# APPROXIMATE VERSIONS OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS 

By<br>YAO, WANG

A dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Graduate Program in Operations Research
written under the direction of Jonathan Eckstein
and approved by
$\qquad$
$\qquad$
$\qquad$
$\qquad$
$\qquad$

New Brunswick, New Jersey
October, 2016

# ABSTRACT OF THE DISSERTATION 

# Approximate versions of the alternating direction method of multipliers 

By YAO, WANG

## Dissertation Director:

Jonathan Eckstein

Convex optimization is at the core of many of today's analysis tools for large datasets, and in particular machine learning methods. This thesis will develop approximate versions of the alternating directrion of multipliers (ADMM) for the general setting of minimizing the sum of two convex functions.

The alternating direction method of multipliers is a form of augmented Lagrangian algorithm that has experienced a renaissance in recent years due to its applicability to optimization problems arising from "big data" and image processing applications, and the relative ease with which it may be implemented in parallel and distributed computational environments.

There are two fundamental approaches for proving the convergence of the ADMM, each based on a different form of two-way splitting, that is, expressing a mapping as the sum of two simpler mappings. The first approach is based on Douglas-Rachford operator splitting theory, and yields considerable insight into the convergence of the

ADMM. The second convergence proof approach is at its core based on the Lagrangian splitting analysis. We present three new approximate versions of ADMM based on both convergence analyses, all of which require only knowledge of subgradients of the subproblem objectives, rather bounds on the distance to the exact subproblem solution. One version, which applies only to certain common special cases, is based on combining the operator splitting analysis of the ADMM with a relative-error proximal point algorithm of Solodov and Svaiter. A byproduct of this analysis is a new, relativeerror version of the Douglas-Rachford splitting algorithm for monotone operators. The other two approximate versions of the ADMM are more general and based on the Lagrangian splitting analysis of the ADMM: one uses a summable absolute error criterion, and the other uses a relative error criterion and an auxiliary iterate sequence.

We experimentally compare our new algorithms to an essentially exact form of the ADMM and to an inexact form that can be easily derived from prior theory (but again applies only to certain common special cases). These experiments show that our methods can significantly reduce total computational effort when iterative methods are used to solve ADMM subproblems.

## Acknowledgments

First and foremost I would like to express the deepest appreciation to my academic father, Professor Jonathan Eckstein for his constant support and encouragement in the past six years. I am truly grateful for his contributions of time, ideas and funding to finish my Ph.D. Professor Eckstein played the most fundamental role in my study and without his patience and openness, this thesis would never have been possible. He helped me come up with the thesis topic and guided me through the most difficult times in my research. He gave me freedom to pursue various projects, summer internships or personal affairs without objection. He always makes himself available to me no matter how busy he is and provides insightful discussions about my research as well as other related results or ideas that could be helpful. His server is my primary resource for completing my computing tasks, he also makes great effort on proofreading and editing my papers. In addition to immense knowledge on my research topic, he is also an expert in many other areas include linguistics, music and food. It is an pity that I no longer can find the piece of note paper from our very first meeting, on which he drew an illustrative figure and said something like "...(gradient iteration, the special case of proximal point method) it seems impossible to solve, right? ...", and that was exactly the moment when this journey started.

I gratefully acknowledge the funding sources that made my Ph.D. work. I was funded by university fellowships for my first two years and by National Science Foundation grant CCF-1115638 for three years, once again thanks to my advisor Jonathan Eckstein. I also have to thank the members of my Ph.D. committee, Professors Andrzej Ruszczyński, Adi Ben-Israel, Farid Alizadeh and Kristin Dana for their helpful advice and suggestions about my research.

I am very thankful to Professors András Prékopa and Endre Boros. Professor Prékopa first introduce me to the world of operations research, and I still think fondly of my time as an student in his class. His enthusiasm and love for teaching is very impressive. Professor Boros was very helpful in providing advice on my graduate school career many times during my first two years at RUTCOR. I also thank RUTCOR staff Lynn Agre, Clare Smietana, Terry Hart and Katie D'Agosta for their kind support. They have all been so friendly and personable, making me feel like a family member of RUTCOR. Dozens of people have helped and taught me immensely at Rutgers, I would like to express my appreciation to them. Special acknowledgments go to my college mathematics instructors: Professors Richard Falk, Abbas Bahri, Michael Beals, Butler Terence, Lisa Carbone, Michael Saks, Eugene Speer, Michael Vogelius, Richard Wheeden, Jeffry Kahn, Daniel Ocone, Van Vu, Ted Petrie, Wolmer Vasconcelos, Richard Gundy, Regina Liu, and William Edward Strawderman. They gave me a good foundation in both theoretical and applied mathematics, which are essential for building my research background.

My time at RUTCOR was made enjoyable in large part due to many friends. I am grateful for time spent in the old RUTCOR building with Emre Yamangil, Jinwook Lee, Tsvetan Asamov, Aritanan Gruber, Mariya Naumova, Anh Ninh, Matthew Oster, Kunikazu Yoda, Minh Pham, Svetlana Soloveva, Yu Du, Mohammad Ranjbar, Ai Kagawa and Gyorgy Matyasfalvi (or just George). We raised about two hundreds dollar to buy a Pingpong table and played many games in the lounge. I will forever remember our times of study and discussion. It was such a pleasurable time when we had lunch together at Busch campus center or Noodle Gourmet, and when we went to Quick Check for coffee and snacks late at night after a long study sessions.

I also thank my friends for providing support and friendship that I needed. I especially thank Tong Jin, Yun Jiang, Ou Liu, Bin Guo, Ping Lu, Lu Wang,

Qibing Zhou, Jian Li, Ziqing Duan, Jinjun Zhuge for being supportive throughout my time at Rutgers and for discussing various interesting subjects that are related to their own research. I own a great debt of gratitude to George Brown and his mother Francis Brown. George helped me to settle in when I first came to U.S. He educated me on almost everything about his country and took me to many places. Francis was a nurse of U.S. Army during World War II. I learned form her why they are said to come from "the greatest generation".

I would like to thank my mother Lina Yao for raising me with lots of love and hard work. My mother sacrificed so much of her life for me. She taught me the most important things in my life and has been an outstanding inspiration to me. As a single mother, she experienced tremendous difficulties in the past thirty years. She came to U.S. with extraordinary courage, and guided me to where I am today. Without her motivation and strength I would have never had the courage to overcome the adversities I have faced. She is the main reason for many things I have done, to make her proud.

I am deeply thankful to my dear wife Yiming Luo for her love, encouragement and sacrifice. She worked very hard during my Ph.D. years, and gave birth to my son and daughter. In addition, Yiming has been a true and great supporter and has unconditionally loved me during my good and bad times. She had faith in me and my intellect when I struggled. These past several years have not been an easy ride, both academically and personally. What we both have learned about life, strengthens our commitment to each other. I truly thank Yiming for standing by my side, even when I was irritable and depressed. Yiming, my son and daughter have cherished with me every great moment, they are the most basic sources of my life energy. Special thanks go to my parents-in-law Xiaomei Luo and Benhong Luo: their support has been unconditional all these years.

I finish this acknowledgments with my maternal grandmother Shumin Zhang, whose role in my life was and remains immense. There are no words to convey how
much I love her. She and my maternal grandfather Chengwen Yao never abandoned me. Only when being with them, could I enjoy the innermost ultimate peace.

Thank you all.

Yao, Wang
Rutgers University
April, 2016

## Dedication

I dedicate this thesis to my family, my wife, Yiming, my mother, Lina, my beloved Vincent and Kaileen, and loving memory of my maternal grandmother, Shumin Zhang (1930-2007) for their constant support and unconditional love.

I love you all dearly.

## Table of Contents

Abstract ..... ii
Acknowledgments ..... iv
Dedication ..... viii

1. Introduction ..... 1
1.1. Our main contributions ..... 3
1.2. Thesis outline ..... 6
2. Literature Review ..... 7
2.1. Proximal alternating direction methods ..... 8
2.2. Logarithmic-quadratic proximal ADMM ..... 11
3. Formalizing Approximate Subproblem Solution ..... 14
4. Approximate ADMM Algorithms Derived through Operator Split-
ting Analysis ..... 17
4.1. A subgradient-based application of [29, Theorem 8] ..... 18
4.2. Background: a relative-error proximal point algorithm ..... 20
4.3. A relative-error variant of Douglas-Rachford splitting ..... 23
4.4. Deriving a partially inexact ADMM from the partially inexact DRsplitting method30
5. Approximate ADMM Algorithms Derived from Lagrangian Splitting ..... 37
5.1. A parametric conjugate duality framework ..... 37
5.2. Analyzing the exact ADMM by Lagrangian splitting ..... 43
5.3. Common elements of the Lagrangian splitting analyses ..... 53
5.4. Approximate ADMM with absolute summable error criteria ..... 62
5.5. An approximate ADMM with relative error criteria ..... 72
5.6. Partially inexact ADMM with relative error criteria ..... 79
5.7. Complete form of relative-error algorithm with both minimizations in-exact81
6. Numerical Tests ..... 88
6.1. Comparison algorithms ..... 88
6.2. Termination criteria and algorithm parameters ..... 90
6.2.1. LASSO regression ..... 91
6.2.2. $\quad L_{1}$-regularized logistic regression ..... 96
6.2.3. Sparse inverse covariance selection ..... 98
7. Conclusion ..... 106
References ..... 108

## Chapter 1

## Introduction

Many machine learning models and signal processing problems are finding parameters through minimizing the objectieriantve function. Convex optimization plays the critical role in developing statistical computing algorithms. Let $x \in \mathbb{R}^{n}$ denote the parameter, $A \in \mathbb{R}^{m \times n}$ be the data matrix and $b \in \mathbb{R}^{m}$ represents the response vector, a general convex model fitting problem very often can be written in the form of minimizing the sum of two terms:

$$
\begin{equation*}
\min l(x ; A, b)+\lambda r(x) \tag{1.1}
\end{equation*}
$$

where $l: \mathbb{R}^{m} \rightarrow \mathbb{R}$ is a convex loss function, such as $l_{1}, l_{2}$ hinge or logistic loss, while $r$ is the convex regularization function which enforces a certain property in $x$ and $\lambda>0$ is a positive regularization parameter. A common example of $l$ is the squared Euclidean distance and if $r(x)=\|x\|_{1}$ or the lasso penalty, which usually is used to prompt the sparsity in parameter $x$, then immediately we obtain the LASSO 79 (least absolute shrinkage and selection operator)

$$
\begin{equation*}
\min _{x \in \mathbb{R}^{n}}\|A x-b\|^{2}+\lambda\|x\|_{1} . \tag{1.2}
\end{equation*}
$$

Replacing the lasso penalty with Tikhonov regularization or ridge penalty, that is $r(x)=\|x\|_{2}^{2}$, the consequence is called ridge regression 2, 24, 77

$$
\begin{equation*}
\min _{x \in \mathbb{R}^{n}}\|A x-b\|^{2}+\lambda\|x\|_{2}^{2} \tag{1.3}
\end{equation*}
$$

Other common examples for $r$ include nuclear norm [84, 89, mixed-norm (91 and total variation [16, 63, 82].

In many modern applications, the datasets are usually in high dimensions and contain huge number of training examples and it is challenging to directly solve the optimization problem like 1.1. Another characteristic these large scale applications share is that the data is often stored in a distributed manner, thus it is very important to develop algorithms that capable of handling the complexity of data and suitable to process huge datasets in parallelized or distributed manner. The alternating direction method of multipliers (ADMM) [34, 37] is a simple and yet powerful decentralized algorithms that is well suited for distributed convex optimization problems [15].

To illustrate how ADMM works, consider a general convex optimization problem

$$
\begin{equation*}
\min _{x \in \mathbb{R}^{n}} f(x)+g(M x) \tag{1.4}
\end{equation*}
$$

where $f: \mathbb{R}^{n} \rightarrow(-\infty,+\infty]$ and $g: \mathbb{R}^{m} \rightarrow(-\infty,+\infty]$ are closed proper convex functions and $M$ is a $m \times n$ matrix. An equivalent formulation is

$$
\begin{array}{ll}
\min & f(x)+g(z) \\
\text { s. t. } & M x=z  \tag{P}\\
& x \in \mathbb{R}^{n}, z \in \mathbb{R}^{m}
\end{array}
$$

Let $p \in \mathbb{R}^{m}$ denote the Lagrangian multiplier attached to the constraint $M x=z$, the augmented Lagrangian of $P$ is

$$
\begin{equation*}
L_{c}(x, z, p)=f(x)+g(z)+\langle p, M x-z\rangle+\frac{c}{2}\|M x-z\|^{2} \tag{1.5}
\end{equation*}
$$

where $c>0$ is a positive scalar. The ADMM for solving $(\mathbb{P})$ consists of the recursions

$$
\begin{align*}
& x^{k+1} \in \underset{x \in \mathbb{R}^{n}}{\operatorname{Arg} \min }\left\{f(x)+g\left(z^{k}\right)+\left\langle p^{k}, M x-z^{k}\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}  \tag{1.6}\\
& z^{k+1} \in \underset{z \in \mathbb{R}^{m}}{\operatorname{Arg} \min }\left\{f\left(x^{k+1}\right)+g(z)+\left\langle p^{k}, M x^{k+1}-z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}  \tag{1.7}\\
& p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right) \tag{1.8}
\end{align*}
$$

In words, ADMM works as follows. It takes the form of block decomposition procedure, so that the solutions of smaller subproblems (1.6) and (1.7) are used to find
global solution of the larger problem. At iteration $k$, for fixed multiplier $p^{k}$ and $z^{k}$, the new point $x^{k+1}$ is obtained as the exact minimizer of the augmented Lagrangian with respect to $x$. Then in a similar fashion, the $x$ component is fixed at $x^{k+1}$, and the augmented Lagrangian is minimized with respect to $z$ to obtain $z^{k+1}$. Finally the Lagrangian multipliers $p^{k+1}$ is updated in a simple manner. The process is repeated until the overall convergence. A key feature of ADMM is that the variables $x$ and $z$ are updated in a Gauss-Seidel style, i.e., $x$ is updated while $z$ is fixed then the new value of $x$ is used to find values for $z$.

Removing constant terms from (1.6) to (1.8), we get the following equivalent recursion

$$
\begin{align*}
& x^{k+1} \in \underset{x}{\operatorname{Arg} \min }\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}  \tag{1.9}\\
& z^{k+1} \in \underset{z}{\operatorname{Arg} \min }\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}  \tag{1.10}\\
& p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right) . \tag{1.11}
\end{align*}
$$

Throughout this paper, we refer ADMM to the recursion (1.9) to (1.11). To avoid pathological special cases, we make the following standing assumption:

Assumption 1. Problem ( P ) possesses a KKT point, that is, there exists at least one $\left(x^{*}, z^{*}, p^{*}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ such that

$$
-M^{\top} p^{*} \in \partial f\left(x^{*}\right) \quad p^{*} \in \partial g\left(z^{*}\right) \quad M x^{*}=z^{*}
$$

It is easily seen that for any $\operatorname{KKT}$ point $\left(x^{*}, z^{*}, p^{*}\right)$, then $x^{*}$ is a solution of $(\mathrm{P})$.

### 1.1 Our main contributions

This paper concerns situations in which at least one of the subproblems (1.9) or (1.10) is solved by some kind of iterative method, either because no simple solution formula exists, or the problem is too high-dimensional for such a formula to be used practically (for example, if a solution formula exists, but it involves factoring a very large matrix).

In such cases, it seems wasteful to expend great effort solving (1.9) to high precision early in the solution process, when the values of $p^{k}$ and $z^{k}$ may be far from their final values; a similar observation applies to (1.10). It is therefore natural to ask whether it is possible, without disrupting the convergence properties of the algorithm, to solve the subproblems approximately and gradually increase their precision. It has been known since [29] that the answer to this question is positive. However, the subproblem approximation criteria in that result, which we will here refer to as the $E B$ criteria, involve the distances between the approximate solutions and the respective true solutions to each subproblem. In the case that $M$ has full column rank (that is, $\operatorname{ker} M=\{0\}$ ), the EB criteria can be used to derive approximation criteria based on more readily testable quantities, namely subgradients of the objective functions of the subproblems at the current trial solutions; see Section 4.1 below. In general, however, upper bounds on the distances to the optimal subproblem solutions may not be available. In this paper, we develop approximate versions of the ADMM whose approximation criteria use only subgradients, but apply in the general case.

Another potential drawback of the EB criteria is that they are based on allowable error sequences that appear in the analysis as external parameters. The associated convergence theory does not provide any direct guidance as to how to select these infinite sequences of parameters, other than requiring that they be nonnegative and summable. While some approximate proximal algorithms use such exogenous error sequences, others instead use "relative" error criteria which have only a single parameter that controls the subproblem error proportionally to other quantities occurring naturally in the algorithms. For abstract proximal methods, this idea began with 75 and was followed by [76] and a variety of generalizations. For the classical (non-alternating-direction) method of multipliers, a similar error criterion was developed in [30]. In this paper, two of the error criteria we develop are relative, using ideas from [76] and 30, respectively. The third new method we develop uses absolute
summable error criteria with formally exogenous parameter sequences, as in the original EB criteria. However, our new absolute approximation criteria do not require $M$ to have full column rank, or need any other form of strong convexity for the $x$ minimization, and are therefore easier to verify in general. Their analysis uses techniques inspired by [28].

Convergence of the ADMM has traditionally been proven in two related but different ways. One approach, dating back to [34], uses the monotonicity of the (convexconcave) subgradient of the Lagrangian function of $(\mathrm{P})$, splitting the Lagrangian into the sum of two convex-concave functions; we refer to this approach as Lagrangian splitting. The other approach, dating back to [37], expresses the subgradient of the dual function of $(\overline{\mathrm{P}})$ as the sum of two monotone operators and shows that the ADMM is equivalent to applying a Douglas-Rachford operator splitting method [57] to this pair operators. The derivation of our first new approximate ADMM algorithm is based on this operator-splitting analysis: we start by reformulating Douglas-Rachford operator splitting as an application of the proximal point algorithm [71] (PPA) as shown possible in [29]; see also [51. We then apply the relative-error proximal algorithm of 76 to this reformulation to obtain a new relative-error version of DR splitting, which in turn leads to a relative-error variant of the ADMM. This analysis, covered in Chapter 4, approaches the problem of creating an approximate ADMM by assembling existing theoretical building blocks in a novel way.

In order to be practical, the algorithm we derive through operator splitting requires that it be possible to solve the second ADMM subproblem 1.10 quickly and exactly, and therefore it is applicable when only the $x$ minimization (1.9) requires an iterative solution method. This situation is extremely common, however. A second potential drawback of our operator-splitting-derived algorithm is that it essentially requires that the matrix $M$ be the identity, which is also common. The two other new methods we derive in this paper, however, have neither of these potential drawbacks: there is no requirement that $M$ be the identity, and both subproblems may
be solved approximately if need be. Rather than assembling "modules" from prior results as in our operator splitting analysis, we derive these methods by modifying the original Lagrangian-splitting convergence proof of 34] to incorporate ideas used to develop approximate versions of the classical (non-alternating-direction) method of multipliers. For one method, we augment the Lagrangian-splitting ADMM convergence proof using techniques developed in [28]: this approach results in a method with absolute summable error criteria, but based on subproblem subgradients rather than the distance to the exact solution. For our second method, we instead use techniques developed in [30], resulting in a method with a relative error criterion which incorporates an (easily maintained) auxiliary sequence not present in the exact version of the algorithm.

### 1.2 Thesis outline

The thesis is organized as follows. We begin in Chapter 2 with a brief review of existed inexact traditional (non-alternating) Lagrangian methods and inexact alternating directions algorithms as well as related error criteria. This chapter is intended mainly for background and can be skimmed. Chapter 3 is dedicated to mathematically model the approximate solution processes for the ADMM subproblems. Chapter 4 is mainly used to derive approximate ADMM through exploiting the relationship between ADMM, Douglas-Rachford splitting, and the proximal point algorithm. In Section 4.1, we first present an inexact ADMM that is a direct application of [29, Theorem 8], then a partially inexact ADMM will be given in 4.4. In Chapter 5 we develop two new approximate ADMM algorithms by modifying the Lagrangian splitting analysis. These methods are given in Sections 5.4 and 5.5, respectively, with background material in Section 5.1. Chapter 6 presents some numerical experiments on representative $l_{1}$ minimization problems, establishing the potential utility of our new algorithms.

## Chapter 2 <br> Literature Review

ADMM has been extensively studied in recent years due to its ease of applicability and empirical performance. ADMM was first proposed in 1970's by Glowinski and Marroco [38, p. 69] and Gabay and Mercier [36]. It stemmed from the augmented Lagrangian method (also known as the method of multipliers) 47, 67 dating back to late 1960's, and its global convergence was soon established in the literature 29, 34, 37; also see a more recent tutorial paper [31] for a relatively accessible version of the convergence proof. As reviewed in the comprehensive paper [15], ADMM can be a natural fit in the field of large scale distributed optimization. ADMM has been shown to have an $O(1 / k)$ rate of convergence for convex problems 44, 45, 60, 80], where $k$ stands for the number of iterations. A rate of $O\left(1 / k^{2}\right)$ has been shown for a modified version [39]. When the objective functions are strongly convex and Lipschitz, ADMM has been show to have a linear convergence rate [21, 48].

The main topic of this dissertation is approximate versions of ADMM. The first approximate ADMM dates back to [29], in which Eckstein and Bertsekas developed an approximate version of ADMM with absolute summable error criteria. Specifically, let $\left\{\epsilon_{k}\right\}_{k=0}^{\infty} \subset[0, \infty)$ and $\left\{\tau_{k}\right\}_{k=0}^{\infty} \subset[0, \infty)$ be two infinite sequences of error tolerances such that $\sum_{k=0}^{\infty} \epsilon_{k}<\infty$ and $\sum_{k=0}^{\infty} \tau_{k}<\infty$. Then [29] shows that one may replace the exact minimizations 1.6 and 1.7 with the approximation criteria

$$
\begin{align*}
& \left\|x^{k+1}-\arg \min \left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\right\| \leq \epsilon_{k+1}  \tag{2.1}\\
& \left\|z^{k+1}-\arg \min \left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\right\| \leq \tau_{k+1} \tag{2.2}
\end{align*}
$$

In practice, these conditions may be difficult to verify since they involve bounding the
distance to the true minimizer of the subproblems, which are unknown. The other drawback of these error criteria is that there is no explicit guidance for selecting the error tolerances sequences $\left\{\epsilon_{k}\right\}$ and $\left\{\tau_{k}\right\}$.

### 2.1 Proximal alternating direction methods

In [18], Chen and Teboulle gives an inexact predictor corrector proximal multiplier with summable error criteria that involves a prediction step 2.3 and a correction step (2.6). The algorithm they proposeded can be written as

$$
\begin{align*}
& q^{k+1}=p^{k}+c_{k}\left(M x^{k}-z^{k}\right)  \tag{2.3}\\
& x^{k+1}=\arg \min \left\{f(x)+\left\langle q^{k+1}, M x\right\rangle+\frac{1}{2 c_{k}}\left\|x-x^{k}\right\|^{2}\right\}  \tag{2.4}\\
& z^{k+1}=\arg \min \left\{g(z)-\left\langle q^{k+1}, z\right\rangle+\frac{1}{2 c_{k}}\left\|z-z^{k}\right\|^{2}\right\}  \tag{2.5}\\
& p^{k+1}=p^{k}+c_{k}\left(M x^{k+1}-z^{k+1}\right) . \tag{2.6}
\end{align*}
$$

Here $\left\{p^{k}\right\}$ and $\left\{q^{k}\right\}$ are two different sequences of estimates for the Lagrange multipliers. To solve (2.4) and (2.5) inexactly, the following error criteria are proposed, for all $k$, let $\epsilon_{k}, \tau_{k} \geq 0$ and $\sum_{0}^{\infty} \sqrt{\epsilon_{k}}<\infty, \sum_{0}^{\infty} \sqrt{\tau_{k}}<\infty$ such that

$$
\begin{align*}
& \left\|x^{k+1}-\arg \min \left\{f(x)+\left\langle q^{k+1}, M x\right\rangle+\frac{1}{2 c_{k}}\left\|x-x^{k}\right\|^{2}\right\}\right\| \leq \epsilon_{k+1}  \tag{2.7}\\
& \left\|z^{k+1}-\arg \min \left\{g(z)-\left\langle q^{k+1}, z\right\rangle+\frac{1}{2 c_{k}}\left\|z-z^{k}\right\|^{2}\right\}\right\| \leq \tau_{k+1} \tag{2.8}
\end{align*}
$$

Since this method also uses absolute summable error criteria, it suffers from the same drawbacks we have pointed out for such criteria (however, the subproblem objectives are necessarily strongly convex, which aids in bounding the distance to the true solution). Li 55 also adopted this kind of two-step structure and developed a more complicated algorithm. The first step is called the prediction step, and uses the method developed in (18 to produce a trial triplet (predictor) $\left(\tilde{x}^{k}, \tilde{z}^{k}, \tilde{p}^{k}\right)$ of the new iterate, and then the second step corrects the predictor with a steepest-descent-class
step to generate the new iterate $\left(x^{k+1}, z^{k+1}, p^{k+1}\right)$. However, there is no computational evidence showing that this algorithm actually outperforms the aforementioned algorithm in [18]. In [19, 40, 65], similar methods that consisting of projection and contraction steps are proposed. For the sake of simplicity, we assume both $f$ and $g$ are differentiable and their gradients are denoted by $\nabla f$ and $\nabla g$. We take the algorithm found in 19 for example, for given $x^{k} \in \mathcal{X}, z^{k} \in \mathcal{Z}$, the prediction step computes a trial iterate $\left(\tilde{x}^{k}, \tilde{z}^{k}, \tilde{p}^{k}\right)$ by

$$
\begin{align*}
& \tilde{x}^{k}=P_{\mathcal{X}}\left(x^{k}-\beta_{k}\left[\nabla f\left(x^{k}\right)+M^{\top} p^{k}-c M^{\top}\left(M x-z^{k}\right)\right]\right)  \tag{2.9}\\
& \tilde{z}^{k}=P_{\mathcal{Z}}\left(z^{k}-\beta_{k}\left[\nabla g\left(z^{k}\right)-p^{k}+c\left(M \tilde{x}^{k}-z\right)\right]\right)  \tag{2.10}\\
& \tilde{p}^{k}=p^{k}+c\left(M \tilde{x}^{k}-\tilde{z}^{k}\right) \tag{2.11}
\end{align*}
$$

with the error criteria:

$$
\begin{aligned}
\beta_{k}\left\|\nabla f\left(x^{k}\right)-\nabla f\left(\tilde{x}^{k}\right)\right\| & \leq \nu\left\|x^{k}-\tilde{x}^{k}\right\| \\
\beta_{k}\left\|\nabla g\left(z^{k}\right)-\nabla g\left(\tilde{z}^{k}\right)\right\| & \leq \nu\left\|z^{k}-\tilde{z}^{k}\right\|,
\end{aligned}
$$

where $0<\nu<1$ and $\left\{\beta_{k}\right\}>0$ is bounded. When $\mathcal{X}=\mathbb{R}^{n}$ and $\mathcal{Z}=\mathbb{R}^{m}$, then the subproblems (2.9) and 2.10) are equivalent to

$$
\begin{aligned}
& \tilde{x}^{k} \approx \arg \min \left\{f(x)+\left\langle M x, p^{k}\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}+\frac{1}{2 \beta_{k}}\left\|x-x^{k}\right\|^{2}\right\} \\
& \tilde{z}^{k} \approx \arg \min \left\{g(z)-\left\langle z, p^{k}\right\rangle+\frac{c}{2}\left\|M \tilde{x}^{k}-z\right\|+\frac{1}{2 \beta_{k}}\left\|z-z^{k}\right\|\right\}
\end{aligned}
$$

The subsequent correction step moves along a descent direction towards the set of optimal solutions, but since our main concern here is with approximation criteria, we do not give the details of the correction step here. This kind of algorithm appears to require extra computational effort, but still uses traditional error conditions.

Among the family of alternating directions algorithms, and related to the algorithms just discussed, the proximal ADMM was first proposed by Eckstein in [27]. This method regularizes the ADMM subproblems with primal proximal terms, and
takes the form

$$
\begin{align*}
& x^{k+1}=\arg \min \left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}+\frac{r}{2}\left\|x-x^{k}\right\|^{2}\right\}  \tag{2.12}\\
& z^{k+1}=\arg \min \left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|+\frac{s}{2}\left\|z-z^{k}\right\|^{2}\right\}  \tag{2.13}\\
& p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k}\right) \tag{2.14}
\end{align*}
$$

where $c, r, s>0$. We denote by $x_{\text {exact }}^{k+1}$ (resp., $z_{\text {exact }}^{k+1}$ ) the exact minimizer of (2.12) (resp., 2.13), and use $x^{k+1}$ (resp., $z^{k+1}$ ) for an approximate solution to (2.12) (resp., (2.13). Based on (2.12)-(2.14), He [43], proposed an inexact proximal ADMM in which the proximal parameters $r, s$ and the penalty parameter $c$ are replaced with some sequences of both upper and lower bounded positive definite matrices $\left\{R_{k}\right\}$, $\left\{S_{k}\right\}$ and $H_{k}$ respectively. The approximate solutions $x^{k+1}, z^{k+1}$ are obtained by absolute summable error criteria

$$
\begin{equation*}
\left\|x^{k+1}-x_{\text {exact }}^{k+1}\right\| \leq \nu_{k} \quad\left\|z^{k+1}-z_{\text {exact }}^{k+1}\right\| \leq \nu_{k} \tag{2.15}
\end{equation*}
$$

where $\left\{\nu_{k}\right\}$ is a nonnegative sequence such that $\sum \nu_{k}<\infty$. Yuan 86] further relaxed these error criteria to

$$
\begin{align*}
\left\|x^{k+1}-x_{\text {exact }}^{k+1}\right\| & \leq \nu_{k}\left\|x^{k}-x^{k+1}\right\|  \tag{2.16}\\
\left\|z^{k+1}-z_{\text {exact }}^{k+1}\right\| & \leq \nu_{k}\left\|z^{k}-z^{k+1}\right\| \tag{2.17}
\end{align*}
$$

where $\sum \nu_{k}^{2}<\infty$. Compared with (2.15), the error criteria (2.17) is relative in a sense that the terms $\left\|x^{k}-x^{k+1}\right\|$ and $\left\|z^{k}-z^{k+1}\right\|$ are "self-adapted". However, these conditions still involves the unknown exact minimizers, so they are not always practical. Furthermore, the empirical performance of the proximal ADMM is not as good as the standard ADMM. For applications of the proximal ADMM, see for example 62,82,90]. Recently, Shen [74] further developed similar error conditions for proximal ADMM using the gradients of the (2.12) and (2.13). Suppose we define

$$
\begin{aligned}
& \xi_{x}^{k+1} \in \partial\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}+\frac{r}{2}\left\|x-x^{k}\right\|^{2}\right\}_{x=x^{k+1}} \\
& \xi_{z}^{k+1} \in \partial\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|+\frac{s}{2}\left\|z-z^{k}\right\|^{2}\right\}_{z=z^{k+1}}
\end{aligned}
$$

Then the absolute error criteria proposed in [74] can be expressed by

$$
\left\|\xi_{x}^{k+1}\right\| \leq \mu_{k+1} \quad\left\|\xi_{z}^{k+1}\right\| \leq \nu_{k+1}
$$

with $\sum_{0}^{\infty} \mu_{k+1}<\infty$ and $\sum_{0}^{\infty} \nu_{k+1}<\infty$, and the relative error criteria in 74 are

$$
\left\|\xi_{x}^{k+1}\right\| \leq \mu_{k+1}\left\|x^{k+1}-x^{k}\right\| \quad\left\|\xi_{z}^{k+1}\right\| \leq \nu_{k+1}\left\|z^{k+1}-z^{k}\right\|
$$

with $\sum_{0}^{\infty} \mu_{k+1}^{p}<\infty$ and $\sum_{0}^{\infty} \nu_{k+1}^{p}<\infty$, and $p=1$ or 2 . These criteria are somewhat similar to those proposed in this disseration, but apply to the proximal ADMM, not the standard ADMM.

### 2.2 Logarithmic-quadratic proximal ADMM

The combination of the ADMM and logarithmic-quadratic proximal regularization was first proposed in [5] and has been extensively studied in many papers, such as $[3,4,42,53,78,88]$. The LQP ADMM takes forms of

$$
\begin{align*}
x^{k+1} & =\arg \min \left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}+r d_{k}\left(x, x^{k}\right)\right\}  \tag{2.18}\\
z^{k+1} & =\arg \min \left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}+s d_{k}\left(z, z^{k}\right)\right\}  \tag{2.19}\\
p^{k+1} & =p^{k}+c\left(M x^{k+1}-z^{k+1}\right) \tag{2.20}
\end{align*}
$$

where $c, r, s$ are positive scalars and for any $z \in \mathcal{R}_{++}^{N}, d_{k}(\cdot, \cdot)$ is defined by

$$
d_{k}\left(u^{\prime}, u\right)= \begin{cases}\sum_{j=1}^{N}\left[\frac{1}{2}\left(u_{j}^{\prime}-u_{j}\right)^{2}+\mu_{k}\left(u_{j}^{2} \log \frac{u_{j}}{u_{j}^{\prime}}+u_{j}^{\prime} u_{j}-u_{j}^{2}\right)\right], & \text { if } u^{\prime} \in \mathbb{R}_{++}^{N} \\ +\infty & \text { otherwise }\end{cases}
$$

and for any $u^{\prime} \in \mathbb{R}_{++}^{N}$, we have

$$
\nabla_{u^{\prime}} d_{k}\left(u^{\prime}, u\right)=\left(u^{\prime}-u\right)+\mu_{k}\left[u-U^{2}\left(u^{\prime}\right)^{-1}\right]
$$

where $\left\{\mu_{k}\right\} \subset(0,1), U:=\operatorname{diag}\left(u_{1}, u_{2} \ldots, u_{N}\right) \in \mathbb{R}^{N \times N}$ and $\left(u^{\prime}\right)^{-1}$ is a vector whose $j$-th element is $1 / z_{j}^{\prime}$. For any $u^{\prime}, u \in \mathbb{R}_{++}^{N}$, we have $d\left(u^{\prime}, u\right) \geq \frac{1}{2}\left\|u^{\prime}-u\right\|^{2}$ [5] and
$d\left(u^{\prime}, u\right)=0$ if and only if $u^{\prime}=u$. A worst-case $O(1 / k)$ convergence rate for LQP ADMM was established in (78, and more recent convergence analyses can be found in [54]. Inexact versions of LQP ADMM with similar error criteria to those of the proximal ADMM have also been developed. We use $x^{k+1}$ and $z^{k+1}$ to denote the approximate solutions of subproblems (2.18) and (2.19) respectively, with exact minimizers being denoted by $x_{\text {exact }}^{k+1}$ and $z_{\text {exact }}^{k+1}$. The inexact LQP ADMM proposed in 11 uses the following error criteria:

$$
\left\|x^{k+1}-x_{\text {exact }}^{k+1}\right\| \leq\left\|x^{k}-x^{k+1}\right\| \quad\left\|z^{k+1}-z_{\text {exact }}^{k+1}\right\| \leq\left\|z^{k}-z^{k+1}\right\|
$$

where $\left\{\nu_{k}\right\}$ is a nonnegative sequence such that $\sum_{k=0}^{\infty} \nu_{k}^{2}<\infty$. One approximate LQP ADMM proposed in (17) uses the following scheme: obtain $x^{k+1}$ and $\xi_{x}^{k+1}$ through solving the variation inequality

$$
\begin{aligned}
& f(x)-f\left(x^{k+1}\right) \\
& \quad+\left(x-x^{k+1}\right)^{\top}\left\{-M^{\top}\left[p^{k}-c\left(M x^{k+1}-z^{k}\right)\right]+r \nabla d\left(x^{k+1}, x^{k}\right)+\xi_{x}^{k+1}\right\} \geq 0
\end{aligned}
$$

and obtain $z^{k+1}$ and $\xi_{z}^{k+1}$ by solving

$$
\begin{aligned}
& g(z)-g\left(z^{k+1}\right) \\
& \quad+\left(z-z^{k+1}\right)^{\top}\left\{-\left[p^{k}-c\left(M^{k+1}-z^{k+1}\right)\right]+s \nabla d\left(z^{k+1}, z^{k}\right)+\xi_{z}^{k+1}\right\} \geq 0
\end{aligned}
$$

where $\xi_{x}^{k+1}$ and $\xi_{z}^{k+1}$ satisfy the approximation criteria

$$
\begin{aligned}
\left\|\xi_{x}^{k+1}\right\| & \leq \nu_{k} r \sqrt{2\left(1+\mu_{k}\right)}\left\|x^{k+1}-x^{k}\right\| \\
\left\|\xi_{z}^{k+1}\right\| & \leq \nu_{k} s \sqrt{2\left(1+\mu_{k}\right)}\left\|z^{k+1}-z^{k}\right\|
\end{aligned}
$$

and the requirements for $\left\{\mu_{k}\right\}$ and $\left\{\nu_{k}\right\}$ are that $0<\mu_{k}<1-2 \nu_{k}, \nu_{k} \geq 0$ and $\sum_{k=0}^{\infty} \nu_{k}<\infty$. In 9, 10, 12 14, approximate LQP ADMM methods with predictioncorrection steps for solving variational inequalities with separable structure were proposed. Although their correction steps are different, these algorithms' error criterion
for 2.18 and 2.19 are same: let

$$
\begin{aligned}
\xi_{x}^{k+1} & :=\partial_{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}+r d_{k}\left(x, x^{k}\right)\right\}_{x=x^{k+1}} \\
\xi_{z}^{k+1} & :=\partial_{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}+s d_{k}\left(z, z^{k}\right)\right\}_{z=z^{k+1}}
\end{aligned}
$$

Then the error criteria can be expressed as

$$
\left\|\left(1+\mu_{k} r\right) \xi_{x}^{k+1}+\left(1+\mu_{k}\right) s \xi_{z}^{k+1}\right\|^{2} \leq \frac{1-\mu_{k}}{1+\mu_{k}} \eta^{2}\left\|\left(x^{k}, z^{k}\right)-\left(x^{k+1}, z^{k+1}\right)\right\|^{2}
$$

where $\left\{\mu_{k}\right\} \subset(0,1)$ and $\eta \in(0,1)$.

