

# Approximate Versions of the Alternating Direction Method of Multipliers\*

Jonathan Eckstein<sup>†</sup>      Wang Yao<sup>‡</sup>

## Abstract

We present three new approximate versions of the alternating direction method of multipliers (ADMM), all of which require only knowledge of subgradients of the subproblem objectives, rather than 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, relative-error 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.

## 1 Introduction

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) + g(Mx) \tag{1}$$

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

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

---

\*This work was funded in part by the National Science Foundation under grant CCF-1115638. The authors would also like to thank Heinz Bauschke for suggesting the proof of Lemma 4.

<sup>†</sup>Department of Management Science and Information Systems and RUTCOR, Rutgers University, 100 Rockafeller Road, Piscataway, NJ 08854

<sup>‡</sup>RUTCOR, Rutgers University, 100 Rockafeller Road, Piscataway, NJ 08854

The alternating direction method of multipliers (ADMM) [14, 16] for solving (P) consists of the recursions

$$x^{k+1} = \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\} \quad (2)$$

$$z^{k+1} = \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\} \quad (3)$$

$$p^{k+1} = p^k + c (Mx^{k+1} - z^{k+1}) \quad (4)$$

This method has recently become popular for numerous applications; see for example [3]. This paper concerns situations in which at least one of the subproblems (2) or (3) 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 it involves factoring a very large matrix). In such cases, it seems wasteful to expend great effort in solving (2) 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 (3). 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 [10] that the answer to this question is positive. However, the subproblem approximation criteria in that result, which we will here refer to as the *EB 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,  $\ker M = \{0\}$ ), 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 3.1 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 to 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 [36] and was followed by [37] and a variety of generalizations. For the classical (non-alternating-direction) method of multipliers, a similar error criterion was developed in [11]. In this paper, two of the error criteria we develop are relative, using ideas from [37] and [11], 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, nor do they need any other form of strong convexity for the subproblems, and are therefore easier to verify in general. Their analysis uses techniques inspired by [9].

Convergence of the ADMM has traditionally been proven in two related but different ways. One approach, dating back to [14], uses the monotonicity of the (convex-concave)

subgradient of the Lagrangian function of (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 [16], expresses the subgradient of the dual function of (P) as the sum of two monotone operators and shows that the ADMM is equivalent to applying a Douglas-Rachford operator splitting method [24] 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 [32] (PPA) as shown possible in [10]; see also [21]. We then apply the relative-error proximal algorithm of [37] 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 Section 3, 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 (3) quickly and exactly, and therefore it is applicable when only the  $x$  minimization (2) 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 [14] to incorporate ideas used to develop approximate versions of the classical method of multipliers. For one method, we augment the Lagrangian-splitting ADMM convergence proof using techniques developed in [9]: 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 [11], 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. We develop these methods in Sections 4.3 and 4.4, respectively, with background material in Section 4.1. Section 5 presents some numerical experiments establishing the potential utility of our algorithms.

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  $(x^*, z^*, p^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  such that*

$$-M^\top p^* \in \partial f(x^*) \qquad p^* \in \partial g(z^*) \qquad Mx^* = z^*.$$

It is easily seen that if  $(x^*, z^*, p^*)$  is a KKT point, then  $x^*$  must be a solution of (P).

## 2 Formalizing Approximate Subproblem Solution

In order to formalize our methods clearly, we present mathematical models of the approximate solution processes for the subproblems (2) and (3). All of our methods will use the following assumption:

**Assumption 2.** *To approximately solve (2), we assume the existence of some mapping  $\mathcal{F} : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R}^n \times \mathbb{N} \rightarrow \mathbb{R}^n \times \mathbb{R}^n$  such that if  $(x^l, y_1^l) = \mathcal{F}(p, z, c, \bar{x}, l)$  for all  $l \in \mathbb{N}$ , then*

$$\lim_{l \rightarrow \infty} y_1^l = 0 \quad (\forall l \in \mathbb{N}) \quad y_1^l \in \partial_x [f(x) + \langle p, Mx \rangle + \frac{c}{2} \|Mx - z\|^2]_{x=x^l}.$$

The idea behind this definition is that  $\mathcal{F}(p, z, c, \bar{x}, l)$  is the  $l^{\text{th}}$  iterate produced 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 (2), we may take iterates of the form  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$  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 practice 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 (3) may be solved exactly. However, some of our algorithms also allow for the case that both (2) and (3) are sufficiently difficult to merit iterative solution, in which case we also need the following assumption:

**Assumption 3.** *To approximately solve (3), 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  $(z^l, y_2^l) = \mathcal{G}(p, x, c, \bar{z}, l)$  for all  $l \in \mathbb{N}$ , then*

$$\lim_{l \rightarrow \infty} y_2^l = 0 \quad (\forall l \in \mathbb{N}) \quad y_2^l \in \partial_z [g(z) - \langle p, z \rangle + \frac{c}{2} \|Mx - z\|^2]_{z=z^l}.$$

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  $g(z) - \langle p, z \rangle + \frac{c}{2} \|Mx - z\|^2$ , 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 (2) or (3) is easily solved exactly, we may simply take  $\mathcal{F}$  or  $\mathcal{G}$  as respectively “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 (3), for example, we may take

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

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  $\{x^l\}$  and  $\{z^l\}$  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  $\{x^l\}, \{y^l\}$  be sequences in  $\mathbb{R}^n$  such that  $y^l \in \partial h(x^l)$  for all  $l$  and  $y^l \rightarrow 0$ . Then if the set of minimizers of  $h$  is nonempty and bounded,  $\{x^l\}$  must be bounded, with all its limit points being minimizers of  $h$ . If  $h$  has a unique minimizer, then  $\{x^l\}$  converges to that minimizer.*

*Proof.* By [30, Theorem 27.1(d)],  $h$  having a nonempty bounded set of minimizers is equivalent to  $0 \in \text{int dom } h^*$ , where  $h^*$  denotes the convex conjugate of  $h$ . By [30, Theorem 23.4], we then have  $0 \in \text{int dom } \partial h^*$ . By the Rockafellar-Vesely theorem [29], the maximal monotone point-to-set map  $\partial h^*$  must then be locally bounded at 0, meaning that for some  $\epsilon > 0$ , the set  $S(\epsilon) = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^n : x \in \partial h^*(y), \|y\| < \epsilon\}$  is bounded. By [30, Theorem 23.5],  $S(\epsilon) = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^n : y \in \partial h(x), \|y\| < \epsilon\}$ , so the convergence of  $\{y^l\}$  to zero implies that  $\{x^l\}$  is bounded. If we consider any limit point  $x^\infty$  of  $\{x^l\}$ , 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  $0 \in \partial h(x^\infty)$ , and  $x^\infty$  must be a minimizer of  $h$ . In the case that  $h$  has a unique minimizer  $\bar{x}$ , we then have that  $\{x^l\}$  is a bounded sequence whose only possible limit point is  $\bar{x}$ , so it must converge to  $\bar{x}$ .  $\square$

This lemma has the following immediate consequences.

**Lemma 5.** *If the set of minimizers of  $f(x) + \langle p, Mx \rangle + \frac{c}{2} \|Mx - z\|^2$  is bounded, then the sequence  $\{x^l\}$  generated by  $(x^l, y_1^l) = \mathcal{F}(p, z, c, \bar{x}, l)$  with  $\mathcal{F}$  as in Assumption 2 must be bounded. If the minimizer of  $f(x) + \langle p, Mx \rangle + \frac{c}{2} \|Mx - z\|^2$  is unique,  $\{x^l\}$  must converge to it.*

*Proof.* Immediate from Lemma 4, setting  $h(x) = f(x) + \langle p, Mx \rangle + \frac{c}{2} \|Mx - z\|^2$ .  $\square$

**Lemma 6.** *The sequence generated by  $(z^l, y_2^l) = \mathcal{G}(p, x, c, \bar{z}, l)$ , with  $\mathcal{G}$  as in Assumption 3, always converges to the unique minimizer over  $z$  of  $g(z) - \langle p, z \rangle + \frac{c}{2} \|Mx - z\|^2$ .*

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

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

### 3 Approximate ADMM Algorithms Derived through Operator Splitting Analysis

This section 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 distance-based approximation criteria in the approximate ADMM in [10, 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 [10, Theorem 8].

The proof of [10, Theorem 8] is based on the relationship of the ADMM to Douglas-Rachford (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 takes 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.

### 3.1 A subgradient-based application of [10, Theorem 8]

Let  $\{\epsilon_k\}_{k=1}^\infty, \{\tau_k\}_{k=1}^\infty \subset \mathbb{R}_{++}$  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 3.1.1** Inexact ADMM with absolute summable error criteria

---

**initialization:** Pick  $c > 0$  and initial points  $p^0, z^0 \in \mathbb{R}^m$   
**repeat** {**for**  $k = 0, 1, 2, \dots$ }  
    **repeat** {**for**  $l = 1, 2, \dots$ }  
        Improve the solution to  $x^{k+1} \approx \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\}$  by taking  
        ing  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$   
        **until**  $\|y_1^{k,l}\| \leq \epsilon_{k+1}$   
         $x^{k+1} = x^{k,l}$   
         $y_1^{k+1} = y_1^{k,l}$   
    **repeat** {**for**  $l = 1, 2, \dots$ }  
        Improve the solution to  $z^{k+1} \approx \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\}$  by taking  
        ing  $(z^{k,l}, y_2^{k,l}) = \mathcal{G}(p^k, x^{k+1}, c, z^k, l)$   
        **until**  $\|y_2^{k,l}\| \leq \tau_{k+1}$   
         $z^{k+1} = z^{k,l}$   
         $y_2^{k+1} = y_2^{k,l}$   
         $p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$   
    **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

$$\|Mx^{k+1} - z^{k+1}\| \leq \delta_1 \quad (5)$$

$$\|y_2^{k+1}\| \leq \delta_2 \quad (6)$$

$$\exists \bar{y}_1^{k+1} \in \partial_x \left[ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^{k+1}\|^2 \right]_{x=x^{k+1}} : \|\bar{y}_1^{k+1}\| \leq \delta_3, \quad (7)$$

where  $\delta_1, \delta_2, \delta_3$  are small positive scalars. One possible choice of  $\bar{y}_1^{k+1}$  in (7) is  $y_1^k$ .

