium for a competitive market. *Econometrica: journal of the Econometric Society*, pages 54–71, 1959.

[67] B.T. Polyak. *Introduction to optimization*. Translations series in mathematics and engineering. Optimization Software, Publications Division, 1987.

[68] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97 – 116, 1976.

[69] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[70] R.T. Rockafellar. *Convex Analysis*. Convex Analysis. Princeton University Press, 1997.

[71] P.J.S. Silva and J. Eckstein. Double-regularization proximal methods, with complementarity applications. *Computational Optimization and Applications*, 33(2-3):115–156, 2006.

[72] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of bregman functions. *Mathematics of Operations Research*, 25(2):214–230, 2000.

[73] G. Stephanopoulos and A.W. Westerberg. The use of Hestenes' method of multipliers to resolve dual gaps in engineering system optimization. *Journal of Optimization Theory and Applications*, 15(3):285–309, 1975.

[74] B. Svaiter. On weak convergence of the douglas-rachford method. *SIAM Journal on Control and Optimization*, 49(1):280–287, 2011.

[75] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[76] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1):119–138, 1991.

[77] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

[78] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[79] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.

[80] L. Walras. *Éléments d'économie politique pure; ou, Théorie de la richesse sociale*. F. Rouge, 1896.

[81] X. Wang and X. Yuan. The linearized alternating direction method of multipliers for dantzig selector. *SIAM Journal on Scientific Computing*, 34(5):A2792–A2811, 2012.

[82] M.H. Xu. Proximal alternating directions method for structured variational inequalities. *Journal of Optimization Theory and Applications*, 134(1):107–117, 2007.

[83] X. Yuan and M. Li. An lqp-based decomposition method for solving a class of variational inequalities. *SIAM Journal on Optimization*, 21(4):1309–1318, 2011.

[84] L.A. Zadeh, L.W. Neustadt, A.V. Balakrishnan, Society for Industrial, Applied Mathematics, Berkeley University of California, and University of Southern California. *Computing methods in optimization problems: 2; papers*. Number v. 1 in Lecture notes in operations research and mathematical economics. Academic Press, 1969.

[85] X. Zhang, Martin B., and S. Osher. A unified primal-dual algorithm framework based on bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011.

[86] H. Zou and M. Yuan. The fo-norm support vector machine. *Statistica Sinica*, 18:379–398, 2008.