## Chapter 3 <br> Formalizing Approximate Subproblem Solution

In order to formalize our methods clearly, we present a mathematical model of the approximate solution processes for the subproblems (1.9) and 1.10. All of our methods will use the following assumption:

Assumption 2. To approximately solve (1.9), we assume the existence of some mapping $\mathcal{F}: \mathbb{R}^{m} \times \mathbb{R}^{m} \times \mathbb{R}_{++} \times \mathbb{R}^{n} \times \mathbb{N} \rightarrow \mathbb{R}^{n} \times \mathbb{R}^{n}$ such that if we let $\left(x^{l}, y_{1}^{l}\right)=\mathcal{F}(p, z, c, \bar{x}, l)$, we have that

$$
\lim _{l \rightarrow \infty} y_{1}^{l}=0 \quad(\forall l \in \mathbb{N}) y_{1}^{l} \in \partial_{x}\left[f(x)+\langle p, M x\rangle+\frac{c}{2}\|M x-z\|^{2}\right]_{x=x^{l}}
$$

The idea behind this definition is that $\mathcal{F}(p, z, c, \bar{x}, l)$ is the $l^{\text {th }}$ iterate produces by the $x$-subproblem solution procedure with penalty parameter $c$, the Lagrange multiplier estimate $p^{k}$ equal to $p$, and $z^{k}=z$, starting from the solution estimate $\bar{x}$. So, to solve the subproblem (1.9), we would take iterates of the form $\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)$ for increasing $l$ until obtaining a suitably small $y_{1}^{k, l}$. The "starting point" argument $\bar{x}$ is intended to model the customary computational practive of "warm-starting" iterative subroutines from the value obtained at the previous iteration, but we do not require this information to be used in any specific way. For example, $\mathcal{F}$ is free to simply ignore the $\bar{x}$ argument.

Assumption 2 will suffice in cases for which we assume that the second subproblem (1.10) may be solved exactly. However, we also allow for the case that both 1.9 ) and 1.10 are sufficiently difficult to merit iterative solution, in which case we also need the following assumption:

Assumption 3. To approximately solve (1.10), we assume the existence of some mapping $\mathcal{G}: \mathbb{R}^{m} \times \mathbb{R}^{n} \times \mathbb{R}_{++} \times \mathbb{R}^{m} \times \mathbb{N} \rightarrow \mathbb{R}^{m} \times \mathbb{R}^{m}$ such that if we let $\left(z^{l}, y_{2}^{l}\right)=$ $\mathcal{G}(p, x, c, \bar{z}, l)$, we have that

$$
\lim _{l \rightarrow \infty} y_{2}^{l}=0 \quad(\forall l \in \mathbb{N}) y_{2}^{l} \in \partial_{z}\left[g(z)-\langle p, z\rangle+\frac{c}{2}\|M x-z\|^{2}\right]_{z=z^{l}}
$$

Similarly to the previous assumption, $\mathcal{G}(p, x, c, \bar{z}, l)$ models the $l^{\text {th }}$ iterative approximate solution to the problem of minimizing $g(x)-\langle p, z\rangle+\frac{c}{2}\|M x-z\|^{2}$, with the starting point $\bar{z}$. The function $\mathcal{G}$ may use the starting point information $\bar{z}$ in an arbitrary way, which may include simply ignoring it.

In cases for which either (1.9) or (1.10) is easily solved exactly, we may respectively simply take $\mathcal{F}$ or $\mathcal{G}$ as "jumping" immediately to the exact solution and a zero subgradient for $l=1$, and simply returning the same information for larger values of $l$. In the case that it is easy to compute an exact solution of 1.10 , for example, we may take

$$
\mathcal{G}(p, x, c, \bar{z}, l)=\left(\underset{z \in \mathbb{R}^{m}}{\arg \min }\left\{g(z)-\langle p, z\rangle+\frac{c}{2}\|M x-z\|^{2}\right\}, 0\right) \quad \forall l \in \mathbb{N}
$$

although in practice it should not be necessary to evaluate $\mathcal{G}(p, x, c, \bar{z}, l)$ for $l>1$, an exact solution to the subproblem already having been calculated.

We close this section by establishing some properties of the sequences $\left\{x^{l}\right\}$ and $\left\{z^{l}\right\}$ generated by $\mathcal{F}$ and $\mathcal{G}$ respectively. To do so, we first prove a convex-analytic lemma:

Lemma 4. Let $h: \mathbb{R}^{n} \rightarrow \mathbb{R} \cup\{+\infty\}$ be closed proper convex, and let $\left\{x^{l}\right\}$, $\left\{y^{l}\right\}$ be sequences in $\mathbb{R}^{n}$ such that $y^{l} \in \partial h\left(x^{l}\right)$ for all $l$ and $y^{l} \rightarrow 0$. Then if the set of minimizers of $h$ is nonempty and bounded, $\left\{x^{l}\right\}$ must be bounded, with all its limit points being minimizers of $h$. If h has a unique minimizer, then $\left\{x^{l}\right\}$ converges to that minimizer.

Proof. By 69, Theorem 27.1(d)], $h$ having a nonempty bounded set of minimizers is equivalent to $0 \in \operatorname{int} \operatorname{dom} h^{*}$, where $h^{*}$ denotes the convex conjugate of $h$. By 69, Theorem 23.4], we than have $0 \in \operatorname{int} \operatorname{dom} \partial h^{*}$. By the Rockafellar-Veselý theorem [68], the
maximal monotone point-to-set map $\partial h^{*}$ must then be locally bounded at 0 , meaning that for some $\epsilon>0$, the set $S(\epsilon)=\left\{x \in \mathbb{R}^{n} \mid \exists y \in \mathbb{R}^{n}: x \in \partial h^{*}(y),\|y\|<\epsilon\right\}$ is bounded. By [69, Theorem 23.5], $S(\epsilon)=\left\{x \in \mathbb{R}^{n} \mid \exists y \in \mathbb{R}^{n}: y \in \partial h(x),\|y\|<\epsilon\right\}$, so the convergence of $\left\{y^{l}\right\}$ to zero implies that $\left\{x^{l}\right\}$ is bounded. If we consider any limit point $x^{\infty}$ of $\left\{x^{l}\right\}$, with $x^{l} \rightarrow_{\mathcal{L}} x^{\infty}$ for some infinite index sequence $\mathcal{L}$, then we have $y^{l} \rightarrow_{\mathcal{L}} 0$ and hence by the closure property of the maximal monotone operator $\partial h$, we have $0 \in \partial h\left(x^{\infty}\right)$, and $x^{\infty}$ must be a minimizer of $h$. In the case that that $h$ has a unique minimizer $\bar{x}$, we then have that $\left\{x^{l}\right\}$ is a bounded sequence whose only possible limit point is $\bar{x}$, so it must converge to $\bar{x}$.

This lemma has the following immediate consequences.

Lemma 5. If the set of minimizers of $f(x)+\langle p, M x\rangle+\frac{c}{2}\|M x-z\|^{2}$ is bounded, then the sequence $\left\{x^{l}\right\}$ generated by $\left(x^{l}, y_{1}^{l}\right)=\mathcal{F}(p, z, c, \bar{x}, l)$ with $\mathcal{F}$ as in Assumption 2 must be bounded. If the minimizer of $f(x)+\langle p, M x\rangle+\frac{c}{2}\|M x-z\|^{2}$ is unique, $\left\{x^{l}\right\}$ must converge to it.

Proof. Immediate from Lemma 4, setting $h(x)=f(x)+\langle p, M x\rangle+\frac{c}{2}\|M x-z\|^{2}$.
Lemma 6. The sequence generated by $\left(z^{l}, y_{2}^{l}\right)=\mathcal{G}(p, x, c, \bar{z}, l)$, with $\mathcal{G}$ as in Assumption 3. always converges to the unique minimizer over $z$ of $g(z)-\langle p, z\rangle+\frac{c}{2}\|M x-z\|^{2}$.

Proof. We observe that $g(z)-\langle p, z\rangle+\frac{c}{2}\|M x-z\|^{2}$ is strongly convex as a function of $z$, and therefore has a unique minimizer. The result then follows immediately from Lemma 4

With regard to Lemma 5, one condition sufficient for the minimizer of $f(x)+$ $\langle p, M x\rangle+\frac{c}{2}\|M x-z\|^{2}$ to be unique is that $M$ have full column rank, or that the minimum exist with $f$ being strictly convex.

## Chapter 4 Approximate ADMM Algorithms Derived through Operator Splitting Analysis

This chapter presents two approximate ADMM algorithms that may be derived through the operator-splitting analysis of the ADMM. The first applies in special cases for which subgradients supply sufficient information to guarantee the distancebased approximation criteria in the approximate ADMM in [29, Theorem 8]. It uses absolute error criteria with formally exogenous summable error sequence parameters, and its analysis is very brief due to the results already present in [29, Theorem 8].

The proof of [29, Theorem 8] is based on the relationship of the ADMM to DouglasRachford (DR) splitting, the equivalence of DR splitting to the proximal point algorithm (PPA), and the application of an approximate PPA using an absolute error criterion. The remainder of this section considers taking a similar "modular" approach to deriving an approximate ADMM, but using a relative-error version of the PPA. The result is a new, relative-error variant of the ADMM. A byproduct of the analysis is a new, relative-error variant of DR splitting.

The remainder of this chapter derives different subgradient-based approximate ADMM through a "modular" approach that exploits the relationship between ADMM, Douglas-Rachford splitting, and the proximal point algorithm, in combination with a suitable inexact proximal point algorithm (PPA). The advantage of using such a modular framework is that convergence follows fairly directly from the convergence of the inexact proximal point algorithm; the main effort in the analysis is in deriving the correct form of the method from its analytical "building blocks".

### 4.1 A subgradient-based application of [29, Theorem 8]

Let $\left\{\epsilon_{k}\right\}_{k=1}^{\infty},\left\{\tau_{k}\right\}_{k=1}^{\infty} \subset \mathbb{R}_{++}$be positive scalar sequences such that $\sum_{k=1}^{\infty} \epsilon_{k}<\infty$ and $\sum_{k=1}^{\infty} \tau_{k}<\infty$. Using these sequences as parameters, one of the simplest imaginable way to construct an approximate ADMM based on the mappings hypothesized in Assumptions 2 and 3 is as follows:

```
Algorithm 4.1.1 Inexact ADMM with absolute summable error criteria
    initialization: Pick \(c>0\) and initial points \(p^{0}, z^{0} \in \mathbb{R}^{m}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\)
            by taking \(\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until \(\left\|y_{1}^{k, l}\right\| \leq \epsilon_{k+1}\)
        \(x^{k+1}=x^{k, l}\)
        \(y_{1}^{k+1}=y_{1}^{k, l}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(z^{k+1} \approx \arg \min _{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\)
            by taking \(\left(z^{k, l}, y_{2}^{k, l}\right)=\mathcal{G}\left(p^{k}, x^{k+1}, c, z^{k}, l\right)\)
        until \(\left\|y_{2}^{k, l}\right\| \leq \tau_{k+1}\)
        \(z^{k+1}=z^{k, l}\)
        \(y_{2}^{k+1}=y_{2}^{k, l}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
until Overall convergence
```

Throughout this paper, we will leave the "overall convergence" termination criterion for the outer loops of our algorithms of the algorithm abstract, since the best choice
may be application-dependent. A reasonable generic choice, however, would be

$$
\left.\begin{array}{rl}
\left\|M x^{k+1}-z^{k+1}\right\| & \leq \delta_{1} \\
\exists y_{2}^{k+1} \| & \leq \delta_{2} \\
1 \tag{4.3}
\end{array}\right] \partial_{x}\left[f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k+1}\right\|^{2}\right]_{x=x^{k+1}}:\left\|\bar{y}_{1}^{k+1}\right\| \leq \delta_{3}, ~ \$
$$

where $\delta_{1}, \delta_{2}, \delta_{3}$ are small positive scalars. One possible choice of $\bar{y}_{1}^{k+1}$ in 4.2) is $y_{1}^{k}$. If one uses an overall convergence test that does not require the sequence $\left\{y_{2}^{k}\right\}$, then one can omit the assignment $y_{2}^{k+1}=y_{2}^{k, l}$ from the implementation of the algorithm, and similarly for $\left\{y_{1}^{k}\right\}$.

While we cannot guarantee convergence of Algorithm 4.1.1 in the general case, there are several special cases in which the condition $\left\|y_{1}^{k, l}\right\| \leq \epsilon_{k+1}$ guarantees a bound on the distance to the exact solution of the subproblem 1.9 , which in turn means that the algorithm is a special case of the algorithm proved to converge in [29, Theorem 8]. Essentially, we require that the minimand in (1.9) be strongly convex:

Proposition 7. Under Assumptions 2a and 3, the inner loops (over l) in Algorithm 4.1.1 always terminate finitely. Under Assumption 1, if either $f$ is strongly convex or $M$ has full column rank, then $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ converges to a KKT point.

Proof. The assertion about finite convergence of the inner loops follows immediately from Assumptions 2 and 3, combined with the positivity of $\epsilon_{k}$ and $\tau_{k}$ for all $k$.

Let $\alpha$ be the modulus of strong convexity of $f$ and let $\kappa\left(M^{\top} M\right)$ denote the smallest eigenvalue of the symmetric matrix $M^{\top} M$. For each $k$, let $\bar{f}_{k}(x)=f(x)+\left\langle p^{k}, M x\right\rangle+$ $\frac{c}{2}\left\|M x-z^{k}\right\|^{2}$, the minimand in (1.9) expressed as a function of $x$. Under the hypotheses, we have $\alpha>0$ or $\kappa\left(M^{\top} M\right)>0$, so $\bar{f}_{k}$ is strongly convex with modulus $\bar{\alpha}=\alpha+c \kappa\left(M^{\top} M\right)>0$. It follows that $\bar{f}_{k}$ 's subdifferential map $\partial \bar{f}_{k}$ is strongly monotone with modulus $\bar{\alpha}$. Letting $\bar{x}^{k+1}$ denote the exact minimizer of (1.9), the Cauchy-Schwarz inequality and strong monotonicity of $\partial \bar{f}_{k}$ combine to yield

$$
\left\|y_{1}^{k, l}\right\|\left\|x^{k, l}-\bar{x}^{k+1}\right\| \geq\left\langle y_{1}^{k, l}-0, x^{k, l}-\bar{x}^{k+1}\right\rangle \geq \bar{\alpha}\left\|x^{k, l}-\bar{x}^{k+1}\right\|^{2}
$$

and hence

$$
\left\|y_{1}^{k, l}\right\| \geq \bar{\alpha}\left\|x^{k, l}-\bar{x}^{k+1}\right\|
$$

for all $k$ and $l$ encountered in the algorithm. Combining this result with the termination condition for the approximate $x$ minimization, we obtain

$$
\left\|x^{k+1}-\bar{x}^{k+1}\right\| \leq \epsilon_{k+1} / \bar{\alpha}
$$

for all $k$. Thus, the distance of $x^{k}$ to the exact $x$-subproblem solution is bounded above by a summable sequence, namely $\left\{\epsilon_{k} / \bar{\alpha}\right\}$.

We next consider the $z$ subproblem (1.10). Its minimand is always strongly convex with modulus $c$, so a similar analysis shows that the distance from $z^{k}$ to the exact $z$-subproblem solution is bounded above by the summable sequence $\left\{\tau_{k} / c\right\}$. All the remaining claims of the proposition now follow immediately from [29, Theorem 8].

Remark: Although this form is rarely needed in practice, the ADMM is sometimes presented in the context of the more general problem

$$
\min _{x, z}\{f(x)+g(z) \mid M x+N z=b\}
$$

where $N$ is an additional constraint matrix and $b$ is some given vector; see for example 15. Generalizing the above result to this case would require the assumption that $g$ be strongly convex or $N$ have full column rank, in addition to the assumptions on $f$ and $M$.

### 4.2 Background: a relative-error proximal point algorithm

We now embark on the derivation of an algorithm with similar theoretical underpinnings to Algorithm 4.1.1, but with a relative error criterion. Our analysis employs the inexact relative-error proximal point algorithm developed by Solodov and Svaiter in [76]. This relative-error algorithm allows for general Bregman distance kernels, but here we use only the special case of the standard squared Euclidean distance kernel $D(x, y)=\frac{1}{2}\|x-y\|^{2}$ derived from the canonical Bregman function $h(x)=\frac{1}{2}\|x\|^{2}$.

An operator $T$ on $\mathbb{R}^{n}$ is called monotone if

$$
\langle y-v, x-u\rangle \geq 0
$$

for all $v \in T(u)$ and $y \in T(x)$. For any real number $c$ and operator $T$, we let $c T$ be the operator $\{(x, c y) \mid y \in T(x)\}$ and if $A$ and $B$ are any operators, then

$$
A+B=\{(x, y+z) \mid y \in A(x), z \in B(x)\}
$$

A monotone operator $T$ is maximal monotone if there is no other monotone operator that properly contain $T$. We do not distinguish an operator $T$ and its graph, thus expressions of $y \in T(x)$ and $(x, y) \in T$ are equivalent. Let $T$ be maximal monotone, a fundamental problem here is finding a zero of $T$, i.e.

$$
\begin{equation*}
0 \in T(z), z \in \mathbb{R}^{n} \tag{4.4}
\end{equation*}
$$

The classical algorithm for solving this problem is the proximal point method. It was first introduced by Martinet [59] and further developed by Rockafellar [71].

The proximal point algorithm for solving the generic inclusion $0 \in T(z)$, where $T: \mathbb{R}^{n} \rightrightarrows \mathbb{R}^{n}$ is a maximal monotone operator, involves generating a sequence $\left\{z^{k}\right\} \subset$ $\mathbb{R}^{n}$ such that $\left\{z^{k+1}\right\}$ is the solution of

$$
\begin{equation*}
0 \in \lambda_{k} T(z)+z-z^{k} \tag{4.5}
\end{equation*}
$$

for all $k \geq 0$, where $\left\{\lambda_{k}\right\}$ is a sequence of scalar parameters with $\inf _{k \geq 0}\left\{\lambda_{k}\right\}>0$. Equivalent problems are to find a pair $(u, v)$ such that

$$
\begin{equation*}
v \in T(u) \quad \lambda_{k} v+u-z^{k}=0 \tag{4.6}
\end{equation*}
$$

or

$$
\begin{equation*}
v \in T(u) \quad u=z^{k}-\lambda_{k} v \tag{4.7}
\end{equation*}
$$

We now define a notion of an inexact solution of (4.6), specializing [76, Definition 3.1] to the case of the squared Euclidean distance kernel:

Definition 8. Let $\lambda_{k}>0$ and $\sigma \in[0,1)$. We say that a pair $(u, v)$ is an inexact solution with tolerance $\sigma$ for the proximal subproblem (4.6) if

$$
\begin{equation*}
v \in T(u) \quad\left\|u+\lambda_{k} v-z^{k}\right\| \leq \sigma\left\|u-z^{k}\right\| \tag{4.8}
\end{equation*}
$$

Observe that when $\sigma=0$, we must have $u=z^{k}-\lambda_{k} v$, meaning that the proximal subproblem (4.6) must be solved exactly. We will base our approximate ADMM on the following inexact proximal point algorithm, which is [76, Algorithm 1] specialized to the case of the squared Euclidean distance kernel:

## Algorithm 4.2.1 Inexact generalized proximal point method <br> Initialization:

Choose some $\lambda>0$, error tolerance parameter $\sigma \in[0,1)$, and starting point $z^{0} \in$ $\mathbb{R}^{n}$.
for $k=0,1, \ldots$ do
Select any $\lambda_{k} \geq \lambda$, and let $\left(u^{k}, v^{k}\right)$ be some inexact solution with tolerance $\sigma$ of the subproblem $0 \in \lambda_{k} T(z)+z-z^{k}$, that is, find some $\left(u^{k}, v^{k}\right)$ such that

$$
\begin{equation*}
v^{k} \in T\left(u^{k}\right) \quad\left\|u^{k}+\lambda_{k} v^{k}-z^{k}\right\| \leq \sigma\left\|u^{k}-z^{k}\right\| \tag{4.9}
\end{equation*}
$$

Set

$$
\begin{equation*}
z^{k+1}:=z^{k}-\lambda_{k} v^{k} \tag{4.10}
\end{equation*}
$$

end for

We may summarize the recursions of Algorithm 4.2.1 as follows:

$$
\begin{aligned}
v^{k} & \in T\left(u^{k}\right) \\
\lambda_{k} v^{k}+z^{k+1}-z^{k} & =0 \\
\left\|u^{k}-z^{k+1}\right\| & \leq \sigma\left\|u^{k}-z^{k}\right\|
\end{aligned}
$$

The behavior of Algorithm 4.2.1 is already established in [76]:

Proposition 9. If a solution to $0 \in T(z)$ exists, then $\left\{z^{k}\right\}$ generated by Algorithm 4.2.1 converges to such a solution. In addition, $\left\{u^{k}\right\}$ also converges to this solution and $v^{k} \rightarrow 0$.

Proof. The result follows directly by specializing [76, Proposition 4.4] and [76, Corollary 4.3 ] to the case of the Euclidean squared distance kernel, whose zone of definition is $C=\mathbb{R}^{n}$.

### 4.3 A relative-error variant of Douglas-Rachford splitting

We will apply the relative-error proximal point algorithm stated in the last section to the ADMM through a two-step process: first, capitalizing on the analysis in [29], we will use Algorithm 4.2.1 to derive a relative-error variant of Douglas-Rachford (DR) splitting method for pairs of maximal monotone operators. In the next section, we will use this result to derive a relative-error version of the ADMM, using that the ADMM is a special case of DR splitting as first established in [34].

We now derive a relative-error DR method. The original DR splitting method of [57] is a method for solving the problem $A(x)+B(x) \ni 0$, where $A, B: \mathcal{H} \rightrightarrows \mathcal{H}$ are maximal monotone operators on a real Hilbert space $\mathcal{H}$. Here, we will consider only the case $\mathcal{H}=\mathbb{R}^{n}$. The goal of the method is to converge to a solution to $A(x)+B(x) \ni 0$ through a process that evaluates only resolvents $(I+\gamma A)^{-1}$ and $(I+\gamma B)^{-1}$ of the respective individual operators $A$ and $B$, rather than working directly with the operator $A+B$.

Given two set-valued operators $A$ and $B$ on $\mathbb{R}^{n}$ and a scalar $\gamma>0$, the analysis in (29) defines the splitting operator $S_{\gamma, A, B}: \mathbb{R}^{n} \rightrightarrows \mathbb{R}^{n}$ to be the set-valued map

$$
\begin{equation*}
S_{\gamma, A, B}=\{(r+\gamma b, s-r) \mid(s, b) \in B,(r, a) \in A, r+\gamma a=s-\gamma b\} . \tag{4.11}
\end{equation*}
$$

Here, we do not distinguish between an operator and its graph: a set-valued map $T$ is considered to be the set of ordered pairs $(x, y)$ such that $y \in T(x)$. If both $A$
and $B$ are (maximal) monotone then $S_{\gamma, A, B}$ is (maximal) monotone for any scalar $\gamma>0$ [29, Theorem 4]. There is also an important relationship between the zeros of $S_{\gamma, A, B}$ and those of $A+B$ : letting $\operatorname{zer}(T)$ denote the set of all zeros of operator $T$, then by [29, Theorem 5] we have

$$
\begin{aligned}
\operatorname{zer}\left(S_{\gamma, A, B}\right) & =\{r+\gamma b \mid b \in B(r),-b \in A(r)\} \\
& \subseteq\{r+\gamma b \mid r \in \operatorname{zer}(A+B), b \in B(r)\}
\end{aligned}
$$

Thus, given that $A$ and $B$ are maximal monotone, one may attempt to use the proximal point algorithm on $S_{\gamma, A, B}$ to find a zero of $S_{\gamma, A, B}$, from which one may easily calculate a zero of $A+B$. It is shown in [29, Theorem 6] that the DR splitting method is equivalent to applying the proximal point algorithm to $S_{\gamma, A, B}$ with the proximal parameter $\lambda_{k}$ always set to 1 , that is,

$$
\begin{equation*}
z^{k+1}=\left(I+S_{\gamma, A, B}\right)^{-1}\left(z^{k}\right) \tag{4.12}
\end{equation*}
$$

This viewpoint is exploited in [29] to develop approximate versions of the DR splitting method, by applying an approximate rather than exact version of the PPA to 4.12). Since the approximate PPA employed in this analysis used an absolute summable error criterion, the resulting approximate DR method inherited the same kind of error criteria. Here, we instead consider applying the relative-error inexact PPA in Algorithm 4.2.1 to $S_{\gamma, A, B}$, obtaining a relative-error inexact variant of DR splitting.

The recursion 4.12) consists of repeatedly applying the mapping

$$
\begin{equation*}
\left(I+S_{\gamma, A, B}\right)^{-1}=\{(s+\gamma b, r+\gamma b) \mid(s, b) \in B,(r, a) \in A, r+\gamma a=s-\gamma b\} . \tag{4.13}
\end{equation*}
$$

Repeated application of this mapping may be carried out through the following steps:

DR1. Given some $r^{k}, b^{k} \in \mathbb{R}^{n}$, find $\left(s^{k+1}, b^{k+1}\right) \in B$ such that $s^{k+1}+\gamma b^{k+1}=$ $r^{k}+\gamma b^{k}$. Note that this calculation is equivalent to finding $s^{k+1}=(I+$ $\gamma B)^{-1}\left(r^{k}+\gamma b^{k}\right)$ and setting $b^{k+1}=b^{k}+\frac{1}{\gamma}\left(r^{k}-w^{k+1}\right)$.

DR2. Find $\left(r^{k+1}, a^{k+1}\right) \in A$ such that $r^{k+1}+\gamma a^{k+1}=s^{k+1}-\gamma b^{k}$. Similarly to the previous step, this calculation is equivalent to finding $r^{k+1}=(I+$ $\gamma A)^{-1}\left(s^{k+1}-\gamma b^{k}\right)$ and setting $a^{k+1}=\frac{1}{\gamma}\left(s^{k+1}-r^{k+1}\right)-b^{k}$.

DR3. Increment $k$ and return to step DR1.

This procedure is one of the standard formulations of DR splitting. We will henceforth assume that step DR1 is the more difficult of the two calculations, meaning that for a given $u \in \mathbb{R}^{n}$ some iterative process is required to solve systems of the form

$$
\begin{equation*}
b \in B(s) \quad s+\gamma b=u \tag{4.14}
\end{equation*}
$$

We model this iterative process using the following generalization of the techniques proposed in Section 3:

Assumption 10. There exists a mapping $\mathcal{B}: \mathbb{R}^{n} \times \mathbb{R}_{+} \times \mathbb{R}^{n} \times \mathbb{R}^{n} \times \mathbb{N} \rightarrow B$ such that if one defines $\left(s^{l}, b^{l}\right)=\mathcal{B}(u, \gamma, \bar{s}, \bar{b}, l)$ for all $l \geq 1$, then the sequence $\left\{\left(s^{l}, b^{l}\right)\right\}_{l=1}^{\infty}$ is convergent and $\lim _{l \rightarrow \infty} s^{l}+\gamma b^{l}=u$.

Intuitively, we intend $\left(s^{l}, b^{l}\right)=\mathcal{B}(u, \gamma, \bar{s}, \bar{b}, l) \in B$ to be the $l^{\text {th }}$ trial approximate solution to (4.14) starting from some initial guess $(\bar{s}, \bar{b})$; however, we do not specify exactly how the starting point information $(\bar{s}, \bar{b})$ is incorporated into the calculation, and it is possible for it to be ignored.

On the other hand, we assume that a set of conditions similar to (4.14) for the operator $A$ may be solved rapidly, and therefore step DR2 of the above sequence is relatively easy to carry out exactly. Under this assumption, if one is given any point $(s, b) \in B$, one can quickly determine a pair in the operator $S_{\gamma, A, B}$ by finding $(r, a) \in A$ such that $r+\gamma a=s-\gamma b$. It then follows from (4.11) that $(r+\gamma b, s-r) \in S_{\gamma, A, B}$.

To apply Algorithm 4.2.1 to the operator $S_{\gamma, A, B}$, we need to find inexact solutions of the conditions 4.9 for the case $T=S_{\gamma, A, B}$ and $\lambda_{k}=1$. Let us call the iterative procedure abstracted in Assumption 10 the " $B$-procedure", and let $l$ be an
"inner" iteration index associated with this procedure, we may attempt to execute each iteration of Algorithm 4.2.1 as applied to $S_{\gamma, A, B}$ as follows, starting with $l=1$ :

S1. Execute one step of the $B$-procedure, yielding $\left(s^{k, l}, b^{k, l}\right)=\mathcal{B}\left(z^{k}, \gamma, s^{k}, b^{k}, l\right) \in B$ with $s^{k, l}+\gamma b^{k, l} \approx z^{k}$. We call $\left(s^{k, l}, b^{k, l}\right)$ a trial point.

S2. Find a corresponding $\left(r^{k, l}, a^{k, l}\right) \in A$ such that $r^{k, l}+\gamma a^{k, l}=s^{k, l}-\gamma b^{k, l}$. It follows immediately from (4.11) that $\left(u^{k, l}, v^{k, l}\right) \stackrel{\text { def }}{=}\left(r^{k, l}+\gamma b^{k, l}, s^{k, l}-r^{k, l}\right) \in S_{\gamma, A, B}$.

S3. Test whether the $\left(u^{k, l}, v^{k, l}\right)$ satisfies the conditions on $\left(u^{k}, v^{k}\right)$ specified in 4.9). If not, increment $l \leftarrow l+1$ and return to step S1 to execute additional steps of the $B$-procedure to produce a more accurate trial point. Otherwise, accept $\left(u^{k}, v^{k}\right)=\left(u^{k, l}, v^{k, l}\right) \in S_{\gamma, A, B}$ as a pair satisfying 4.9).

S4. Once we have accepted $\left(u^{k}, v^{k}\right)=\left(u^{k, l}, v^{k, l}\right)$, set $z^{k+1}=z^{k}-v^{k}$, that is, 4.10) with $\lambda_{k}=1$.

We now make step S3 more concrete: substituting $\left(u^{k, l}, v^{k, l}\right) \stackrel{\text { def }}{=}\left(r^{k, l}+\gamma b^{k, l}, s^{k, l}-r^{k, l}\right)$ for $\left(u^{k}, v^{k}\right)$ in (4.9), along with and $\lambda_{k} \equiv 1$ from (4.12), we obtain the condition

$$
\left\|r^{k, l}+\gamma b^{k, l}+s^{k, l}-r^{k, l}-z^{k}\right\| \leq \sigma\left\|r^{k, l}+\gamma b^{k, l}-z^{k}\right\| .
$$

Canceling $r^{k, l}$ from the left-hand side, we obtain

$$
\begin{equation*}
\left\|s^{k, l}+\gamma b^{k, l}-z^{k}\right\| \leq \sigma\left\|r^{k, l}+\gamma b^{k, l}-z^{k}\right\| \tag{4.15}
\end{equation*}
$$

Next, we consider the extragradient update $z^{k+1}=z^{k}-v^{k}$ in step S4 of the above sequence. Given some $k$, let us suppose that we have $z^{k}=r^{k}+\gamma b^{k}$, as is the case in step DR1 of the DR splitting method. If we wish this same relation to hold for $k+1$ as well as $k$, the update $z^{k+1}=z^{k}-v^{k}$ takes the form

$$
r^{k+1}+\gamma b^{k+1}=r^{k}+\gamma b^{k}-v^{k}=r^{k}+\gamma b^{k}-\left(s^{k, l}-r^{k, l}\right)
$$

If we take $r^{k+1}=r^{k, l}$, the above equation becomes

$$
r^{k+1}+\gamma b^{k+1}=r^{k}+\gamma b^{k}-\left(s^{k, l}-r^{k+1}\right)
$$

from which we may cancel $r^{k+1}$ from both sides to yield

$$
\gamma b^{k+1}=r^{k}+\gamma b^{k}-s^{k, l} \quad \Leftrightarrow \quad b^{k+1}=b^{k}+\frac{1}{\gamma}\left(r^{k}-s^{k, l}\right) .
$$

Letting $s^{k+1}=s^{k, l}$ for completeness, one possible way to implement the update $z^{k+1}=z^{k}-v^{k}$ is therefore

$$
\begin{equation*}
s^{k+1}=s^{k, l} \quad r^{k+1}=r^{k, l} \quad b^{k+1}=b^{k}+\frac{1}{\gamma}\left(r^{k}-s^{k, l}\right) . \tag{4.16}
\end{equation*}
$$

If we update the iterates in this manner and start with an arbitrary $z^{0}=r^{0}+\gamma b^{0}$, then by induction we maintain $z^{k}=r^{k}+\gamma b^{k}$ for all $k$. Substituting $z^{k}=r^{k}+\gamma b^{k}$ into (4.15), we obtain

$$
\begin{equation*}
\left\|s^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\| \leq \sigma\left\|r^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\| . \tag{4.17}
\end{equation*}
$$

Similarly substituting $z^{k}=r^{k}+\gamma b^{k}$ throughout steps S1/S4 and in (4.17) above, we arrive at the following algorithm:

```
Algorithm 4.3.1 A partially inexact primal Douglas-Rachford splitting algorithm
    initialization: Choose \(\gamma>0, \sigma \in[0,1)\). Initialize \(s^{0}, b^{0}, r^{0} \in \mathbb{R}^{n}\) arbitrarily
    for \(k=0,1,2, \ldots\) do
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(\left(s^{k, l}, b^{k, l}\right) \in B\) and \(s^{k, l}+\gamma b^{k, l} \approx r^{k}+\gamma b^{k}\) by setting
            \(\left(s^{k, l}, b^{k, l}\right)=\mathcal{B}\left(r^{k}+\gamma b^{k}, \gamma, s^{k}, b^{k}, l\right)\) (thus incrementally executing a step of the
            \(B\)-procedure)
            Exactly find \(\left(r^{k, l}, a^{k, l}\right) \in A\) such that \(r^{k, l}+\gamma a^{k, l}=s^{k, l}-\gamma b^{k, l}\)
        until \(\left\|s^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\| \leq \sigma\left\|r^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\|\)
        \(s^{k+1}=s^{k, l}\)
        \(r^{k+1}=r^{k, l}\)
        \(b^{k+1}=b^{k}-\frac{1}{\gamma}\left(s^{k+1}-r^{k}\right)\)
    end for
```

We summarize the convergence properties of this algorithm as follows.