If one uses an overall convergence test that does not require the sequence  $\{y_2^k\}$ , then one can omit the assignment  $y_2^{k+1} = y_2^{k,l}$  from the implementation of the algorithm, and similarly for  $\{y_1^k\}$ .

While there is currently no known proof that Algorithm 3.1.1 converges in the general case, there are several special cases in which the condition  $\|y_1^{k,l}\| \leq \epsilon_{k+1}$  guarantees a bound on the distance to the exact solution of the subproblem (2), which in turn means that the algorithm is a special case of the algorithm proved to converge in [10, Theorem 8]. Essentially, we require that the minimand in (2) be strongly convex:

**Proposition 7.** *Under Assumptions 2 and 3, the inner loops (over  $l$ ) in Algorithm 3.1.1 always terminate finitely. Under Assumption 1, if either  $f$  is strongly convex or  $M$  has full column rank, then  $\{(x^k, z^k, p^k)\}$  converges to a KKT point of (P).*

*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  $\kappa(M^\top M)$  denote the smallest eigenvalue of the symmetric matrix  $M^\top M$ . For each  $k$ , let  $\bar{f}_k(x) = f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2$ , the minimand in (2) expressed as a function of  $x$ . Under the hypotheses, we have  $\alpha > 0$  or  $\kappa(M^\top M) > 0$ , so  $\bar{f}_k$  is strongly convex with modulus  $\bar{\alpha} = \alpha + c\kappa(M^\top M) > 0$ . 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 (2), the Cauchy-Schwarz inequality and strong monotonicity of  $\partial \bar{f}_k$  combine to yield

$$\|y_1^{k,l}\| \|x^{k,l} - \bar{x}^{k+1}\| \geq \langle y_1^{k,l} - 0, x^{k,l} - \bar{x}^{k+1} \rangle \geq \bar{\alpha} \|x^{k,l} - \bar{x}^{k+1}\|^2,$$

and hence

$$\|y_1^{k,l}\| \geq \bar{\alpha} \|x^{k,l} - \bar{x}^{k+1}\|$$

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

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

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

We next consider the  $z$  subproblem (3). 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  $\{\tau_k/c\}$ . All the remaining claims of the proposition now follow immediately from [10, Theorem 8].  $\square$

**Remark:** Although such a form rarely appears needed in applications, the ADMM is sometimes presented for the more general problem  $\min \{f(x) + g(z) \mid Mx + Nz = b\}$ , where  $N$  is an additional constraint matrix and  $b$  is some given vector; see for example [3]. 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  $M$ .

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

We now embark on the derivation of an algorithm with similar theoretical underpinnings to Algorithm 3.1.1, but with a relative error criterion. Our analysis employs the inexact relative-error proximal point algorithm developed by Solodov and Svaiter in [35].

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  $\{z^k\} \subset \mathbb{R}^n$  such that  $\{z^{k+1}\}$  is the solution (over  $z$ ) of

$$0 \in \lambda_k T(z) + z - z^k \tag{8}$$

for each  $k \geq 0$ , where  $\{\lambda_k\}$  is a sequence of scalar parameters with  $\inf_{k \geq 0} \{\lambda_k\} > 0$ . Equivalent problems are to find a pair  $(u, v)$  such that

$$v \in T(u) \quad \lambda_k v + u - z^k = 0 \quad (9)$$

or

$$v \in T(u) \quad u = z^k - \lambda_k v. \quad (10)$$

Concisely, the algorithm may be written  $z^{k+1} = (I + \lambda_k T)^{-1}(z^k)$ , and the (necessarily single-valued) mapping  $(I + \lambda T)^{-1}$  is called a *resolvent* of  $T$  of any scalar  $\lambda > 0$ .

We now define a notion of an inexact solution of (9), specializing [35, Definition 2.1] to the case  $\epsilon = 0$ :

**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 (9) if*

$$v \in T(u) \quad \|u + \lambda_k v - z^k\| \leq \sigma \|u - z^k\|. \quad (11)$$

Observe that when  $\sigma = 0$ , we must have  $u = z^k - \lambda_k v$ , meaning that the proximal subproblem (9) must be solved exactly. We will base our approximate ADMM on the following inexact proximal point algorithm, which is [35, Algorithm 3.1] specialized to the case  $\epsilon_k \equiv 0$ :

---

**Algorithm 3.2.1** A relative-error proximal point algorithm

---

**Initialization:**

Choose some  $\lambda > 0$ , error tolerance parameter  $\sigma \in [0, 1)$ , and starting point  $z^0 \in \mathbb{R}^n$ .

**for**  $k = 0, 1, \dots$  **do**

Select any  $\lambda_k \geq \lambda$ , and let  $(u^k, v^k)$  be some inexact solution with tolerance  $\sigma$  of the subproblem  $0 \in \lambda_k T(z) + z - z^k$ , that is, find some  $(u^k, v^k)$  such that

$$v^k \in T(u^k) \quad \|u^k + \lambda_k v^k - z^k\| \leq \sigma \|u^k - z^k\|. \quad (12)$$

Set

$$z^{k+1} = z^k - \lambda_k v^k. \quad (13)$$

**end for**

---

We may summarize the recursions of Algorithm 3.2.1 as follows:

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

The behavior of Algorithm 3.2.1 has already been established in [35]:

**Proposition 9.** *If a solution to  $0 \in T(z)$  exists, then  $\{z^k\}$  generated by Algorithm 3.2.1 converges to such a solution. In addition,  $\{u^k\}$  also converges to this solution and  $v^k \rightarrow 0$ .*



*Proof.* We appeal to the analysis in [35] in the special case that  $\mathcal{H} = \mathbb{R}^n$  and  $\epsilon_k = 0$  for all  $k \geq 0$ : specifically, convergence of  $\{z^k\}$  is established in Theorem 3.1, convergence of  $\{u^k\}$  follows immediately from Corollary 4.2(2), and convergence of  $\{v^k\}$  follows immediately from Corollary 4.2(4).  $\square$

The proposition also may also be established by using the Bregman-distance results in [37, Proposition 4.4 and Corollary 4.3] in the special case of the Euclidean squared distance kernel, whose zone of definition is  $C = \mathbb{R}^n$ .

### 3.3 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 [10], we will use Algorithm 3.2.1 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 [14].

The original DR splitting method [24] is a method for solving  $A(x) + B(x) \ni 0$ , where  $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$  are maximal monotone operators on a real Hilbert space  $\mathcal{H}$ . Here, we only consider 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 [10] defines the *splitting operator*  $S_{\gamma, A, B} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  to be the set-valued map

$$S_{\gamma, A, B} = \{(r + \gamma b, s - r) \mid (s, b) \in B, (r, a) \in A, r + \gamma a = s - \gamma b\}. \quad (14)$$

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  $\gamma > 0$  [10, Theorem 4]. There is also an important relationship between the zeros of  $S_{\gamma, A, B}$  and those of  $A + B$ : letting  $\text{zer}(T)$  denote the set of all zeros of operator  $T$ , then by [10, Theorem 5] we have

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

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 [10, 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,

$$z^{k+1} = (I + S_{\gamma, A, B})^{-1}(z^k). \quad (15)$$

This viewpoint is exploited in [10] to develop approximate versions of the DR splitting method, specifically by applying an approximate version of the PPA to (15). 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 3.2.1 to  $S_{\gamma,A,B}$ , obtaining a relative-error inexact variant of DR splitting.

The recursion (15) consists of repeatedly applying the mapping

$$(I + S_{\gamma,A,B})^{-1} = \{ (s + \gamma b, r + \gamma b) \mid (s, b) \in B, (r, a) \in A, r + \gamma a = s - \gamma b \}. \quad (16)$$

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  $(s^{k+1}, b^{k+1}) \in B$  such that  $s^{k+1} + \gamma b^{k+1} = r^k + \gamma b^k$ . This calculation is equivalent to finding  $s^{k+1} = (I + \gamma B)^{-1}(r^k + \gamma b^k)$  and setting  $b^{k+1} = b^k + \frac{1}{\gamma}(r^k - w^{k+1})$ .

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

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

$$b \in B(s) \quad s + \gamma b = u, \quad (17)$$

We model this iterative process using the following generalization of the models proposed in Section 2:

**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  $(s^l, b^l) = \mathcal{B}(u, \gamma, \bar{s}, \bar{b}, l)$  for all  $l \in \mathbb{N}$ , then the sequence  $\{(s^l, b^l)\}_{l=1}^\infty$  is convergent and  $\lim_{l \rightarrow \infty} s^l + \gamma b^l = u$ .*

Intuitively, we intend  $(s^l, b^l) = \mathcal{B}(u, \gamma, \bar{s}, \bar{b}, l) \in B$  to be the  $l^{\text{th}}$  trial approximate solution to (17) 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.

As opposed to the situation with  $B$ , we assume that a set of conditions similar to (17) 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 (14) that  $(r + \gamma b, s - r) \in S_{\gamma,A,B}$ .

To apply Algorithm 3.2.1 to the operator  $S_{\gamma,A,B}$ , we need to find inexact solutions of the conditions (12) for the case  $T = S_{\gamma,A,B}$  and  $\lambda_k = 1$ . Calling the iterative procedure abstracted in Assumption 10 the “ $B$ -procedure”, and letting  $l$  be an “inner” iteration index associated with this procedure, we may attempt to execute each iteration of Algorithm 3.2.1 as applied to  $S_{\gamma,A,B}$  as follows, starting with  $l = 1$ :

- S1. Execute one step of the  $B$ -procedure, yielding  $(s^{k,l}, b^{k,l}) = \mathcal{B}(z^k, \gamma, s^k, b^k, l) \in B$  with  $s^{k,l} + \gamma b^{k,l} \approx z^k$ . We call  $(s^{k,l}, b^{k,l})$  a *trial point*.
- S2. Find a corresponding  $(r^{k,l}, a^{k,l}) \in A$  such that  $r^{k,l} + \gamma a^{k,l} = s^{k,l} - \gamma b^{k,l}$ . It follows immediately from (14) that  $(u^{k,l}, v^{k,l}) \stackrel{\text{def}}{=} (r^{k,l} + \gamma b^{k,l}, s^{k,l} - r^{k,l}) \in S_{\gamma,A,B}$ .
- S3. Test whether the  $(u^{k,l}, v^{k,l})$  satisfies the conditions on  $(u^k, v^k)$  specified in (12). If not, increment  $l \leftarrow l + 1$  and go back to step S1 to execute an additional step of the  $B$ -procedure, with the aim of producing a more accurate trial point. Otherwise, accept  $(u^k, v^k) = (u^{k,l}, v^{k,l}) \in S_{\gamma,A,B}$  as a pair satisfying (12).
- S4. Once we have accepted  $(u^k, v^k) = (u^{k,l}, v^{k,l})$ , set  $z^{k+1} = z^k - v^k$ , that is, (13) with  $\lambda_k = 1$ .