Proposition 11. Suppose that the inclusion $0 \in A(x)+B(x)$ has a solution. Then there are two possible execution sequences for Algorithm 4.3.1:

1. The outer loop (over $k$ ) executes an infinite number of times, with each inner loop (over l) terminating in a finite number of iterations. Then $\left\{s^{k}\right\}$ and $\left\{r^{k}\right\}$ both converge to some $x^{*}$ for which $0 \in A\left(x^{*}\right)+B\left(x^{*}\right)$, and $b^{k}$ converges to some $b^{*} \in B\left(x^{*}\right)$ such that $-b^{*} \in A\left(x^{*}\right)$.
2. The outer loop executes only a finite number of times, ending with $k=\bar{k}$, with the last invocation of the inner loop executing indefinitely. In this case, $\lim _{l \rightarrow \infty} s^{\bar{k}, l}=\lim _{l \rightarrow \infty} r^{\bar{k}, l}=x^{*}$ for some $x^{*}$ for which $0 \in A\left(x^{*}\right)+B\left(x^{*}\right)$, while $\lim _{l \rightarrow \infty} b^{\bar{k}, l}=b^{*}$ for some $b^{*} \in B\left(x^{*}\right)$ such that $-b^{*} \in A\left(x^{*}\right)$ and $\lim _{l \rightarrow \infty} a^{\bar{k}, l}=$ $-b^{*}$.

Proof. We begin by considering the second case, in which the inner loop fails to terminate for some $k=\bar{k}$. By the assumed properties of the $B$-procedure modeled by the mapping $\mathcal{B}$, we have that $\left\{\left(s^{\bar{k}, l}, b^{\bar{k}, l}\right)\right\}_{l=1}^{\infty} \in B$ converges to some limit $\left(x^{*}, b^{*}\right)$ with $x^{*}+\gamma b^{*}=z^{\bar{k}}=r^{\bar{k}}+\gamma b^{\bar{k}}$ as $l \rightarrow \infty$. By the closedness properties of maximal monotone operators, we also have $b^{*} \in B\left(x^{*}\right)$. From the construction of the points $\left(r^{\bar{k}, l}, a^{\bar{k}, l}\right)$, we have for all $l$ that

$$
\left(r^{\bar{k}, l}, a^{\bar{k}, l}\right) \in A \quad a^{\bar{k}, l}=\frac{1}{\gamma}\left(s^{\bar{k}, l}-r^{\bar{k}, l}\right)-b^{\bar{k}, l} \quad r^{\bar{k}, l}=J_{\gamma A}\left(s^{\bar{k}, l}-\gamma b^{\bar{k}, l}\right),
$$

where $J_{\gamma A}$ denotes the resolvent map of the maximal monotone operator $A$. Taking limits and using that resolvent maps are continuous, we obtain that ( $\left.r^{\bar{k}, l}, a^{\bar{k}, l}\right)$ converges to the limit $\left(r^{*}, a^{*}\right)=\left(J_{\gamma A}\left(x^{*}-\gamma b^{*}\right), \frac{1}{\gamma}\left(x^{*}-r^{*}\right)-b^{*}\right)$, and from the closedness of the maximal monotone operator $A$, we also have $\left(r^{*}, a^{*}\right) \in A$. Now consider the innerloop termination condition $\left\|s^{\bar{k}, l}+\gamma b^{\bar{k}, l}-\left(r^{k}+\gamma b^{k}\right)\right\| \leq \sigma\left\|r^{\bar{r}, l}+\gamma b^{\bar{k}, l}-\left(r^{k}+\gamma b^{k}\right)\right\|$ for iteration $\bar{k}$. By the assumed properties of the $B$-procedure, its left-hand side converges to zero as $l \rightarrow \infty$. Its right-hand side is nonnegative, so in order for the inner loop to execute an infinite number of times, the right-hand side must also converge to
zero. Taking limits, this convergence means that $r^{*}+\gamma b^{*}=r^{\bar{k}}+\gamma b^{\bar{k}}=z^{\bar{k}}=x^{*}+\gamma b^{*}$, where the last equality was established above. Canceling $\gamma b^{*}$ from both sides, we conclude that $r^{*}=x^{*}$ and consequently that $a^{*}=\frac{1}{\gamma}\left(x^{*}-r^{*}\right)-b^{*}=-b^{*}$. Thus, we have $b^{*} \in B\left(x^{*}\right)$ and $-b^{*}=a^{*} \in A\left(r^{*}\right)=A\left(x^{*}\right)$. Therefore, $A\left(x^{*}\right)+B\left(x^{*}\right) \ni-b^{*}+b^{*}=0$ and all the claims in the second case of the proposition have been established.

To complete the proof, we must consider the first case, in which the inner loop always terminates and the outer loop executes an infinite number of times. In this situation, let $l(k)$ be the index of inner iteration that first meets the inner-loop termination condition for outer iteration $k$. According to the algorithm's update rule, we have $s^{k+1}=s^{k, l(k)}$ and $r^{k+1}=r^{k, l(k)}$ for all $k$. In view of the derivation immediately preceding Algorithm 4.3.1, we have in this case that $\left\{z^{k}\right\}=\left\{r^{k}+\gamma b^{k}\right\}$ is identical to the sequence generated by Algorithm4.2.1 with $T=S_{\gamma, A, B}$ and $\lambda_{k} \equiv 1$. Proposition 9 then implies that $z^{k} \rightarrow z^{*}$ such that $0 \in S_{\gamma, A, B}\left(z^{*}\right)$. Next, in view of step S2 above, we note that the point $\left(u^{k}, v^{k}\right) \in T$ of Algorithm 4.2.1 may be expressed as

$$
\left(r^{k, l(k)}+\gamma b^{k, l(k)}, s^{k, l(k)}-r^{k, l(k)}\right) \in S_{\gamma, A, B} .
$$

Proposition 9 asserts that $\left(u^{k}, v^{k}\right) \rightarrow\left(z^{*}, 0\right)$, so we conclude that

$$
\lim _{k \rightarrow \infty} r^{k, l(k)}+\gamma b^{k, l(k)}=z^{*} \quad \lim _{k \rightarrow \infty} s^{k, l(k)}-r^{k, l(k)}=0
$$

We next observe that

$$
s^{k, l(k)}+\gamma b^{k, l(k)}=s^{k, l(k)}-r^{k, l(k)}+r^{k, l(k)}+\gamma b^{k, l(k)},
$$

and therefore that

$$
\lim _{k \rightarrow \infty} s^{k, l(k)}+\gamma b^{k, l(k)}=\lim _{k \rightarrow \infty} s^{k, l(k)}-r^{k, l(k)}+\lim _{k \rightarrow \infty} r^{k, l(k)}+\gamma b^{k, l(k)}=0+z^{*}=z^{*}
$$

By construction, $b^{k, l(k)} \in B\left(s^{k, l(k)}\right)$ for all $k$ so $s^{k, l(k)} \in J_{\gamma B}\left(s^{k, l(k)}+\gamma b^{k, l(k)}\right)$, where $J_{\gamma B}$ denotes the resolvent map of the maximal monotone operator $B$. Once again using that resolvent maps are continuous, we obtain that

$$
\lim _{k \rightarrow \infty} s^{k, l(k)}=J_{\gamma B}\left(\lim _{k \rightarrow \infty} s^{k, l(k)}+\gamma b^{k, l(k)}\right)=J_{\gamma B}\left(z^{*}\right)
$$

where the first limit must exist. Define $x^{*}=\lim _{k \rightarrow \infty} s^{k, l(k)}$. Since

$$
b^{k, l(k)}=\frac{1}{\gamma}\left(\left(s^{k, l(k)}+\gamma b^{k, l(k)}\right)-s^{k, l(k)}\right) \rightarrow \frac{1}{\gamma}\left(z^{*}-x^{*}\right),
$$

we ascertain that $b^{*}=\lim _{k \rightarrow \infty} b^{k, l(k)}$ exists and $z^{*}=x^{*}+\gamma b^{*}$. By the closure property of maximal monotone operators, we may take the limit in the inclusion $b^{k, l(k)} \in B\left(s^{k, l(k)}\right)$ to obtain $b^{*} \in B\left(x^{*}\right)$. Since $s^{k, l(k)}-r^{k, l(k)} \rightarrow 0$, we deduce that $\lim _{k \rightarrow \infty} r^{k, l(k)}=x^{*}$ and also, using the equation $r^{k, l}+\gamma a^{k, l}=s^{k, l}-\gamma b^{k, l}$ from the algorithm, that $\lim _{k \rightarrow \infty} a^{k, l(k)}=\lim _{k \rightarrow \infty}\left(\frac{1}{\gamma}\left(s^{k, l(k)}-r^{k, l(k)}\right)-b^{k, l(k)}\right)=-b^{*}$. Since we have $a^{k, l(k)} \in A\left(y^{k, l(k)}\right)$ by construction, we may take the limit to obtain $-b^{*} \in A\left(x^{*}\right)$. In conclusion, we have $b^{*} \in B\left(x^{*}\right),-b^{*} \in A\left(x^{*}\right)$ and therefore $0 \in A\left(x^{*}\right)+B\left(x^{*}\right)$, and all claims in the first case of the proposition are established.

To summarize, the proposition states that Algorithm 4.3.1 must converge to a solution to $A(x)+B(x) \ni 0$ in one of two ways, either through convergence of its outer loop with finite termination of each inner loop, or by finite termination of its outer loop combined with convergence of the last instance of its inner loop. The former case follows (with some technical manipulation) from convergence of the relative-error proximal point algorithm on $S_{\gamma, A, B}$, while the latter involves some additional analysis.

### 4.4 Deriving a partially inexact ADMM from the partially inexact DR splitting method

We now derive an inexact version of the ADMM from the Algorithm 4.3.1. The standard dual formulation of the problem (1.4) is

$$
\begin{equation*}
\min _{p \in \mathbb{R}^{m}} f^{*}\left(-M^{\top} p\right)+g^{*}(p) \tag{4.18}
\end{equation*}
$$

where $f^{*}$ and $g^{*}$ are the convex conjugate of $f$ and $g$, respectively. In [37], Gabay showed that ADMM can be derived by applying the Douglas-Rachford splitting method to 4.18 with $A=\partial\left[f^{*} \circ\left(-M^{\top}\right)\right]$ and $B=\partial g^{*}$. Unfortunately, these
choices of $A$ and $B$ are inconvenient for Algorithm 4.3.1, because verifying the condition $b \in B(s)$ would require an exact minimization involving the function $g$, precisely the kind of operation one is trying to avoid.

Instead, we consider the primal splitting approach in which one lets let $A=$ $\partial[g \circ M]$ and $B=\partial f$. As shown in [26, Section 3.5.6], applying Douglas-Rachford splitting to this choice of $A$ and $B$ results in the algorithm

$$
\begin{align*}
& s^{k+1}=\underset{s}{\arg \min }\left\{f(s)+\frac{1}{2 \gamma}\left\|s-\left(r^{k}+\gamma b^{k}\right)\right\|^{2}\right\}  \tag{4.19}\\
& r^{k+1}=\underset{r}{\arg \min }\left\{g(M r)+\frac{1}{2 \gamma}\left\|r-\left(s^{k+1}-\gamma b^{k}\right)\right\|^{2}\right\}  \tag{4.20}\\
& b^{k+1}=b^{k}+\frac{1}{\gamma}\left(r^{k+1}-s^{k+1}\right) . \tag{4.21}
\end{align*}
$$

This ADMM-like method is appropriate when the composition of $g$ and $M$ is convenient to work with, so that the minimization (4.20) is not too hard to perform. Since redefining $g \leftarrow g \circ M$ and then $M \leftarrow I$ in problem (1.4) results in exactly the same algorithm when applying (4.19)-(4.21), we may without loss of generality take $M=I$. Furthermore, when when $M=I$, Proposition 3.43 of [26] shows that (4.19)-4.21) is identical to the ADMM if one sets $c=1 / \gamma$; these ideas are developed somewhat further in 83].

Fixing $M=I$, we now consider applying the partially exact Douglas-Rachford splitting method of Algorithm 4.3.1 with this same choice of $A=\partial[g \circ M]=\partial g$ and $B=\partial f$. We now develop an analysis similar to [26, Proposition 3.43], but in the context of Algorithm 4.3.1: this algorithm requires that resolvent operation in step S 2 be carried out exactly, while the resolvent calculation in step 51 may be approximate. Therefore, the subproblems associated with $\mathrm{S1}$ and S2 are respectively

$$
\begin{align*}
& s^{k, l} \approx \underset{s}{\arg \min }\left\{f(s)+\frac{1}{2 \gamma}\left\|s-\left(r^{k}+\gamma b^{k}\right)\right\|^{2}\right\}  \tag{4.22}\\
& r^{k, l}=\underset{r}{\arg \min }\left\{g(r)+\frac{1}{2 \gamma}\left\|r-\left(s^{k, l}-\gamma b^{k, l}\right)\right\|^{2}\right\} . \tag{4.23}
\end{align*}
$$

Expanding squares and dropping constant terms from the minimands in these calculations, we equivalently obtain

$$
\begin{align*}
& s^{k, l} \approx \underset{s}{\arg \min }\left\{f(s)-\left\langle b^{k}, s\right\rangle+\frac{1}{2 \gamma}\left\|s-r^{k}\right\|^{2}\right\}  \tag{4.24}\\
& r^{k, l}=\underset{r}{\arg \min }\left\{g(r)+\left\langle b^{k, l}, r\right\rangle+\frac{1}{2 \gamma}\left\|r-s^{k, l}\right\|^{2}\right\} . \tag{4.25}
\end{align*}
$$

Next, we define some parallel notation for Algorithm 4.3.1 by letting $p^{k}=-b^{k}$, $p^{k, l}=-b^{k, l}, x^{k}=s^{k}, x^{k, l}=s^{k, l}, z^{k}=r^{k}$, and $z^{k, l}=r^{k, l}$. Using this alternative notation and letting $c=1 / \gamma$, the calculations 4.22)-4.23 may be expressed as

$$
\begin{align*}
x^{k, l} & \approx \underset{x}{\arg \min }\left\{f(x)+\left\langle p^{k}, x\right\rangle+\frac{c}{2}\left\|x-z^{k}\right\|^{2}\right\}  \tag{4.26}\\
z^{k, l} & =\underset{z}{\arg \min }\left\{g(z)-\left\langle p^{k, l}, z\right\rangle+\frac{c}{2}\left\|z-x^{k, l}\right\|^{2}\right\} . \tag{4.27}
\end{align*}
$$

We now state the precise form of our proposed algorithm, in which we make the meaning of the " $\approx$ " in 4.26 more specific. The following convergence proof is based on relating the " $\mathcal{F}$-procedure" assumed to exist in Assumption 2 to the more abstract $\mathcal{B}$-procedure of Assumption 10 in the case $B=\partial f$ and $\gamma=1 / c$.

## Algorithm 4.4.1 ADMM variant derived from partially exact Douglas-Rachford splitting

```
    initialization: Choose \(c>0, \sigma \in[0,1)\). Initialize \(x^{0}, p^{0}, z^{0}\).
```

    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, x\right\rangle+\frac{c}{2}\left\|x-z^{k}\right\|^{2}\right\}\) by
                \(\operatorname{taking}\left(x^{k, l}, y^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
            \(p^{k, l}=p^{k}+c\left(x^{k, l}-z^{k}\right)-y_{1}^{k, l}\)
            \(z^{k, l}=\arg \min _{z}\left\{g(z)-\left\langle p^{k, l}, z\right\rangle+\frac{c}{2}\left\|x^{k, l}-z\right\|^{2}\right\}\)
        until \(\left\|y_{1}^{k, l}\right\| \leq \sigma\left\|p^{k, l}-p^{k}-c\left(z^{k, l}-z^{k}\right)\right\|\)
        \(x^{k+1}=x^{k, l}\)
        \(z^{k+1}=z^{k, l}\)
        \(p^{k+1}=p^{k}+c\left(x^{k, l}-z^{k}\right)\)
    until Overall convergence

Proposition 12. Suppose that there exists $x^{*}$ such that $0 \in \partial f\left(x^{*}\right)+\partial g\left(x^{*}\right)$ and $\mathcal{F}$ meets the conditions in Assumption 2 for $M=I$. Then there are two possible execution sequences for Algorithm 4.4.1:

1. The outer loop (over $k$ ) executes an infinite number of times, with each invocation of the inner loop (over l) terminating in a finite number of iterations. Then $\left\{x^{k}\right\}$ and $\left\{z^{k}\right\}$ both converge to some $x^{*}$ for which $0 \in \partial g\left(x^{*}\right)+\partial f\left(x^{*}\right)$, and $\left\{p^{k}\right\}$ converges to some $p^{*} \in \partial g\left(x^{*}\right)$ such that $-p^{*} \in \partial f\left(x^{*}\right)$.
2. The outer loop executes only a finite number of times, ending with $k=\bar{k}$, with the last invocation of the inner loop executing indefinitely. In this case, we have $\lim _{l \rightarrow \infty} x^{\bar{k}, l}=\lim _{l \rightarrow \infty} z^{\bar{k}, l}=x^{*}$ for some $x^{*}$ such that $0 \in \partial f\left(x^{*}\right)+\partial g\left(x^{*}\right)$, while $\lim _{l \rightarrow \infty} p^{\bar{k}, l}=p^{*}$ for some $p^{*} \in \partial g\left(x^{*}\right)$ such that $-p^{*} \in \partial f\left(x^{*}\right)$.

Proof. We claim that $p^{k}=-b^{k}, x^{k}=s^{k}$, and $z^{k}=r^{k}$ for all $k \geq 0$ and $p^{k, l}=-b^{k, l}$, $x^{k, l}=s^{k, l}$, and $z^{k, l}=r^{k, l}$ for all $k \geq 0$ and $l \geq 1$, for Algorithm 4.3.1 as applied
to $A=\partial g, B=\partial f, \gamma=1 / c$, and a valid form of the $\mathcal{B}$-procedure hypothesized in Assumption 10.

To establish the claim, we start by setting $\gamma=1 / c, b^{0}=-p^{0}$ and $r^{0}=z^{0}$ and then proceed by induction. Take any $k \geq 0$ and assume that $p^{k}=-b^{k}$ and $r^{k}=z^{k}$. First, consider the corresponding inner loop over $l$. Substituting $M=I$ into Assumption 2 and using [69, Theorem 23.8], we obtain that for each $l$ in the inner loop, we have

$$
\begin{aligned}
& y_{1}^{k, l} \\
\Leftrightarrow & \left.-p^{k}-c\left(x^{k, l}-z^{k}\right)+y^{k, l}\right)+p^{k}+c\left(x^{k, l}-z^{k}\right) \\
\Leftrightarrow & -p^{k, l} \in \partial f\left(x^{k, l}\right) \\
\Leftrightarrow &
\end{aligned}
$$

Therefore, if we set $s^{k, l}=x^{k, l}$ and $b^{k, l}=-p^{k, l}$, we have $\left(s^{k, l}, b^{k, l}\right) \in \partial f=B$. Furthermore,

$$
\begin{aligned}
s^{k, l}+\gamma b^{k, l} & =x^{k, l}-\frac{1}{c} p^{k, l} \\
& =x^{k, l}-\frac{1}{c}\left(p^{k}+c\left(x^{k, l}-z^{k}\right)-y_{1}^{k, l}\right) \\
& =-\frac{1}{c} p^{k}+z^{k}+\frac{1}{c} y_{1}^{k, l} \\
& =\gamma b^{k}+r^{k}+\frac{1}{c} y_{1}^{k, l} .
\end{aligned}
$$

Since $\lim _{l \rightarrow \infty} y_{1}^{k, l}=0$, it follows that $\lim _{l \rightarrow \infty} s^{k, l}+\gamma b=r^{k}+\gamma b^{k}$. This means that the procedure of taking $\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)$ followed by $p^{k, l}=p^{k}+c\left(x^{k, l}-z^{k}\right)-$ $y_{1}^{k, l}$ has exactly the same properties hypothesized for the $\mathcal{B}$-procedure in Assumption 10, for $u=r^{k}+\gamma b^{k}=z^{k}-(1 / c) p^{k}$ (we may take the starting point arguments for $\mathcal{B}$ to be $\bar{s}=s^{k}=x^{k}$ and $\left.\bar{b}=b^{k}=-p^{k}\right)$.

Next, from $z^{k, l}=\arg \min _{z}\left\{g(z)-\left\langle p^{k, l}, z\right\rangle+\frac{c}{2}\left\|x^{k, l}-z\right\|^{2}\right\}$ and 69. Theorem 23.8], we must have $0 \in \partial g\left(z^{k, l}\right)-p^{k, l}+c\left(z^{k, l}-x^{k, l}\right)$, and hence $p^{k, l}+c\left(x^{k, l}-z^{k, l}\right) \in \partial g\left(z^{k, l}\right)$. Therefore, we take $a^{k, l}=p^{k, l}+c\left(x^{k, l}-z^{k, l}\right)$ and $r^{k, l}=z^{k, l}$, and then have $\left(r^{k, l}, a^{k, l}\right) \in A$ and

$$
r^{k, l}+\gamma a^{k, l}=z^{k, l}+\frac{1}{c}\left(p^{k, l}+c\left(x^{k, l}-z^{k, l}\right)\right)=x^{k, l}+\frac{1}{c} p^{k, l}=s^{k, l}-\gamma b^{k, l} .
$$

Therefore, we have $\left(r^{k, l}, a^{k, l}\right) \in A$ and $r^{k, l}+\gamma a^{k, l}=s^{k, l}-\gamma b^{k, l}$, which are the unique determining conditions for $\left(r^{k, l}, a^{k, l}\right)$ in Algorithm 4.3.1. Thus all the steps within the inner loop of Algorithm 4.4.1 are equivalent to the steps in the inner loop of Algorithm 4.3.1. We next turn to the termination condition for the inner loop of Algorithm 4.3.1. $\left\|s^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\| \leq \sigma\left\|r^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)\right\|$. Substituting

$$
s^{k, l}=x^{k, l} \quad \gamma=1 / c \quad b^{k, l}=-p^{k, l}=-p^{k}-c\left(x^{k, l}-z^{k}\right)+y_{1}^{k, l} \quad r^{k}=z^{k} \quad b^{k}=-p^{k}
$$

into the expression within the norm on the left-hand side of this condition, we obtain

$$
s^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)=x^{k, l}+\frac{1}{c}\left(-p^{k}-c\left(x^{k, l}-z^{k}\right)+y_{1}^{k, l}\right)-z^{k}-\frac{1}{c}\left(-p^{k}\right)=\frac{1}{c} y_{1}^{k, l} .
$$

Performing similar substitutions in the expression within the norm on the right-hand side of Algorithm 4.3.1 s inner-loop termination condition, we obtain

$$
r^{k, l}+\gamma b^{k, l}-\left(r^{k}+\gamma b^{k}\right)=z^{k, l}-\frac{1}{c} p^{k, l}-z^{k}+\frac{1}{c} p^{k}=z^{k, l}-z^{k}-\frac{1}{c}\left(p^{k, l}-p^{k}\right)
$$

Thus we obtain the following inner-loop termination condition equivalent to that of Algorithm 4.3.1:

$$
\left\|\frac{1}{c} y_{1}^{k, l}\right\| \leq\left\|z^{k, l}-z^{k}-\frac{1}{c}\left(p^{k, l}-p^{k}\right)\right\|
$$

Multiplying through by $c$, we obtain exactly the same inner-loop termination condition as in Algorithm 4.4.1. Therefore, if one enters iteration $k$ with $p^{k}=-b^{k}$ and $r^{k}=z^{k}$, Algorithm 4.4.1 will execute exactly the same number of inner-loop iterations as Algorithm 4.3.1 with the $\mathcal{B}$-procedure constructed as described above.

Once the inner loop has terminated, Algorithm4.4.1 performs the updates $x^{k+1}=$ $x^{k, l}, z^{k+1}=z^{k, l}$, and $p^{k+1}=p^{k}+c\left(x^{k, l}-z^{k}\right)$. If we let $s^{k+1}=x^{k+1}$, then we have $s^{k+1}=x^{k, l}=s^{k, l}$, the same update as performed by Algorithm 4.3.1. Similarly, setting $r^{k+1}=z^{k+1}=z^{k, l}=r^{k, l}$ yields the same value of $r^{k+1}$ as in Algorithm 4.3.1. Finally, if we let $b^{k+1}=-p^{k+1}$, we have

$$
b^{k+1}=-\left(p^{k}+c\left(x^{k, l}-z^{k}\right)\right)=-p^{k}-\frac{1}{\gamma}\left(s^{k, l}-r^{k}\right)=b^{k}-\frac{1}{\gamma}\left(s^{k, l}-r^{k}\right),
$$

which is exactly the same value of $b^{k+1}$ computed by Algorithm 4.3.1. Thus the induction is complete and the claim is verified.

The claim having been established, the conclusions of the Proposition now follow directly from Proposition 11 .

## Chapter 5 <br> Approximate ADMM Algorithms Derived from Lagrangian Splitting

We now develop approximate ADMM algorithms by modifying the Lagrangian splitting analysis pioneered in [34]. While the derivations and algorithms are more complicated, the resulting methods are not subject to the restrictive assumptions applying to the methods derived from operator splitting. In contrast to the methods derived in Chapter 4, neither the absolute-error algorithm nor the relative-error algorithm we derive here imposes any restriction on $M$, and the relative-error algorithm allows both the $x$ and $z$ minimizations to be approximate.

### 5.1 A parametric conjugate duality framework

We now introduce a parametric conjugate duality framework specializing the one described in [69, Chapters 28-30] and [70] to the case of (P). All results, except those proved explicitly here, follow immediately from results in those references. First, we define the functions $F_{1}, F_{2}: \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow \mathbb{R} \cup\{+\infty\}$ as follows:

$$
\begin{align*}
& F_{1}\left(x, z, u_{1}\right)= \begin{cases}f(x), & \text { if } u_{1}+M x=0 \\
+\infty, & \text { otherwise }\end{cases}  \tag{5.1}\\
& F_{2}\left(x, z, u_{2}\right)= \begin{cases}g(z), & \text { if } u_{2}-z=0 \\
+\infty, & \text { otherwise }\end{cases} \tag{5.2}
\end{align*}
$$

If $f$ and $g$ are closed and convex, it is easily seen that $F_{1}$ and $F_{2}$ are closed and convex. We next define the parametric objective function $F: \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow(-\infty,+\infty]$
of $(\overline{\mathrm{P}})$ to be the infimal convolution $\left[69\right.$, page 34] of $F_{1}$ and $F_{2}$ with respect to the last argument, that is

$$
\begin{align*}
F(x, z, u) & =\inf _{\substack{u_{1}, u_{2} \\
u_{1}+u_{2}=u}}\left\{F_{1}\left(x, z, u_{1}\right)+F_{2}\left(x, z, u_{2}\right)\right\} \\
& = \begin{cases}f(x)+g(z), & \text { if } M x-z+u=0 \\
+\infty, & \text { otherwise }\end{cases} \tag{5.3}
\end{align*}
$$

where $u \in \mathbb{R}^{m}$ represents a perturbation of the constraints $M x=z$. It is also easily seen that $F$ is closed and convex [69, Theorem 5.4]. Within this framework, the original problem $(\mathbb{P})$ is equivalent to the primal problem

$$
\begin{equation*}
\min _{x \in \mathbb{R}^{n}} F(x, z, 0) \tag{5.4}
\end{equation*}
$$

One obtains the Lagrangian $L: \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow[-\infty,+\infty]$ of (5.4) by taking the concave conjugate [69, p. 111] of $F$ with respect to the perturbation argument $u$ :

$$
\begin{align*}
L(x, z, p) & =\inf _{u \in \mathbb{R}^{m}}\{F(x, z, u)-\langle p, u\rangle\} \\
& =\inf _{u=z-M x}\{f(x)+g(z)-\langle p, u\rangle\} \\
& =f(x)+g(z)+\langle p, M x-z\rangle \tag{5.5}
\end{align*}
$$

This derivation coincides with the usual Lagrangian for (P). Lagrangians $L(x, z, p)$ derived in this manner are convex with respect to $(x, z)$ and concave with respect to $p$ (and in this particular case, $L$ is linear with respect to $p$ ). Let $\partial L$ denote its convex-concave subgradient map, that is, $\partial L(x, z, p)$ is the set consisting of all $\left(y_{1}, y_{2}, u\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ such that

$$
\begin{aligned}
L\left(x^{\prime}, z^{\prime}, p\right) & \geq L(x, z, p)+\left\langle y_{1}, x^{\prime}-x\right\rangle+\left\langle y_{2}, z^{\prime}-z\right\rangle & & \forall\left(x^{\prime}, z^{\prime}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \\
L\left(x, z, p^{\prime}\right) & \leq L(x, z, p)-\left\langle u, p^{\prime}-p\right\rangle & & \forall p^{\prime} \in \mathbb{R}^{m} .
\end{aligned}
$$

For any $p \in \mathbb{R}^{m}$, the function $(x, z) \mapsto L(x, z, p)=(f(x)+\langle p, M x\rangle)+(g(z)-\langle p, z\rangle)$ is separable with respect to $x$ and $z$, so we obtain

$$
\begin{align*}
\partial L(x, z, p) & =\partial_{(x, z)} L(x, z, p) \times \partial_{p}(-L(x, z, p)) \\
& =\partial_{x} L(x, z, p) \times \partial_{z} L(x, z, p) \times \partial_{p}(-L(x, z, p)) \\
& =\left\{\partial f(x)+M^{\top} p\right\} \times\{\partial g(z)-p\} \times\{z-M x\} . \tag{5.6}
\end{align*}
$$

The following result is now immediate:

Lemma 13. A point $\left(x^{*}, z^{*}, p^{*}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{n}$ is a $K K T$ point as defined in Assumption 1 if and only if it is a saddle point of $L$, that is, $(0,0,0) \in \partial L\left(x^{*}, z^{*}, p^{*}\right)$.

We obtain $L_{1}, L_{2}: \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow(0, \infty]$, the Lagrangian functions corresponding to (5.1) and (5.2), respectively, by taking their concave conjugates with respect to the perturbation arguments $u_{1}$ and $u_{2}$ :

$$
\begin{aligned}
& L_{1}(x, z, p)=\inf _{u_{1} \in \mathbb{R}^{m}}\left\{F_{1}\left(x, z, u_{1}\right)-\left\langle p, u_{1}\right\rangle\right\}=f(x)+\langle p, M x\rangle \\
& L_{2}(x, z, p)=\inf _{u_{2} \in \mathbb{R}^{m}}\left\{F_{2}\left(x, z, u_{2}\right)-\left\langle p, u_{2}\right\rangle\right\}=g(z)-\langle p, z\rangle
\end{aligned}
$$

From this derivation, it is immediate that $L_{1}, L_{2}$ are concave in $p$ and convex in $x$ and $z$. Furthermore, we observe that $L=L_{1}+L_{2}$. Letting $\partial L_{1}$ and $\partial L_{2}$ denote the respective convex-concave subdifferential maps of these two functions, we obtain

$$
\begin{align*}
& \partial L_{1}(x, z, p)=\left\{\partial f(x)+M^{\top} p\right\} \times\{0\} \times\{-M x\}  \tag{5.7}\\
& \partial L_{2}(x, z, p)=\{0\} \times\{\partial g(z)-p\} \times\{z\}, \tag{5.8}
\end{align*}
$$

and we have $\partial L_{1}+\partial L_{2}=\partial L$. The point-to-set maps $\partial L_{1}, \partial L_{2}$ are the respective partial inverses of the subgradient maps of the closed convex functions $F_{1}$ and $F_{2}$, so they are maximal monotone operators. We call this technique Lagrangian splitting: we have expressed the maximal monotone operator $\partial L$ as the sum of two simpler maximal monotone operators $\partial L_{1}$ and $\partial L_{2}$. Furthermore, $z$ is only a nominal argument whose choice has no effect on the value of $L_{1}(x, z, p)$, and similarly $x$ is only a nominal, ignored argument to $L_{2}(x, z, p)$.

The parametric dual function $Q: \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow[-\infty,+\infty]$ of $(\mathrm{P})$ can be defined in two equivalent ways: either as the concave conjugate of $F$ jointly with respect to $(x, z)$ and $u$, or as the concave conjugate of $L$ with respect to $x$ and $z$. Proceeding in the latter manner, we obtain

$$
\begin{align*}
Q\left(y_{1}, y_{2}, p\right) & =\inf _{x \in \mathbb{R}^{n}, z \in \mathbb{R}^{m}}\left\{L(x, z, p)-\left\langle y_{1}, x\right\rangle-\left\langle y_{2}, z\right\rangle\right\} \\
& =\inf _{x \in \mathbb{R}^{n}, z \in \mathbb{R}^{m}}\left\{f(x)+g(z)+\langle p, M x-z\rangle-\left\langle y_{1}, x\right\rangle-\left\langle y_{2}, z\right\rangle\right\} \\
& =\inf _{x \in \mathbb{R}^{n}, z \in \mathbb{R}^{m}}\left\{f(x)+\langle p, M x\rangle-\left\langle y_{1}, x\right\rangle+g(z)-\langle p, z\rangle-\left\langle y_{2}, z\right\rangle\right\} \\
& =\inf _{x \in \mathbb{R}^{n}}\left\{f(x)+\langle p, M x\rangle-\left\langle y_{1}, x\right\rangle\right\}+\inf _{z \in \mathbb{R}^{m}}\left\{g(z)-\langle p, z\rangle-\left\langle y_{2}, z\right\rangle\right\} \\
& =\inf _{x \in \mathbb{R}^{n}}\left\{L_{1}(x, 0, p)-\left\langle y_{1}, x\right\rangle\right\}+\inf _{z \in \mathbb{R}^{m}}\left\{L_{2}(0, z, p)-\left\langle y_{2}, z\right\rangle\right\} \tag{5.9}
\end{align*}
$$

Here, we arbitrarily use 0 as the $z$ argument to $L_{1}$, since $z$ is only a nominal argument that does not affect the value of the function. Similarly, $x$ does not affect the value of $L_{2}$, so we arbitrarily use 0 for its $x$ argument. Using "*" to denote the convex conjugate, we define, for all $y_{1} \in \mathbb{R}^{n}, y_{2} \in \mathbb{R}^{m}$, and $p \in \mathbb{R}^{m}$,

$$
\begin{align*}
& Q_{1}\left(y_{1}, y_{2}, p\right)=-F_{1}^{*}\left(y_{1}, y_{2}, p\right)=\inf _{x \in \mathbb{R}^{m}}\left\{L_{1}(x, 0, p)-\left\langle y_{1}, x\right\rangle\right\}=Q_{1}\left(y_{1}, 0, p\right)  \tag{5.10}\\
& Q_{2}\left(y_{1}, y_{2}, p\right)=-F_{2}^{*}\left(y_{1}, y_{2}, p\right)=\inf _{z \in \mathbb{R}^{m}}\left\{L_{2}(0, z, p)-\left\langle y_{2}, z\right\rangle\right\}=Q_{2}\left(0, y_{2}, p\right) \tag{5.11}
\end{align*}
$$

and note that it follows from (5.9) that

$$
\begin{equation*}
Q\left(y_{1}, y_{2}, p\right)=Q_{1}\left(y_{1}, y_{2}, p\right)+Q_{2}\left(y_{1}, y_{2}, p\right) \tag{5.12}
\end{equation*}
$$

Note that $y_{2}$ is a nominal (ignored) argument to $Q_{1}$ and $y_{1}$ is similarly an ignored argument to $Q_{2}$.

By construction, $Q_{1}$ and $Q_{2}$ are closed concave functions. Letting $\partial Q_{1}$ and $\partial Q_{2}$ be the subgradient maps of the respective convex functions $-Q_{1}$ and $-Q_{2}$, we have

$$
\begin{aligned}
& \left(y_{1}, y_{2}, p\right) \in \partial F_{1}(x, z, u) \Leftrightarrow\left(y_{1}, y_{2}, u\right) \in \partial L_{1}(x, z, p) \Leftrightarrow(x, z, u) \in \partial Q_{1}\left(y_{1}, y_{2}, p\right) \\
& \left(y_{1}, y_{2}, p\right) \in \partial F_{2}(x, z, u) \Leftrightarrow\left(y_{1}, y_{2}, u\right) \in \partial L_{2}(x, z, p) \Leftrightarrow(x, z, u) \in \partial Q_{2}\left(y_{1}, y_{2}, p\right)
\end{aligned}
$$

We have that $\left(x, z, u_{1}\right) \in \partial Q_{1}\left(y_{1}, y_{2}, p\right)$ if only if

$$
\begin{equation*}
Q_{1}\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \leq Q_{1}\left(y_{1}, y_{2}, p\right)-\left\langle x, y_{1}^{\prime}-y_{1}\right\rangle-\left\langle z, y_{2}^{\prime}-y_{2}\right\rangle-\left\langle u_{1}, p^{\prime}-p\right\rangle \tag{5.13}
\end{equation*}
$$

and similarly that $\left(x, z, u_{2}\right) \in \partial Q_{2}\left(y_{1}, y_{2}, p\right)$ if only if

$$
\begin{equation*}
Q_{2}\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \leq Q_{2}\left(y_{1}, y_{2}, p\right)-\left\langle x, y_{1}^{\prime}-y_{1}\right\rangle-\left\langle z, y_{2}^{\prime}-y_{2}\right\rangle-\left\langle u_{2}, p^{\prime}-p\right\rangle . \tag{5.14}
\end{equation*}
$$

The point-to-set mappings $\partial F, \partial L$ and $\partial Q$ are all maximal monotone 69, p. 240] and for all $(x, z),\left(y_{1}, y_{2}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m}$ and $u, p \in \mathbb{R}^{m}$,

$$
\left(y_{1}, y_{2}, p\right) \in \partial F(x, z, u) \Leftrightarrow\left(y_{1}, y_{2}, u\right) \in \partial L(x, z, p) \Leftrightarrow(x, z, u) \in \partial Q\left(y_{1}, y_{2}, p\right) .
$$

The dual function $Q_{0}: \mathbb{R}^{m} \rightarrow[-\infty,+\infty)$ of $(\mathbb{P})$ is the parametric dual function evaluated at $\left(y_{1}, y_{2}\right)=0$, that is,

$$
\begin{align*}
Q_{0}(p) & =Q(0,0, p) \\
& =Q_{1}(0,0, p)+Q_{2}(0,0, p) \\
& =\inf _{x \in \mathbb{R}^{n}}\{f(x)+\langle p, M x\rangle\}+\inf _{z \in \mathbb{R}^{m}}\{g(z)-\langle p, z\rangle\} \\
& =\left(-f^{*}\left(-M^{\top} p\right)\right)+\left(-g^{*}(p)\right), \tag{5.15}
\end{align*}
$$

where "**" again denotes the convex conjugate. The dual problem corresponding to (5.4) is that of maximizing $Q_{0}(p)$ over $p \in \mathbb{R}^{m}$, that is,

$$
\begin{equation*}
\max _{p \in \mathbb{R}^{m}} Q_{1}(0,0, p)+Q_{2}(0,0, p) \tag{D}
\end{equation*}
$$

In view of (5.15), problem (D) is identical to the usual dual problem (4.18) of (P). As an application of Fenchel's inequality [69, Theorem 23.5], we have the weak duality relation $Q(0,0, p) \leq F(x, z, 0)$ for all $x \in \mathbb{R}^{n}, z \in \mathbb{R}^{m}$, and $p \in \mathbb{R}^{m}$.

Now suppose we are using an iterative method to solve subproblem (1.9), and let $x^{k+1}$ denote some approximate solution (rather than an exact one), with

$$
\begin{equation*}
y_{1}^{k+1} \in \partial_{x}\left[f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right]_{x=x^{k+1}} . \tag{5.16}
\end{equation*}
$$

Note that if we were to exactly solve the subproblem, 0 would be a possible value of $y_{1}^{k+1}$. Employing 69, Theorem 23.8], we have

$$
\begin{array}{rr}
\Leftrightarrow & y_{1}^{k+1} \in \partial f\left(x^{k+1}\right)+M^{\top} p^{k}+c M^{\top}\left(M x^{k+1}-z^{k}\right) \\
\Leftrightarrow & \left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in\left(\partial f\left(x^{k+1}\right)+M^{\top}\left(p^{k}+c\left(M x^{k+1}-z^{k}\right)\right)\right) \\
& \times\{0\} \times\left\{-M x^{k+1}\right\} \\
\Leftrightarrow & \left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k}, p^{k}+c\left(M x^{k+1}-z^{k}\right)\right) \tag{5.18}
\end{array}
$$

Now suppose we have some inexact solution $z^{k+1}$ to 1.10 and

$$
\begin{equation*}
y_{2}^{k+1} \in \partial\left[g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|z-M x^{k+1}\right\|^{2}\right]_{z=z^{k+1}} \tag{5.19}
\end{equation*}
$$

Much as for (5.16, if $z^{k+1}$ were an exact solution to (1.10), it would be possible to choose $y_{2}^{k+1}=0$ in 5.19 . Following a similar development to that of 5.18), we have

$$
\begin{array}{cc}
\Leftrightarrow & y_{2}^{k+1} \in \partial g\left(z^{k+1}\right)-p^{k}+c\left(z^{k+1}-M x^{k+1}\right) \\
\Leftrightarrow & \left(0, y_{2}^{k+1}, z^{k+1}\right) \in\{0\} \times\left\{\partial g\left(z^{k+1}\right)-\left(p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\right)\right\} \\
& \times\left\{z^{k+1}\right\} \\
\Leftrightarrow & \left(0, y_{2}^{k+1}, z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\right) \tag{5.21}
\end{array}
$$

Using the standard multiplier update

$$
\begin{equation*}
p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right) \tag{5.22}
\end{equation*}
$$

and letting

$$
\begin{equation*}
\mu^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k}\right) \tag{5.23}
\end{equation*}
$$

we observe that $\mu^{k+1}-p^{k+1}=c\left(z^{k+1}-z^{k}\right)$ and we obtain

$$
\begin{align*}
5.16 & \Leftrightarrow & \left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k}, \mu^{k+1}\right)  \tag{5.24}\\
5.19 & \Leftrightarrow & \left(0, y_{2}^{k+1}, z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k+1}\right) \tag{5.25}
\end{align*}
$$

### 5.2 Analyzing the exact ADMM by Lagrangian splitting

In this section, we present a convergence proof for exact ADMM using Lagrangian splitting. This proof is largely equivalent to Fortin's [34, Theorem 5.1] in the case without overrelaxation, but uses the more modern constructs in Section 5.1. The exact ADMM that is considered in this section consists of recursions (1.9), 1.10) and the standard multiplier update step (1.11). Similar analyses can be found in [8, 15. In [34], besides the assumptions that we have made on $f$ and $g, M$ is also assumed to have full column rank. In the modified and strengthened proof given here, however, this assumption is not required for most of our conclusions. We begin with a result by Alves and Svaiter [1] that is needed in several places in this Chapter.

Lemma 14. [1, Lemma 2] Let $W \subseteq \mathbb{R}^{p_{1}}$ and $V \subseteq \mathbb{R}^{p_{2}}$ be nonempty sets and suppose that $\left\{s_{k}=\left(\alpha_{k}, \beta_{k}\right)\right\}_{k \geq 0} \subset \mathbb{R}^{p_{1}} \times \mathbb{R}^{p_{2}}$ is a sequence such that:

1. $\left\{\left\|s_{k}-s\right\|\right\}$ is nonincreasing for all $s \in W \times V$ and
2. Every limit point of $\left\{\beta_{k}\right\}$ belongs to $V$.

Then $\left\{\beta_{k}\right\}$ converges to some element in $V$.

The main convergence properties of exact ADMM are now given by next two propositions.

Proposition 15. Under Assumption 1, let $\left\{x^{k}\right\},\left\{z^{k}\right\}$ and $\left\{p^{k}\right\}$ obey the recursions (1.9), 1.10 and 1.11), then

- $\left\{M x^{k}\right\},\left\{z^{k}\right\}$ and $\left\{p^{k}\right\}$ are all bounded sequences.
- $\sum_{0}^{\infty}\left\|M x^{k}-z^{k}\right\|^{2}<\infty, \sum_{0}^{\infty}\left\|z^{k+1}-z^{k}\right\|^{2}<\infty, \sum_{0}^{\infty}\left\|p^{k+1}-p^{k}\right\|^{2}<\infty$.
- For any KKT point $\left(x^{*}, z^{*}, p^{*}\right)$ of ( P$)$, the sequence $\left\{\left\|\left(\frac{1}{c} p^{k}, z^{k}\right)-\left(\frac{1}{c} p^{*}, z^{*}\right)\right\|\right\}$ is convergent.
- $\left\{\left(\frac{1}{c} p^{k}, z^{k}\right)\right\}$ is Fejér monotone to $\left\{\left.\left(\frac{1}{c} p^{*}, z^{*}\right) \right\rvert\,\left(x^{*}, z^{*}, p^{*}\right)\right.$ is KKT point $\}$.

Proof. By Assumption (1), there exists a KKT point $\left(x^{*}, z^{*}, p^{*}\right)$ of (P), such that $(0,0,0) \in \partial L\left(x^{*}, z^{*}, p^{*}\right)$ and $M x^{*}=z^{*}$. Then

$$
(0,0,0) \in \partial L\left(x^{*}, z^{*}, p^{*}\right)=\left\{\partial f\left(x^{*}\right)+M^{\top} p^{*}\right\} \times\left\{\partial g\left(z^{*}\right)-p^{*}\right\} \times\left\{z^{*}-M x^{*}\right\}
$$

and from $L=L_{1}+L_{2}$ and (5.7)-(5.8) we have

$$
\begin{aligned}
\left(0,0,-M x^{k+1}\right) & \in \partial L_{1}\left(x^{k+1}, z^{k}, p^{k}+c\left(M x^{k+1}-z^{k}\right)\right) \\
\left(0,0,-M x^{*}\right) & \in \partial L_{1}\left(x^{*}, z^{*}, p^{*}+c\left(M x^{*}-z^{*}\right)\right) \\
\left(0,0, z^{k+1}\right) & \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\right) \\
\left(0,0, z^{*}\right) & \in \partial L_{2}\left(x^{*}, z^{*}, p^{*}+c\left(M x^{*}-z^{*}\right)\right)
\end{aligned}
$$

since $\partial L_{1}$ is a monotone operator, we obtain

$$
\left\langle-\left(M x^{k+1}-M x^{*}\right), p^{k}-p^{*}+c\left(M x^{k+1}-M x^{*}\right)-c\left(z^{k}-z^{*}\right)\right\rangle \geq 0
$$

Expanding this inequality produces

$$
\begin{align*}
&\left\langle M x^{k+1}-M x^{*}, p^{k}-p^{*}\right\rangle+c\left\|M x^{k+1}-M x^{*}\right\|^{2} \\
&-c\left\langle M x^{k+1}-M x^{*}, z^{k}-z^{*}\right\rangle \leq 0 . \tag{5.26}
\end{align*}
$$

Using the monotonicity of $\partial L_{2}$, we have

$$
\left\langle z^{k+1}-z^{*}, p^{k}-p^{*}+c\left(M x^{k+1}-M x^{*}\right)-c\left(z^{k+1}-z^{*}\right)\right\rangle \geq 0
$$

which we may rearrange into

$$
\begin{align*}
c\left\|z^{k+1}-z^{*}\right\|^{2}-c\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{*}\right\rangle & \\
& -\left\langle z^{k+1}-z^{*}, p^{k}-p^{*}\right\rangle \leq 0 \tag{5.27}
\end{align*}
$$

Adding 5.26 and 5.27, we obtain that

$$
\begin{align*}
& c\left\|M x^{k+1}-M x^{*}\right\|-c\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{*}\right\rangle+c\left\|z^{k+1}-z^{*}\right\| \\
&+\left\langle M x^{k+1}-z^{k+1}, p^{k}-p^{*}\right\rangle-c\left\langle M x^{k+1}-M x^{*}, z^{k}-z^{*}\right\rangle \leq 0 \tag{5.28}
\end{align*}
$$

which may be regrouped into

$$
\begin{aligned}
& c\left\|M x^{k+1}-M x^{*}-\left(z^{k+1}-z^{*}\right)\right\|^{2}+c\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{*}-\left(z^{k}-z^{*}\right)\right\rangle \\
&+\left\langle M x^{k+1}-z^{k+1}, p^{k}-p^{*}\right\rangle \leq 0
\end{aligned}
$$

that is

$$
\begin{align*}
c\left\|M x^{k+1}-z^{k+1}\right\|^{2}+c\left\langle M x^{k+1}-M x^{*}, z^{k+1}\right. & \left.-z^{k}\right\rangle \\
& +\left\langle M x^{k+1}-z^{k+1}, p^{k}-p^{*}\right\rangle \leq 0 . \tag{5.29}
\end{align*}
$$

Rewriting the standard multiplier update step (5.22) as

$$
\begin{equation*}
p^{k+1}-p^{*}=p^{k}-p^{*}+c\left(M x^{k+1}-z^{k+1}\right) \tag{5.30}
\end{equation*}
$$

and then squaring both sides yields, after some rearrangement,

$$
\begin{align*}
& 2 c\left\langle M x^{k+1}-z^{k+1}, p^{k}-p^{*}\right\rangle= \\
& \left\|p^{k+1}-p^{*}\right\|^{2}-\left\|p^{k}-p^{*}\right\|^{2}-c^{2}\left\|M x^{k+1}-z^{k+1}\right\|^{2} \tag{5.31}
\end{align*}
$$

By combining (5.29) and (5.31), we obtain

$$
\begin{align*}
\frac{1}{c^{2}}\left(\left\|p^{k}-p^{*}\right\|^{2}-\left\|p^{k+1}-p^{*}\right\|^{2}\right) \geq & \\
& +2\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{k}\right\rangle+\left\|M x^{k+1}-z^{k+1}\right\|^{2} \tag{5.32}
\end{align*}
$$

Next, using that $M x^{*}=z^{*}$, we rewrite $M x^{k+1}-M x^{*}$ as

$$
M x^{k+1}-M x^{*}=\left(M x^{k+1}-M x^{k}\right)+\left(M x^{k}-z^{k}\right)+\left(z^{k}-z^{*}\right)
$$

so that

$$
\begin{align*}
\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{k}\right\rangle & =\left\langle M x^{k+1}-M x^{k}, z^{k+1}-z^{k}\right\rangle \\
+ & \left\langle M x^{k}-z^{k}, z^{k+1}-z^{k}\right\rangle+\left\langle z^{k}-z^{*}, z^{k+1}-z^{k}\right\rangle . \tag{5.33}
\end{align*}
$$

Applying the identity $\langle a, b\rangle=\frac{1}{2}\left[\|a+b\|^{2}-\|a\|^{2}-\|b\|^{2}\right]$ to the last term in (5.33), we arrive at

$$
\begin{aligned}
\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{k}\right\rangle= & \left\langle M x^{k+1}-M x^{k}, z^{k+1}-z^{k}\right\rangle+\left\langle M x^{k}-z^{k}, z^{k+1}-z^{k}\right\rangle \\
& +\frac{1}{2}\left[\left\|z^{k+1}-z^{*}\right\|^{2}-\left\|z^{k}-z^{*}\right\|^{2}-\left\|z^{k+1}-z^{k}\right\|^{2}\right] .
\end{aligned}
$$

Rearranging terms in this equation, we obtain

$$
\begin{align*}
\left\langle M x^{k+1}-M x^{k}, z^{k+1}-z^{k}\right\rangle & =\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{k}\right\rangle-\left\langle M x^{k}-z^{k}, z^{k+1}-z^{k}\right\rangle \\
- & \frac{1}{2}\left[\left\|z^{k+1}-z^{*}\right\|^{2}-\left\|z^{k}-z^{*}\right\|^{2}-\left\|z^{k+1}-z^{k}\right\|^{2}\right] . \tag{5.34}
\end{align*}
$$

From (5.21) we know that

$$
\begin{aligned}
\left(0,0, z^{k}\right) & \in \partial L_{2}\left(x^{k}, z^{k}, p^{k-1}+c M x^{k}-c z^{k}\right) \\
\left(0,0, z^{k+1}\right) & \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k}+c M x^{k+1}-c z^{k+1}\right)
\end{aligned}
$$

Again, using the monotonicity of $\partial L_{2}$, we have

$$
\left\langle z^{k+1}-z^{k}, p^{k}-p^{k-1}+c\left(M x^{k+1}-M x^{k}\right)-c\left(z^{k+1}-z^{k}\right)\right\rangle \geq 0
$$

Since $p^{k}-p^{k-1}=c\left(M x^{k}-z^{k}\right)$, we have

$$
\begin{equation*}
\left\langle M x^{k+1}-M x^{k}, z^{k+1}-z^{k}\right\rangle \geq\left\|z^{k+1}-z^{k}\right\|^{2}-\left\langle M x^{k}-z^{k}, z^{k+1}-z^{k}\right\rangle . \tag{5.35}
\end{equation*}
$$

Substituting (5.34) into (5.35), we obtain

$$
\begin{equation*}
2\left\langle M x^{k+1}-M x^{*}, z^{k+1}-z^{k}\right\rangle \geq\left\|z^{k+1}-z^{*}\right\|^{2}-\left\|z^{k}-z^{*}\right\|^{2}+\left\|z^{k+1}-z^{k}\right\|^{2} \tag{5.36}
\end{equation*}
$$

Adding (5.32) and (5.36), we obtain

$$
\begin{align*}
\frac{1}{c^{2}}\left\|p^{k}-p^{*}\right\|^{2}+\left\|z^{k}-z^{*}\right\|^{2}-\left(\frac{1}{c^{2}}\left\|p^{k+1}-p^{*}\right\|^{2}+\left\|z^{k+1}-z^{*}\right\|^{2}\right) \\
\geq\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k+1}-z^{k}\right\|^{2} \tag{5.37}
\end{align*}
$$

Summing (5.37) for $k=0, \ldots, K$ and canceling "telescoping" terms gives that, for any $K \geq 0$,

$$
\begin{array}{r}
\sum_{0}^{K}\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\sum_{0}^{K}\left\|z^{k+1}-z^{k}\right\|^{2}+\frac{1}{c^{2}}\left\|p^{K+1}-p^{*}\right\|^{2}+\left\|z^{K+1}-z^{*}\right\|^{2} \\
\leq \frac{1}{c^{2}}\left\|p^{0}-p^{*}\right\|^{2}+\left\|z^{0}-z^{*}\right\|^{2} \tag{5.38}
\end{array}
$$

From 5.37) and (5.38 we may immediately conclude the following:

- Both $\left\{z^{k}\right\}$ and $\left\{p^{k}\right\}$ are bounded.
- $\left\{\left\|\left(\frac{1}{c}\left(p^{k}-p^{*}\right),\left(z^{k}-z^{*}\right)\right)\right\|^{2}\right\}=\left\{\left\|\left(\frac{1}{c} p^{k}, z^{k}\right)-\left(\frac{1}{c} p^{*}, z^{*}\right)\right\|^{2}\right\}$ is non-increasing.
- Since $\left(x^{*}, z^{*}, p^{*}\right)$ was arbitrarily chosen, $\left\{\left(\frac{1}{c} p^{k}, z^{k}\right)\right\}$ is Fejér monotone to the set $\left\{\left.\left(\frac{1}{c} p^{*}, z^{*}\right) \right\rvert\,\left(x^{*}, z^{*}, p^{*}\right)\right.$ is a KKT point $\}$.
- Since $\left\{\left\|\left(\frac{1}{c} p^{k}, z^{k}\right)-\left(\frac{1}{c} p^{*}, z^{*}\right)\right\|\right\}$ is bounded below by 0 and decreasing, it converges to a finite limit.
- Letting $K \rightarrow \infty$, it follows from (5.38) that

$$
\sum_{0}^{\infty}\left\|M x^{k}-z^{k}\right\|^{2}<\infty \quad \sum_{0}^{\infty}\left\|z^{k+1}-z^{k}\right\|^{2}<\infty
$$

Because $p^{k+1}-p^{k}=c\left(M x^{k+1}-z^{k+1}\right)$, we also have $\sum_{0}^{\infty}\left\|p^{k+1}-p^{k}\right\|^{2}<\infty$.

- It immediately follows from $\sum_{0}^{\infty}\left\|M x^{k}-z^{k}\right\|^{2}<\infty$ that $M x^{k}-z^{k} \rightarrow 0$ and $\left\{z^{k}\right\}$ is bounded, so therefore $\left\{M x^{k}\right\}$ must be bounded.

Proposition 16. If Assumption (1) is satisfied, let $\left\{x^{k}\right\},\left\{z^{k}\right\}$ and $\left\{p^{k}\right\}$ obey the recursions (1.9), (1.10) and (1.11), we have the following convergence results:

- Every limit point of $\left\{x^{k}, z^{k}, p^{k}\right\}$ (if any such points exist) is a KKT point of (P).
- $\left\{p^{k}\right\}$ converges to an optimal solution $p^{*}$ of the dual problem (D).
- $\lim _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}$ exists and equals the optimal value of (1.4) and (P).

Furthermore, if $M$ has full column rank, then $\left\{x^{k}\right\}$ converges to an optimal solution $x^{*}$ of the original problem (1.4) and $z^{k} \rightarrow M x^{*}$.

Proof. Since $x^{k+1}$ and $z^{k+1}$ are the minimizers of the subproblems (1.9) and 1.10 respectively, it follows that

$$
\begin{align*}
-M^{\top} p^{k}-c M^{\top}\left(M x^{k+1}-z^{k}\right) & \in \partial f\left(x^{k+1}\right)  \tag{5.39}\\
p^{k}+c\left(M x^{k+1}-z^{k+1}\right) & \in \partial g\left(z^{k+1}\right) \tag{5.40}
\end{align*}
$$

Let $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ denote any limit point of $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$, and $\mathcal{K} \in \mathbb{N}$ be a infinite index set such that $\left(x^{k+1}, z^{k+1}, p^{k+1}\right) \rightarrow_{\mathcal{K}}\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$. From Proposition 15, we know that $M x^{k+1}-z^{k+1} \rightarrow 0$ and $z^{k+1}-z^{k} \rightarrow 0$, from which it follows that $M x^{k+1}-$ $z^{k} \rightarrow 0$. Taking the limit over $\mathcal{K}$, we have $M x^{\infty}=z^{\infty}$. We can further assert that $\left\{p^{k}\right\}$ has the same limit over $\mathcal{K}$ as $\left\{p^{k+1}\right\}$ does, since $c\left(M x^{k+1}-z^{k+1}\right)=p^{k+1}-p^{k} \rightarrow 0$. By taking the limit over $\mathcal{K}$ in (5.39) and (5.40) and using that subdifferentials of closed functions are closed, we obtain that

$$
\begin{equation*}
-M^{\top} p^{\infty} \in \partial f\left(x^{\infty}\right) \quad p^{\infty} \in \partial g\left(z^{\infty}\right) \tag{5.41}
\end{equation*}
$$

Thus the limit point $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ is a KKT point for $(\overline{\mathrm{P}})$.
Although its limit points must all be KKT points, $\left\{x^{k}, z^{k}, p^{k}\right\}$ does not necessarily have any limit points because $\left\{x^{k}\right\}$ might not be bounded unless we impose some restrictions on $M$. Nevertheless, we now show that every limit point of $\left\{p^{k}\right\}$ is a optimal solution for (D) whether or not $\left\{x^{k}\right\}$ is bounded. We have from (5.7) that

$$
\begin{equation*}
\left(0,0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k}, \mu^{k+1}\right) \tag{5.42}
\end{equation*}
$$

where $\mu^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k}\right)$, as defined in (5.23). Since $\partial Q_{1}$ is the partial inverse of $\partial L_{1}$, we therefore have

$$
\left(x^{k+1}, z^{k},-M x^{k+1}\right) \in \partial Q_{1}\left(0,0, \mu^{k+1}\right) .
$$

From (5.13), we have for all $\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ that

$$
Q_{1}\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \leq Q_{1}\left(0,0, \mu^{k+1}\right)-\left\langle x^{k+1}, y_{1}^{\prime}\right\rangle-\left\langle z^{k}, y_{2}^{\prime}\right\rangle-\left\langle-M x^{k+1}, p^{\prime}-\mu^{k+1}\right\rangle .
$$

Let $\left\{x^{*}, z^{*}, p^{*}\right\}$ be a KKT point that is hypothesized to exist by Assumption 1. Setting $\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right)=\left(0,0, p^{*}\right)$ in the previous inequality, we have

$$
\begin{equation*}
Q_{1}\left(0,0, p^{*}\right) \leq Q_{1}\left(0,0, \mu^{k+1}\right)-\left\langle-M x^{k+1}, p^{*}-\mu^{k+1}\right\rangle \tag{5.43}
\end{equation*}
$$

Similarly, from (5.8) and $z^{k+1}$ being exact minimizer of (1.10), we know that

$$
\begin{equation*}
\left(0,0, z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k+1}\right) \tag{5.44}
\end{equation*}
$$

and $\partial Q_{2}$ is the partial inverse of $\partial L_{2}$, so it follows that

$$
\left(x^{k+1}, z^{k+1}, z^{k+1}\right) \in \partial Q_{2}\left(0,0, p^{k+1}\right)
$$

and therefore

$$
\begin{equation*}
Q_{2}\left(0,0, p^{*}\right) \leq Q_{2}\left(0,0, p^{k+1}\right)-\left\langle z^{k+1}, p^{*}-p^{k+1}\right\rangle \tag{5.45}
\end{equation*}
$$

According to (5.12), we know that $Q\left(0,0, p^{*}\right)=Q_{1}\left(0,0, p^{*}\right)+Q_{2}\left(0,0, p^{*}\right)$. Substituting this equation into the inequality obtained by adding (5.43) and (5.45), we arrive at

$$
\begin{align*}
Q\left(0,0, p^{*}\right) \leq Q_{1}\left(0,0, \mu^{k+1}\right)+ & Q_{2}\left(0,0, p^{k+1}\right) \\
& -\left\langle-M x^{k+1}, p^{*}-\mu^{k+1}\right\rangle-\left\langle z^{k+1}, p^{*}-p^{k+1}\right\rangle \tag{5.46}
\end{align*}
$$

Using that $\mu^{k+1}=p^{k+1}-c\left(z^{k}-z^{k+1}\right)$, we may rewrite 5.46) as follows:

$$
\begin{align*}
& Q\left(0,0, p^{*}\right) \leq Q_{1}\left(0,0, \mu^{k+1}\right)+Q_{2}\left(0,0, p^{k+1}\right) \\
& \quad \quad-\left\langle-M x^{k+1}+z^{k+1}, p^{*}-p^{k+1}\right\rangle-c\left\langle-M x^{k+1}, z^{k}-z^{k+1}\right\rangle \\
& =Q_{1}\left(0,0, \mu^{k+1}\right)+Q_{2}\left(0,0, p^{k+1}\right) \\
& \quad \quad+\left\langle M x^{k+1}-z^{k+1}, p^{*}-p^{k+1}\right\rangle+c\left\langle M x^{k+1}, z^{k}-z^{k+1}\right\rangle . \tag{5.47}
\end{align*}
$$

Since Proposition 15 asserts that $\left\{p^{k}\right\}$ is bounded and $M x^{k+1}-z^{k+1} \rightarrow 0$, we conclude that $\left\langle M x^{k+1}-z^{k+1}, p^{*}-p^{k+1}\right\rangle \rightarrow 0$. Once again by Proposition 15, we know that $\left\{M x^{k}\right\}$ is bounded and $z^{k+1}-z^{k} \rightarrow 0$, so $\left\langle M x^{k+1}, z^{k}-z^{k+1}\right\rangle \rightarrow 0$. Thus the last two terms in (5.47) both converge to 0 .

Consider any limit point $p^{\infty}$ of $\left\{p^{k}\right\}$, and suppose $\mathcal{K} \subseteq \mathbb{N}$ is an infinite sequence of indices such that $p^{k} \rightarrow_{\mathcal{K}} p^{\infty}$. Since $z^{k+1}-z^{k} \rightarrow 0$ by Proposition 15 , we have $\mu^{k}-p^{k} \rightarrow 0$, so we also have $\mu^{k} \rightarrow_{\mathcal{K}} p^{\infty}$. Thus, taking limits over $\mathcal{K}$ in (5.47) and using that $Q_{1}$ and $Q_{2}$ are closed (upper semicontinuous) concave functions, we have

$$
\begin{aligned}
f\left(x^{*}\right)+g\left(z^{*}\right) & =Q\left(0,0, p^{*}\right) \\
& \leq \liminf _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}\left\{Q_{1}\left(0,0, \mu^{k+1}\right)+Q_{2}\left(0,0, p^{k+1}\right)\right\} \\
& \leq \limsup _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}\left\{Q_{1}\left(0,0, \mu^{k+1}\right)+Q_{2}\left(0,0, p^{k+1}\right)\right\} \\
& \leq \limsup _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} Q_{1}\left(0,0, \mu^{k+1}\right)+\underset{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}{\limsup } Q_{2}\left(0,0, p^{k+1}\right) \\
& \leq Q_{1}\left(0,0, \lim _{\substack{\rightarrow \rightarrow \infty \\
k \in \mathcal{K}}} \mu^{k+1}\right)+Q_{2}\left(0,0, \lim _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} p^{k+1}\right) \\
& =Q_{1}\left(0,0, p^{\infty}\right)+Q_{2}\left(0,0, p^{\infty}\right) \\
& =Q\left(0,0, p^{\infty}\right) \\
& \leq f\left(x^{*}\right)+g\left(z^{*}\right),
\end{aligned}
$$

where the last inequality is a consequence of $f\left(x^{*}\right)+g\left(z^{*}\right)$ being the maximum possible value of $Q$. Therefore, we obtain $Q\left(0,0, p^{\infty}\right)=Q\left(0,0, p^{*}\right)=f\left(x^{*}\right)+g\left(z^{*}\right)$, meaning that $p^{\infty}$ is also a dual optimal solution. Now let

$$
\begin{aligned}
V & =\left\{\left.\frac{1}{c} p \right\rvert\, p \text { is an optimal solution of }(\mathrm{D})\right\} \\
W & =\{M x \mid x \text { is an optimal solution of (1.4) }\}
\end{aligned}
$$

By [69, Corollary 30.5.1], which essentially states that the set of KKT points of a convex program is the Cartesian product of the sets of primal and dual solutions, we may infer that

$$
W \times V=\left\{\left.\left(z, \frac{1}{c} p\right) \right\rvert\,(x, z, p) \text { is a KKT point of of }(\mathrm{P})\right\}
$$

From Proposition 15, we therefore know that $\left\|\left(z^{k}, \frac{1}{c} p^{k}\right)-s\right\|$ is nonincreasing for every $s \in W \times V$. We have also established that any limit point $p^{\infty}$ of $\left\{p^{k}\right\}$ must be
an optimal dual solution, that is, that all limit points of $\left\{\frac{1}{c} p^{k}\right\}$ are in $V$. Lemma 14 therefore implies that $\left\{\frac{1}{c} p^{k}\right\}$ converges to a member of $V$, that is, that $\left\{p^{k}\right\}$ converges to an optimal dual solution.

To show convergence of objective values, we first add $c\left(z^{k}-z^{k+1}\right)$ to both sides of the inclusion (5.40), regrouping terms to obtain
$c\left(z^{k}-z^{k+1}\right) \in \partial g\left(z^{k+1}\right)-p^{k}+c\left(z^{k+1}-M x^{k+1}\right)+c\left(z^{k}-z^{k+1}\right)=\partial g\left(z^{k+1}\right)-\mu^{k+1}$,

From the definition for $\partial L_{2}$, which is given in (5.8), we then can ascertain that

$$
\begin{equation*}
\left(0, c\left(z^{k}-z^{k+1}\right), z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, \mu^{k+1}\right) \tag{5.48}
\end{equation*}
$$

Adding (5.42) and (5.48), and using that $\partial L_{1}+\partial L_{2}=\partial L$, we obtain

$$
\left(0, c\left(z^{k}-z^{k+1}\right), z^{k+1}-M x^{k+1}\right) \in \partial L\left(x^{k+1}, z^{k+1}, \mu^{k+1}\right)
$$

Since $\partial L$ is a partial inverse of $\partial F$ as defined in (5.3), we know that

$$
\begin{equation*}
\left(0, c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \in \partial F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right) \tag{5.49}
\end{equation*}
$$

Recalling that $Q$ is the negative of the convex conjugate of $F$, combining Fenchel's equality 69, Theorem 23.5] with 5.49) produces

$$
\begin{aligned}
F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right)-Q & \left(0, c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \\
& =c\left\langle z^{k+1}, z^{k}-z^{k+1}\right\rangle+\left\langle z^{k+1}-M x^{k+1}, \mu^{k+1}\right\rangle .
\end{aligned}
$$

Using the definition (5.3) of $F$, we have $F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right)=f\left(x^{k+1}\right)+$ $g\left(z^{k+1}\right)$. Substituting this identity into the above equation and rearranging, we obtain

$$
\begin{align*}
f\left(x^{k+1}\right)+g\left(z^{k+1}\right)=Q & \left(0, c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right)  \tag{5.50}\\
& +c\left\langle z^{k+1}, z^{k}-z^{k+1}\right\rangle+\left\langle z^{k+1}-M x^{k+1}, \mu^{k+1}\right\rangle . \tag{5.51}
\end{align*}
$$

The first term in (5.51) converges to zero because Proposition 15 asserts that $\left\{z^{k}\right\}$ is bounded and $z^{k+1}-z^{k} \rightarrow 0$. Since $\left\{p^{k}\right\}$ is bounded and $z^{k+1}-z^{k} \rightarrow 0$, it follows
that $\left\{\mu^{k+1}\right\}$ is bounded, because $\mu^{k+1}=p^{k+1}-c\left(z^{k}-z^{k+1}\right)$. From Proposition 15 we have $M x^{k+1}-z^{k+1} \rightarrow 0$, so the second term in 5.51) also converges to zero. Let $\mathcal{K}_{1} \subseteq \mathbb{N}$ be any infinite sequence of indices for which $\lim _{k \in \mathcal{K}_{1}} f\left(x^{k+1}\right)+g\left(z^{k+1}\right)=$ $\limsup _{k \rightarrow \infty} f\left(x^{k}\right)+g\left(z^{k}\right)$. Since $\left\{\mu^{k}\right\}$ is bounded, there exists some infinite subsequence $\mathcal{K}_{1}^{\prime} \subseteq \mathcal{K}_{1}$ over which $\left\{\mu^{k+1}\right\}$ converges to some limit $p^{\infty} \in \mathbb{R}^{m}$. Taking the limit over $\mathcal{K}_{1}^{\prime}$ in (5.50)-(5.51), we obtain, since we have established that all the terms in (5.51) converge to zero, that

$$
\begin{equation*}
\limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=\lim _{k \in \mathcal{K}_{1}^{\prime}} Q\left(0, c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \tag{5.52}
\end{equation*}
$$

Because $Q$ is an upper semicontinuous function, $z^{k+1}-z^{k} \rightarrow 0$, and $\mu^{k} \rightarrow_{\mathcal{K}_{1}^{\prime}} p^{\infty}$, we must have

$$
\begin{equation*}
\lim _{k \in \mathcal{K}_{1}^{\prime}} Q\left(0, c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \leq Q\left(0,0, p^{\infty}\right) \tag{5.53}
\end{equation*}
$$

By weak duality, we also have $Q\left(0,0, p^{\infty}\right) \leq \inf _{x \in \mathbb{R}^{n}}\{f(x)+g(z)\}=f\left(x^{*}\right)+g\left(z^{*}\right)$. Combining this observation with 5.52 and (5.53), we obtain

$$
\begin{equation*}
\limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq f\left(x^{*}\right)+g\left(z^{*}\right) \tag{5.54}
\end{equation*}
$$

On the other hand, $\left(x^{*}, z^{*}\right)$ minimizes the ordinary Lagrangian

$$
L(x, z, p)=f(x)+g(z)+\langle p, M x-z\rangle
$$

of (P) with respect to $(x, z)$ for $p=p^{*}$, so for any $k \geq 0$ we have $L\left(x^{*}, z^{*}, p^{*}\right) \leq$ $L\left(x^{k}, z^{k}, p^{*}\right)$, which is equivalent to

$$
f\left(x^{*}\right)+g\left(z^{*}\right) \leq f\left(x^{k}\right)+g\left(z^{k}\right)+\left\langle p^{*}, M x^{k}-z^{k}\right\rangle
$$

which with a minor rearrangement is in turn equivalent to

$$
\begin{equation*}
f\left(x^{*}\right)+g\left(z^{*}\right)-\left\langle p^{*}, M x^{k}-z^{k}\right\rangle \leq f\left(x^{k}\right)+g\left(z^{k}\right) \tag{5.55}
\end{equation*}
$$

Let $\mathcal{K}_{2} \subseteq \mathbb{N}$ be any infinite sequence of indices for which

$$
\lim _{k \in \mathcal{K}_{2}}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=\liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}
$$

Since the Proposition 15 ensures that $\left\langle p^{*}, M x^{k}-z^{k}\right\rangle \rightarrow 0$, we obtain by taking the limit over $\mathcal{K}_{2}$ in (5.55) that

$$
\begin{equation*}
f\left(x^{*}\right)+g\left(z^{*}\right) \leq \liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \tag{5.56}
\end{equation*}
$$

Combining (5.54) and 5.56, we have

$$
f\left(x^{*}\right)+g\left(z^{*}\right) \leq \liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq \limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq f\left(x^{*}\right)+g\left(z^{*}\right)
$$

which is equivalent to $\lim _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=f\left(x^{*}\right)+g\left(z^{*}\right)$.
Finally, if $M$ has full column rank, the boundedness $\left\{z^{k}\right\}$ and $M x^{k}-z^{k} \rightarrow 0$ imply that $\left\{x^{k}\right\}$ is bounded. Therefore, $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ must have a limit point, which must be a KKT point. This proves the final assertion.