We now make step S3 more concrete: substituting  $(u^{k,l}, v^{k,l}) \stackrel{\text{def}}{=} (r^{k,l} + \gamma b^{k,l}, s^{k,l} - r^{k,l})$  for  $(u^k, v^k)$  in (12), along with and  $\lambda_k \equiv 1$  from (15), we obtain the condition

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

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

$$\|s^{k,l} + \gamma b^{k,l} - z^k\| \leq \sigma \|r^{k,l} + \gamma b^{k,l} - z^k\|. \quad (18)$$

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

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

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

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

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

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

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

$$s^{k+1} = s^{k,l} \quad r^{k+1} = r^{k,l} \quad b^{k+1} = b^k + \frac{1}{\gamma}(r^k - s^{k,l}). \quad (19)$$

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 (18), we obtain

$$\|s^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\| \leq \sigma \|r^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\|. \quad (20)$$

Similarly substituting  $z^k = r^k + \gamma b^k$  throughout steps S1-S4 and in (20) above, we arrive at the following algorithm:

---

**Algorithm 3.3.1** 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**  $k = 0, 1, 2, \dots$  **do**  
    **repeat** {**for**  $l = 1, 2, \dots$  }  
        Improve the the solution to  $(s^{k,l}, b^{k,l}) \in B$  and  $s^{k,l} + \gamma b^{k,l} \approx r^k + \gamma b^k$  by setting  
         $(s^{k,l}, b^{k,l}) = \mathcal{B}(r^k + \gamma b^k, \gamma, s^k, b^k, l)$  (thus incrementally executing a step of the  $B$ -  
        procedure)  
        Exactly find  $(r^{k,l}, a^{k,l}) \in A$  such that  $r^{k,l} + \gamma a^{k,l} = s^{k,l} - \gamma b^{k,l}$   
    **until**  $\|s^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\| \leq \sigma \|r^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\|$   
     $s^{k+1} = s^{k,l}$   
     $r^{k+1} = r^{k,l}$   
     $b^{k+1} = b^k - \frac{1}{\gamma} (s^{k,l} - r^k)$   
**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 3.3.1:*

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  $\{s^k\}$  and  $\{r^k\}$  both converge to some  $x^*$  for which  $0 \in A(x^*) + B(x^*)$ , and  $b^k$  converges to some  $b^* \in B(x^*)$  such that  $-b^* \in A(x^*)$ .*
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  $\lim_{l \rightarrow \infty} s^{\bar{k},l} = \lim_{l \rightarrow \infty} r^{\bar{k},l} = x^*$  for some  $x^*$  for which  $0 \in A(x^*) + B(x^*)$ , while  $\lim_{l \rightarrow \infty} b^{\bar{k},l} = b^*$  for some  $b^* \in B(x^*)$  such that  $-b^* \in A(x^*)$  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  $\{(s^{\bar{k},l}, b^{\bar{k},l})\}_{l=1}^\infty \in B$  converges to some limit  $(x^*, b^*)$  with  $x^* + \gamma b^* = z^{\bar{k}} = r^{\bar{k}} + \gamma b^{\bar{k}}$  as  $l \rightarrow \infty$ . By the closedness property of maximal monotone operators, we also have  $b^* \in B(x^*)$ . From the construction of the points  $(r^{\bar{k},l}, a^{\bar{k},l})$ , we have for all  $l$  that

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

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  $(r^{\bar{k},l}, a^{\bar{k},l})$  converges to the limit  $(r^*, a^*) = (J_{\gamma A}(x^* - \gamma b^*), \frac{1}{\gamma}(x^* - r^*) - b^*)$ , and from the closedness of the maximal monotone operator  $A$ , we also have  $(r^*, a^*) \in A$ . Now consider the inner-loop termination condition  $\|s^{\bar{k},l} + \gamma b^{\bar{k},l} - (r^{\bar{k}} + \gamma b^{\bar{k}})\| \leq \sigma \|r^{\bar{k},l} + \gamma b^{\bar{k},l} - (r^{\bar{k}} + \gamma b^{\bar{k}})\|$  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  $r^* + \gamma b^* = r^k + \gamma b^k = z^k = x^* + \gamma b^*$ , where the last equality was established above. Canceling  $\gamma b^*$  from both sides, we conclude that  $r^* = x^*$  and consequently that  $a^* = \frac{1}{\gamma}(x^* - r^*) - b^* = -b^*$ . Thus, we have  $b^* \in B(x^*)$  and  $-b^* = a^* \in A(r^*) = A(x^*)$ . Therefore,  $A(x^*) + B(x^*) \ni -b^* + b^* = 0$  and all the claims in the second case of the proposition have been established.

We now consider the first case of the proposition, 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 3.3.1, we have in this case that  $\{z^k\} = \{r^k + \gamma b^k\}$  is identical to the sequence generated by Algorithm 3.2.1 with  $T = S_{\gamma,A,B}$  and  $\lambda_k \equiv 1$ . Proposition 9 then implies that  $z^k \rightarrow z^*$  such that  $0 \in S_{\gamma,A,B}(z^*)$ . Next, in view of step S2 above, we note that the point  $(u^k, v^k) \in T$  of Algorithm 3.2.1 may be expressed as

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

Proposition 9 asserts that  $(u^k, v^k) \rightarrow (z^*, 0)$ , 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,  $b^{k,l(k)} \in B(s^{k,l(k)})$  for all  $k$ , so  $s^{k,l(k)} \in J_{\gamma B}(s^{k,l(k)} + \gamma b^{k,l(k)})$ , 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

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

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

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

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  $b^{k,l(k)} \in B(s^{k,l(k)})$  to obtain  $b^* \in B(x^*)$ . 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

$$\lim_{k \rightarrow \infty} a^{k,l(k)} = \lim_{k \rightarrow \infty} \left\{ \frac{1}{\gamma} (s^{k,l(k)} - r^{k,l(k)}) - b^{k,l(k)} \right\} = -b^*.$$

Since we have  $a^{k,l(k)} \in A(y^{k,l(k)})$  by construction, we may take the limit to obtain  $-b^* \in A(x^*)$ . In conclusion, we have  $b^* \in B(x^*)$ ,  $-b^* \in A(x^*)$  and therefore  $0 \in A(x^*) + B(x^*)$ , and all claims in the first case of the proposition are established.  $\square$

To summarize, the proposition states that Algorithm 3.3.1 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.

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

We now derive an inexact version of the ADMM from the Algorithm 3.3.1. The standard dual formulation of the problem (1) is

$$\min_{p \in \mathbb{R}^m} f^*(-M^\top p) + g^*(p), \quad (21)$$

where  $f^*$  and  $g^*$  are the convex conjugate of  $f$  and  $g$ , respectively. In [16], Gabay showed that ADMM can be derived by applying the Douglas-Rachford splitting method to (21) with  $A = \partial[f^* \circ (-M^\top)]$  and  $B = \partial g^*$ . Unfortunately, these choices of  $A$  and  $B$  are inconvenient for Algorithm 3.3.1, because verifying the condition  $a \in A(r)$  would require an exact minimization involving the function  $f$ , 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 [8, Section 3.5.6], applying Douglas-Rachford splitting to this choice of  $A$  and  $B$  results in the algorithm

$$s^{k+1} = \arg \min_s \left\{ f(s) + \frac{1}{2\gamma} \|s - (r^k + \gamma b^k)\|^2 \right\} \quad (22)$$

$$r^{k+1} = \arg \min_r \left\{ g(Mr) + \frac{1}{2\gamma} \|r - (s^{k+1} - \gamma b^k)\|^2 \right\} \quad (23)$$

$$b^{k+1} = b^k + \frac{1}{\gamma}(r^{k+1} - s^{k+1}). \quad (24)$$

This ADMM-like method is appropriate when the composition of  $g$  and  $M$  is convenient to work with, so that the minimization (23) is not too hard to perform. Since redefining  $g \leftarrow g \circ M$  and then  $M \leftarrow I$  in problem (1) results in exactly the same algorithm when applying (22)-(24), we may without loss of generality take  $M = I$ . Furthermore, when  $M = I$ , Proposition 3.43 of [8] shows that (22)-(24) is identical to the ADMM if one sets  $c = 1/\gamma$ ; these ideas are developed somewhat further in [40].