### 5.3 Common elements of the Lagrangian splitting analyses

We will use the above tools and notation to develop two different approximate versions of the ADMM. This section develops some common analysis underlying both versions. Some of these results isolate or generalize specific arguments in the preceeding section.

Lemma 17. Suppose that $\left\{x^{k}\right\},\left\{y_{1}^{k}\right\} \subset \mathbb{R}^{n}$, and $\left\{z^{k}\right\},\left\{p^{k}\right\},\left\{y_{2}^{k}\right\} \subset \mathbb{R}^{m}$ obey the recursions (5.16), (5.19), and 5.22). If it is also true that

$$
M x^{k}-z^{k} \rightarrow 0 \quad M x^{k+1}-z^{k} \rightarrow 0 \quad y_{1}^{k} \rightarrow 0 \quad y_{2}^{k} \rightarrow 0
$$

then all limit points of the sequence $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ are KKT points of (P).
Proof. Consider any limit point $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ of $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$, along with an infinite index set $\mathcal{K} \subseteq \mathbb{N}$ such that $\left(x^{k+1}, z^{k+1}, p^{k+1}\right) \rightarrow_{\mathcal{K}}\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$. By rearranging (5.17) and (5.20), which are respectively equivalent to they hypotheses (5.16) and (5.19), we arrive at

$$
\begin{align*}
y_{1}^{k+1}-M^{\top}\left(p^{k}+c\left(M x^{k+1}-z^{k}\right)\right) & \in \partial f\left(x^{k+1}\right)  \tag{5.57}\\
y_{2}^{k+1}+\left(p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\right) & \in \partial g\left(z^{k+1}\right) \tag{5.58}
\end{align*}
$$

From (5.22) and the hypothesis that $M x^{k}-z^{k} \rightarrow 0$, we conclude that $\left\{p^{k}\right\}$ has the same limit over $\mathcal{K}$ as $\left\{p^{k+1}\right\}$ does. Using this information and the hypotheses that $y_{1}^{k}, y_{2}^{k} \rightarrow 0$, we obtain by taking limits over $\mathcal{K}$ in 5.57) and (5.58 and using that subdifferentials of closed functions have closed graphs that

$$
\begin{equation*}
-M^{\top} p^{\infty} \in \partial f\left(x^{\infty}\right) \quad p^{\infty} \in \partial g\left(z^{\infty}\right) \tag{5.59}
\end{equation*}
$$

Since we have assumed that $M x^{k}-z^{k} \rightarrow 0$, it also follows by taking limits over $\mathcal{K}$ that $M x^{\infty}=z^{\infty}$. In view of (5.59), we conclude that $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ is a KKT point of (P).

Note that the lemma does not address the question of whether any limit points of $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ exist. One set of conditions sufficient to guarantee that such limit points exist is that $\left\{x^{k}\right\}$ possess at least one limit point and that $\left\{z^{k}\right\}$ and $\left\{p^{k}\right\}$ be bounded.

Clearly, any two of the conditions

$$
M x^{k}-z^{k} \rightarrow 0 \quad M x^{k+1}-z^{k} \rightarrow 0 \quad z^{k+1}-z^{k} \rightarrow 0
$$

are sufficient to imply the remaining one, so we may substitute any two of these conditions for the assumptions $M x^{k+1}-z^{k+1} \rightarrow$ and $M x^{k+1}-z^{k} \rightarrow 0$ in the above lemma.

We now give another result, similar to Lemma 17, that will prove useful for cases in which various "inner loops" of the algorithms to be proposed below do not terminate:

Lemma 18. Suppose $\bar{x} \in \mathbb{R}^{n}$ and $\bar{p} \in \mathbb{R}^{m}$, and that the sequences $\left\{\hat{x}^{i}\right\}_{i=1}^{\infty},\left\{\hat{y}_{1}^{i}\right\}_{i=1}^{\infty} \subset$ $\mathbb{R}^{n}$ and $\left\{\hat{z}^{i}\right\}_{i=1}^{\infty},\left\{\hat{p}^{i}\right\}_{i=1}^{\infty},\left\{\hat{y}_{2}^{i}\right\}_{i=1}^{\infty} \subset \mathbb{R}^{m}$ conform for all $i$ to the recursions

$$
\begin{aligned}
& \hat{y}_{2}^{i} \in \partial\left[g(z)-\langle\bar{p}, z\rangle+\frac{c}{2}\|M \bar{x}-z\|^{2}\right]_{z=\hat{z}^{i}} \\
& \hat{p}^{i}=\bar{p}+c\left(M \bar{x}-\hat{z}^{i}\right) \\
& \hat{y}_{1}^{i} \in \partial\left[f(x)+\left\langle\hat{p}^{i}, M x\right\rangle+\frac{c}{2}\left\|M x-\hat{z}^{i}\right\|^{2}\right]_{x=\hat{x}^{i}} .
\end{aligned}
$$

If it is also true that

$$
\begin{equation*}
\lim _{i \rightarrow \infty} \hat{y}_{1}^{i}=0 \quad \lim _{i \rightarrow \infty} \hat{y}_{2}^{i}=0 \quad \lim _{i \rightarrow \infty} M \hat{x}^{i}-\hat{z}^{i}=0 \tag{5.60}
\end{equation*}
$$

then all limit points of $\left\{\left(\hat{x}^{i}, \hat{z}^{i}, \hat{p}^{i}\right)\right\}$ are KKT points of ( P$)$.

Proof. Let $\left(\hat{x}^{\infty}, \hat{z}^{\infty}, \hat{p}^{\infty}\right)$ be any limit point of $\left\{\left(\hat{x}^{i}, \hat{z}^{i}, \hat{p}^{i}\right)\right\}$ and $\mathcal{K} \subseteq \mathbb{N}$ be some infinite index set such that $\left(\hat{x}^{i}, \hat{z}^{i}, \hat{p}^{i}\right) \rightarrow_{\mathcal{K}}\left(\hat{x}^{\infty}, \hat{z}^{\infty}, \hat{p}^{\infty}\right)$. For all $i$, we then have

$$
\begin{equation*}
\bar{y}_{2}^{i}+\hat{p}^{i} \in \partial g\left(\hat{z}^{i}\right) \quad \bar{y}_{1}^{i}-M^{\top} \hat{p}^{i}-c\left(M \hat{x}^{i}-\hat{z}^{i}\right) \in \partial f\left(\hat{x}^{i}\right) \tag{5.61}
\end{equation*}
$$

Taking limits over $\mathcal{K}$ in (5.61), applying the hypotheses in (5.60), and using that subdifferentials of closed functions are closed and $M \hat{x}^{i}-\hat{z}^{i} \rightarrow 0$, we obtain

$$
\begin{equation*}
\hat{p}^{\infty} \in \partial g\left(\hat{z}^{\infty}\right) \quad-M^{\top} \hat{p}^{\infty} \in \partial f\left(\hat{x}^{\infty}\right) \quad M \hat{x}^{\infty}=\hat{z}^{\infty} \tag{5.62}
\end{equation*}
$$

that is, that $\left(\hat{x}^{\infty}, \hat{z}^{\infty}, \hat{p}^{\infty}\right)$ is a KKT point of $(\mathrm{P})$.

The following special case of Lemma 18 will prove useful in cases in which the $g$ subproblems is easy to solve exactly:

Corollary 19. Suppose $\bar{x} \in \mathbb{R}^{n}$ and $\bar{p}, \hat{p}, \hat{z} \in \mathbb{R}^{m}$ satisfy

$$
\begin{aligned}
& \hat{z}=\underset{z}{\arg \min }\left\{g(z)-\langle\bar{p}, z\rangle+\frac{c}{2}\|M \bar{x}-z\|^{2}\right\} \\
& \hat{p}=\bar{p}+c(M \bar{x}-\hat{z})
\end{aligned}
$$

and that $\left\{\hat{x}^{i}\right\}_{i=1}^{\infty},\left\{\hat{y}_{1}^{i}\right\}_{i=1}^{\infty} \subset \mathbb{R}^{n}$ satisfy the inclusion

$$
\hat{y}_{1}^{i} \in \partial\left[f(x)+\langle\hat{p}, M x\rangle+\frac{c}{2}\|M x-\hat{z}\|^{2}\right]_{x=\hat{x}^{i}}
$$

for all $i$. If it is also true that

$$
\lim _{i \rightarrow \infty} \hat{y}_{1}^{i}=0 \quad \lim _{i \rightarrow \infty} M \hat{x}^{i}-\hat{z}=0
$$

then for every limit point $\hat{x}^{\infty}$ of $\left\{\hat{x}^{i}\right\}$, we have that $\left(\hat{x}^{\infty}, \hat{z}, \hat{p}\right)$ is a KKT point of ( P ).

Proof. The result follows immediately by applying Lemma 18 in the case that $\hat{z}^{i}=\hat{z}$, $\hat{y}_{2}^{i}=0$, and $\hat{p}^{i}=\hat{p}$ for all $i$.

We now use the monotonicity of subdifferential mappings such as $\partial L, \partial L_{1}$ and $\partial L_{2}$ to derive crucial inequalities that will prove useful in analyzing our proposed algorithms:

Lemma 20. Suppose that $\left(y_{2}^{k+1}, z^{k+1}\right)$ satisfies (5.19), $\left(y_{2}^{k}, z^{k}\right)$ satisfies (5.19) with $k$ replaced by $k-1$, and $p^{k+1}=p^{k}+c\left(M x^{k}-z^{k}\right)$. Then

$$
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \leq\left\|M x^{k+1}-z^{k}\right\|^{2}+\frac{2}{c}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle
$$

Proof. Applying (5.25) and its equivalent with $k$ replaced by $k-1$, we have

$$
\begin{aligned}
\left(0, y_{2}^{k}, z^{k}\right) & \in \partial L_{2}\left(x^{k}, z^{k}, p^{k}\right) \\
\left(0, y_{2}^{k+1}, z^{k+1}\right) & \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k+1}\right)
\end{aligned}
$$

Since $\partial L_{2}$ is monotone [69, Corollary 37.5.2], the above two inclusions imply that

$$
\begin{aligned}
&\left\langle p^{k}-p^{k+1}, z^{k}-z^{k+1}\right\rangle+\left\langle y_{2}^{k}-y_{2}^{k+1} z^{k}-z^{k+1}\right\rangle \geq 0 \\
& \Leftrightarrow\left\langle-c\left(M x^{k+1}-z^{k+1}\right), z^{k}-z^{k+1}\right\rangle+\left\langle y_{2}^{k}-y_{2}^{k+1}, z^{k}-z^{k+1}\right\rangle \\
& \geq 0 \\
& \Leftrightarrow c\left\langle M x^{k+1}-z^{k+1}, z^{k}-z^{k+1}\right\rangle-\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle \leq 0
\end{aligned}
$$

where the first equivalence uses $p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)$. Multiplying by $2 / c$ and expanding the first inner product using the identity $\langle a, b\rangle=\|a\|^{2}+\|b\|^{2}-\|a-b\|^{2}$, we obtain the equivalent inequality

$$
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2}-\left\|M x^{k+1}-z^{k}\right\|^{2}-\frac{2}{c}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle \leq 0
$$

which we may rearrange into

$$
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \leq\left\|M x^{k+1}-z^{k}\right\|^{2}+\frac{2}{c}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle
$$

The next two propositions assert that under certain conditions, the sequences produced by algorithms conforming to (5.16), (5.19), and (5.22) are asymptotically optimal even when $\left\{x^{k}\right\}$ does not have any limit points.

Proposition 21. Suppose that $\left\{x^{k}\right\},\left\{y_{1}^{k}\right\} \subset \mathbb{R}^{n}$ and $\left\{z^{k}\right\},\left\{p^{k}\right\},\left\{y_{2}^{k}\right\} \subset \mathbb{R}^{m}$ obey the recursions (5.16), (5.19), and (5.22), and Assumption 1 holds. If the sequences $\left\{M x^{k}\right\}$ and $\left\{p^{k}\right\}$ are bounded and we also have

$$
\begin{equation*}
z^{k+1}-z^{k} \rightarrow 0 \quad M x^{k}-z^{k} \rightarrow 0 \quad y_{1}^{k} \rightarrow 0 \quad y_{2}^{k} \rightarrow 0 \quad\left\langle y_{1}^{k}, x^{k}\right\rangle \rightarrow 0 \quad\left\langle y_{2}^{k}, z^{k}\right\rangle \rightarrow 0, \tag{5.63}
\end{equation*}
$$

then every limit point of $\left\{p^{k}\right\}$ is an optimal solution to the dual problem (D).
Proof. Let $\left(x^{*}, z^{*}, p^{*}\right)$ be a saddle point for $(\mathrm{P})$. Defining $\mu^{k}$ as in (5.23), we have from (5.24) that $\left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k}, \mu^{k+1}\right)$. Since $\partial Q_{1}$ is the partial inverse of $\partial L_{1}$, we therefore have

$$
\left(x^{k+1}, z^{k},-M x^{k+1}\right) \in \partial Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)
$$

From (5.13), we have for all $\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \in \mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ that

$$
\begin{aligned}
Q_{1}\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right) \leq & Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right) \\
& \quad-\left\langle x^{k+1}, y_{1}^{\prime}-y_{1}^{k+1}\right\rangle-\left\langle z^{k}, y_{2}^{\prime}-0\right\rangle-\left\langle-M x^{k+1}, p^{\prime}-\mu^{k+1}\right\rangle
\end{aligned}
$$

Let $p^{*}$ be some optimal dual solution, which must exist by Assumption 1. Setting $\left(y_{1}^{\prime}, y_{2}^{\prime}, p^{\prime}\right)=\left(0,0, p^{*}\right)$ in the previous inequality, we have

$$
\begin{equation*}
Q_{1}\left(0,0, p^{*}\right) \leq Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)-\left\langle x^{k+1}, 0-y_{1}^{k+1}\right\rangle-\left\langle-M x^{k+1}, p^{*}-\mu^{k+1}\right\rangle \tag{5.64}
\end{equation*}
$$

Similarly, 5.25 and $\partial Q_{2}$ being the partial inverse of $\partial L_{2}$ imply that

$$
\left(x^{k+1}, z^{k+1}, z^{k+1}\right) \in \partial Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)
$$

and therefore

$$
\begin{equation*}
Q_{2}\left(0,0, p^{*}\right) \leq Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)-\left\langle z^{k+1}, 0-y_{2}^{k+1}\right\rangle-\left\langle z^{k+1}, p^{*}-p^{k+1}\right\rangle \tag{5.65}
\end{equation*}
$$

From (5.12), we know that $Q\left(0,0, p^{*}\right)=Q_{1}\left(0,0, p^{*}\right)+Q_{2}\left(0,0, p^{*}\right)$. Substituting this equation into the inequality obtained by adding (5.64) and (5.65), we arrive at

$$
\begin{align*}
Q\left(0,0, p^{*}\right) \leq Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right) & +Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)+\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}, y_{2}^{k+1}\right\rangle \\
& -\left\langle-M x^{k+1}, p^{*}-\mu^{k+1}\right\rangle-\left\langle z^{k+1}, p^{*}-p^{k+1}\right\rangle \tag{5.66}
\end{align*}
$$

Since $p^{k+1}=\mu^{k+1}+c\left(z^{k}-z^{k+1}\right)$, we can rewrite (5.66) as follows:

$$
\begin{array}{r}
Q\left(0,0, p^{*}\right) \leq Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)+Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)+\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}, y_{2}^{k+1}\right\rangle \\
\quad-\left\langle-M x^{k+1}+z^{k+1}, p^{*}-p^{k+1}\right\rangle-c\left\langle-M x^{k+1}, z^{k}-z^{k+1}\right\rangle \\
=Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)+Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)+\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}, y_{2}^{k+1}\right\rangle \\
\quad+\left\langle M x^{k+1}-z^{k+1}, p^{*}-p^{k+1}\right\rangle+c\left\langle M x^{k+1}, z^{k}-z^{k+1}\right\rangle . \tag{5.67}
\end{array}
$$

From the hypotheses that $\left\{p^{k}\right\}$ is bounded and $M x^{k+1}-z^{k+1} \rightarrow 0$, we conclude that $\left\langle M x^{k+1}-z^{k+1}, p^{*}-p^{k+1}\right\rangle \rightarrow 0$. Similarly, the hypotheses that $\left\{M x^{k}\right\}$ is bounded and $z^{k+1}-z^{k} \rightarrow 0$, imply that $\left\langle M x^{k+1}, z^{k}-z^{k+1}\right\rangle \rightarrow 0$. Since we have also assumed that $\left\langle y_{1}^{k+1}, x^{k+1}\right\rangle \rightarrow 0$ and $\left\langle y_{2}^{k+1}, z^{k+1}\right\rangle \rightarrow 0$, the last four terms in (5.67) all converge to 0 .

Consider any limit point $p^{\infty}$ of $\left\{p^{k}\right\}$, and suppose $\mathcal{K} \subseteq \mathbb{N}$ is an infinite sequence of indices such that $p^{k} \rightarrow_{\mathcal{K}} p^{\infty}$. The assumption that $z^{k+1}-z^{k} \rightarrow 0$ implies that $\mu^{k}-p^{k} \rightarrow 0$, so we also have $\mu^{k} \rightarrow \mathcal{K} p^{\infty}$. Thus, taking limits over $\mathcal{K}$ in (5.67) and using the fact that $Q_{1}$ and $Q_{2}$ are closed (upper semicontinuous) concave functions, we have

$$
\begin{aligned}
f\left(x^{*}\right)+g\left(z^{*}\right) & =Q\left(0,0, p^{*}\right) \\
& \leq \underset{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}{\liminf ^{\prime}}\left\{Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)+Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)\right\} \\
& \leq \limsup _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}\left\{Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)+Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right)\right\} \\
& \leq \limsup _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} Q_{1}\left(y_{1}^{k+1}, 0, \mu^{k+1}\right)+\underset{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}}{\limsup } Q_{2}\left(0, y_{2}^{k+1}, p^{k+1}\right) \\
& \leq Q_{1}\left(\lim _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} y_{1}^{k+1}, 0, \lim _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} \mu^{k+1}\right)+Q_{2}\left(0, \lim _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} y_{2}^{k+1}, \lim _{\substack{k \rightarrow \infty \\
k \in \mathcal{K}}} p^{k+1}\right) \\
& =Q_{1}\left(0,0, p^{\infty}\right)+Q_{2}\left(0,0, p^{\infty}\right) \\
& =Q\left(0,0, p^{\infty}\right) \\
& \leq f\left(x^{*}\right)+g\left(z^{*}\right),
\end{aligned}
$$

where the last inequality is a consequence of $f\left(x^{*}\right)+g\left(z^{*}\right)$ being the maximum possible value of $Q$. Therefore, we obtain $Q\left(0,0, p^{\infty}\right)=Q\left(0,0, p^{*}\right)=f\left(x^{*}\right)+g\left(z^{*}\right)$, meaning that $p^{\infty}$ is also a dual solution.

Proposition 22. Suppose that $\left\{x^{k}\right\},\left\{y_{1}^{k}\right\} \subset \mathbb{R}^{n}$ and $\left\{z^{k}\right\},\left\{p^{k}\right\},\left\{y_{2}^{k}\right\} \subset \mathbb{R}^{m}$ obey the recursions (5.16), (5.19), and (5.22). If the same limits in (5.63) hold and the sequences $\left\{p^{k}\right\}$ and $\left\{z^{k}\right\}$ are bounded, then

$$
\begin{equation*}
\limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq \inf _{x \in \mathbb{R}^{n}}\{f(x)+g(M x)\} \tag{5.68}
\end{equation*}
$$

If Assumption 1 is satisfied, then the stronger condition

$$
\begin{equation*}
\lim _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=\inf _{x \in \mathbb{R}^{n}} f(x)+g(M x)=f\left(x^{*}\right)+g\left(M x^{*}\right) \tag{5.69}
\end{equation*}
$$

holds, where $x^{*}$ is any solution of (1.4).
Proof. Adding $c\left(z^{k}-z^{k+1}\right)$ to both sides of the inclusion (5.20), which we know holds from the hypothesis (5.19), we obtain
$y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right) \in \partial g\left(z^{k+1}\right)-p^{k}+c\left(z^{k+1}-M x^{k+1}\right)+c\left(z^{k}-z^{k+1}\right)=\partial g\left(z^{k+1}\right)-\mu^{k+1}$, where $\mu^{k+1}$ is defined as in 5.23). From the formula for $\partial L_{2}$ given in (5.8), we then ascertain that

$$
\begin{equation*}
\left(0, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, \mu^{k+1}\right) \tag{5.70}
\end{equation*}
$$

Next, consider (5.24), which is a consequence of the hypothesis (5.16). Since $L_{1}$ is independent of its second $(z)$ argument, we may replace $z^{k}$ by $z^{k+1}$ in (5.24) and obtain

$$
\begin{equation*}
\left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k+1}, \mu^{k+1}\right) \tag{5.71}
\end{equation*}
$$

Adding (5.71) and (5.70), and using that $\partial L_{1}+\partial L_{2}=\partial L$, we obtain

$$
\left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), z^{k+1}-M x^{k+1}\right) \in \partial L\left(x^{k+1}, z^{k+1}, \mu^{k+1}\right) .
$$

Since $\partial L$ is a partial inverse of $\partial F$ as defined in (5.3), we know that

$$
\begin{equation*}
\left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \in \partial F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right) . \tag{5.72}
\end{equation*}
$$

Recalling that $Q$ is the negative of the convex conjugate of $F$, combining Fenchel's equality [69, Theorem 23.5] with 5.72) produces

$$
\begin{aligned}
& F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right)-Q\left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \\
& \quad=\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}, y_{2}^{k+1}\right\rangle+c\left\langle z^{k+1}, z^{k}-z^{k+1}\right\rangle+\left\langle z^{k+1}-M x^{k+1}, \mu^{k+1}\right\rangle .
\end{aligned}
$$

Using the definition (5.3) of $F$, we have $F\left(x^{k+1}, z^{k+1}, z^{k+1}-M x^{k+1}\right)=f\left(x^{k+1}\right)+$ $g\left(z^{k+1}\right)$. Substituting this identity into the above equation and rearranging, we obtain

$$
\begin{align*}
f\left(x^{k+1}\right)+g\left(z^{k+1}\right)=Q & \left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right)  \tag{5.73}\\
& +\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}, y_{2}^{k+1}\right\rangle  \tag{5.74}\\
& +c\left\langle z^{k+1}, z^{k}-z^{k+1}\right\rangle+\left\langle z^{k+1}-M x^{k+1}, \mu^{k+1}\right\rangle . \tag{5.75}
\end{align*}
$$

The hypotheses (5.63) directly ensure that the terms on line (5.74) converge to zero. The first term in also converges to zero, because we have assumed that $\left\{z^{k}\right\}$ is bounded and (5.63) contains the assumption that $z^{k+1}-z^{k} \rightarrow 0$. Since we have assumed that $\left\{p^{k}\right\}$ is bounded and $z^{k+1}-z^{k} \rightarrow 0$, it follows that $\left\{\mu^{k}\right\}$ is bounded. Furthermore, (5.63) contains the assumption that $M x^{k+1}-z^{k} \rightarrow 0$, so the second term in (5.75) also converges to zero. Let $\mathcal{K}_{1} \subseteq \mathbb{N}$ be any infinite sequence of indices for which $\lim _{k \in \mathcal{K}_{1}} f\left(x^{k+1}\right)+g\left(z^{k+1}\right)=\lim \sup _{k \rightarrow \infty} f\left(x^{k}\right)+g\left(z^{k}\right)$. Since we have assumed that $\left\{p^{k}\right\}$ is bounded, which implies that $\left\{\mu^{k}\right\}$ is bounded, there exists some infinite subsequence $\mathcal{K}_{1}^{\prime} \subseteq \mathcal{K}_{1}$ over which $\left\{\mu^{k+1}\right\}$ converges to some limit $p^{\infty} \in \mathbb{R}^{m}$. Taking the limit over $\mathcal{K}_{1}^{\prime}$ in (5.73)-5.75), we obtain, since we have established that all the terms in (5.74)-5.75) converge to zero, that

$$
\begin{equation*}
\limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=\lim _{k \in \mathcal{K}_{1}^{\prime}} Q\left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \tag{5.76}
\end{equation*}
$$

Furthermore, since $Q$ is an upper semicontinuous function and $y_{1}^{k} \rightarrow 0, y_{2}^{k} \rightarrow 0$, $z^{k+1}-z^{k} \rightarrow 0$, and $\mu^{k} \rightarrow_{\mathcal{K}_{1}^{\prime}} p^{\infty}$, we must have

$$
\begin{equation*}
\lim _{k \in \mathcal{K}_{1}^{\prime}} Q\left(y_{1}^{k+1}, y_{2}^{k+1}+c\left(z^{k}-z^{k+1}\right), \mu^{k+1}\right) \leq Q\left(0,0, p^{\infty}\right) \tag{5.77}
\end{equation*}
$$

By weak duality, we also have $Q\left(0,0, p^{\infty}\right) \leq \inf _{x \in \mathbb{R}^{n}}\{f(x)+g(M x)\}$. Combining this observation with (5.76) and (5.77), we obtain (5.68).

Now assume Assumption 1 holds, and let $\left(x^{*}, z^{*}, p^{*}\right)$ be any KKT point. Then (5.68) immediately becomes

$$
\begin{equation*}
\limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq \inf _{x \in \mathbb{R}^{n}}\{f(x)+g(M x)\}=f\left(x^{*}\right)+g\left(z^{*}\right) \tag{5.78}
\end{equation*}
$$

The point $\left(x^{*}, z^{*}\right)$ minimizes the ordinary Lagrangian $L(x, z, p)=f(x)+g(z)+$ $\langle p, M x-z\rangle$ of $(\mathrm{P})$ with respect to $(x, z)$ for $p=p^{*}$, so for any $k \geq 0$ we have $L\left(x^{*}, z^{*}, p^{*}\right) \leq L\left(x^{k}, z^{k}, p^{*}\right)$, which is equivalent to

$$
f\left(x^{*}\right)+g\left(z^{*}\right) \leq f\left(x^{k}\right)+g\left(z^{k}\right)+\left\langle p^{*}, M x^{k}-z^{k}\right\rangle,
$$

which with a minor rearrangement is in turn equivalent to

$$
\begin{equation*}
f\left(x^{*}\right)+g\left(z^{*}\right)-\left\langle p^{*}, M x^{k}-z^{k}\right\rangle \leq f\left(x^{k}\right)+g\left(z^{k}\right) \tag{5.79}
\end{equation*}
$$

Let $\mathcal{K}_{2} \subseteq \mathbb{N}$ be any infinite sequence of indices for which

$$
\lim _{k \in \mathcal{K}_{2}}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=\liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}
$$

Since the hypotheses (5.63) include $\left\langle p^{*}, M x^{k}-z^{k}\right\rangle \rightarrow 0$, we obtain by taking the limit over $\mathcal{K}_{2}$ in 5.79 that

$$
\begin{equation*}
f\left(x^{*}\right)+g\left(z^{*}\right) \leq \liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} . \tag{5.80}
\end{equation*}
$$

Combining (5.78) and (5.80, we have

$$
f\left(x^{*}\right)+g\left(z^{*}\right) \leq \liminf _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq \limsup _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\} \leq f\left(x^{*}\right)+g\left(z^{*}\right)
$$

which, in view of $M x^{*}=z^{*}$, is equivalent to (5.69).

### 5.4 Approximate ADMM with absolute summable error criteria

We are now in a position to develop an approximate ADMM algorithms whose convergence analysis is based on Lagrangian splitting. Our first method uses two given absolute error sequences, one for the $x$ subproblem, and one for the $z$ subproblem, and takes the form of a slightly more complicated form of Algorithm 4.1.1 developed above. However, we are able to prove its convergence without any strong convexity or matrix rank assumptions. Unlike the method from 29 underlying Algorithm 4.1.1, the convergence analysis uses only residual subgradient information and does not require bounds on the distance to the exact subproblem solutions.

Broadly speaking, the analysis here blends the original ADMM convergence proof in (34 with the techniques developed in 28 for approximate solution of subproblems in the standard (non-alternating-direction) augmented Lagrangian method. Just as in Section 4.1, let $\left\{\epsilon_{k}\right\}_{k=1}^{\infty},\left\{\tau_{k}\right\}_{k=1}^{\infty} \subset \mathbb{R}_{++}$be scalar sequences such that $\sum_{k=1}^{\infty} \epsilon_{k}<\infty$ and $\sum_{k=1}^{\infty} \tau_{k}<\infty$. We are free to construct these sequences to fit a particular class of problems of problem instances. For example, for a relatively difficult $x$ subproblem we might choose relatively large and slowly decreasing values of $\epsilon_{k}$, whereas if the $z$ subproblem is easily solved exactly, we might even take $\tau_{k} \equiv 0$. We also take $\beta_{1}>0$ and $\beta_{2}>0$ to be two arbitrary positive scalars. Our proposed algorithm is as follows:

```
Algorithm 5.4.1 Inexact ADMM with absolutely summable error criteria
    initialization: Pick \(c>0\) and initial points \(p^{0}, z^{0} \in \mathbb{R}^{m}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\)
            by taking \(\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until \(\left\|y_{1}^{k, l}\right\| \leq \frac{\epsilon_{k+1}}{\max \left\{\beta_{1},\left\|x^{k, l}\right\|\right\}}\)
        \(x^{k+1}=x^{k, l}\)
        \(y_{1}^{k+1}=y_{1}^{k, l}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(z^{k+1} \approx \arg \min _{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\)
            by taking \(\left(z^{k, l}, y_{2}^{k, l}\right)=\mathcal{G}\left(p^{k}, x^{k+1}, c, z^{k}, l\right)\)
        until \(\left\|y_{2}^{k, l}\right\| \leq \frac{\tau_{k+1}}{\max \left\{\beta_{2},\left\|z^{k, l}\right\|\right\}}\)
        \(z^{k+1}=z^{k, l}\)
        \(y_{2}^{k+1}=y_{2}^{k, l}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
```

until Overall convergence

Remark: When the sequence $\left\{y_{1}^{k}\right\}$ is not needed for the "overall convergence" test, the assignment $y_{1}^{k+1}=y_{1}^{k, l}$ can be omitted from the algorithm's implementation, and similarly for $\left\{y_{2}^{k}\right\}$, just as in Algorithm 4.1.1. However, the sequences $\left\{y_{1}^{k}\right\}$ and $\left\{y_{2}^{k}\right\}$ figure prominently in the convergence analysis of Algorithm 5.4.1, so we make sure to define them above.

Remark: Generally speaking, the parameters $\beta_{1}$ and $\beta_{2}$ should be chosen to be large numbers since they play the roles of "safe radii" in which the iterates are ordinarily expected to be contained. If the iterates stay contained within these radii, that is $\left\|x^{k, l}\right\| \leq \beta_{1}$ and $\left\|z^{k, l}\right\| \leq \beta_{2}$ for all combinations of $k$ and $l$ encountered in the
course of the algorithm, then the sequence of iterates produced by the algorithm are indistinguishable from those produced by Algorithm 4.1.1, with $\epsilon_{k}$ replaced by $\epsilon_{k} / \beta_{1}$ and $\tau_{k}$ replaced by $\tau_{k} / \beta_{2}$. The sequences $\left\{\epsilon_{k} / \beta_{1}\right\}$ and $\left\{\tau_{k} / \beta_{2}\right\}$ remain summable and thus meet the assumptions of Algorithm 4.1.1, so in this case Algorithm 5.4.1 essentially coincides with Algorithm 4.1.1.

We begin by noting that it is not possible for Algorithm 5.4.1 to become "trapped" in one of its inner loops, subject to a mild condition of the approximation procedure modeled by $\mathcal{F}$ :

Lemma 23. Suppose that the sequence $\left\{\left(x^{l}, y_{1}^{l}\right)\right\}=\{\mathcal{F}(p, z, c, \bar{x}, l)\}$ must be bounded for any $(p, z, c, \bar{x})$. Then the inner loops (over l) of Algorithm 5.4.1 always terminate in a finite number of iterations.

Proof. Fix any $k$. We begin by considering the first inner loop. Note that we must have $y_{1}^{k, l} \rightarrow 0$ by Assumption 2, so that the real impact of the boundedness assumption is to assert that $\left\{x^{k, l}\right\}_{l=1}^{\infty}$ is bounded. Therefore, the right-hand side of the innerloop termination condition $\left\|y_{1}^{k, l}\right\| \leq \epsilon_{k+1} / \max \left\{\beta_{1},\left\|x^{k, l}\right\|\right\}$ is bounded below by some positive quantity. Since its left-hand side converges to zero by Assumption 2, the condition must eventually be satisfied for some finite $l$.

Now consider the second inner loop. From assumption 3, we have $\lim _{l \rightarrow \infty} y_{2}^{k, l}=0$. By Lemma 6, $\left\{z^{k, l}\right\}_{l=1}^{\infty}$ must converge to the unique solution to the $z$ subproblem, so it must be bounded. Since this sequence is bounded, an argument similar to that for the first inner loop asserts that the second inner loop must also terminate finitely.

Remark: By Lemma 5, the boundedness assumption on $\mathcal{F}$ is automatically satisfied and thus redundant whenever the solution set of the $x$ subproblem is bounded, and in particular when the $x$ subproblem solution is unique, for example when $M$ has full column rank.

We now prove the convergence of this method. Many of the techniques are adapted from [28, Sections 3 and 4], but simplified to the special case of the standard Euclidean distance kernel, as opposed to the more general Bregman distances treated in [28.

Lemma 24. [66, Section 2.2] Suppose $\left\{\alpha_{k}\right\}_{k=0}^{\infty},\left\{\gamma_{k}\right\}_{k=0}^{\infty} \subset \mathbb{R}$ are sequences such that $\left\{\alpha_{k}\right\}$ is bounded below, $\sum_{k=0}^{\infty} \gamma_{k}$ exists and is finite, and the recursion $\alpha_{k+1} \leq \alpha_{k}+\gamma_{k}$ holds for all $k \geq 1$. Then $\left\{\alpha_{k}\right\}$ converges to a finite limit.

With an analysis similar to [28, Lemma 5], we prove the following result:
Lemma 25. In Algorithm 5.4.1, we have

$$
\begin{equation*}
\sum_{k=0}^{\infty}\left\|y_{1}^{k}\right\|<\infty \quad \sum_{k=0}^{\infty}\left\|y_{2}^{k}\right\|<\infty \quad \sum_{k=0}^{\infty}\left\langle x^{k}, y_{1}^{k}\right\rangle<\infty \quad \sum_{k=0}^{\infty}\left\langle z^{k}, y_{2}^{k}\right\rangle<\infty \tag{5.81}
\end{equation*}
$$

all these limits being guaranteed to exist.

Proof. The termination condition of the algorithm's first inner loop guarantees that for all $k \geq 0$,

$$
\begin{equation*}
\left\|y_{1}^{k+1}\right\| \leq \frac{\epsilon_{k+1}}{\max \left\{\beta_{1},\left\|x^{k+1}\right\|\right\}} \leq \frac{\epsilon_{k+1}}{\beta_{1}} \tag{5.82}
\end{equation*}
$$

Recalling that $\sum_{k=1}^{\infty} \epsilon_{k}<\infty$, we conclude that $\sum_{k=0}^{\infty}\left\|y_{1}^{k}\right\|<\infty$, establishing the first claim. From the first inequality in 5.82, we also have

$$
\begin{equation*}
\left\|y_{1}^{k+1}\right\| \leq \frac{\epsilon_{k+1}}{\max \left\{\beta_{1},\left\|x^{k+1}\right\|\right\}} \leq \frac{\epsilon_{k+1}}{\left\|x^{k+1}\right\|} \tag{5.83}
\end{equation*}
$$

Therefore,

$$
\left|\left\langle x^{k+1}, y_{1}^{k+1}\right\rangle\right| \leq\left\|x^{k+1}\right\|\left\|y_{1}^{k+1}\right\| \leq\left\|x^{k+1}\right\| \frac{\epsilon_{k+1}}{\left\|x^{k+1}\right\|}=\epsilon_{k+1}
$$

From the summability of $\left\{\epsilon_{k}\right\}$, we immediately deduce that $\sum_{k=0}^{\infty}\left|\left\langle x^{k}, y_{1}^{k}\right\rangle\right|<\infty$ and therefore we have that $\sum_{k=0}^{\infty}\left\langle x^{k}, y_{1}^{k}\right\rangle$ exists and is finite. The proofs of $\sum_{k=0}^{\infty}\left\|y_{2}^{k}\right\|$ and $\sum_{k=0}^{\infty}\left\langle z^{k}, y_{2}^{k}\right\rangle$ are nearly identical to those just presented.

Lemma 26. If $\left\{z^{k}\right\}_{k=1}^{\infty}$ is bounded in Algorithm 5.4.1, then

$$
\sum_{k=0}^{\infty}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle<\infty
$$

Proof. By the Cauchy-Schwartz inequality,

$$
\left|\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle\right| \leq\left\|y_{2}^{k+1}-y_{2}^{k}\right\|\left\|z^{k+1}-z^{k}\right\|
$$

Since $\left\{y_{2}^{k}\right\}$ is summable by Lemma 25, $\left\{y_{2}^{k+1}-y_{2}^{k}\right\}$ is also summable. Since $\left\{z^{k}\right\}$ is assumed to be bounded, it follows that the sequence $\left\{\left\|z^{k+1}-z^{k}\right\|\right\}$ is also bounded. We may then deduce from the above inequality that the sequence $\left\{\left|\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle\right|\right\}$ is summable, and therefore that the sequence $\left\{\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle\right\}$ is summable.

Proposition 27. If Assumption 1 holds, then the sequences $\left\{z^{k}\right\}$, $\left\{p^{k}\right\}$ and $\left\{M x^{k}\right\}$ generated by Algorithm 5.4.1, are all bounded, and $\sum_{k=0}^{\infty}\left\|M x^{k+1}-z^{k}\right\|^{2}<\infty$. Furthermore, for any KKT point $\left(x^{*}, z^{*}, p^{*}\right)$ of (P), the sequence $\left\{\left\|\left(c z^{k}, p^{k}\right)-\left(c z^{*}, p^{*}\right)\right\|\right\}$ converges to a finite limit.

Proof. Let $\left(x^{*}, z^{*}, p^{*}\right)$ be any a KKT point of $(\mathbb{P})$, hypothesized to exist by Assumption 1. Then

$$
(0,0,0) \in \partial L\left(x^{*}, z^{*}, p^{*}\right)=\left\{\partial f\left(x^{*}\right)+M^{\top} p^{*}\right\} \times\left\{\partial g\left(z^{*}\right)-p^{*}\right\} \times\left\{z^{*}-M x^{*}\right\}
$$

and from $L=L_{1}+L_{2}$ and (5.7)-(5.8) we have

$$
\begin{aligned}
\left(y_{1}^{k+1}, 0,-M x^{k+1}\right) & \in \partial L_{1}\left(x^{k+1}, z^{k}, p^{k}+c M x^{k+1}-c z^{k}\right) \\
\left(0,0,-M x^{*}\right) & \in \partial L_{1}\left(x^{*}, z^{*}, p^{*}+c M x^{*}-c z^{*}\right) \\
\left(0, y_{2}^{k+1}, z^{k+1}\right) & \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k}+c M x^{k+1}-c z^{k+1}\right) \\
\left(0,0, z^{*}\right) & \in \partial L_{2}\left(x^{*}, z^{*}, p^{*}+c M x^{*}-c z^{*}\right)
\end{aligned}
$$

Since $\partial L_{1}$ is a monotone operator, we have

$$
\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle+\left\langle p^{k}-p^{*}+c M\left(x^{k+1}-x^{*}\right)-c\left(z^{k}-z^{*}\right),-M\left(x^{k+1}-x^{*}\right)\right\rangle \geq 0
$$

Rearranging this inequality, we obtain

$$
\begin{equation*}
\left\|M\left(x^{k+1}-x^{*}\right)\right\|^{2} \leq \frac{1}{c}\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle-\left\langle\frac{1}{c}\left(p^{k}-p^{*}\right)-\left(z^{k}-z^{*}\right), M\left(x^{k+1}-x^{*}\right)\right\rangle . \tag{5.84}
\end{equation*}
$$

Using the monotonicity of $\partial L_{2}$, we have

$$
\begin{aligned}
\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle+\left\langle p^{k}-p^{*}+c M\left(x^{k+1}-x^{*}\right)-c\left(z^{k+1}-z^{*}\right), z^{k+1}-z^{*}\right\rangle \geq 0 \\
\Leftrightarrow \quad\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle+\left\langle p^{k}-p^{*}+c M\left(x^{k+1}-x^{*}\right), z^{k+1}-z^{*}\right\rangle-c\left\|z^{k+1}-z^{*}\right\|^{2} \geq 0
\end{aligned}
$$

which we may rearrange into

$$
\begin{equation*}
\left\|z^{k+1}-z^{*}\right\|^{2} \leq \frac{1}{c}\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle+\left\langle\frac{1}{c}\left(p^{k}-p^{*}\right)+M\left(x^{k+1}-x^{*}\right), z^{k+1}-z^{*}\right\rangle \tag{5.85}
\end{equation*}
$$

Adding (5.84) and (5.85), we obtain

$$
\begin{align*}
& \left\|M\left(x^{k+1}-x^{*}\right)\right\|^{2}+\left\|z^{k+1}-z^{*}\right\|^{2} \\
& \begin{aligned}
\leq \frac{1}{c}\left(\left\langlex^{k+1}-\right.\right. & \left.\left.x^{*}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle\right) \\
& +\left\langle M\left(x^{k+1}-x^{*}\right), z^{k}-z^{*}\right\rangle+\left\langle M\left(x^{k+1}-x^{*}\right), z^{k+1}-z^{*}\right\rangle \\
& \quad+\frac{1}{c}\left\langle p^{k}-p^{*}, z^{k+1}-M x^{k+1}\right\rangle
\end{aligned} \\
& \begin{array}{r}
=\frac{1}{c}\left(\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle\right) \\
\quad+\left\langle M\left(x^{k+1}-x^{*}\right), z^{k}-z^{*}\right\rangle+\left\langle M\left(x^{k+1}-x^{*}\right), z^{k+1}-z^{*}\right\rangle \\
\\
\quad+\frac{1}{c^{2}}\left\langle p^{k}-p^{*}, p^{k}-p^{k+1}\right\rangle
\end{array} \\
& \left.\quad+\frac{1}{2}\left(\| x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle+\left\langle x^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle\right)  \tag{5.86}\\
& \left.\quad+\frac{1}{2}\left(\left\|M\left(x^{*}\right)\right\|^{2}+\left\|z^{k}-z^{*}\right\|^{2}-x^{*}\right)\left\|^{2}+\right\| x^{k+1}-z^{*}\left\|^{2}-\frac{1}{c^{2}}\right\| z^{k+1}-p^{k} \|^{2}\right) \\
& \quad+\frac{1}{2 c^{2}}\left(\left\|p^{k}-p^{*}\right\|^{2}+\left\|p^{k}-p^{k+1}\right\|^{2}-\left\|p^{k+1}-p^{*}\right\|^{2}\right)
\end{align*}
$$

where we use the multiplier update $p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)$ to obtain the term (5.87), we apply the identity $\langle a, b\rangle=\frac{1}{2}\left(\|a\|^{2}+\|b\|^{2}-\|a-b\|^{2}\right)$ to obtain (5.88)(5.90), and both 5.86) and the last term in (5.88) use that $M x^{*}=z^{*}$. Multiplying the resulting inequality by $2 c^{2}$ and rearranging, we obtain

$$
c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|M x^{k+1}-z^{k}\right\|^{2}
$$

$$
\begin{equation*}
\leq c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}+2 c\left(\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle+\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle\right) \tag{5.91}
\end{equation*}
$$

We now bound the last term in (5.91), using the termination conditions for the algorithm's inner loops:

$$
\begin{aligned}
\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}\right\rangle & +\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}\right\rangle \leq\left\|x^{k+1}-x^{*}\right\|\left\|y_{1}^{k+1}\right\|+\left\|z^{k+1}-z^{*}\right\|\left\|y_{2}^{k+1}\right\| \\
& \leq\left\|x^{k+1}\right\|\left\|y_{1}^{k+1}\right\|+\left\|x^{*}\right\|\left\|y_{1}^{k+1}\right\|+\left\|z^{k+1}\right\|\left\|y_{2}^{k+1}\right\|+\left\|z^{*}\right\|\left\|y_{2}^{k+1}\right\| \\
& \leq\left\|x^{k+1}\right\| \frac{\epsilon_{k+1}}{\left\|x^{k+1}\right\|}+\left\|x^{*}\right\| \frac{\epsilon_{k+1}}{\beta_{1}}+\left\|z^{k+1}\right\| \frac{\tau_{k+1}}{\left\|z^{k+1}\right\|}+\left\|z^{*}\right\| \frac{\tau_{k+1}}{\beta_{2}} \\
& \leq\left(1+\frac{\left\|x^{*}\right\|}{\beta_{1}}\right) \epsilon_{k+1}+\left(1+\frac{\left\|z^{*}\right\|}{\beta_{2}}\right) \tau_{k+1} .
\end{aligned}
$$

Substituting this inequality into (5.91), we obtain

$$
\begin{align*}
& c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|M x^{k+1}-z^{k}\right\|^{2} \\
& \leq c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}+2 c\left[\left(1+\frac{\left\|x^{*}\right\|}{\beta_{1}}\right) \epsilon_{k+1}+\left(1+\frac{\left\|z^{*}\right\|}{\beta_{2}}\right) \tau_{k+1}\right] . \tag{5.92}
\end{align*}
$$

Adding this inequality for $k=0, \ldots, K$ and canceling "telescoping" terms yields that for any $K \geq 0$,

$$
\begin{align*}
& c^{2}\left\|z^{K+1}-z^{*}\right\|^{2}+\left\|p^{K+1}-p^{*}\right\|^{2}+c^{2} \sum_{k=0}^{K}\left\|M x^{k+1}-z^{k}\right\|^{2} \\
& \leq c^{2}\left\|z^{0}-z^{*}\right\|^{2}+\left\|p^{0}-p^{*}\right\|^{2}+2 c\left(1+\frac{\left\|x^{*}\right\|}{\beta_{1}}\right) \sum_{k=0}^{K} \epsilon_{k+1}+2 c\left(1+\frac{\left\|z^{*}\right\|}{\beta_{2}}\right) \sum_{k=0}^{K} \tau_{k+1} . \tag{5.93}
\end{align*}
$$

By assumption, both $\left\{\epsilon_{k}\right\}$ and $\left\{\tau_{k}\right\}$ are summable, so the right-hand side of (5.93) converges to a finite limit as $K \rightarrow \infty$. From this boundedness, (5.93) allows us to draw the following conclusions:

- $\left\{p^{k}\right\}$ and $\left\{z^{k}\right\}$ are bounded
- $\sum_{k=0}^{\infty}\left\|M x^{k+1}-z^{k}\right\|^{2}<\infty$
- Consequently, $M x^{k+1}-z^{k} \rightarrow 0$
- Because $M x^{k+1}-z^{k} \rightarrow 0$ and $\left\{z^{k}\right\}$ is bounded, $\left\{M x^{k}\right\}$ must be bounded.

From (5.92), since $\left\|M x^{k+1}-z^{k}\right\|^{2}$ is always nonnegative, we have for all $k$ that

$$
\begin{align*}
& c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|p^{k+1}-p^{*}\right\|^{2} \\
& \leq c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}+2 c\left[\left(1+\frac{\left\|x^{*}\right\|}{\beta_{1}}\right) \epsilon_{k+1}+\left(1+\frac{\left\|z^{*}\right\|}{\beta_{2}}\right) \tau_{k+1}\right] . \tag{5.94}
\end{align*}
$$

Let $\alpha_{k}=c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}$, which is bounded below by 0 . Also let

$$
\gamma_{k}=2 c\left[\left(1+\frac{\left\|x^{*}\right\|}{\beta_{1}}\right) \epsilon_{k+1}+\left(1+\frac{\left\|z^{*}\right\|}{\beta_{2}}\right) \tau_{k+1}\right],
$$

from which we can rewrite (5.94) as $\alpha_{k+1} \leq \alpha_{k}+\gamma_{k}$. Since $\left\{\epsilon_{k}\right\}$ and $\left\{\gamma_{k}\right\}$ are summable, we know that $\sum_{k=0}^{\infty} \gamma_{k}<\infty$, so by Lemma 24 , we conclude that

$$
\left\{\alpha_{k}\right\}=\left\{c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}\right\}=\left\{\left\|\left(c z^{k}, p^{k}\right)-\left(c z^{*}, p^{*}\right)\right\|^{2}\right\}
$$

converges to a finite limit, and consequently so does $\left\{\left\|\left(c z^{k}, p^{k}\right)-\left(c z^{*}, p^{*}\right)\right\|\right\}$.
Remark: If we take $y_{1}^{k+1}=0$ and $y_{2}^{k+1}=0$ for all $k$, then the inequality (5.91) reduces to

$$
c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|M x^{k+1}-z^{k}\right\|^{2} \leq c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|p^{k}-p^{*}\right\|^{2}
$$

Under this conditions on $y_{1}^{k+1}$ and $y_{2}^{k+1}$, it holds that by Lemma 20

$$
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \leq\left\|M x^{k+1}-z^{k}\right\|^{2}
$$

Immediately, it follows that

$$
\begin{align*}
& \frac{1}{c^{2}}\left\|p^{k}-p^{*}\right\|^{2}+\left\|z^{k}-z^{*}\right\|^{2}-\left(\frac{1}{c^{2}}\left\|p^{k+1}-p^{*}\right\|^{2}+\left\|z^{k+1}-z^{*}\right\|^{2}\right) \geq \\
&\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k+1}-z^{k}\right\|^{2} \tag{5.95}
\end{align*}
$$

which is exactly the inequality (5.37) in the convergence analysis for exact ADMM.

Proposition 28. If Assumption 1 is satisfied, then the sequences generated by Algorithm 5.4.1 have the properties

$$
\sum_{k=0}^{\infty}\left\|z^{k+1}-z^{k}\right\|^{2}<\infty \quad \sum_{k=0}^{\infty}\left\|M x^{k+1}-z^{k+1}\right\|^{2}<\infty \quad \sum_{k=0}^{\infty}\left\|p^{k+1}-p^{k}\right\|^{2}<\infty
$$

Proof. By Lemma 20, we know that

$$
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \leq\left\|M x^{k+1}-z^{k}\right\|^{2}+\frac{2}{c}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle
$$

Adding this inequality for $k=1, \ldots, K$, we obtain that for any $K \geq 0$ that

$$
\begin{align*}
\sum_{k=0}^{K}\left\|M x^{k+1}-z^{k+1}\right\|^{2} & +\sum_{k=0}^{K}\left\|z^{k}-z^{k+1}\right\|^{2} \\
& \leq \sum_{k=0}^{K}\left\|M x^{k+1}-z^{k}\right\|^{2}+\frac{2}{c} \sum_{k=0}^{K}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle \tag{5.96}
\end{align*}
$$

Now consider the limit as $K \rightarrow \infty$. Proposition 27 asserts that the first term on the right-hand side of the above inequality converges to a finite limit, while Lemma 26 guarantees that the same holds for the last term. Therefore we conclude that

$$
\sum_{k=0}^{\infty}\left\|M x^{k+1}-z^{k+1}\right\|^{2}<\infty \quad \sum_{k=0}^{\infty}\left\|z^{k+1}-z^{k}\right\|^{2}<\infty
$$

Since the multiplier update formula is equivalent to $p^{k+1}-p^{k}=c\left(M x^{k+1}-z^{k+1}\right)$, it follows immediately from the first of these inequalities that $\sum_{k=0}^{\infty}\left\|p^{k+1}-p^{k}\right\|^{2}<$ $\infty$.

The following two propositions summarize, respectively, the dual and primal behavior of Algorithm 5.4.1.

Proposition 29. Under Assumption 11, the sequences $\left\{p^{k}\right\}$ and $\left\{z^{k}\right\}$ generated by Algorithm 5.4.1 are bounded and all limit points of $\left\{p^{k}\right\}$ are solutions to the dual problem (D). Furthermore, if $\left\{x^{k}\right\}$ has at least one limit point, then $\left\{p^{k}\right\}$ converges to a particular dual solution.

Proof. We will use Proposition 21, whose assumptions are that $\left\{M x^{k}\right\}$ and $\left\{p^{k}\right\}$ are bounded, along with

$$
z^{k+1}-z^{k} \rightarrow 0 \quad M x^{k}-z^{k} \rightarrow 0 \quad y_{1}^{k} \rightarrow 0 \quad y_{2}^{k} \rightarrow 0 \quad\left\langle y_{1}^{k}, x^{k}\right\rangle \rightarrow 0 \quad\left\langle y_{2}^{k}, z^{k}\right\rangle \rightarrow 0
$$

The boundedness of $\left\{M x^{k}\right\}$ and $\left\{p^{k}\right\}$ was established in Proposition 27. The first two limit conditions above follow immediately from stronger results established in Proposition 28, while the remaining limit conditions similarly follow from stronger results already established in Lemma 25. Therefore, the hypotheses of Proposition 21 are satisfied, and we conclude that all limit points of $\left\{p^{k}\right\}$ must be dual solutions.

Now suppose that $\left\{x^{k}\right\}$ has at least one limit point $x^{\infty}$. Since Proposition 27 asserts that $\left\{p^{k}\right\}$ and $\left\{z^{k}\right\}$ are bounded, there exist $z^{\infty}, p^{\infty} \in \mathbb{R}^{m}$ and an infinite set of indices $\mathcal{K}$ for which $\left(x^{k}, z^{k}, p^{k}\right) \rightarrow_{\mathcal{K}}\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$. We will use Lemma 17 to assert that $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ is KKT point. The hypotheses of Lemma 17 are

$$
M x^{k}-z^{k} \rightarrow 0 \quad M x^{k+1}-z^{k} \rightarrow 0 \quad y_{1}^{k} \rightarrow 0 \quad y_{2}^{k} \rightarrow 0
$$

The condition $M x^{k}-z^{k} \rightarrow 0$ is a consequence of Proposition 28, while $M x^{k+1}-z^{k} \rightarrow 0$ follows from Proposition 27. The last two limit conditions above were established in Lemma 25. Therefore, Lemma 17 applies and $\left(x^{\infty}, z^{\infty}, p^{\infty}\right)$ must be KKT point. Proposition 27 then asserts that $\left\{\left\|\left(c z^{k}, p^{k}\right)-\left(c z^{\infty}, p^{\infty}\right)\right\|\right\}$ must converge to a finite limit, but since one of its limit points must be zero, the entire sequence converges and we deduce that $p^{k} \rightarrow p^{\infty}$.

Note that since $\left\{M x^{k}\right\}$ must be bounded by Proposition 27, one possible sufficient condition for $\left\{x^{k}\right\}$ to have a limit point is that $M$ has full column rank.

Proposition 30. For sequences $\left\{x^{k}\right\}$ and $\left\{z^{k}\right\}$ generated by Algorithm 5.4.1 under Assumption 1, we have

$$
M x^{k}-z^{k} \rightarrow 0 \quad \lim _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=f\left(x^{*}\right)+g\left(z^{*}\right)
$$

Proof. The first (feasibility) claim is, as in the proof of the previous proposition, an immediate consequence of Proposition 28, In view of Lemma 25, Proposition 27, and Proposition 28, this second claim is a direct application of Proposition 22,

### 5.5 An approximate ADMM with relative error criteria

One drawback to Algorithm 5.4.1 above is that it is not inherently clear how to select the parameter sequences $\left\{\epsilon_{k}\right\}$ and $\left\{\tau_{k}\right\}$. In the general study of proximal algorithms, it has become common starting with the work of Solodov and Svaiter in 75 to replace such absolute, formally exogenous error sequences with relative error criteria that require a single scalar parameter controlling the ratios of various algorithmic quantities. Such techniques have the advantage of adapting automatically to individual problem instances. This section proposes an algorithm similar to the one in the last section, but using such a relative error criterion and having only a few parameters. The analysis has much in common with the previous analysis, but combines the fundamental Lagrangian splitting proof approach dating back to 34 with the relative error non-alternating augmented Lagrangian techniques in [30], as opposed to the absolute error techniques in [28]. As in [30], we introduce an auxiliary sequence into the iterative process, in this case $\left\{\left(w_{1}^{k}, w_{2}^{k}\right)\right\}_{k=0}^{\infty} \subset \mathbb{R}^{n} \times \mathbb{R}^{m}$. The other sequences generated by the algorithm, $\left\{x^{k}\right\},\left\{y_{1}^{k}\right\} \subset \mathbb{R}^{n}$ and $\left\{z^{k}\right\},\left\{y_{2}^{k}\right\},\left\{p^{k}\right\} \subset \mathbb{R}^{m}$, play similar roles similar to the corresponding sequences in Algorithm 5.4.1. To simplify the statement of the algorithm, we define the following notation for all $k \geq 0$ :

$$
\boldsymbol{w}^{k}=\left[\begin{array}{c}
w_{1}^{k} \\
w_{2}^{k}
\end{array}\right] \quad \boldsymbol{y}^{k}=\left[\begin{array}{c}
y_{1}^{k} \\
y_{2}^{k}
\end{array}\right] \quad \boldsymbol{x}^{k}=\left[\begin{array}{c}
x^{k} \\
z^{k}
\end{array}\right]
$$

Succinctly, using a single scalar parameter $\sigma \in[0,1)$, the algorithm starts from arbitrary points $p^{0}, z^{0} \in \mathbb{R}^{m}$ and $\boldsymbol{w}^{0} \in \mathbb{R}^{n} \times \mathbb{R}^{m}$ and develops sequences conforming to
the following recursive conditions for all $k \geq 0$ :

$$
\begin{gather*}
y_{1}^{k+1} \in \partial_{x}\left[f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right]_{x=x^{k+1}}  \tag{5.97}\\
y_{2}^{k+1} \in \partial_{z}\left[g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|z-M x^{k+1}\right\|^{2}\right]_{z=z^{k+1}}  \tag{5.98}\\
\frac{2}{c}\left|\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle\right|+\left\|\boldsymbol{y}^{k+1}\right\|^{2} \leq \sigma\left\|M x^{k+1}-z^{k}\right\|^{2}  \tag{5.99}\\
p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)  \tag{5.100}\\
\boldsymbol{w}^{k+1}=\boldsymbol{w}^{k}-c \boldsymbol{y}^{k} . \tag{5.101}
\end{gather*}
$$

Here, (5.97), 5.98, and 5.100 are essentially the same as in Algorithm 5.4.1, while (5.99) and (5.101) are essentially the same approximation criterion and update proposed in [30]. In fact, applying the algorithm of 30 to $(\mathrm{P}$ with a constant penalty parameter would result in the same algorithm, except that the sequential conditions (5.97)-5.98) would be replace by the single coupled condition

$$
\boldsymbol{y}^{k+1} \in \partial_{(x, z)}\left[f(x)+g(z)+\left\langle p^{k}, M x-z\right\rangle+\frac{c}{2}\|M x-z\|^{2}\right]_{(x, z)=\left(x^{k+1}, y^{k+1}\right)}
$$

We will show that the recursions (5.97)-(5.101) converge if (P) has a KKT point, essentially establishing that if one keeps the penalty parameter constant in the algorithm of [30] applied to $(\mathrm{P})$, one can weaken the condition of approximately minimizing the augmented Lagrangian jointly with respect to $x$ and $z$ to an approximate minimization with respect to $x$ with $z$ fixed, followed by an approximate minimization with respect to $z$ with $x$ fixed.

Proposition 31. If there exists any saddle point of $L$, that is, some $\left(x^{*}, z^{*}, p^{*}\right) \in$ $\mathbb{R}^{n} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ such that $(0,0,0) \in \partial L\left(x^{*}, z^{*}, p^{*}\right)$, and $\sigma \in[0,1)$. Suppose that $\left\{\boldsymbol{x}^{k}\right\},\left\{\boldsymbol{y}^{k}\right\},\left\{p^{k}\right\}$ and $\left\{\boldsymbol{w}^{k}\right\}$ obey for all $k \geq 1$ the recursion (5.97)-(5.101), then the following hold:

- The sequences $\left\{p^{k}\right\},\left\{z^{k}\right\}$ and $\left\{\boldsymbol{w}^{k}\right\}$ are bounded.
- $\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)$ is Fejér monotone to $\left(\boldsymbol{x}^{*}, c z^{*}, p^{*}\right)$.

$$
\begin{aligned}
& -\left\{\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}^{*}, c z^{*}, p^{*}\right)\right\|^{2}\right\} \text { is convergent. } \\
& -M x^{k+1}-z^{k} \rightarrow 0, \boldsymbol{y}^{k} \rightarrow 0 \text { and }\left\langle\boldsymbol{x}^{k}, \boldsymbol{y}^{k}\right\rangle \rightarrow 0 \\
& -\left\|\left(\boldsymbol{w}^{k+1}, c z^{k+1}, p^{k+1}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\| \leq\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\|
\end{aligned}
$$

Proof. As in the previous analysis, we let $\mu^{k+1}=p^{k}+c M x^{k+1}-c z^{k}$. Then for any $k \geq 1$,

$$
\begin{aligned}
& \left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2} \\
= & \left\|p^{k}-p^{k+1}+p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{k+1}+z^{k+1}-z^{*}\right\|^{2} \\
= & \left\|p^{k}-p^{k+1}\right\|^{2}+2\left\langle p^{k}-p^{k+1}, p^{k+1}-p^{*}\right\rangle+\left\|p^{k+1}-p^{*}\right\|^{2} \\
& \quad+c^{2}\left\|z^{k}-z^{k+1}\right\|^{2}+2 c^{2}\left\langle z^{k}-z^{k+1}, z^{k+1}-z^{*}\right\rangle+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
= & \left\|p^{k}-p^{k+1}\right\|^{2}+2 c\left\langle z^{k+1}-M x^{k+1}, p^{k+1}-p^{*}\right\rangle+\left\|p^{k+1}-p^{*}\right\|^{2} \\
& \quad+c^{2}\left\|z^{k}-z^{k+1}\right\|^{2}+2 c\left\langle p^{k+1}-\mu^{k+1}, z^{k+1}-z^{*}\right\rangle+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
= & \left\|p^{k}-p^{k+1}\right\|^{2}+\left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{k+1}\right\|^{2}+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
& +2 c\left\langle z^{k+1}-z^{*}-M x^{k+1}+M x^{*}, p^{k+1}-p^{*}\right\rangle+2 c\left\langle p^{k+1}-\mu^{k+1}, z^{k+1}-z^{*}\right\rangle
\end{aligned}
$$

where the last equality uses $M x^{*}=z^{*}$. Thus,

$$
\begin{align*}
& \left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
& =\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}-\left\|p^{k}-p^{k+1}\right\|^{2}-c^{2}\left\|z^{k}-z^{k+1}\right\|^{2} \\
& -2 c\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle-2 c\left\langle p^{k+1}-\mu^{k+1}, z^{k+1}-z^{*}\right\rangle \\
& -2 c\left\langle-M x^{k+1}+M x^{*}, p^{k+1}-p^{*}\right\rangle . \tag{5.102}
\end{align*}
$$

Rewriting the last term in 5.102) as

$$
\left\langle-M x^{k+1}+M x^{*}, p^{k+1}-p^{*}\right\rangle=\left\langle-M x^{k+1}+M x^{*}, p^{k+1}-\mu^{k+1}+\mu^{k+1}-p^{*}\right\rangle
$$

we next obtain, using $M x^{*}=z^{*}$ that

$$
\begin{aligned}
& \left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
& =\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}-\left\|p^{k}-p^{k+1}\right\|^{2}-c^{2}\left\|z^{k}-z^{k+1}\right\|^{2} \\
& \begin{aligned}
&-2 c\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle-2 c\left\langle-M x^{k+1}+M x^{*}, p^{k+1}-\mu^{k+1}+\mu^{k+1}-p^{*}\right\rangle \\
& \quad-2 c\left\langle p^{k+1}-\mu^{k+1}, z^{k+1}-z^{*}\right\rangle
\end{aligned} \\
& \begin{array}{r}
\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2} \quad-2 c\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle-2 c\left\langle-M x^{k+1}+M x^{*}, \mu^{k+1}-p^{*}\right\rangle \\
\quad-\left\|p^{k}-p^{k+1}\right\|^{2}-c^{2}\left\|z^{k}-z^{k+1}\right\|^{2}+2 c\left\langle M x^{k+1}-z^{k+1}, p^{k+1}-\mu^{k+1}\right\rangle \\
=\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2} \\
\quad-2 c\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle-2 c\left\langle-M x^{k+1}+M x^{*}, \mu^{k+1}-p^{*}\right\rangle \\
\quad-c^{2}\left[\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2}-2\left\langle M x^{k+1}-z^{k+1}, z^{k}-z^{k+1}\right\rangle\right] .
\end{array}
\end{aligned}
$$

The final quantity in brackets is simply $\left\|M x^{k+1}-z^{k}\right\|^{2}$, so it follows that

$$
\begin{align*}
\left\|p^{k+1}-p^{*}\right\|^{2} & +c^{2}\left\|z^{k+1}-z^{*}\right\|^{2} \\
= & \left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}-c^{2}\left\|M x^{k+1}-z^{k}\right\|^{2} \\
& -2 c\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle-2 c\left\langle-M x^{k+1}-\left(-M x^{*}\right), \mu^{k+1}-p^{*}\right\rangle \tag{5.103}
\end{align*}
$$

Now we consider $\left\|\boldsymbol{w}^{k+1}-\boldsymbol{x}^{*}\right\|^{2}$. Since $\boldsymbol{w}^{k+1}=\boldsymbol{w}^{k}-c \boldsymbol{y}^{k+1}$, it follows that

$$
\begin{align*}
& \left\|\boldsymbol{w}^{k+1}-\boldsymbol{x}^{*}\right\|^{2}  \tag{5.104}\\
= & \left\|\boldsymbol{w}^{k}-c \boldsymbol{y}^{k+1}-\boldsymbol{x}^{*}\right\|^{2} \\
= & \left\|\boldsymbol{w}^{k}-\boldsymbol{x}^{*}\right\|^{2}-2 c\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{*}, \boldsymbol{y}^{k+1}\right\rangle+c^{2}\left\|\boldsymbol{y}^{k+1}\right\|^{2} \\
= & \left\|\boldsymbol{w}^{k}-\boldsymbol{x}^{*}\right\|^{2}-2 c\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle-2 c\left\langle\boldsymbol{x}^{k+1}-\boldsymbol{x}^{*}, \boldsymbol{y}^{k+1}\right\rangle+c^{2}\left\|\boldsymbol{y}^{k+1}\right\|^{2} \tag{5.105}
\end{align*}
$$

Adding (5.103) to 5.105 and grouping terms yields

$$
\begin{align*}
& \left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k+1}-\boldsymbol{x}^{*}\right\|^{2} \\
& =\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k}-\boldsymbol{x}^{*}\right\|^{2} \\
& -2 c\left[\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle+\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}-0\right\rangle+\left\langle 0, x^{k+1}-x^{*}\right\rangle\right]  \tag{A}\\
& -2 c\left[\left\langle-M x^{k+1}-\left(-M x^{*}\right), \mu^{k+1}-p^{*}\right\rangle+\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}-0\right\rangle+\left\langle 0, z^{k}-z^{*}\right\rangle\right]  \tag{B}\\
& \quad-2 c\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle  \tag{C}\\
& \quad+c^{2}\left\|\boldsymbol{y}^{k+1}\right\|^{2}-c^{2}\left\|M x^{k+1}-z^{k}\right\|^{2} \tag{5.106}
\end{align*}
$$

Since

$$
\left(y_{1}^{k+1}, 0,-M x^{k+1}\right) \in \partial L_{1}\left(x^{k+1}, z^{k}, \mu^{k+1}\right) \quad\left(0,0,-M x^{*}\right) \in \partial L_{1}\left(x^{*}, z^{*}, p^{*}\right)
$$

it follows from the monotonicity of $\partial L_{1}$ that

$$
\left\langle-M x^{k+1}-\left(-M x^{*}\right), \mu^{k+1}-p^{*}\right\rangle+\left\langle x^{k+1}-x^{*}, y_{1}^{k+1}-0\right\rangle+\left\langle 0, z^{k}-z^{*}\right\rangle \geq 0
$$

Similarly, from

$$
\left(0, y_{2}^{k+1}, z^{k+1}\right) \in \partial L_{2}\left(x^{k+1}, z^{k+1}, p^{k+1}\right) \quad\left(0,0, z^{*}\right) \in \partial L_{2}\left(x^{*}, z^{*}, p^{*}\right)
$$

the monotonicity of $\partial L_{2}$ yields

$$
\left\langle z^{k+1}-z^{*}, p^{k+1}-p^{*}\right\rangle+\left\langle z^{k+1}-z^{*}, y_{2}^{k+1}-0\right\rangle+\left\langle 0, x^{k+1}-x^{*}\right\rangle \geq 0
$$

For (C), it is always true that

$$
-2 c\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle \leq 2 c\left|\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle\right|
$$

Applying $\left(\overline{\mathrm{B}^{\prime}}\right)\left(\overline{\mathrm{A}^{\prime}}\right)$ and $\left(\overline{\mathrm{C}^{\prime}}\right)$ to 5.106$)$, we obtain

$$
\begin{align*}
\left\|p^{k+1}-p^{*}\right\|^{2}+ & c^{2}
\end{align*} \quad\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k+1}-\boldsymbol{x}^{*}\right\|^{2} .
$$

Multiplying the relative error condition (5.99) by $c^{2}$ yields

$$
\begin{equation*}
2 c\left|\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle\right|+c^{2}\left\|\boldsymbol{y}^{k+1}\right\|^{2} \leq c^{2} \sigma\left\|M x^{k+1}-z^{k}\right\|^{2} \tag{5.108}
\end{equation*}
$$

Combining 5.108 and 5.106, we obtain the inequality that is the key to the convergence analysis:

$$
\begin{align*}
& \left\|p^{k+1}-p^{*}\right\|^{2}+c^{2}\left\|z^{k+1}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k+1}-\boldsymbol{x}^{*}\right\|^{2} \\
& \leq\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k}-\boldsymbol{x}^{*}\right\|^{2} \\
& \quad-c^{2}(1-\sigma)\left\|M x^{k+1}-z^{k}\right\|^{2} \tag{5.109}
\end{align*}
$$

Since (5.109) is true for all $k \geq 1$, the following conclusions can be drawn:

- The sequence $\left\{\left\|p^{k}-p^{*}\right\|^{2}+c^{2}\left\|z^{k}-z^{*}\right\|^{2}+\left\|\boldsymbol{w}^{k}-\boldsymbol{x}^{*}\right\|^{2}\right\}$ is non-increasing and convergent because it is bounded below by 0. Thus Sequences $\left\{p^{k}\right\},\left\{z^{k}\right\}$ and $\left\{\boldsymbol{w}^{k}\right\}$ are bounded and $\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)$ is Fejér monotone to $\left(\boldsymbol{x}^{*}, c z^{*}, p^{*}\right)$.
- By [7, Proposition 5.4], we have that $\left\{\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}^{*}, c z^{*}, p^{*}\right)\right\|^{2}\right\}$ is convergent.
- Since $(1-\sigma)>0$, the inequality (5.109) implies that

$$
\left\|\left(\boldsymbol{w}^{k+1}, c z^{k+1}, p^{k+1}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\| \leq\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\|
$$

- By summing (5.109) over $k$, we deduce that $\left\{\left\|M x^{k+1}-z^{k}\right\|^{2}\right\}$ is a summable sequence and hence that $M x^{k+1}-z^{k} \rightarrow 0$. From the relative error condition (5.99), we know that $\left\{\left\|\boldsymbol{y}^{k}\right\|^{2}\right\}$ and $\left\{\left|\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle\right|\right\}$ are both summable, thus $\left\{\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle\right\}$ is also summable and thus $\boldsymbol{y}^{k} \rightarrow 0,\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle \rightarrow$ 0 .
- Writing

$$
\left\langle\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle=\left\langle\boldsymbol{w}^{k}, \boldsymbol{y}^{k+1}\right\rangle-\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k}\right\rangle,
$$

we may reason as follows: since $\left\{\boldsymbol{w}^{k}\right\}$ is bounded and $\boldsymbol{y}^{k} \rightarrow 0$, we may conclude that $\left\langle\boldsymbol{w}^{k}, \boldsymbol{y}^{k+1}\right\rangle \rightarrow 0$. Since we have already established the limit $\left\langle\boldsymbol{w}^{k}-\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}\right\rangle \rightarrow 0$, it follows that $\left\langle\boldsymbol{x}^{k}, \boldsymbol{y}^{k}\right\rangle \rightarrow 0$.