Fixing  $M = I$ , we now consider applying the partially exact Douglas-Rachford splitting method of Algorithm 3.3.1 with this same choice of  $A = \partial[g \circ M] = \partial g$  and  $B = \partial f$ . We now develop an analysis similar to [8, Proposition 3.43], but in the context of Algorithm 3.3.1: this algorithm requires that the resolvent operation in step S2 be carried out exactly, while the

resolvent calculation in step S1 may be approximate. Therefore, the subproblems associated with S1 and S2 are respectively

$$s^{k,l} \approx \arg \min_s \left\{ f(s) + \frac{1}{2\gamma} \|s - (r^k + \gamma b^k)\|^2 \right\} \quad (25)$$

$$r^{k,l} = \arg \min_r \left\{ g(r) + \frac{1}{2\gamma} \|r - (s^{k,l} - \gamma b^{k,l})\|^2 \right\}. \quad (26)$$

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

$$s^{k,l} \approx \arg \min_s \left\{ f(s) - \langle b^k, s \rangle + \frac{1}{2\gamma} \|s - r^k\|^2 \right\} \quad (27)$$

$$r^{k,l} = \arg \min_r \left\{ g(r) + \langle b^{k,l}, r \rangle + \frac{1}{2\gamma} \|r - s^{k,l}\|^2 \right\}. \quad (28)$$

Next, we define some parallel notation for Algorithm 3.3.1 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 (25)-(26) may be expressed as

$$x^{k,l} \approx \arg \min_x \left\{ f(x) + \langle p^k, x \rangle + \frac{c}{2} \|x - z^k\|^2 \right\} \quad (29)$$

$$z^{k,l} = \arg \min_z \left\{ g(z) - \langle p^{k,l}, z \rangle + \frac{c}{2} \|z - x^{k,l}\|^2 \right\}, \quad (30)$$

which are identical to the minimization operations of the  $M = I$  case of the ADMM. We now state the precise form of our proposed algorithm, making the meaning of the “ $\approx$ ” in (29) 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 3.4.1** 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**  $k = 0, 1, 2, \dots$ }

**repeat** {**for**  $l = 1, 2, \dots$ }

        Improve the solution to  $x^{k+1} \approx \arg \min_x \{f(x) + \langle p^k, x \rangle + \frac{c}{2} \|x - z^k\|^2\}$  by

        taking  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$

$p^{k,l} = p^k + c(x^{k,l} - z^k) - y_1^{k,l}$

$z^{k,l} = \arg \min_{z \in \mathbb{R}^m} \{g(z) - \langle p^{k,l}, z \rangle + \frac{c}{2} \|x^{k,l} - z\|^2\}$

**until**  $\|y_1^{k,l}\| \leq \sigma \|p^{k,l} - p^k - c(z^{k,l} - z^k)\|$

$x^{k+1} = x^{k,l}$

$z^{k+1} = z^{k,l}$

$p^{k+1} = p^k + c(x^{k,l} - z^k)$

**until** Overall convergence

---

**Proposition 12.** *Suppose that there exists some  $x^*$  such that  $0 \in \partial f(x^*) + \partial g(x^*)$  and  $\mathcal{F}$  meets the conditions in Assumption 2 for  $M = I$ . Then there are two possible execution sequences for Algorithm 3.4.1:*

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  $\{x^k\}$  and  $\{z^k\}$  both converge to some  $x^*$  for which  $0 \in \partial g(x^*) + \partial f(x^*)$ , and  $\{p^k\}$  converges to some  $p^* \in \partial g(x^*)$  such that  $-p^* \in \partial f(x^*)$ .*
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  $0 \in \partial f(x^*) + \partial g(x^*)$ , while  $\lim_{l \rightarrow \infty} p^{\bar{k},l} = p^*$  for some  $p^* \in \partial g(x^*)$  such that  $-p^* \in \partial f(x^*)$ .*

*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 3.3.1 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 [30, Theorem 23.8], we obtain that for each  $l$  in the inner loop, we have

$$\begin{aligned} & y_1^{k,l} \in \partial f(x^{k,l}) + p^k + c(x^{k,l} - z^k) \\ \Leftrightarrow & -p^k - c(x^{k,l} - z^k) + y_1^{k,l} \in \partial f(x^{k,l}) \\ \Leftrightarrow & -p^{k,l} \in \partial f(x^{k,l}) \end{aligned}$$

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

$$\begin{aligned} s^{k,l} + \gamma b^{k,l} &= x^{k,l} - \frac{1}{c} p^{k,l} \\ &= x^{k,l} - \frac{1}{c} (p^k + c(x^{k,l} - z^k) - y_1^{k,l}) \\ &= -\frac{1}{c} p^k + z^k + \frac{1}{c} y_1^{k,l} \\ &= \gamma b^k + r^k + \gamma y_1^{k,l}. \end{aligned}$$

Since  $\lim_{l \rightarrow \infty} y_1^{k,l} = 0$ , it follows that  $\lim_{l \rightarrow \infty} s^{k,l} + \gamma b^{k,l} = r^k + \gamma b^k$ . This means that the procedure of taking  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^{k,l})$  followed by  $p^{k,l} = p^k + c(x^{k,l} - z^k) - y_1^{k,l}$  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  $\bar{b} = b^k = -p^k$ ).

Next, from  $z^{k,l} = \arg \min_{z \in \mathbb{R}^m} \{g(z) - \langle p^{k,l}, z \rangle + \frac{c}{2} \|x^{k,l} - z\|^2\}$  and [30, Theorem 23.8], we must have  $0 \in \partial g(z^{k,l}) - p^{k,l} + c(z^{k,l} - x^{k,l})$ , and hence  $p^{k,l} + c(x^{k,l} - z^{k,l}) \in \partial g(z^{k,l})$ . Therefore, we take  $a^{k,l} = p^{k,l} + c(x^{k,l} - z^{k,l})$  and  $r^{k,l} = z^{k,l}$ , and then have  $(r^{k,l}, a^{k,l}) \in A$  and

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

Therefore, we have  $(r^{k,l}, a^{k,l}) \in A$  and  $r^{k,l} + \gamma a^{k,l} = s^{k,l} - \gamma b^{k,l}$ , which are the unique determining conditions for  $(r^{k,l}, a^{k,l})$  in Algorithm 3.3.1. Thus all the steps within the inner loop of Algorithm 3.4.1 are equivalent to the steps in the inner loop of Algorithm 3.3.1. We next turn to



the termination condition for the inner loop of Algorithm 3.3.1,  $\|s^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\| \leq \sigma \|r^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k)\|$ . Substituting

$$s^{k,l} = x^{k,l} \quad \gamma = 1/c \quad b^{k,l} = -p^{k,l} = -p^k - c(x^{k,l} - z^k) + y_1^{k,l} \quad r^k = z^k \quad b^k = -p^k$$

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

$$s^{k,l} + \gamma b^{k,l} - (r^k + \gamma b^k) = x^{k,l} + \frac{1}{c}(-p^k - c(x^{k,l} - z^k) + y_1^{k,l}) - z^k - \frac{1}{c}(-p^k) = \frac{1}{c}y_1^{k,l}.$$

Performing similar substitutions in the expression within the norm on the right-hand side of Algorithm 3.3.1's inner-loop termination condition, we obtain

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

Thus we obtain that the following inner-loop termination condition is in this case equivalent to that of Algorithm 3.3.1:

$$\left\| \frac{1}{c}y_1^{k,l} \right\| \leq \|z^{k,l} - z^k - \frac{1}{c}(p^{k,l} - p^k)\|.$$

Multiplying through by  $c$ , we obtain exactly the same inner-loop termination condition as in Algorithm 3.4.1. Therefore, if one enters iteration  $k$  with  $p^k = -b^k$  and  $r^k = z^k$ , Algorithm 3.4.1 will execute exactly the same number of inner-loop iterations as Algorithm 3.3.1 with the  $\mathcal{B}$ -procedure constructed as described above.

Once the inner loop has terminated, Algorithm 3.4.1 performs the updates  $x^{k+1} = x^{k,l}$ ,  $z^{k+1} = z^{k,l}$ , and  $p^{k+1} = p^k + c(x^{k,l} - z^k)$ . 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 3.3.1. 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 3.3.1. Finally, if we let  $b^{k+1} = -p^{k+1}$ , we have

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

which is exactly the same value of  $b^{k+1}$  computed by Algorithm 3.3.1. 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.  $\square$

## 4 Approximate ADMM Algorithms Derived from Lagrangian Splitting

We now develop two new approximate ADMM algorithms by modifying the Lagrangian splitting analysis pioneered in [14]. 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 Section 3, the algorithms we derive here place no restrictions on  $M$  and allow both the  $x$  and  $z$  minimizations to be approximate.

## 4.1 A parametric conjugate duality framework

We now introduce a parametric conjugate duality framework specializing the one described in [30, Chapters 28-30] and [31] 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:

$$F_1(x, z, u_1) = \begin{cases} f(x), & \text{if } u_1 + Mx = 0 \\ +\infty, & \text{otherwise} \end{cases} \quad (31)$$

$$F_2(x, z, u_2) = \begin{cases} g(z), & \text{if } u_2 - z = 0 \\ +\infty, & \text{otherwise.} \end{cases} \quad (32)$$

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 (P) to be the infimal convolution [30, page 34] of  $F_1$  and  $F_2$  with respect to the last argument, that is

$$\begin{aligned} F(x, z, u) &= \inf_{\substack{u_1, u_2 : \\ u_1 + u_2 = u}} \{F_1(x, z, u_1) + F_2(x, z, u_2)\} \\ &= \begin{cases} f(x) + g(z), & \text{if } Mx - z + u = 0 \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \quad (33)$$

Here,  $u \in \mathbb{R}^m$  may be considered to represent a perturbation of the constraints  $Mx = z$ . It is also easily seen that  $F$  is closed and convex [30, Theorem 5.4]. Within this framework, the original problem (P) is equivalent to the primal problem

$$\min_{x \in \mathbb{R}^n} F(x, z, 0). \quad (34)$$

One obtains the Lagrangian  $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow [-\infty, +\infty]$  of (34) by taking the concave conjugate [30, p. 111] of  $F$  with respect to the perturbation argument  $u$ :

$$\begin{aligned} L(x, z, p) &= \inf_{u \in \mathbb{R}^m} \{F(x, z, u) - \langle p, u \rangle\} \\ &= \inf_{u = z - Mx} \{f(x) + g(z) - \langle p, u \rangle\} \\ &= f(x) + g(z) + \langle p, Mx - z \rangle. \end{aligned} \quad (35)$$

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  $(y_1, y_2, u) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  such that

$$\begin{aligned} L(x', z', p) &\geq L(x, z, p) + \langle y_1, x' - x \rangle + \langle y_2, z' - z \rangle & \forall (x', z') \in \mathbb{R}^n \times \mathbb{R}^m \\ L(x, z, p') &\leq L(x, z, p) - \langle u, p' - p \rangle & \forall p' \in \mathbb{R}^m. \end{aligned}$$

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

$$\begin{aligned} \partial L(x, z, p) &= \partial_{(x, z)} L(x, z, p) \times \partial_p (-L(x, z, p)) \\ &= \partial_x L(x, z, p) \times \partial_z L(x, z, p) \times \partial_p (-L(x, z, p)) \\ &= \{\partial f(x) + M^\top p\} \times \{\partial g(z) - p\} \times \{z - Mx\}. \end{aligned} \quad (36)$$

The following result is now immediate:

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

We obtain  $L_1, L_2 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow (0, \infty]$ , the Lagrangian functions respectively corresponding to  $F_1$  and  $F_2$ , by taking their concave conjugates with respect to the perturbation arguments  $u_1$  and  $u_2$ :

$$L_1(x, z, p) = \inf_{u_1 \in \mathbb{R}^m} \{F_1(x, z, u_1) - \langle p, u_1 \rangle\} = f(x) + \langle p, Mx \rangle$$

$$L_2(x, z, p) = \inf_{u_2 \in \mathbb{R}^m} \{F_2(x, z, u_2) - \langle p, u_2 \rangle\} = g(z) - \langle p, z \rangle$$

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

$$\partial L_1(x, z, p) = \{\partial f(x) + M^\top p\} \times \{0\} \times \{-Mx\} \quad (37)$$

$$\partial L_2(x, z, p) = \{0\} \times \{\partial g(z) - p\} \times \{z\}, \quad (38)$$

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 in  $L_2(x, z, p)$ .

The *parametric dual function*  $Q : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow [-\infty, +\infty]$  of (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

$$\begin{aligned} Q(y_1, y_2, p) &= \inf_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} \{L(x, z, p) - \langle y_1, x \rangle - \langle y_2, z \rangle\} \\ &= \inf_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} \{f(x) + g(z) + \langle p, Mx - z \rangle - \langle y_1, x \rangle - \langle y_2, z \rangle\} \\ &= \inf_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} \{f(x) + \langle p, Mx \rangle - \langle y_1, x \rangle + g(z) - \langle p, z \rangle - \langle y_2, z \rangle\} \\ &= \inf_{x \in \mathbb{R}^n} \{f(x) + \langle p, Mx \rangle - \langle y_1, x \rangle\} + \inf_{z \in \mathbb{R}^m} \{g(z) - \langle p, z \rangle - \langle y_2, z \rangle\} \\ &= \inf_{x \in \mathbb{R}^n} \{L_1(x, 0, p) - \langle y_1, x \rangle\} + \inf_{z \in \mathbb{R}^m} \{L_2(0, z, p) - \langle y_2, z \rangle\}. \end{aligned} \quad (39)$$

In (39), 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  $p \in \mathbb{R}^m$ ,

$$Q_1(y_1, y_2, p) = -F_1^*(y_1, y_2, p) = \inf_{x \in \mathbb{R}^m} \{L_1(x, 0, p) - \langle y_1, x \rangle\} = Q_1(y_1, 0, p) \quad (40)$$

$$Q_2(y_1, y_2, p) = -F_2^*(y_1, y_2, p) = \inf_{z \in \mathbb{R}^m} \{L_2(0, z, p) - \langle y_2, z \rangle\} = Q_2(0, y_2, p), \quad (41)$$

and note that it follows from (39) that

$$Q(y_1, y_2, p) = Q_1(y_1, y_2, p) + Q_2(y_1, y_2, p). \quad (42)$$

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

$$\begin{aligned} (y_1, y_2, p) \in \partial F_1(x, z, u) &\Leftrightarrow (y_1, y_2, u) \in \partial L_1(x, z, p) \Leftrightarrow (x, z, u) \in \partial Q_1(y_1, y_2, p) \\ (y_1, y_2, p) \in \partial F_2(x, z, u) &\Leftrightarrow (y_1, y_2, u) \in \partial L_2(x, z, p) \Leftrightarrow (x, z, u) \in \partial Q_2(y_1, y_2, p). \end{aligned}$$

We have that  $(x, z, u_1) \in \partial Q_1(y_1, y_2, p)$  if only if

$$Q_1(y'_1, y'_2, p') \leq Q_1(y_1, y_2, p) - \langle x, y'_1 - y_1 \rangle - \langle z, y'_2 - y_2 \rangle - \langle u_1, p' - p \rangle, \quad (43)$$

and similarly that  $(x, z, u_2) \in \partial Q_2(y_1, y_2, p)$  if only if

$$Q_2(y'_1, y'_2, p') \leq Q_2(y_1, y_2, p) - \langle x, y'_1 - y_1 \rangle - \langle z, y'_2 - y_2 \rangle - \langle u_2, p' - p \rangle. \quad (44)$$

The point-to-set mappings  $\partial F$ ,  $\partial L$  and  $\partial Q$  are all maximal monotone [30, p. 240] and for all  $(x, z), (y_1, y_2) \in \mathbb{R}^n \times \mathbb{R}^m$  and  $u, p \in \mathbb{R}^m$ ,

$$(y_1, y_2, p) \in \partial F(x, z, u) \Leftrightarrow (y_1, y_2, u) \in \partial L(x, z, p) \Leftrightarrow (x, z, u) \in \partial Q(y_1, y_2, p).$$

The *dual function*  $Q_0 : \mathbb{R}^m \rightarrow [-\infty, +\infty)$  of (P) is the parametric dual function evaluated at  $(y_1, y_2) = 0$ , that is,

$$\begin{aligned} Q_0(p) &= Q(0, 0, p) \\ &= Q_1(0, 0, p) + Q_2(0, 0, p) \\ &= \inf_{x \in \mathbb{R}^n} \{f(x) + \langle p, Mx \rangle\} + \inf_{z \in \mathbb{R}^m} \{g(z) - \langle p, z \rangle\} \\ &= (-f^*(-M^\top p)) + (-g^*(p)), \end{aligned} \quad (45)$$

where “\*” again denotes the convex conjugate. The dual problem corresponding to (34) is that of maximizing  $Q_0(p)$  over  $p \in \mathbb{R}^m$ , that is,

$$\max_{p \in \mathbb{R}^m} Q_1(0, 0, p) + Q_2(0, 0, p). \quad (D)$$

In view of (45), problem (D) is identical to the usual dual problem (21) of (P). As an application of Fenchel's inequality [30, 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 (2), and let  $x^{k+1}$  denote some approximate solution to (2) with

$$y_1^{k+1} \in \partial_x \left[ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right]_{x=x^{k+1}}. \quad (46)$$

Note that if we were to exactly solve the subproblem, 0 would be a possible value of  $y_1^{k+1}$ . Employing [30, Theorem 23.8], we have

$$(46) \Leftrightarrow y_1^{k+1} \in \partial f(x^{k+1}) + M^\top p^k + cM^\top (Mx^{k+1} - z^k) \quad (47)$$

$$\Leftrightarrow (y_1^{k+1}, 0, -Mx^{k+1}) \in \{\partial f(x^{k+1}) + M^\top(p^k + c(Mx^{k+1} - z^k))\} \times \{0\} \times \{-Mx^{k+1}\}$$

$$\Leftrightarrow (y_1^{k+1}, 0, -Mx^{k+1}) \in \partial L_1(x^{k+1}, z^k, p^k + c(Mx^{k+1} - z^k)). \quad (48)$$

Now suppose we have some inexact solution  $z^{k+1}$  to (3) and

$$y_2^{k+1} \in \partial_z \left[ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|z - Mx^{k+1}\|^2 \right]_{z=z^{k+1}}. \quad (49)$$

Much as for (46), if  $z^{k+1}$  were an exact solution to (3), it would be possible to choose  $y_2^{k+1} = 0$  in (49). Following a similar development to that of (48), we have

$$(49) \Leftrightarrow y_2^{k+1} \in \partial g(z^{k+1}) - p^k + c(z^{k+1} - Mx^{k+1}) \quad (50)$$

$$\Leftrightarrow (0, y_2^{k+1}, z^{k+1}) \in \{0\} \times \{\partial g(z^{k+1}) - (p^k + c(Mx^{k+1} - z^{k+1}))\} \times \{z^{k+1}\}$$

$$\Leftrightarrow (0, y_2^{k+1}, z^{k+1}) \in \partial L_2(x^{k+1}, z^{k+1}, p^k + c(Mx^{k+1} - z^{k+1})). \quad (51)$$

Using the standard multiplier update

$$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1}) \quad (52)$$

and letting

$$\mu^{k+1} = p^k + c(Mx^{k+1} - z^k), \quad (53)$$

we observe that  $\mu^{k+1} - p^{k+1} = c(z^{k+1} - z^k)$ , along with

$$(46) \Leftrightarrow (y_1^{k+1}, 0, -Mx^{k+1}) \in \partial L_1(x^{k+1}, z^k, \mu^{k+1}) \quad (54)$$

$$(49) \Leftrightarrow (0, y_2^{k+1}, z^{k+1}) \in \partial L_2(x^{k+1}, z^{k+1}, p^{k+1}). \quad (55)$$

## 4.2 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.

**Lemma 14.** *Suppose that  $\{x^k\}, \{y_1^k\} \subset \mathbb{R}^n$  and  $\{z^k\}, \{p^k\}, \{y_2^k\} \subset \mathbb{R}^m$  obey the recursions (46), (49), and (52). If it is also true that*

$$Mx^k - z^k \rightarrow 0 \quad Mx^{k+1} - z^k \rightarrow 0 \quad y_1^k \rightarrow 0 \quad y_2^k \rightarrow 0,$$

*then all limit points of the sequence  $\{(x^k, z^k, p^k)\}$  are KKT points of (P).*

*Proof.* Consider any limit point  $(x^\infty, z^\infty, p^\infty)$  of  $\{(x^k, z^k, p^k)\}$ , along with an infinite index set  $\mathcal{K} \subseteq \mathbb{N}$  such that  $(x^{k+1}, z^{k+1}, p^{k+1}) \rightarrow_{\mathcal{K}} (x^\infty, z^\infty, p^\infty)$ . By rearranging (47) and (50), which are respectively equivalent to they hypotheses (46) and (49), we arrive at

$$y_1^{k+1} - M^\top (p^k + c(Mx^{k+1} - z^k)) \in \partial f(x^{k+1}) \quad (56)$$

$$y_2^{k+1} + (p^k + c(Mx^{k+1} - z^{k+1})) \in \partial g(z^{k+1}) \quad (57)$$

From (52) and the hypothesis that  $Mx^k - z^k \rightarrow 0$ , we conclude that  $\{p^k\}$  has the same limit over  $\mathcal{K}$  as  $\{p^{k+1}\}$  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 (56) and (57) and using that subdifferentials of closed convex functions have closed graphs that

$$-M^\top p^\infty \in \partial f(x^\infty) \quad p^\infty \in \partial g(z^\infty). \quad (58)$$