The role of the sequence $\left\{\boldsymbol{w}^{k}\right\}$ is similar to that of the sequence $\left\{w^{k}\right\}$ in 30 : it may be considered as accumulating the total subgradient error "drift" over the course of the algorithm.

Proposition 32. If the sequences $\left\{x^{k}\right\},\left\{z^{k}\right\},\left\{p^{k}\right\}$, and $\left\{\boldsymbol{w}^{k}\right\}$ conform to the recursion (5.97)-(5.101), then

$$
M x^{k+1}-z^{k+1} \rightarrow 0 \quad z^{k+1}-z^{k} \rightarrow 0 \quad p^{k+1}-p^{k} \rightarrow 0
$$

Proof. By Lemma 20, we have

$$
\begin{equation*}
\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \leq\left\|M x^{k+1}-z^{k}\right\|^{2}+\frac{2}{c}\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle \tag{5.110}
\end{equation*}
$$

Expanding the inner product in 5.110, we write

$$
\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle=\left\langle y_{2}^{k+1}, z^{k+1}\right\rangle-\left\langle y_{2}^{k+1}, z^{k}\right\rangle-\left\langle y_{2}^{k}, z^{k+1}\right\rangle+\left\langle y_{2}^{k}, z^{k}\right\rangle
$$

By Proposition 31, $\left\{z^{k}\right\}$ is bounded, $M x^{k+1}-z^{k} \rightarrow 0$ and $\left\{y_{2}^{k}\right\} \rightarrow 0$ so we may assert that

$$
\left\|M x^{k+1}-z^{k}\right\|^{2}+\left\langle y_{2}^{k+1}-y_{2}^{k}, z^{k+1}-z^{k}\right\rangle \rightarrow 0
$$

It follows from (5.110) that $\left\|M x^{k+1}-z^{k+1}\right\|^{2}+\left\|z^{k}-z^{k+1}\right\|^{2} \rightarrow 0$. The first two claims follow immediately, and the last claim is then a consequence of $p^{k+1}-p^{k}=$ $c\left(M x^{k+1}-z^{k+1}\right)$.

Following pattern established in the previous subsection, next we show that all limit points of $\left\{p^{k}\right\}$ are dual solutions.

Proposition 33. All accumulation points of the sequence $\left\{p^{k}\right\}$ generated by the recursions (5.97)-(5.101) are solutions to the dual problem (D), and all limit points of $\left\{x^{k}\right\}$ are solutions to the primal problem $(\overline{\mathrm{P}})$. Furthermore, if $\left(x^{*}, z^{*}\right)$ is an optimal solution to $(\mathrm{P})$, we have

$$
\lim _{k \rightarrow \infty}\left\{f\left(x^{k}\right)+g\left(z^{k}\right)\right\}=f\left(x^{*}\right)+g\left(z^{*}\right)
$$

Proof. By Propositions 31 and 32, we know that the sequences $\left\{M x^{k}\right\}$ and $\left\{p^{k}\right\}$ generated by the recursions (5.97)-(5.101) are bounded and

$$
z^{k+1}-z^{k} \rightarrow 0 \quad M x^{k+1}-z^{k+1} \rightarrow 0 \quad\left\langle\boldsymbol{y}^{k+1}, \boldsymbol{x}^{k+1}\right\rangle \rightarrow 0
$$

We may then apply Proposition 21 to obtain that every limit point of $\left\{p^{k}\right\}$ is a dual solution.

We now consider the primal sequence $\left\{x^{k}\right\}$. From Lemma 17, we conclude that all its limit points are primal solutions. The final claimed inequality follows from Proposition 22 ,

We close this section by showing that the sequence $\left\{p^{k}\right\}$ defined by the recursions (5.97)-5.101) converges to a dual solution.

Proposition 34. If the set of KKT points for (P) is nonempty, then the sequence $\left\{p^{k}\right\}$ defined by the recursions (5.97)-(5.101) converges to a solution of (D).

Proof. From Proposition 31, we have

$$
\left\|\left(\boldsymbol{w}^{k+1}, c z^{k+1}, p^{k+1}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\| \leq\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\|
$$

so $\left\{\left\|\left(\boldsymbol{w}^{k}, c z^{k}, p^{k}\right)-\left(\boldsymbol{x}, c z^{*}, p^{*}\right)\right\|\right\}$ is nonincreasing. For all $k \geq 0$ let $\alpha_{k}=\left(\boldsymbol{w}^{k+1}, c z^{k+1}\right)$ and $\beta_{k}=p^{k}$, and define $V$ to be the set all solutions of (D). By Proposition 33, we know that every limit point of $\left\{p^{k}\right\}$ belongs to $V$. Therefore, Lemma 14 implies that $\left\{p^{k}\right\}$ converges to a solution of (D).

Propositions 34 and 33 together summarize the convergence properties of the recursions (5.97)-5.101). We next consider to the implementation of concrete algorithms that conform to these recursions.

### 5.6 Partially inexact ADMM with relative error criteria

In many applications of the ADMM, subproblem (1.10) may be solved exactly. For instance, $l_{1}$-regularized regression is a very common technique for high-dimensional
statistical learning problems, and in problems with $l_{1}$ regularization one typically has $g(z)=\nu\|z\|_{1}$ for some $\nu>0$. In this case, the exact solution to the $g$ subproblem 1.10 can be easily computed by the soft thresholding operator given by

$$
\begin{equation*}
z_{i}^{k+1}=\operatorname{sgn}\left(M_{i}^{\top} x_{i}^{k+1}+\frac{1}{c} p_{i}^{k}\right) \max \left\{0,\left|M_{i}^{\top} x_{i}^{k+1}+\frac{1}{c} p_{i}^{k}\right|-\frac{\nu}{c}\right\} \quad i=1, \ldots, m, \tag{5.111}
\end{equation*}
$$

where the vector $M_{i}$ denotes the $i^{\text {th }}$ row of the matrix $M$. In such applications, inexact computation need not be applied to the second subproblem. We now give a concrete algorithm that applies (5.97)-(5.101) to such situations:

```
Algorithm 5.6.1 Partially inexact ADMM with relative error criteria
    initialization: Pick scalar parameters \(c>0\), and \(\sigma \in[0,1)\), along with initial
    points \(w^{0}, p^{0}, z^{0} \in \mathbb{R}^{m}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\)
            by taking \(\left(x^{k, l}, y^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until \(\frac{2}{c}\left|\left\langle w^{k}-x^{k, l}, y^{k, l}\right\rangle\right|+\left\|y^{k, l}\right\|^{2} \leq \sigma\left\|M x^{k, l}-z^{k}\right\|^{2}\)
        \(x^{k+1}=x^{k, l}, y^{k+1}=y^{k, l}\)
        \(z^{k+1}=\arg \min _{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
        \(w^{k+1}=w^{k}-c y^{k}\)
```

until Overall convergence

Proposition 35. Suppose Assumption 1 holds. If the inner loop of (overl) of Algorithm 5.6.1 always terminates in a finite number of iterations, all limit points of the sequence $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ generated by the algorithm are the KKT points of $(\overline{\mathrm{P}}),\left\{p^{k}\right\}$ converges to an optimal solution of the dual problem (D), and $\lim _{k \rightarrow \infty} f\left(x^{k}\right)+g\left(z^{k}\right)=$ $f\left(x^{*}\right)+g\left(z^{*}\right)$, where $\left(x^{*}, z^{*}\right)$ is any optimal solution of $(\mathrm{P})$. If the inner loop cycles infinitely over $l$ for some iteration $k=\bar{k}$, then for all limit points $x^{\infty}$ of $\left\{x^{\bar{k}, l}\right\}_{l=1}^{\infty}$, we
have that $\left(x^{\infty}, z^{\bar{k}}, p^{\bar{k}}\right)$ is a KKT point of $(\mathrm{P})$.

Proof. Consider the case that inner loop always terminates finitely. It is easily verified that Algorithm 5.6.1 generates sequences conforming to the recursions (5.97)-5.101) with $y_{2}^{k}=0$ for all $k$. The relevant conclusions then follow from Propositions 33 and 34 .

The remaining case is that the first loop executes an infinite number of times at outer iteration $k=\bar{k}$, in which we must have

$$
\lim _{l \rightarrow \infty} M x^{\bar{k}, l}-z^{\bar{k}}=0 \quad \lim _{l \rightarrow \infty} y^{\bar{k}, l}=0
$$

and by Corollary 19 all limit points of $\left\{\left(x^{\bar{k}, l}, z^{\bar{k}}, p^{\bar{k}}\right)\right\}$ are the KKT points.
Regarding the existence of limit points of $\left\{x^{\bar{k}, l}\right\}_{l=1}^{\infty}$ in the case of the inner loop running indefinitely, we may appeal to Lemma 5s specifically, if the solution set of the $x$ subproblem is bounded, then $\left\{x^{\bar{k}, l}\right\}_{l=1}^{\infty}$ is bounded and hence must have limit points. If the solution of the $x$ subproblem is unique, then $\left\{x^{\bar{k}, l}\right\}_{l=1}^{\infty}$ must converge to it, so the unique limit $x^{\infty}$ of $\left\{x^{\bar{k}, l}\right\}_{l=1}^{\infty}$ is a solution of (P) by Proposition 35 .

### 5.7 Complete form of relative-error algorithm with both minimizations inexact

While the partially inexact Algorithm 5.6.1 covers many applications, it could still be desirable to consider the possibility of solving both subproblems inexactly. For such cases, we propose the following algorithm:

```
Algorithm 5.7.1 Inexact ADMM with relative error criteria
    initialization: Pick scalar parameters \(c>0, \sigma \in[0,1), \tau, \alpha \in(0,1)\), and \(\beta, \gamma>0\),
    along with initial points \(x^{1}, w_{1}^{1} \in \mathbb{R}^{m}\) and \(p^{0}, z^{0}, w_{2}^{0} \in \mathbb{R}^{m}\), with \(M x^{1} \neq z^{0}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        \(l \leftarrow 0\)
        repeat \(\{\) for \(t=0,1, \ldots\}\)
            repeat
\[
\begin{aligned}
& \quad l \leftarrow l+1 \\
& \qquad\left(z^{k, l}, y_{2}^{k, l}\right)=\mathcal{G}\left(p^{k}, x^{k+1}, c, z^{k}, l\right) \\
& \text { until } \frac{2}{c}\left|\left\langle w_{2}^{k}-z^{k, l}, y_{2}^{k, l}\right\rangle\right|+\left\|y_{2}^{k, l}\right\|^{2} \leq \alpha^{t}(1-\tau) \sigma\left\|M x^{k+1}-z^{k}\right\|^{2} \\
& p^{k, l}=p^{k}+c\left(M x^{k+1}-z^{k, l}\right) \\
& w_{2}^{k, l}=w_{2}^{k}-c y_{2}^{k, l} \\
& \bar{x}^{k, 0}=x^{k+1} \\
& \text { repeat }\left\{\text { for } l^{\prime}=0,1, \ldots\right\}
\end{aligned}
\]
\[
\begin{aligned}
& \quad\left(x^{k, t, l^{\prime}}, y_{1}^{k, t, l^{\prime}}\right)=\mathcal{F}\left(p^{k, l}, z^{k, l}, c, \bar{x}^{k, t-1}, l^{\prime}\right) \\
& \text { accept } \leftarrow\left(\frac{2}{c}\left|\left\langle w_{1}^{k}-x^{k, t, l^{\prime}}, y_{1}^{k, t, l^{\prime}}\right\rangle\right|+\left\|y_{1}^{k, t, l^{\prime}}\right\|^{2}<\tau \sigma\left\|M x^{k, t, l^{\prime}}-z^{k, l}\right\|^{2}\right) \\
& \text { until accept or }\left(\left\|y_{1}^{k, t, l^{\prime}}\right\|^{2} \leq \beta\left\|y_{2}^{k, l}\right\|^{2} \text { and }\left\|M x^{k, t, l^{\prime}}-z^{k, l}\right\|^{2} \leq \gamma\left\|y_{2}^{k, l}\right\|^{2}\right) \\
& \bar{x}^{k, t}=x^{k, t, l^{\prime}} ; \bar{z}^{k, t}=z^{k, l} ; \bar{p}^{k, t}=p^{k, l} ; \bar{y}_{1}^{k, t}=y_{1}^{k, t, l^{\prime}} ; \bar{y}_{2}^{k, t}=y_{2}^{k, l}
\end{aligned}
\]
```

until accept

$$
x^{k+2}=x^{k, t, l^{\prime}} ; z^{k+1}=z^{k, l} ; p^{k+1}=p^{k, l} ; y_{1}^{k+1}=y_{1}^{k, t, l^{\prime}} ; y_{2}^{k+1}=y_{2}^{k, l} ; w_{2}^{k+1}=w_{2}^{k, l}
$$

$$
w_{1}^{k+2}=w_{1}^{k+1}-c y_{1}^{k+1}
$$

until Overall convergence

Compared to Algorithm 5.4.1, this algorithm has a somewhat complicated structure. The reason is that if one simply modifies Algorithm 5.6.1 so that it includes a second approximation loop in the manner of Algorithm 5.4.1, there are some seemingly unlikely situations in which one of the inner loops can run indefinitely without converging
to a KKT point. In particular, if we were to try to generalize the Algorithm 5.6.1 to include an approximate $z$ minimization, one might at first consider the following algorithm:

```
Algorithm 5.7.2 Insert caption here
    initialization: Pick scalar parameters \(c>0\), and \(\sigma \in[0,1)\), along with initial
    points \(w^{0}, p^{0}, z^{0} \in \mathbb{R}^{m}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            \(\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until \(\frac{2}{c}\left|\left\langle w^{k}-x^{k, l}, y_{1}^{k, l}\right\rangle\right|+\left\|y_{1}^{k, l}\right\|^{2} \leq \sigma\left\|M x^{k, l}-z^{k}\right\|^{2}\)
        \(x^{k+1}=x^{k, l}, y_{1}^{k+1}=y_{1}^{k, l}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            \(\left(z^{k, l}, y_{2}^{k, l}\right)=\mathcal{G}\left(p^{k}, x^{k+1}, c, z^{k}, l\right)\)
        until \(\frac{2}{c}\left|\left\langle w_{2}^{k}-z^{k, l}, y_{2}^{k, l}\right\rangle\right|+\left\|y_{2}^{k, l}\right\|^{2} \leq(1-\tau) \sigma\left\|M x^{k+1}-z^{k, l}\right\|^{2}\)
        \(z^{k+1}=z^{k, l}, y_{2}^{k}=y_{2}^{k, l}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
        \(\boldsymbol{w}^{k+1}=\boldsymbol{w}^{k}-c \boldsymbol{y}^{k}\)
    until Overall convergence
```

Unfortunately, it is conceivable that the second loop over $l$ in the above sequence could become "trapped" in a situation in an infinite loop if both sides of the termination condition $\frac{2}{c}\left|\left\langle w_{2}^{k}-z^{k, l}, y_{2}^{k, l}\right\rangle\right|+\left\|y_{2}^{k, l}\right\|^{2} \leq(1-\tau) \sigma\left\|M x^{k+1}-z^{k, l}\right\|^{2}$ converge to zero. If the prior $x$ minimization were exact, this situation would essentially result in the infinite inner loop converging to a solution. Unfortunately, however, if the prior minimization over $x$ may not have been exact. Our strategy in this case is return to the prior $x$ minimization and tighten its accuracy, and then revisit the $z$ minimization; if this procedure is managed properly, we are able to show convergence in all cases. However, to be able to express the algorithm in a conventional block structure, we
reorder its components so that it appears to start with the $z$-minimization step, resulting in Algorithm 5.7.1. We now establish the convergence properties of the algorithm:

Lemma 36. If the inner loops of Algorithm 5.7.1 always terminate finitely, the algorithm produces sequences conforming to the recursions (5.97)-(5.101).

Proof. Consider the loop termination conditions

$$
\begin{align*}
& \frac{2}{c}\left|\left\langle w_{1}^{k}-x^{k+1}, y_{1}^{k+1}\right\rangle\right|+\left\|y_{1}^{k+1}\right\|^{2} \leq \tau \sigma\left\|M x^{k+1}-z^{k}\right\|^{2}  \tag{5.112}\\
& \frac{2}{c}\left|\left\langle w_{2}^{k}-z^{k+1}, y_{2}^{k+1}\right\rangle\right|+\left\|y_{2}^{k+1}\right\|^{2} \leq \alpha^{t}(1-\tau) \sigma\left\|M x^{k+1}-z^{k}\right\|^{2} \tag{5.113}
\end{align*}
$$

Using that $\tau \in[0,1], \alpha \in(0,1)$, and $t$ is a nonnegative integer, (5.113) implies that

$$
\begin{equation*}
\frac{2}{c}\left|\left\langle w_{2}^{k}-z^{k+1}, y_{2}^{k+1}\right\rangle\right|+\left\|y_{2}^{k+1}\right\|^{2} \leq(1-\tau) \sigma\left\|M x^{k+1}-z^{k}\right\|^{2} \tag{5.114}
\end{equation*}
$$

Adding (5.112) and (5.114), we obtain the relative error condition (5.99). Considering the updates of $p^{k, l}$ and $w_{2}^{k, l}$ after the completion of the $\mathcal{G}$ loop, and the updates to $p^{k+1}$ and $\boldsymbol{w}^{k+1}$ near the end of the outer loop, it may be verified that the algorithm produces sequences conforming to (5.97)-5.101).

The best choice of $\tau$ is clearly application-dependent and depends on the relative difficulty of solving the $x$ and $z$ subproblems. In the case that the $z$ subproblem is easily solved exactly, one could choose $\tau=1$, but in this case it would be much simpler to just use Algorithm 6.1.2 instead.

In addition to Assumption 2, we make the following mild assumption, as in Lemma 23 ,

Assumption 37. For any $p, z \in \mathbb{R}^{m} \times \mathbb{R}^{m}, c>0, x \in \mathbb{R}^{n}$, the sequence $\left\{\mathcal{F}_{1}\left(p^{\tau}, z^{\tau}, c, x, l\right)\right\}_{l=1}^{\infty}$ is bounded.

Note that all sequences $\left\{z^{l}\right\}$ generated by the $g$ approximation scheme $\mathcal{G}$ must be bounded, because they must converge to the exact solution of the subproblem.

Lemma 38. In Algorithm 5.7.1, the loop over l always terminates finitely.

Proof. By the initialization step of the algorithm, $M x^{k+1}-z^{k}$ is nonzero for $k=$ 0 . By assumption it is nonzero for $k=0$. In subsequent iterations $k \geq 1$, the previous iteration must have completed with accept becoming true. Due to the strict inequality in the definition of accept, we must also have $M x^{k+1}-z^{k}$ in this case. Therefore, we have $\left\|M x^{k+1}-z^{k}\right\|>0$ for all $k$, and since the $\mathcal{G}$-procedure guarantees $\lim _{l \rightarrow \infty} y_{2}^{k, l}=0$, the termination condition for the $l$ loop must always hold eventually.

While the $l$ loop must always terminate finitely, it is possible that the loop over $l^{\prime}$ may run indefinitely. In this case, we are in essentially the same situation as in Algorithm 5.6.1, and we can show convergence of the inner loop to a KKT point

Lemma 39. If the loop over $l^{\prime}$ in Algorithm 5.7.1 does not terminate, then ( $\left.x^{k, t, \infty}, p^{k, l}, z^{k, l}\right)$ is a KKT point of ( P ) for any limit point $x^{k, t, \infty}$ of $\left\{x^{k, t, l^{\prime}}\right\}_{l^{\prime}=0}^{\infty}$.

Proof. By the properties of the $\mathcal{F}$-procedure, the $l^{\prime}$ loop can run indefinitely only if $y_{2}^{k, l}=0$ and $\lim _{l^{\prime} \rightarrow \infty} M x^{k, t, l^{\prime}}=z^{k, l}$. In any other situation, the $l^{\prime}$ loop must terminate finitely because the boundedness of $\left\{x^{k, t, l^{\prime}}\right\}_{l^{\prime}=0}^{\infty}$ ensures that the left-hand side of accept converges to zero, while the condition's right-hand side is nonnegative and does not converge to zero. If $\lim _{l^{\prime} \rightarrow \infty} M x^{k, t, l^{\prime}}=z^{k, l}$ but $y_{2}^{k, l} \neq 0$, then the alternative conditions $\left\|y_{1}^{k, t, l^{\prime}}\right\|^{2} \leq \beta\left\|y_{2}^{k, l}\right\|^{2}$ and $\left\|M x^{k, t, l^{\prime}}-z^{k, l}\right\|^{2} \leq \gamma\left\|y_{2}^{k, l}\right\|^{2}$ must eventually hold for sufficiently large $l^{\prime}$.

If $y_{2}^{k, l}=0$, then $z^{k, l}$ is an exact minimizer of last $g$ subproblem. Furthermore, the condition $\lim _{l^{\prime} \rightarrow \infty} M x^{k, t, l^{\prime}}=z^{k, l}$ forces $\lim _{l^{\prime} \rightarrow \infty} y_{1}^{k, t, l^{\prime}}=0$. For any limit point $x^{k, t, \infty}$ of $\left\{x^{k, t, l^{\prime}}\right\}_{l^{\prime}=0}^{\infty}$, Corollary 19 then implies that $\left(x^{k, t, \infty}, p^{k, l}, z^{k, l}\right)$ is a KKT point.

We may refer to Lemma 5 for conditions under which $\left\{x^{k, t, l^{\prime}}\right\}_{l^{\prime}=0}^{\infty}$ must have limit points or must converge.

Next, we consider the scenario that both of the loops over $l$ and $l^{\prime}$ terminate finitely, but the loop over $t$ runs indefinitely. In this case, we again obtain convergence to a KKT point.

Lemma 40. In Algorithm 5.7.1, if the loop over $t$ runs indefinitely for some $k$, then any limit point of $\left\{\bar{x}^{k, t}, \bar{z}^{k, t}, \bar{p}^{k, t}\right\}_{t=0}^{\infty}$ is a KKT point of ( P$)$.

Proof. The loop over $t$ terminates finitely unless accept never holds. However, for the $t$ loop to run indefinitely, then its contained loop over $l^{\prime}$ must always terminate finitely, indicating that the alternative $l^{\prime}$ termination conditions $\left\|y_{1}^{k, t, l^{\prime}}\right\|^{2} \leq$ $\beta\left\|y_{2}^{k, l}\right\|^{2}$ and $\left\|M \bar{x}^{k, t, l^{\prime}}-\bar{z}^{k, l}\right\|^{2} \leq \gamma\left\|y_{2}^{k, l}\right\|^{2}$ hold. By construction, we have $\alpha \in(0,1)$ and $\alpha^{t} \rightarrow 0$, so we know that $\lim _{t \rightarrow \infty} \bar{y}_{2}^{k, t}=0$. Consequently, the alternative condition in loop over $l^{\prime}$ ensures that $\lim _{t \rightarrow \infty} \bar{y}_{1}^{k, t}=0$ and $\lim _{t \rightarrow \infty} M \bar{x}^{k, t}-\bar{z}^{k, t}=0$. Now, we know that $\lim _{t \rightarrow \infty} \bar{y}_{1}^{k, t}=0$ since $\left\{\bar{y}_{1}^{k, t}\right\}_{t=0}^{\infty}$ is a subsequence of $\left\{y_{1}^{k, l}\right\}_{l=1}^{\infty}$, which must converge to 0 . Furthermore, $\left\{\bar{z}^{k, t}\right\}_{t=0}^{\infty}$ must be convergent since it is a subsequence of the convergent sequence $\left\{z^{k, l}\right\}_{l=1}^{\infty}$. The convergence of $\left\{z^{k, l}\right\}_{l=1}^{\infty}$ implies the convergence of $\left\{p^{k, l}\right\}_{l=1}^{\infty}$. The conclusion then holds by Lemma 18 .

Finally we are ready to show the convergence of Algorithm 5.7.1.

Proposition 41. Under Assumptions 37, Algorithm 5.7.1 converges to a KKT point of (P), in one of the following possible ways:

1. The loop over $k$ runs indefinitely, in which case every limit point of the sequence $\left\{\left(x^{k}, z^{k}, p^{k}\right)\right\}$ generated by the algorithm is a KKT points of (P), $\left\{p^{k}\right\}$ converges to an optimal solution of the dual problem (D), and $\lim _{k \rightarrow \infty} f\left(x^{k}\right)+g\left(z^{k}\right)=f\left(x^{*}\right)+g\left(z^{*}\right)$, where $\left(x^{*}, z^{*}\right)$ is any optimal solution of (P).
2. The loop over $l^{\prime}$ runs indefinitely for some $k$, $t$, and $l$. In this case, we have that $\left(x^{k, t, \infty}, p^{k, l}, z^{k, l}\right)$ is a KKT point of (P) for any is any limit point $x^{k, t, \infty}$ of $\left\{x^{k, t, l^{\prime}}\right\}_{l^{\prime}=0}^{\infty}$.
3. The loop over $t$ runs indefinitely for some $k$, in which case every limit point of the sequence $\left\{\bar{x}^{k, t}, \bar{z}^{k, t}, \bar{p}^{k, t}\right\}_{t=0}^{\infty}$ is a KKT point of $(\mathrm{P})$.

Proof. By Lemma 38, the $l$ loop always terminates finitely, so the only possibilities are infinite loops over $k, l^{\prime}$, or $t$. Lemma 39 guarantees the claimed convergence in the case of an infinite $l^{\prime}$ loop. Similarly, Lemma 40 guarantees the claimed convergence if the $t$ loop does not terminate finitely. In the case of an infinite $k$ loop, Lemma 36 asserts that the sequences generated by the algorithm obey the recursions 5.97)(5.101) for all $k \geq 1$ (although not necessarily $k=0$, but that is of no consequence), and therefore we may apply Propositions 33 and 34 to obtain the claimed result.

## Chapter 6 Numerical Tests

### 6.1 Comparison algorithms

In summary, we have developed the following three new approximate versions of the ADMM algorithm:

Version 1. The partially inexact ADMM derived in Chapter 4 by combining operator splitting theory with a relative-error inexact proximal point algorithm. This variant requires that the $g$ subproblem be conveniently exactly solvable, and also essentially requires $M=I$. We use admm_primDR to denote this variant.

Version 2. The algorithm developed in Section 5.4. This version does not require $M=I$ and allows both subproblems to be solved inexactly. We derived this variant by modifying the standard Lagrangian splitting analysis to use absolutely summable error conditions. We will use admm_abssum to denote this version. So long as $\left\|x^{k, l}\right\| \leq \beta_{1}$ and $\left\|z^{k, l}\right\| \leq \beta_{2}$ throughout this algorithm, it coincides with the simpler Algorithm 4.1.1 of Section 4.1. However, the existing convergence theory for that algorithm requires that both subproblems be strongly convex; this is guaranteed in our context, for example, if $M=I$.

Version 3. The algorithm developed in Section 5.5 Like the second version, this variant is also derived from a Lagrangian splitting analysis . It does not require $M=I$ and allow both subproblems to be solved inexactly, but with relative rather than absolute error criteria. We will use admm_relerr to represent this type of algorithm.

On three classes of test problems, we have experimentally studied the performance of these three kinds of algorithms, as well as that of the "exact" ADMM, which we denote by admm_exact. For all three classes of test problems, the $g$ minimization is easy to perform exactly and $M=I$, meaning that all of our algorithms are applicable. For the "exact" ADMM, we still use an iterative solver for the $f$ subproblem, resulting in an algorithm that can be stated as follows:

```
Algorithm 6.1.1 Exact version of ADMM: admm_exact
    initialization: Pick \(c>0\) and initial points \(p^{0}, z^{0} \in \mathbb{R}^{m}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\)
            by taking \(\left(x^{k, l}, y^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until Inner loop convergence
        \(x^{k+1}=x^{k, l}\)
        \(y^{k+1}=y^{k, l}\)
        \(z^{k+1}=\arg \min _{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
until Overall convergence
```

For admm_primDR, we simply use Algorithm 4.4.1. In the case of admm_abssum, we specialize Algorithm 4.1.1 to the case that the $g$ minimization is exact, resulting in the following algorithm:

```
Algorithm 6.1.2 Partially inexact absolute-error ADMM: admm_abssum
    initialization: Pick \(c>0\) and initial points \(p^{0}, z^{0} \in \mathbb{R}^{m}\). Select positive and
    summable sequence \(\left\{\epsilon_{k}\right\}\)
    repeat \(\{\) for \(k=0,1,2, \ldots\}\)
        repeat \(\{\) for \(l=1,2, \ldots\}\)
            Improve the solution to \(x^{k+1} \approx \arg \min _{x}\left\{f(x)+\left\langle p^{k}, M x\right\rangle+\frac{c}{2}\left\|M x-z^{k}\right\|^{2}\right\}\)
            by taking \(\left(x^{k, l}, y_{1}^{k, l}\right)=\mathcal{F}\left(p^{k}, z^{k}, c, x^{k}, l\right)\)
        until \(\left\|y^{k, l}\right\| \leq \epsilon_{k+1}\)
        \(x^{k+1}=x^{k, l}\)
        \(y^{k+1}=y^{k, l}\)
        \(z^{k+1}=\arg \min _{z}\left\{g(z)-\left\langle p^{k}, z\right\rangle+\frac{c}{2}\left\|M x^{k+1}-z\right\|^{2}\right\}\)
        \(p^{k+1}=p^{k}+c\left(M x^{k+1}-z^{k+1}\right)\)
```

until Overall convergence

We are justified in using this version of the absolutely summable algorithm because $M=I$, which has full column rank, making the $f$ subproblem strongly convex.

Finally, for admm_relerr, we simply used Algorithm 5.6.1, since we are only testing the case in which the $g$ minimization is exact.

### 6.2 Termination criteria and algorithm parameters

For all the algorithms and problem classes, we used the same condition for "overall convergence", namely

$$
\begin{equation*}
\operatorname{dist}_{\infty}\left(0, \partial_{x}[f(x)+g(M x)]_{x=x^{k}}\right) \leq \epsilon \tag{6.1}
\end{equation*}
$$

where $\operatorname{dist}_{\infty}(t, S)=\inf \left\{\|t-s\|_{\infty} \mid s \in S\right\}$, and $\epsilon$ is a tolerance parameter we set to $10^{-6}$. We also imposed a maximum of 10,000 outer iterations. In admm_primDR and admm_relerr, we use $\sigma=0.99$, For admm_exact, we used an inner loop convergence
criterion of

$$
\begin{equation*}
\left\|y^{k, l}\right\| \leq \frac{\epsilon}{10} \tag{6.2}
\end{equation*}
$$

which follows customary practice in general-purpose augmented Lagrangian solvers. However, we also set a limit of 200 inner loop $(l)$ iterations.

Once admm_relerr nears the eventual solution, it is possible for its inner-loop termination condition $\frac{2}{c}\left|\left\langle w^{k}-x^{k, l}, y^{k, l}\right\rangle\right|+\left\|y^{k, l}\right\|^{2} \leq \sigma\left\|M x^{k, l}-z^{k}\right\|^{2}$ to be significantly more restrictive than (6.2). This phenomenon can lead to an excessive number of inner iterations and final solutions that are far more accurate than demanded by our overall termination condition (6.1). In our tests, we avoided this behavior by using a "hybrid" approach in which we terminated the inner loop of admm_relerr and admm_primDR as soon as either its relative error criterion or (6.2) holds.

For admm_abssum, we need to select a positive absolutely summable sequence $\left\{\epsilon_{k}\right\}$. After some experimentation, we selected $\epsilon_{k}=k^{-1.5}$, where $k$ is the outer iteration counter.

For each class of test problems, we tried admm_exact with various values of the penalty parameter $c$ and selected the one that appeared to have the best performance over the class of problems. We then used this same value of $c$ when applying the inexact algorithms to the same class of problems. For all algorithms, we kept $c$ constant throughout each run.

### 6.2.1 LASSO regression

A simple and common problem class that fits readily into the form (1.4) is the "LASSO" 79] or "compressed sensing" problem. This problem is an $L_{1}$-regularized version of linear regression, taking the form

$$
\begin{equation*}
\min _{x \in \mathbb{R}^{n}} \frac{1}{2}\|A x-b\|^{2}+\nu\|x\|_{1} \tag{6.3}
\end{equation*}
$$

where $A$ is a $p \times n$ matrix, $b \in \mathbb{R}^{p}$ and $\nu>0$ is a given scalar regularization parameter. The goal of this model is find an approximate solution to the linear equations $A x=b$,
with a preference for making the solution $x \in \mathbb{R}^{n}$ sparse. Letting $f(x)=\frac{1}{2}\|A x-b\|^{2}$, $g(z)=\nu\|z\|_{1}$, and $M=I$ the LASSO problem (6.3) may be written as

$$
\begin{array}{ll}
\min & f(x)+g(z)  \tag{6.4}\\
\text { s.t. } & x=z .
\end{array}
$$

Applying the ADMM to (6.4), the $x$-minimization subproblem (1.9) reduces to solving a system of linear equations involving the matrix $A^{\top} A+c I$ :

$$
\begin{equation*}
x^{k+1}=\left(A^{\top} A+c I\right)^{-1}\left(A^{\top} b+c z^{k}-p^{k}\right) \tag{6.5}
\end{equation*}
$$

Note that the $A^{\top} A+c I$ is always invertible since $c>0$. We use the conjugate gradient method as given in [64, Algorithm 5.2] to solve the linear system (6.5).