Since we have assumed that  $Mx^k - z^k \rightarrow 0$ , it also follows by taking limits over  $\mathcal{K}$  that  $Mx^\infty = z^\infty$ . In view of (58), we conclude that  $(x^\infty, z^\infty, p^\infty)$  is a KKT point of (P).  $\square$

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

Clearly, any two of the conditions

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

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

We now give another result, similar to Lemma 14, that will prove useful for cases in which various “inner loops” of the algorithms to be proposed below do not terminate:

**Lemma 15.** *Suppose  $\bar{x} \in \mathbb{R}^n$  and  $\bar{p} \in \mathbb{R}^m$ , and that the sequences  $\{\hat{x}^i\}_{i=1}^\infty, \{\hat{y}_1^i\}_{i=1}^\infty \subset \mathbb{R}^n$  and  $\{\hat{z}^i\}_{i=1}^\infty, \{\hat{p}^i\}_{i=1}^\infty, \{\hat{y}_2^i\}_{i=1}^\infty \subset \mathbb{R}^m$  conform for all  $i$  to the recursions*

$$\begin{aligned} \hat{y}_2^i &\in \partial \left[ g(z) - \langle \bar{p}, z \rangle + \frac{c}{2} \|M\bar{x} - z\|^2 \right]_{z=\hat{z}^i} \\ \hat{p}^i &= \bar{p} + c(M\bar{x} - \hat{z}^i) \\ \hat{y}_1^i &\in \partial \left[ f(x) + \langle \hat{p}^i, Mx \rangle + \frac{c}{2} \|Mx - \hat{z}^i\|^2 \right]_{x=\hat{x}^i}. \end{aligned}$$

*If it is also true that*

$$\lim_{i \rightarrow \infty} \hat{y}_1^i = 0 \quad \lim_{i \rightarrow \infty} \hat{y}_2^i = 0 \quad \lim_{i \rightarrow \infty} M\hat{x}^i - \hat{z}^i = 0, \quad (59)$$

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

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

$$\bar{y}_2^i + \hat{p}^i \in \partial g(\hat{z}^i) \quad \bar{y}_1^i - M^\top \hat{p}^i - c(M\hat{x}^i - \hat{z}^i) \in \partial f(\hat{x}^i) \quad (60)$$

Taking limits over  $\mathcal{K}$  in (60), applying the hypotheses in (59), and using that subdifferentials of closed functions are closed and  $M\hat{x}^i - \hat{z}^i \rightarrow 0$ , we obtain

$$\hat{p}^\infty \in \partial g(\hat{z}^\infty) \quad -M^\top \hat{p}^\infty \in \partial f(\hat{x}^\infty) \quad M\hat{x}^\infty = \hat{z}^\infty, \quad (61)$$

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

The following special case of Lemma 15 will prove useful in cases in which the  $g$  subproblems is easy to solve exactly:

**Corollary 16.** *Suppose  $\bar{x} \in \mathbb{R}^n$  and  $\bar{p}, \hat{p}, \hat{z} \in \mathbb{R}^m$  satisfy*

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

and that  $\{\hat{x}^i\}_{i=1}^\infty, \{\hat{y}_1^i\}_{i=1}^\infty \subset \mathbb{R}^n$  satisfy the inclusion

$$\hat{y}_1^i \in \partial \left[ f(x) + \langle \hat{p}, Mx \rangle + \frac{c}{2} \|Mx - \hat{z}\|^2 \right]_{x=\hat{x}^i}$$

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

$$\lim_{i \rightarrow \infty} \hat{y}_1^i = 0 \qquad \lim_{i \rightarrow \infty} M\hat{x}^i - \hat{z} = 0,$$

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

*Proof.* The result follows immediately by applying Lemma 15 in the case that  $\hat{z}^i = \hat{z}$ ,  $\hat{y}_2^i = 0$ , and  $\hat{p}^i = \hat{p}$  for all  $i$ .  $\square$

We now use the monotonicity of subdifferential mappings such as  $\partial L$ ,  $\partial L_1$  and  $L_2$  to derive crucial inequalities that will prove useful in analyzing our proposed algorithms:

**Lemma 17.** *Suppose that  $(y_2^{k+1}, z^{k+1})$  satisfies (49),  $(y_2^k, z^k)$  satisfies (49) with  $k$  replaced by  $k-1$ , and  $p^{k+1} = p^k + c(Mx^k - z^k)$ . Then*

$$\|Mx^{k+1} - z^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \leq \|Mx^{k+1} - z^k\|^2 + \frac{2}{c} \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle. \quad (62)$$

*Proof.* Applying (55) and its equivalent with  $k$  replaced by  $k-1$ , we have

$$\begin{aligned}(0, y_2^k, z^k) &\in \partial L_2(x^k, z^k, p^k) \\ (0, y_2^{k+1}, z^{k+1}) &\in \partial L_2(x^{k+1}, z^{k+1}, p^{k+1}).\end{aligned}$$

Since  $\partial L_2$  is monotone [30, Corollary 37.5.2], the above two inclusions imply that

$$\begin{aligned}&\langle p^k - p^{k+1}, z^k - z^{k+1} \rangle + \langle y_2^k - y_2^{k+1}, z^k - z^{k+1} \rangle \geq 0 \\ \Leftrightarrow &\langle -c(Mx^{k+1} - z^{k+1}), z^k - z^{k+1} \rangle + \langle y_2^k - y_2^{k+1}, z^k - z^{k+1} \rangle \geq 0 \\ \Leftrightarrow &c \langle Mx^{k+1} - z^{k+1}, z^k - z^{k+1} \rangle - \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle \leq 0,\end{aligned}$$

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

$$\|Mx^{k+1} - z^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 - \|Mx^{k+1} - z^k\|^2 - \frac{2}{c} \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle \leq 0,$$

which we may rearrange into (62).  $\square$

The next two propositions assert that under certain conditions, the sequences produced by algorithms conforming to (46), (49), and (52) are asymptotically optimal even when  $\{x^k\}$  does not have any limit points.

**Proposition 18.** *Suppose that  $\{x^k\}, \{y_1^k\} \subset \mathbb{R}^n$  and  $\{z^k\}, \{p^k\}, \{y_2^k\} \subset \mathbb{R}^m$  obey the recursions (46), (49), and (52), and Assumption 1 holds. If the sequences  $\{Mx^k\}$  and  $\{p^k\}$  are bounded and we also have*

$$z^{k+1} - z^k \rightarrow 0 \quad Mx^k - z^k \rightarrow 0 \quad y_1^k \rightarrow 0 \quad y_2^k \rightarrow 0 \quad \langle y_1^k, x^k \rangle \rightarrow 0 \quad \langle y_2^k, z^k \rangle \rightarrow 0, \quad (63)$$

*then every limit point of  $\{p^k\}$  is an optimal solution to the dual problem (D).*

*Proof.* Let  $(x^*, z^*, p^*)$  be a saddle point for (P). Defining  $\mu^k$  as in (53), we have from (54) that  $(y_1^{k+1}, 0, -Mx^{k+1}) \in \partial L_1(x^{k+1}, z^k, \mu^{k+1})$ . Since  $\partial Q_1$  is the partial inverse of  $\partial L_1$ , we therefore have

$$(x^{k+1}, z^k, -Mx^{k+1}) \in \partial Q_1(y_1^{k+1}, 0, \mu^{k+1}).$$

From (43), we have for all  $(y'_1, y'_2, p') \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  that

$$Q_1(y'_1, y'_2, p') \leq Q_1(y_1^{k+1}, 0, \mu^{k+1}) - \langle x^{k+1}, y'_1 - y_1^{k+1} \rangle - \langle z^k, y'_2 - 0 \rangle - \langle -Mx^{k+1}, p' - \mu^{k+1} \rangle.$$

Let  $p^*$  be some optimal dual solution, which must exist by Assumption 1. Setting  $(y'_1, y'_2, p') = (0, 0, p^*)$  in the previous inequality, we have

$$Q_1(0, 0, p^*) \leq Q_1(y_1^{k+1}, 0, \mu^{k+1}) - \langle x^{k+1}, 0 - y_1^{k+1} \rangle - \langle -Mx^{k+1}, p^* - \mu^{k+1} \rangle. \quad (64)$$

Similarly, (55) and  $\partial Q_2$  being the partial inverse of  $\partial L_2$  imply that

$$(x^{k+1}, z^{k+1}, z^{k+1}) \in \partial Q_2(0, y_2^{k+1}, p^{k+1}),$$

and therefore

$$Q_2(0, 0, p^*) \leq Q_2(0, y_2^{k+1}, p^{k+1}) - \langle z^{k+1}, 0 - y_2^{k+1} \rangle - \langle z^{k+1}, p^* - p^{k+1} \rangle. \quad (65)$$

From (42), we know that  $Q(0, 0, p^*) = Q_1(0, 0, p^*) + Q_2(0, 0, p^*)$ . Substituting this equation into the inequality obtained by adding (64) and (65), we arrive at

$$Q(0, 0, p^*) \leq Q_1(y_1^{k+1}, 0, \mu^{k+1}) + Q_2(0, y_2^{k+1}, p^{k+1}) + \langle x^{k+1}, y_1^{k+1} \rangle + \langle z^{k+1}, y_2^{k+1} \rangle - \langle -Mx^{k+1}, p^* - \mu^{k+1} \rangle - \langle z^{k+1}, p^* - p^{k+1} \rangle. \quad (66)$$

Since  $p^{k+1} = \mu^{k+1} + c(z^k - z^{k+1})$ , we can rewrite (66) as follows:

$$\begin{aligned} Q(0, 0, p^*) &\leq Q_1(y_1^{k+1}, 0, \mu^{k+1}) + Q_2(0, y_2^{k+1}, p^{k+1}) + \langle x^{k+1}, y_1^{k+1} \rangle + \langle z^{k+1}, y_2^{k+1} \rangle \\ &\quad - \langle -Mx^{k+1} + z^{k+1}, p^* - p^{k+1} \rangle - c \langle -Mx^{k+1}, z^k - z^{k+1} \rangle \\ &= Q_1(y_1^{k+1}, 0, \mu^{k+1}) + Q_2(0, y_2^{k+1}, p^{k+1}) + \langle x^{k+1}, y_1^{k+1} \rangle + \langle z^{k+1}, y_2^{k+1} \rangle \\ &\quad + \langle Mx^{k+1} - z^{k+1}, p^* - p^{k+1} \rangle + c \langle Mx^{k+1}, z^k - z^{k+1} \rangle. \end{aligned} \quad (67)$$

From the hypotheses that  $\{p^k\}$  is bounded and  $Mx^{k+1} - z^{k+1} \rightarrow 0$ , we conclude that  $\langle Mx^{k+1} - z^{k+1}, p^* - p^{k+1} \rangle \rightarrow 0$ . Similarly, the hypotheses that  $\{Mx^k\}$  is bounded and



$z^{k+1} - z^k \rightarrow 0$ , imply that  $\langle Mx^{k+1}, z^k - z^{k+1} \rangle \rightarrow 0$ . Since we have also assumed that  $\langle y_1^{k+1}, x^{k+1} \rangle \rightarrow 0$  and  $\langle y_2^{k+1}, z^{k+1} \rangle \rightarrow 0$ , we conclude that the last four terms in (67) all converge to 0.

Consider any limit point  $p^\infty$  of  $\{p^k\}$ , 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 (67) and using that  $Q_1$  and  $Q_2$  are closed (upper semicontinuous) concave functions, we have

$$\begin{aligned}
f(x^*) + g(z^*) &= Q(0, 0, p^*) \\
&\leq \liminf_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} \{Q_1(y_1^{k+1}, 0, \mu^{k+1}) + Q_2(0, y_2^{k+1}, p^{k+1})\} \\
&\leq \limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} \{Q_1(y_1^{k+1}, 0, \mu^{k+1}) + Q_2(0, y_2^{k+1}, p^{k+1})\} \\
&\leq \limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} Q_1(y_1^{k+1}, 0, \mu^{k+1}) + \limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} Q_2(0, y_2^{k+1}, p^{k+1}) \\
&\leq Q_1\left(\lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} y_1^{k+1}, 0, \lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} \mu^{k+1}\right) + Q_2\left(0, \lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} y_2^{k+1}, \lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} p^{k+1}\right) \\
&= Q_1(0, 0, p^\infty) + Q_2(0, 0, p^\infty) \\
&= Q(0, 0, p^\infty) \\
&\leq f(x^*) + g(z^*),
\end{aligned}$$

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

**Proposition 19.** *Suppose that  $\{x^k\}, \{y_1^k\} \subset \mathbb{R}^n$  and  $\{z^k\}, \{p^k\}, \{y_2^k\} \subset \mathbb{R}^m$  obey the recursions (46), (49), and (52). If the same limits in (63) hold and the sequences  $\{p^k\}$  and  $\{z^k\}$  are bounded, then*

$$\limsup_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} \leq \inf_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\}. \quad (68)$$

If Assumption 1 is satisfied, then the stronger condition

$$\lim_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} = \inf_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\} = f(x^*) + g(Mx^*) \quad (69)$$

holds, where  $x^*$  is any solution of (1).

*Proof.* Adding  $c(z^k - z^{k+1})$  to both sides of the inclusion (50), which we know holds from the hypothesis (49), we obtain

$$y_2^{k+1} + c(z^k - z^{k+1}) \in \partial g(z^{k+1}) - p^k + c(z^{k+1} - Mx^{k+1}) + c(z^k - z^{k+1}) = \partial g(z^{k+1}) - \mu^{k+1},$$

where  $\mu^{k+1}$  is defined as in (53). From the formula for  $\partial L_2$  given in (38), we then ascertain that

$$(0, y_2^{k+1} + c(z^k - z^{k+1}), z^{k+1}) \in \partial L_2(x^{k+1}, z^{k+1}, \mu^{k+1}). \quad (70)$$

Next, consider (54), which is a consequence of the hypothesis (46). Since  $L_1$  is independent of its second ( $z$ ) argument, we may replace  $z^k$  by  $z^{k+1}$  in (54) and obtain

$$(y_1^{k+1}, 0, -Mx^{k+1}) \in \partial L_1(x^{k+1}, z^{k+1}, \mu^{k+1}) \quad (71)$$

Adding (71) and (70), and using that  $\partial L_1 + \partial L_2 = \partial L$ , we obtain

$$(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), z^{k+1} - Mx^{k+1}) \in \partial L(x^{k+1}, z^{k+1}, \mu^{k+1}).$$

Since  $\partial L$  is a partial inverse of  $\partial F$  as defined in (33), we know that

$$(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), \mu^{k+1}) \in \partial F(x^{k+1}, z^{k+1}, z^{k+1} - Mx^{k+1}). \quad (72)$$

Recalling that  $Q$  is the negative of the convex conjugate of  $F$ , combining Fenchel's equality [30, Theorem 23.5] with (72) produces

$$\begin{aligned} F(x^{k+1}, z^{k+1}, z^{k+1} - Mx^{k+1}) - Q(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), \mu^{k+1}) \\ = \langle x^{k+1}, y_1^{k+1} \rangle + \langle z^{k+1}, y_2^{k+1} \rangle + c \langle z^{k+1}, z^k - z^{k+1} \rangle + \langle z^{k+1} - Mx^{k+1}, \mu^{k+1} \rangle. \end{aligned}$$

Using the definition (33) of  $F$ , we have  $F(x^{k+1}, z^{k+1}, z^{k+1} - Mx^{k+1}) = f(x^{k+1}) + g(z^{k+1})$ . Substituting this identity into the above equation and rearranging, we obtain

$$f(x^{k+1}) + g(z^{k+1}) = Q(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), \mu^{k+1}) \quad (73)$$

$$+ \langle x^{k+1}, y_1^{k+1} \rangle + \langle z^{k+1}, y_2^{k+1} \rangle \quad (74)$$

$$+ c \langle z^{k+1}, z^k - z^{k+1} \rangle + \langle z^{k+1} - Mx^{k+1}, \mu^{k+1} \rangle. \quad (75)$$

The hypotheses (63) directly ensure that the terms on line (74) converge to zero. The first term in (75) also converges to zero, because we have assumed that  $\{z^k\}$  is bounded and (63) contains the assumption that  $z^{k+1} - z^k \rightarrow 0$ . Since we have assumed that  $\{p^k\}$  is bounded and  $z^{k+1} - z^k \rightarrow 0$ , it follows that  $\{\mu^k\}$  is bounded. Furthermore, (63) contains the assumption that  $Mx^{k+1} - z^k \rightarrow 0$ , so the second term in (75) also converges to zero. Let  $\mathcal{K}_1 \subseteq \mathbb{N}$  be any infinite sequence of indices for which  $\lim_{k \in \mathcal{K}_1} f(x^{k+1}) + g(z^{k+1}) = \limsup_{k \rightarrow \infty} f(x^k) + g(z^k)$ . Since we have assumed that  $\{p^k\}$  is bounded, which implies that  $\{\mu^k\}$  is bounded, there exists some infinite subsequence  $\mathcal{K}'_1 \subseteq \mathcal{K}_1$  over which  $\{\mu^{k+1}\}$  converges to some limit  $p^\infty \in \mathbb{R}^m$ . Taking the limit over  $\mathcal{K}'_1$  in (73)-(75), we obtain, since we have established that all the terms in (74)-(75) converge to zero, that

$$\limsup_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} = \lim_{k \in \mathcal{K}'_1} Q(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), \mu^{k+1}). \quad (76)$$

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  $\mu^k \rightarrow_{\mathcal{K}'_1} p^\infty$ , we must have

$$\lim_{k \in \mathcal{K}'_1} Q(y_1^{k+1}, y_2^{k+1} + c(z^k - z^{k+1}), \mu^{k+1}) \leq Q(0, 0, p^\infty). \quad (77)$$

By weak duality, we also have  $Q(0, 0, p^\infty) \leq \inf_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\}$ . Combining this observation with (76) and (77), we obtain (68).

Now assume Assumption 1 holds, and let  $(x^*, z^*, p^*)$  be any KKT point. Then (68) immediately becomes

$$\limsup_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} \leq \inf_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\} = f(x^*) + g(z^*). \quad (78)$$

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

$$f(x^*) + g(z^*) \leq f(x^k) + g(z^k) + \langle p^*, Mx^k - z^k \rangle,$$

which with a minor rearrangement is in turn equivalent to

$$f(x^*) + g(z^*) - \langle p^*, Mx^k - z^k \rangle \leq f(x^k) + g(z^k). \quad (79)$$

Let  $\mathcal{K}_2 \subseteq \mathbb{N}$  be any infinite sequence of indices for which

$$\lim_{k \in \mathcal{K}_2} \{f(x^k) + g(z^k)\} = \liminf_{k \rightarrow \infty} \{f(x^k) + g(z^k)\}.$$

Since the hypotheses (63) include  $\langle p^*, Mx^k - z^k \rangle \rightarrow 0$ , we obtain by taking the limit over  $\mathcal{K}_2$  in (79) that

$$f(x^*) + g(z^*) \leq \liminf_{k \rightarrow \infty} \{f(x^k) + g(z^k)\}. \quad (80)$$

Combining (78) and (80), we have

$$f(x^*) + g(z^*) \leq \liminf_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} \leq \limsup_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} \leq f(x^*) + g(z^*),$$

which, in view of  $Mx^* = z^*$ , is equivalent to (69).  $\square$

### 4.3 An approximate ADMM with absolutely summable error criteria

We are now in a position to develop several approximate ADMM algorithms whose convergence analysis is based on Lagrangian splitting. Our first algorithm 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 version of Algorithm 3.1.1 developed above. However, we are able to prove its convergence without any strong convexity or matrix rank assumptions. Unlike the method from [10] underlying Algorithm 3.1.1, 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 [14] with the techniques developed in [9] for approximate solution of subproblems in the standard (non-alternating-direction) augmented Lagrangian method. Just as in Section 3.1, let  $\{\epsilon_k\}_{k=1}^\infty, \{\tau_k\}_{k=1}^\infty \subset \mathbb{R}_{++}$  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 particular classes 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 4.3.1** Inexact ADMM with absolutely summable error criteria