The $z$-minimization subproblem 1.10 reduces to the soft thresholding operator (5.111) with $M=I$ :

$$
\begin{equation*}
z_{i}^{k+1}=\operatorname{sgn}\left(x_{i}^{k+1}+\frac{1}{c} p_{i}^{k}\right) \max \left\{0,\left|x_{i}^{k+1}+\frac{1}{c} p_{i}^{k}\right|-\frac{\nu}{c}\right\}, \quad i=1, \ldots, n . \tag{6.6}
\end{equation*}
$$

This calculation is straightforward and requires a constant amount of time per element $z_{i}^{k+1}$. The multiplier update $p^{k+1}=p^{k}+c\left(x^{k+1}-z^{k+1}\right)$ has a similar property, so the $x$ minimization dominates run time of this application if the ADMM if 6.5 is solved exactly.

As in [15, Section 11.1], we scaled $b$ and the columns of $A$ to have unit $\ell_{2}$ norm and set the regularization parameter $\nu$ to $0.1\left\|A^{\top} b\right\|_{\infty}$. For admm_exact, the termination condition for the inner loop (6.2) becomes

$$
\begin{equation*}
\left\|\left(A^{\top} A+c I\right) x^{k+1}-\left(A^{\top} b+c z^{k}-p^{k}\right)\right\| \leq 10^{-7} \tag{6.7}
\end{equation*}
$$

We performed our LASSO tests on four categories of datasets:

Gene expression: Six standard cancer DNA microarray datasets from 22. These instances have dense, wide, and relatively small matrices $A$, with the number of rows $m \in[42,102]$, and the number of columns $n \in[2000,6033]$.

Figure 6.1: LASSO - number of outer iterations for the cancer datasets.


Single-Pixel Camera: Four dense compressed image sensing problems from [25] with $m \in[410,4770]$ and $n \in[1024,16384]$.

Finance: A single large dense financial dataset [49] with $m=30465$ and $n=216842$.

PEMS: A single large, wide, and dense dataset from the California Department of Transportation [56] with $m=267$ and $n=138672$.

Figures 6.1, 6.2, and 6.3 show the number of outer iterations each method takes for each dataset. Generally speaking, admm_abssum takes the most outer iterations to converge, but in a few cases admm_primDR takes more outer iterations than the other methods. The admm_relerr and admm_exact algorithms tend to take about same number of outer iteration to converge.

Figures 6.4, 6.5, and 6.6 depict the cumulative total number of inner iterations for

Figure 6.2: LASSO - number of outer iterations for the image datasets.


Figure 6.3: LASSO - number of outer iterations for PEMS and finance1000.


Figure 6.4: LASSO - total number of inner iterations for the cancer datasets.


Figure 6.5: LASSO - total number of inner iterations for the image datasets.


Figure 6.6: LASSO - total number of inner iterations for PEMS and finance1000.

each of the four algorithms. This total number of iterations is roughly proportional to the total amount of computational effort and run time. All three inexact methods require significantly fewer total inner iterations than admm_exact, but admm_primDR consistently requires the fewest. For the cancer and PEMS datasets, the superiority of admm_primDR is particularly striking.

### 6.2.2 $\quad L_{1}$-regularized logistic regression

Logistic regression with $L_{1}$ regularization has been proposed as a promising method for feature selection in classification problems [35,61]. Given a training dataset consisting of $m$ pairs $\left(a_{i}, b_{i}\right)$, where $a_{i} \in \mathbb{R}^{n}$ is a feature vector and $b_{i} \in\{-1,+1\}$ is the corresponding label, this problem may be written

$$
\begin{equation*}
\min _{\substack{w \in \mathbb{R}^{m} \\ v \in \mathbb{R}}} \sum_{i=1}^{m} \log \left(1+\exp \left(-b_{i}\left(a_{i}^{\top} w+v\right)\right)\right)+\nu\|w\|_{1} \tag{6.8}
\end{equation*}
$$

Here, $w \in \mathbb{R}^{n}$ represents a weighting of the features vector and $v \in \mathbb{R}$ represents a kind of "bias" or intercept. While the $w$ variables carry an $L_{1}$ regularization penalty, $v$ does
not. We may consider the feature input data as forming a matrix $A=\left[a_{1}, \ldots, a_{m}\right]^{\top}$. We set the problem parameter $\nu$ in the same manner as in [15, Section 11.2]. To apply the ADMM to (6.8), we may formulate it as

$$
\min _{(v, w)} f((v, w))+g((v, w))
$$

where $f((v, w))=\sum_{i=1}^{m} \log \left(1+\exp \left(-b_{i}\left(a_{i}^{\top} w+v\right)\right)\right)$ and $g((v, w))=\nu\|w\|_{1}$. As an alternative, one could also use $M \neq I$ with $M(w, v)=w$ to drop $v$ from the $z$ vector (however, doing so would not conform to the convergence-proof assumptions for admm_primDR). Here, the convex function $f$ is known as the logistic loss function. We may simplify its form slightly by setting $a_{i}^{\prime}=b_{i} a_{i}$ for $i=1, \ldots, m$, yielding

$$
\begin{equation*}
f((v, w))=\sum_{i=1}^{m} \log \left(1+\exp \left(-a_{i}^{\prime \top} w-b_{i} v\right)\right) \tag{6.9}
\end{equation*}
$$

We assemble the vectors $a_{i}^{\prime}$ into a matrix $A^{\prime}=\left[a_{1}^{\prime}, \ldots, a_{m}^{\prime}\right]^{\top}$ which we normalize in the same manner as for LASSO problems. The inner-loop termination conditions of the admm_abssum and admm_relerr algorithms require subgradient information about $f$. Such information is readily available because the logistic loss function is differentiable. Using (6.9), we obtain

$$
\begin{equation*}
\nabla_{v} f((v, w))=-\sum_{i=1}^{m} \frac{\exp \left(-a_{i}^{\prime \top} w-b_{i} v\right)}{1+\exp \left(-a_{i}^{\prime \top} w-b_{i} v\right)} b_{i}=-b^{\top}\left(\frac{\exp \left(-A^{\prime} w-b v\right)}{1+\exp \left(-A^{\prime} w-b v\right)}\right) \tag{6.10}
\end{equation*}
$$

where the exponentiation, addition, and division operations in the final parenthesized expression are interpreted as being applied componentwise. Using a similar notation, we also find that

$$
\begin{equation*}
\nabla_{w} f((v, w))=-\sum_{i=1}^{m} \frac{\exp \left(-a_{i}^{\prime \top} w-b_{i} v\right)}{1+\exp \left(-a_{i}^{\prime \top} w-b_{i} v\right)} a_{i}^{\prime}=-A^{\prime \top}\left(\frac{\exp \left(-A^{\prime} w-b v\right)}{1+\exp \left(-A^{\prime} w-b v\right)}\right) . \tag{6.11}
\end{equation*}
$$

Letting $x=(v, w)$ and defining $z$ and $p$ to have the same dimensions as $x$, we obtain that the gradient (and therefore unique subgradient) $y$ of the $f$ subproblem
at $x=(v, w)$ is

$$
\begin{equation*}
y=\nabla f(x)+p+c(x-z)=\left(\nabla f_{v}((v, w)), \nabla f_{w}((v, w))\right)+p+c(x-z) . \tag{6.12}
\end{equation*}
$$

To approximately solve the $f$ subproblem, we employ the limited-memory BFGS (L-BFGS) method [58] for unconstrained nonlinear optimization. We solve the second (g) subproblem exactly using essentially the same soft thresholding operator 5.111) (applied only to the $w$ component) used in the LASSO problem.

For test data, we selected the cancer datasets from [22] that have $b_{i} \in\{-1,1\}$ for all $i$. In addition, we also used the $a 9 a$ [33] and Arcene [41] datasets, which are both are available from the UCI Machine Learning Repository [56] (where a9a is called adult). Both of these datasets are sparse, and $a 9 a$ has $m=32,561$ and $n=123$, while Arcene has $m=900$ and $n=10,000$.

Figure 6.7 shows the number of outer iterations for each algorithm, revealing a pattern similar to that for the LASSO problem. Figure 6.8 shows the cumulative total number of inner iterations for each algorithm. As with the LASSO problem, the inexact algorithms all perform less total work than the exact method, although the savings for two of the cancer datasets are not dramatic. The comparative behavior of the three inexact algorithms, however, is very different than for LASSO: the admm_primDR method is consistently the slowest method, rather than consistently the fastest, while admm_abssum gives the best results.

### 6.2.3 Sparse inverse covariance selection

The covariance selection problem was first introduced in [20], which suggested that the covariance structure of a multivariate normal population can be simplified by setting elements of the inverse covariance matrix to zero. The graphical interpretation of this covariance selection model is called the Gaussian graphical model [32,50]. It has become a popular statistical tool in reverse engineering of genetic regulatory networks, where individual genes are represented by the nodes of a graph and the conditional

Figure 6.7: Logistic regression - number of outer iterations.


Figure 6.8: Logistic regression - total number of inner iterations.

dependencies between their expression profiles are indicated by graph edges.
In this problem, we are given a dataset of vectors $a_{1}, \ldots, a_{m} \in \mathbb{R}^{n}$ which we model as being samples from a multivariate normal distribution $\mathcal{N}(0, \Sigma)$ for some unknown positive definite covariance matrix $\Sigma$. We believe the inverse of $\Sigma$ to be sparse, but with an unknown sparsity pattern. Letting $S=(1 / m) \sum_{i=1}^{m} a_{i} a_{i}^{\top}$ be the empirical covariance matrix of the sample and using $\mathbb{S}_{++}^{n}$ to denote the cone of positive $n \times n$ matrices, we attempt to estimate $\Sigma^{-1}$ by the solution $X$ of the sparse inverse covariance selection problem

$$
\begin{equation*}
\min _{X \in \mathbb{S}_{++}^{n}} \operatorname{Tr}(S X)-\log \operatorname{det} X+\nu\|X\|_{1} \tag{6.13}
\end{equation*}
$$

where $\|X\|_{1}=\sum_{i=1}^{n} \sum_{j=1}^{n}\left|x_{i j}\right|$. The model's goal is to minimize the negative logliklihood of the sample, combined with an $\ell_{1}$ regularization term promoting sparsity of the solution [6]. Note that $-\log \operatorname{det} X=\log \frac{1}{\operatorname{det} X}=\log \operatorname{det}\left(X^{-1}\right)$.

If one lets $f(X)=\operatorname{Tr}(S X)-\log \operatorname{det} X+\nu\|X\|_{1}$ and $g(X)=\nu\|X\|_{1}$, then (treating the unknown $X$ as a vector) problem (6.13) fits the standard ADMM problem form with $M=I$, that is, $\min _{X}\{f(X)+g(X)\}$. Applying the ADMM, we obtain the the recursions

$$
\begin{align*}
X^{k+1} & =\underset{X \in \mathbb{S}_{++}^{n}}{\arg \min }\left\{\operatorname{Tr}(S X)-\log \operatorname{det} X+\frac{c}{2}\left\|X-Z^{k}+\frac{1}{c} P^{k}\right\|_{F}^{2}\right\}  \tag{6.14}\\
Z_{i j}^{k+1} & =\operatorname{sgn}\left(X_{i j}^{k+1}+\frac{1}{c} P_{i j}^{k}\right) \max \left\{0,\left|X_{i j}^{k+1}+\frac{1}{c} P_{i j}^{k}\right|-\frac{\nu}{c}\right\} \quad i, j=1,2, \ldots, n  \tag{6.15}\\
P^{k+1} & =P^{k}+c\left(X^{k+1}-Z^{k+1}\right) \tag{6.16}
\end{align*}
$$

where $\|\cdot\|_{F}$ denotes the Frobenius norm of a matrix and the Lagrange multiplier estimates $P^{k}$ are members of $\mathbb{S}^{n}$, the vector space of $n \times n$ symmetric real matrices. We uniformly set $\nu=0.5$, as suggested in 52. For some examples of applying ADMM to SICS, see [73] and [87], the latter showing that ADMM outperforms other methods for this problem.

It is possible to develop an analytical solution to subproblem (6.14). First, we
know that

$$
\begin{aligned}
\partial \operatorname{Tr} S X & =\partial\langle S, X\rangle=\{S\} \\
\partial(-\log \operatorname{det} X) & =\left\{-\frac{1}{\operatorname{det} X} \nabla \operatorname{det} X\right\}=\left\{-\frac{1}{\operatorname{det} X} \operatorname{det} X\left(X^{-1}\right)^{\top}\right\}=\left\{-X^{-1}\right\} .
\end{aligned}
$$

Using the first order optimality condition for the convex problem (6.14), whose solution must lie in the interior of the open cone $\mathbb{S}_{++}^{n}$, we obtain

$$
\begin{array}{ll} 
& 0 \in \partial\left\{\operatorname{Tr}(S X)-\log \operatorname{det} X+\frac{c}{2}\left\|X-Z^{k}+\frac{1}{c} P^{k}\right\|_{F}^{2}\right\} \\
\Leftrightarrow & 0=S-X^{-1}+P^{k}+c\left(X-Z^{k}\right) . \tag{6.18}
\end{array}
$$

Rearranging (6.18), we obtain

$$
\begin{equation*}
c X-X^{-1}=c Z^{k}-P^{k}-S \tag{6.19}
\end{equation*}
$$

Thus, $X^{k+1}$ can be obtained by solving (6.19) for $X$. Next we take the orthogonal eigenvalue decomposition of the symmetric matrix on the left-hand side of (6.19), obtaining some diagonal matrix $\Lambda^{k+1}=\operatorname{diag}\left(\lambda_{1}^{k+1}, \ldots, \lambda_{n}^{k+1}\right)$ and orthogonal matrix $Q^{k+1}$ such that

$$
\begin{equation*}
c X^{k+1}-\left(X^{k+1}\right)^{-1}=Q^{k+1} \Lambda^{k+1}\left(Q^{k+1}\right)^{\top} . \tag{6.20}
\end{equation*}
$$

Multiplying by $\left(Q^{k+1}\right)^{\top}$ from the left and by $Q^{k+1}$ from the right on both sides of (6.19), we obtain

$$
\begin{equation*}
c \tilde{X}^{k+1}-\left(\tilde{X}^{k+1}\right)^{-1}=\Lambda^{k+1} \tag{6.21}
\end{equation*}
$$

where $\tilde{X}^{k+1}=\left(Q^{k+1}\right)^{\top} X^{k+1} Q^{k+1}$. We now can construct a diagonal solution $\tilde{X}^{k+1}$ of (6.21). To find each entry $\tilde{X}_{i i}$ on the diagonal of $\tilde{X}^{k+1}$, we need to solve $c \tilde{X}_{i i}^{k+1}-$ $1 / \tilde{X}_{i i}^{k+1}=\lambda_{i}^{k+1}$, and because all the $\tilde{X}_{i i}$ must be nonnegative, we have

$$
\begin{equation*}
\tilde{X}_{i i}^{k+1}=\frac{\lambda_{i}^{k+1}+\sqrt{\left(\lambda_{i}^{k+1}\right)^{2}+4 c}}{2 c} . \tag{6.22}
\end{equation*}
$$

At this point, one can see that the solution of equation (6.19), $X^{k+1}$, is

$$
\begin{equation*}
X^{k+1}=Q^{k+1} \tilde{X}^{k+1}\left(Q^{k+1}\right)^{\top} \tag{6.23}
\end{equation*}
$$

The $g$ minimization (6.15) is the same soft thresholding operation as in (6.6), applied throughout a symmetric matrix. Clearly, the most time-consuming part of (6.14)-(6.16) is the eigenvalue decomposition (6.20) required by the $x$ minimization (6.14). In our numerical experiments, we used the Jacobi iterative method from [72 to perform this calculation. At iteration $(k+1)$, this method produces successively improving (over $l$ ) estimates $Q^{k, l}, H^{k, l}$ of $Q^{k+1}$ and $\Lambda^{k+1}$, respectively. The convergence properties of Jacobi method guarantee that $\lim _{l \rightarrow \infty} Q^{k, l}=Q^{k+1}$ and $\lim _{l \rightarrow \infty} H^{k, l}=\Lambda^{k+1}$. For each $l$, we produce a second approximation $\Lambda^{k, l}$ of $\Lambda^{k+1}$ by taking the projection of $H^{k, l}$ onto the linear subspace of $n \times n$ diagonal matrices, that is, $\Lambda^{k, l}$ is obtained by setting all non-diagonal elements of $H^{k, l}$ to zero. Since this operation is the application of a continuous function that fixes $\Lambda^{k+1}$, it follows that $\lim _{l \rightarrow \infty} \Lambda^{k, l}=\Lambda^{k+1}$. Using the respective estimates $H^{k, l}$ and $\Lambda^{k, l}$ of $H^{k+1}$ and $\Lambda^{k+1}$, we may derive an approximation $X^{k, l}$ of $X^{k+1}$ by solving the following modification of (6.20):

$$
\begin{equation*}
c \tilde{X}^{k, l}-\left(\tilde{X}^{k, l}\right)^{-1} \approx Q^{k, l} \Lambda^{k, l}\left(Q^{k, l}\right)^{\top} \tag{6.24}
\end{equation*}
$$

This equation may be solved using appropriate modifications of of (6.22) and 6.23), namely

$$
\begin{equation*}
\tilde{X}_{i i}^{k, l}=\frac{\lambda_{i}^{k, l}+\sqrt{\left(\lambda_{i}^{k, l}\right)^{2}+4 c}}{2 c} \quad X^{k, l}=Q^{k, l} \tilde{X}^{k, l}\left(Q^{k, l}\right)^{\top} \tag{6.25}
\end{equation*}
$$

where $\lambda_{i}^{k, l}$ denotes the $i^{\text {th }}$ diagonal element of $\Lambda^{k, l}$. The resulting estimate $X^{k, l}$ depends continuously on $Q^{k, l}$ and $\Lambda^{k, l}$, and hence on $Q^{k, l}$ and $H^{k, l}$. Therefore, it converges to the exact value of $X^{k+1}$ given in 6.23). Finally, since $Q^{k, l}$ is orthogonal, it is easy to evaluate the inverse of $X^{k, l}$, and therefore straightforward to compute the gradient of $f$, much as in (6.14).

We tested our implementation on five gene expression network datasets that have been widely used in the model selection and classification literature, as for example in [52]:

Lymph node status (Lymph): This dataset is derived from [23, and preprocessed using the procedure described there. The covariance matrix $S$ has dimension $n=587$ and rank 147 .

Estrogen receptor (ER): This preprocessed dataset is again from [23]. The rank of the covariance matrix $S$ is 157 it has dimension $n=692$.

Arabidopsis thealiana (Arabidopsis): This gene network data set was obtained from [81. $S$ has dimension $n=843$ and its rank is 117 .

Leukemia (Leukemia): A gene expression data set from 85] that has dimension $n=1255$ and rank 71 .

Hereditary breast cancer (Hereditarybc): This data set is from 46]. Its dimension is $n=1869$ and the rank of $S$ is 21 .

The $S$ matrices for all these datasets are dense. The work of Li and Toh [52] and references therein contain more detailed descriptions of datasets, and discussion of how they were selected.

The computational results for the various ADMM algorithms are displayed in Figures 6.9 and 6.10 . Figure 6.9 depicts the number outer iterations, which show very little variation between the various ADMM variants. Figure 6.10 shows the cumulative total iterations of the Jacobi method, which is proportional to overall computational effort. It shows that the inexact methods are all significantly faster than the exact ADMM, with a consistent pattern of admm_primDR being fastest. These results are broadly similar to those for LASSO problems, although the difference between admm_primDR and the other inexact methods is less pronounced.

Figure 6.9: SICS - number of outer iterations.


Figure 6.10: SICS - total number of inner iterations.


## Chapter 7

## Conclusion

ADMM was developed over three decades ago, and it has attracted renewed attention recently due to its applicability to various machine learning and image processing problems. Its applications can be found in distributed and cloud computing systems, massive high-dimensional datasets, and the associated large-scale applied statistical problems. Some characteristics of ADMM include:

- It appears to be well suited to the modern regime, and has the important benefit of being quite general in its scope and applicability.
- The ADMM can take advantage of the structure of those problems which involve optimizing sums of convex functions.
- It is often relatively easy to implement the ADMM in a distributed-memory, parallel manner. This property is important for "big data" problems in which the entire problem dataset may not fit readily into the memory of a single processor.

In practice, it is very often that we need to use some iterative method to solve one or both ADMM subproblems, either because there is no simple analytical solution or the problem is too high-dimensional for such a formula to be used. In such cases, it is wasteful to expend great effort in solving subproblem(s) to high precision in the early stage of solution process. Naturally, a method that allows for solving subproblems inexactly and gradually increasing their precision without disrupting the convergence properties is highly desirable. However, the traditional convergence proofs of ADMM,
both operator splitting and Lagrangian splitting, are based on solving all subproblems exactly. The existence of inexact ADMM methods was unknown until [29], in which Eckstein presented an approximate ADMM with the absolute summable error bound criteria. However, one drawback of this version is that, unless some other condition is satisfied (such as $M$ being full column rank matrix), it is not practical because we have to bound the distance between the approximate solutions and the true solutions of the subproblems.

To address this issue, we developed three new criteria for approximate minimization of ADMM subproblems that only use subgradient information. They do not require any primal regularization terms. The approximate ADMM derived from general inexact proximal point algorithm and operator splitting theory is actually a partially inexact ADMM, since it requires one subproblem to be solved quickly and exactly, which is very common in many $l_{1}$ normalized applications. It also essentially requires that $M$ be the identity matrix. The two new methods that are developed from Lagrangian splitting have neither of these restrictions.

This dissertation tested all three methods with three representative $l_{1}$ normalized models against the classical ADMM. Its numerical results show that all three new approximate versions of ADMM effectively reduce overall computing effort, although the preferred method depends on the application.

## References

[1] M. Marques Alves and B. F. Svaiter. A note on Fejér-monotone sequences in product spaces and its applications to the dual convergence of augmented Lagrangian methods. Math. Program., 155(1-2):613-616, 2016.
[2] Roger W. Hoerl Arthur E. Hoerl, Robert W. Kennard. Practical use of ridge regression: A challenge met. Journal of the Royal Statistical Society. Series C (Applied Statistics), 34(2):114-120, 1985.
[3] Alfred Auslender and Marc Teboulle. Lagrangian duality and related multiplier methods for variational inequality problems. SIAM J. Optim., 10(4):1097-1115 (electronic), 2000.
[4] Alfred Auslender and Marc Teboulle. Entropic proximal decomposition methods for convex programs and variational inequalities. Math. Program., 91(1, Ser. A):33-47, 2001.
[5] Alfred Auslender, Marc Teboulle, and Sami Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. Comput. Optim. Appl., 12(1-3):3140, 1999.
[6] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. J. Mach. Learn. Res., pages 485-516, 2008.
[7] Heinz H. Bauschke and Patrick L. Combettes. Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With a foreword by Hédy Attouch.
[8] Dimitri P Bertsekas and John N Tsitsiklis. Parallel and distributed computation: numerical methods. Prentice-Hall, Inc., 1989.
[9] Abdellah Bnouhachem. On LQP alternating direction method for solving variational inequality problems with separable structure. J. Inequal. Appl., pages 2014:80, 15, 2014.
[10] Abdellah Bnouhachem, Suliman Al-Homidan, and Qamrul Hasan Ansari. New descent LQP alternating direction methods for solving a class of structured variational inequalities. Fixed Point Theory Appl., pages 2015:137, 11, 2015.
[11] Abdellah Bnouhachem, Hafida Benazza, and Mohamed Khalfaoui. An inexact alternating direction method for solving a class of structured variational inequalities. Appl. Math. Comput., 219(14):7837-7846, 2013.
[12] Abdellah Bnouhachem and Abdelouahed Hamdi. Parallel LQP alternating direction method for solving variational inequality problems with separable structure. J. Inequal. Appl., pages 2014:392, 14, 2014.
[13] Abdellah Bnouhachem and Abdelouahed Hamdi. A hybrid LQP alternating direction method for solving variational inequality problems with separable structure. Appl. Math. Inf. Sci., 9(3):1259-1264, 2015.
[14] Abdellah Bnouhachem and M. H. Xu. An inexact LQP alternating direction method for solving a class of structured variational inequalities. Comput. Math. Appl., 67(3):671-680, 2014.
[15] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn., 3(1):1-122, January 2011.
[16] Raymond H. Chan, Min Tao, and Xiaoming Yuan. Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers. SIAM J. Imaging Sci., 6(1):680-697, 2013.
[17] Caihua Chen, Min Li, and Xiaoming Yuan. Further study on the convergence rate of alternating direction method of multipliers with logarithmic-quadratic proximal regularization. J. Optim. Theory Appl., 166(3):906-929, 2015.
[18] Gong Chen and Marc Teboulle. A proximal-based decomposition method for convex minimization problems. Math. Programming, 64(1, Ser. A):81-101, 1994.
[19] Zhongming Chen, Li Wan, and Qingzhi Yang. An inexact alternating direction method for structured variational inequalities. J. Optim. Theory Appl., 163(2):439-459, 2014.
[20] Arthur P. Dempster. Covariance selection. Biometrics, pages 157-175, 1972.
[21] Wei Deng and Wotao Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. Journal of Scientific Computing, pages 1-28, 2015.
[22] Marcel Dettling and Peter Bühlmann. Finding predictive gene groups from microarray data. J. Multivariate Anal., 90(1):106-131, 2004.
[23] Adrian Dobra. Variable selection and dependency networks for genomewide data. Biostatistics, 10(4):621-639, 2009.
[24] Norman R. Draper and R. Craig Van Nostrand. Ridge regression and JamesStein estimation: review and comments. Technometrics, 21(4):451-466, 1979.
[25] Marco F. Duarte, Mark A. Davenport, Dharmpal Takbar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling: Building simpler, smaller, and less-expensive digital cameras. IEEE Sig. Proc. Mag., 25(2):83-91, 2008.
[26] Jonathan Eckstein. Splitting methods for monotone operators with applications to parallel optimization. PhD thesis, Massachusetts Institute of Technology, 1989.
[27] Jonathan Eckstein. Some saddle-function splitting methods for convex programming. Optimization Methods and Software, 4(1):75-83, 1994.
[28] Jonathan Eckstein. A practical general approximation criterion for methods of multipliers based on Bregman distances. Math. Program., 96(1):61-86, 2003.
[29] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program., 55(3):293-318, 1992.
[30] Jonathan Eckstein and Paulo J. S. Silva. A practical relative error criterion for augmented Lagrangians. Math. Program., 141(1-2):319-348, 2013.
[31] Jonathan Eckstein and Wang Yao. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. Pac. J. Optim., 11(4):619-644, 2015.
[32] David Edwards. Introduction to Graphical Modeling. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2000.
[33] Rong-En Fan, Pai-Hsuen Chen, and Chih-Jen Lin. Working set selection using second order information for training support vector machines. J. Mach. Learn. Res., 6:1889-1918, 2005.
[34] Michel Fortin and Roland Glowinski. On decomposition-coordination methods using an augmented Lagrangian. In M. Fortin and R. Glowinski, editors, Augmented Lagrangian methods: Applications to the numerical solution of boundaryvalue problems, volume 15 of Studies in Mathematics and its Applications, pages 97-146. North-Holland Publishing Co., Amsterdam, 1983.
[35] James Franklin. The elements of statistical learning: data mining, inference and prediction. Math. Intelligencer, 27(2):83-85, 2005.
[36] 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.
[37] Daniel Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems, volume 15 of Studies in Mathematics and its Applications, pages 299-331. North-Holland Publishing Co., Amsterdam, 1983.
[38] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér., 9(R-2):41-76, 1975.
[39] Tom Goldstein, Brendan O'Donoghue, Simon Setzer, and Richard Baraniuk. Fast alternating direction optimization methods. SIAM J. Imaging Sci., 7(3):1588-1623, 2014.
[40] Guoyong Gu, Bingsheng He, and Junfeng Yang. Inexact alternating-directionbased contraction methods for separable linearly constrained convex optimization. Journal of Optimization Theory and Applications, 163(1):105-129, 2014.
[41] Isabelle Guyon, Steve Gunn, Asa Ben-Hur, and Gideon Dror. Result analysis of the NIPS 2003 feature selection challenge. In L.K. Saul, Y. Weiss, and L. Bottou, editors, Advances in Neural Information Processing Systems 17, pages 545-552. MIT Press, 2005.
[42] Bing-sheng He, Li-zhi Liao, and Xiao-ming Yuan. A LQP based interior prediction-correction method for nonlinear complementarity problems. J. Comput. Math., 24(1):33-44, 2006.
[43] Bingsheng He, Li-Zhi Liao, Deren Han, and Hai Yang. A new inexact alternating directions method for monotone variational inequalities. Math. Program., 92(1):103-118, 2002.
[44] Bingsheng He and Xiaoming Yuan. On non-ergodic convergence rate of DouglasRachford alternating direction method of multipliers. Numer. Math., 130(3):567577, 2015.
[45] Bingsheng He and Xiaoming Yuan. On the convergence rate of Douglas-Rachford operator splitting method. Math. Program., 153(2):715-722, 2015.
[46] Ingrid Hedenfalk, David Duggan, Yidong Chen, Michael Radmacher, Michael Bittner, Richard Simon, Paul Meltzer, Barry Gusterson, Manel Esteller, Mark Raffeld, et al. Gene-expression profiles in hereditary breast cancer. New Engl. J. Med., 344(8):539-548, 2001.
[47] Magnus R. Hestenes. Multiplier and gradient methods. J. Optimization Theory Appl., 4:303-320, 1969.
[48] Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, pages 3836-3840. IEEE, 2015.
[49] Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. Predicting risk from financial reports with regression. In Proceedings
of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL '09, pages 272-280, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
[50] Steffen L. Lauritzen. Graphical Models, volume 17 of Oxford Statistical Science Series. The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.
[51] Jim Lawrence and Jonathan E. Spingarn. On fixed points of nonexpansive piecewise isometric mappings. Proc. London Math. Soc., 55(3):605-624, 1987.
[52] $\mathrm{Lu} \mathrm{Li} \mathrm{and} \mathrm{Kim-Chuan} \mathrm{Toh} .\mathrm{An} \mathrm{inexact} \mathrm{interior} \mathrm{point} \mathrm{method} \mathrm{for} L_{1}$-regularized sparse covariance selection. Math. Program. Comput., 2(3-4):291-315, 2010.
[53] Min Li. A hybrid LQP-based method for structured variational inequalities. Int. J. Comput. Math., 89(10):1412-1425, 2012.
[54] Min Li, Xinxin Li, and Xiaoming Yuan. Convergence analysis of the generalized alternating direction method of multipliers with logarithmic-quadratic proximal regularization. J. Optim. Theory Appl., 164(1):218-233, 2015.
[55] Min Li and Xiao-ming Yuan. An improved proximal-based decomposition method for structured monotone variational inequalities. Appl. Math. Mech. (English Ed.), 28(12):1659-1668, 2007.
[56] Moshe Lichman. UCI machine learning repository, 2013.
[57] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal., 16(6):964-979, 1979.
[58] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. Math. Program., 45(3):503-528, 1989.
[59] B. Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 4(R3):154-158, 1970.
[60] Renato DC Monteiro and Benar F Svaiter. Iteration-complexity of blockdecomposition algorithms and the alternating direction method of multipliers. SIAM Journal on Optimization, 23(1):475-507, 2013.
[61] Andrew Y. Ng. Feature selection, $L_{1}$ vs. $L_{2}$ regularization, and rotational invariance. In Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, pages 615-622, 2004.
[62] Michael K. Ng, Fan Wang, and Xiaoming Yuan. Inexact alternating direction methods for image recovery. SIAM J. Sci. Comput., 33(4):1643-1668, 2011.
[63] Michael K. Ng, Pierre Weiss, and Xiaoming Yuan. Solving constrained totalvariation image restoration and reconstruction problems via alternating direction methods. SIAM J. Sci. Comput., 32(5):2710-2736, 2010.
[64] Jorge Nocedal and Stephen J. Wright. Numerical optimization. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
[65] Zheng Peng and Wenxing Zhu. A partial inexact alternating direction method for structured variational inequalities. Optimization, 63(7):1043-1055, 2014.
[66] Boris T. Polyak. Introduction to Optimization. Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1987.
[67] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In Optimization (Sympos., Univ. Keele, Keele, 1968), pages 283-298. Academic Press, London, 1969.
[68] R. Tyrrell Rockafellar. Local boundedness of nonlinear, monotone operators. Michigan Math. J., 16:397-407, 1969.
[69] R. Tyrrell Rockafellar. Convex Analysis. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
[70] R. Tyrrell Rockafellar. Conjugate Duality and Optimization. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974.
[71] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. SIAM J. Control Optim., 14(5):877-898, 1976.
[72] Heinz Rutishauser. The Jacobi method for real symmetric matrices. Numer. Math., 9(1):1-10, 1966.
[73] Katya Scheinberg, Shiqian Ma, and Donald Goldfarb. Sparse inverse covariance selection via alternating linearization methods. In John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta, editors, Advances in Neural Information Processing Systems, volume 23, pages 21012109, 2010. Proceedings of "Neural Information Processing Systems 2010".
[74] Li Shen and Shaohua Pan. Inexact indefinite proximal ADMMs for 2-block separable convex programs and applications to 4-block DNNSDPs. ePrint 1505.04519, ArXiv, 2015.
[75] Mikhail V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. J. Convex Anal., 6(1):59-70, 1999.
[76] Mikhail V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. Math. Oper. Res., 25(2):214-230, 2000.
[77] Benee F. Swindel. Geometry of ridge regression illustrated. The American Statistician, 35(1):12-15, 1981.
[78] Min Tao and Xiaoming Yuan. On the $O(1 / t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization. SIAM J. Optim., 22(4):1431-1448, 2012.
[79] Robert Tibshirani. Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B, 58(1):267-288, 1996.
[80] Huahua Wang and Arindam Banerjee. Online alternating direction method (longer version). ePrint 1306.3721, arXiv, 2013.
[81] Anja Wille, Philip Zimmermann, Eva Vranová, Andreas Fürholz, Oliver Laule, Stefan Bleuler, Lars Hennig, Amela Prelic, Peter von Rohr, Lothar Thiele, et al. Sparse graphical Gaussian modeling of the isoprenoid gene network in arabidopsis thaliana. Genome Biol., 5(11):R92, 2004.
[82] Yun-Hai Xiao and Hui-Na Song. An inexact alternating directions algorithm for constrained total variation regularized compressive sensing problems. J. Math. Imaging Vision, 44(2):114-127, 2012.
[83] Ming Yan and Wotao Yin. Self equivalence of the alternating direction method of multipliers. CAM Report 14-59, UCLA, 2014.
[84] Junfeng Yang and Xiaoming Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Math. Comp., 82(281):301-329, 2013.
[85] Ka Yee Yeung, Roger E. Bumgarner, and Adrian E. Raftery. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics, 21(10):2394-2402, 2005.
[86] XiaoMing Yuan. The improvement with relative errors of He et al.'s inexact alternating direction method for monotone variational inequalities. Math. Comput. Modelling, 42(11-12):1225-1236, 2005.
[87] Xiaoming Yuan. Alternating direction method for covariance selection models. J. Sci. Comput., 51(2):261-273, 2012.
[88] Xiaoming Yuan and Min Li. An LQP-based decomposition method for solving a class of variational inequalities. SIAM J. Optim., 21(4):1309-1318, 2011.
[89] Xiaoming Yuan and Junfeng Yang. Sparse and low-rank matrix decomposition via alternating direction method. Pac. J. Optim., 9(1):167-180, 2013.
[90] Yuhua Zeng, Yufei Yang, and Zheng Peng. A line-search-based partial proximal alternating directions method for separable convex optimization. Journal of Applied Mathematics, 2014.
[91] Xin Zhang, Duc-Son Pham, Svetha Venkatesh, Wanquan Liu, and Dinh Phung. Mixed-norm sparse representation for multi view face recognition. Pattern Recognition, 48(9):2935-2946, 2015.