---

**initialization:** Pick  $c, \beta_1, \beta_2 > 0$ , summable sequences  $\{\epsilon_k\}_{k=1}^\infty, \{\tau_k\}_{k=1}^\infty \subset \mathbb{R}_{++}$  and initial points  $p^0, z^0 \in \mathbb{R}^m$

**repeat** {**for**  $k = 0, 1, 2, \dots$ }

**repeat** {**for**  $l = 1, 2, \dots$ }

        Improve the solution to  $x^{k+1} \approx \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\}$  by taking  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$

**until**  $\|y_1^{k,l}\| \leq \frac{\epsilon_{k+1}}{\max\{\beta_1, \|x^{k,l}\|\}}$

$x^{k+1} = x^{k,l}$

$y_1^{k+1} = y_1^{k,l}$

**repeat** {**for**  $l = 1, 2, \dots$ }

        Improve the solution to  $z^{k+1} \approx \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\}$  by taking  $(z^{k,l}, y_2^{k,l}) = \mathcal{G}(p^k, x^{k+1}, c, z^k, l)$

**until**  $\|y_2^{k,l}\| \leq \frac{\tau_{k+1}}{\max\{\beta_2, \|z^{k,l}\|\}}$

$z^{k+1} = z^{k,l}$

$y_2^{k+1} = y_2^{k,l}$

$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$

**until** Overall convergence

---

**Remark:** When the sequence  $\{y_1^k\}$  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  $\{y_2^k\}$ , just as in Algorithm 3.1.1. However, the sequences  $\{y_1^k\}$  and  $\{y_2^k\}$  figure prominently in the convergence analysis, 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,  $\|x^{k,l}\| \leq \beta_1$  and  $\|z^{k,l}\| \leq \beta_2$  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 3.1.1, with  $\epsilon_k$  replaced by  $\epsilon_k/\beta_1$  and  $\tau_k$  replaced by  $\tau_k/\beta_2$ . The sequences  $\{\epsilon_k/\beta_1\}$  and  $\{\tau_k/\beta_2\}$  remain summable and thus meet the assumptions of Algorithm 3.1.1, so in this case Algorithm 4.3.1 essentially coincides with Algorithm 3.1.1.

We begin by proving that it is not possible for Algorithm 4.3.1 to become “trapped” in one of its inner loops, subject to a mild condition of the approximation procedure modeled by  $\mathcal{F}$ :

**Lemma 20.** *Suppose that for any  $(p, z, c, \bar{x})$ , the sequence  $\{(x^l, y_1^l)\} = \{\mathcal{F}(p, z, c, \bar{x}, l)\}$  must be bounded. Then the inner loops (over  $l$ ) of Algorithm 4.3.1 always terminate in a finite number of iterations.*

*Proof.* Fix any  $k$ . We start 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  $\{x^{k,l}\}_{l=1}^\infty$  is bounded. Therefore, the right-hand side of the inner-loop termination condition  $\|y_1^{k,l}\| \leq \epsilon_{k+1}/\max\{\beta_1, \|x^{k,l}\|\}$  is bounded below by some positive quantity as  $l \rightarrow \infty$ . 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,  $\{z^{k,l}\}_{l=1}^\infty$  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.  $\square$

**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.

We now prove the convergence of the algorithm. Many of the techniques are adapted from [9, 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 [9].

**Lemma 21.** [28, Section 2.2] *Suppose  $\{\alpha_k\}_{k=0}^\infty, \{\gamma_k\}_{k=0}^\infty \subset \mathbb{R}$  are sequences such that  $\{\alpha_k\}$  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  $\{\alpha_k\}$  converges to a finite limit.*

With an analysis similar to [9, Lemma 5], we prove the following result:

**Lemma 22.** *In Algorithm 4.3.1, we have*

$$\sum_{k=0}^\infty \|y_1^k\| < \infty \quad \sum_{k=0}^\infty \|y_2^k\| < \infty \quad \sum_{k=0}^\infty \langle x^k, y_1^k \rangle < \infty \quad \sum_{k=0}^\infty \langle z^k, y_2^k \rangle < \infty, \quad (81)$$

*all these limits being guaranteed to exist.*

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

$$\|y_1^{k+1}\| \leq \frac{\epsilon_{k+1}}{\max\{\beta_1, \|x^{k+1}\|\}} \leq \frac{\epsilon_{k+1}}{\beta_1}.$$

Recalling that  $\sum_{k=1}^\infty \epsilon_k < \infty$ , we conclude that  $\sum_{k=0}^\infty \|y_1^k\| < \infty$ , establishing the first claim. Furthermore, whenever  $x^{k+1} \neq 0$  we also have

$$\|y_1^{k+1}\| \leq \frac{\epsilon_{k+1}}{\max\{\beta_1, \|x^{k+1}\|\}} \leq \frac{\epsilon_{k+1}}{\|x^{k+1}\|}.$$

Therefore,

$$|\langle x^{k+1}, y_1^{k+1} \rangle| \leq \|x^{k+1}\| \|y_1^{k+1}\| \leq \|x^{k+1}\| \frac{\epsilon_{k+1}}{\|x^{k+1}\|} = \epsilon_{k+1},$$

which may also easily be seen to hold when  $x^{k+1} = 0$ . From the summability of  $\{\epsilon_k\}$ , we immediately deduce that  $\sum_{k=0}^{\infty} |\langle x^k, y_1^k \rangle| < \infty$  and therefore we have that  $\sum_{k=0}^{\infty} \langle x^k, y_1^k \rangle$  exists and is finite.

The proofs of the boundedness of  $\sum_{k=0}^{\infty} \|y_2^k\|$  and  $\sum_{k=0}^{\infty} \langle z^k, y_2^k \rangle$  are nearly identical to those just presented.  $\square$

**Lemma 23.** *If  $\{z^k\}_{k=1}^{\infty}$  is bounded in Algorithm 4.3.1, then*

$$\sum_{k=0}^{\infty} \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle < \infty.$$

*Proof.* By the Cauchy-Schwartz inequality,

$$|\langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle| \leq \|y_2^{k+1} - y_2^k\| \|z^{k+1} - z^k\|.$$

Since  $\{y_2^k\}$  is summable by Lemma 22,  $\{y_2^{k+1} - y_2^k\}$  is also summable. Since  $\{z^k\}$  is assumed to be bounded, it follows that the sequence  $\{\|z^{k+1} - z^k\|\}$  is also bounded. We may then deduce from the above inequality that  $\{|\langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle|\}$  is summable, and therefore that  $\{\langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle\}$  is summable.  $\square$

**Proposition 24.** *If Assumption 1 holds, then the sequences  $\{z^k\}$ ,  $\{p^k\}$  and  $\{Mx^k\}$  generated by Algorithm 4.3.1, are all bounded and  $\sum_{k=0}^{\infty} \|Mx^{k+1} - z^k\|^2 < \infty$ . Furthermore, for any KKT point  $(x^*, z^*, p^*)$  of (P), the sequence  $\{\|(cz^k, p^k) - (cz^*, p^*)\|\}$  converges to a finite limit.*

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

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

and from  $L = L_1 + L_2$  and (37)-(38) we have

$$\begin{aligned} (y_1^{k+1}, 0, -Mx^{k+1}) &\in \partial L_1(x^{k+1}, z^k, p^k + cMx^{k+1} - cz^k) \\ (0, 0, -Mx^*) &\in \partial L_1(x^*, z^*, p^* + cMx^* - cz^*) \\ (0, y_2^{k+1}, z^{k+1}) &\in \partial L_2(x^{k+1}, z^{k+1}, p^k + cMx^{k+1} - cz^{k+1}) \\ (0, 0, z^*) &\in \partial L_2(x^*, z^*, p^* + cMx^* - cz^*) \end{aligned}$$

Since  $\partial L_1$  is a monotone operator, we have

$$\langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle p^k - p^* + cM(x^{k+1} - x^*) - c(z^k - z^*), -M(x^{k+1} - x^*) \rangle \geq 0$$

Rearranging this inequality, we obtain

$$\|M(x^{k+1} - x^*)\|^2 \leq \frac{1}{c} \langle x^{k+1} - x^*, y_1^{k+1} \rangle - \langle \frac{1}{c}(p^k - p^*) - (z^k - z^*), M(x^{k+1} - x^*) \rangle. \quad (82)$$

Using the monotonicity of  $\partial L_2$ , we have

$$\begin{aligned} & \langle z^{k+1} - z^*, y_2^{k+1} \rangle + \langle p^k - p^* + cM(x^{k+1} - x^*) - c(z^{k+1} - z^*), z^{k+1} - z^* \rangle \geq 0 \\ \Leftrightarrow & \langle z^{k+1} - z^*, y_2^{k+1} \rangle + \langle p^k - p^* + cM(x^{k+1} - x^*), z^{k+1} - z^* \rangle - c \|z^{k+1} - z^*\|^2 \geq 0, \end{aligned}$$

which we may rearrange into

$$\|z^{k+1} - z^*\|^2 \leq \frac{1}{c} \langle z^{k+1} - z^*, y_2^{k+1} \rangle + \langle \frac{1}{c}(p^k - p^*) + M(x^{k+1} - x^*), z^{k+1} - z^* \rangle \quad (83)$$

Adding (82) and (83), we obtain

$$\begin{aligned} & \|M(x^{k+1} - x^*)\|^2 + \|z^{k+1} - z^*\|^2 \\ & \leq \frac{1}{c} (\langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle) \\ & \quad + \langle M(x^{k+1} - x^*), z^k - z^* \rangle + \langle M(x^{k+1} - x^*), z^{k+1} - z^* \rangle \\ & \quad + \frac{1}{c} \langle p^k - p^*, z^{k+1} - Mx^{k+1} \rangle \end{aligned} \quad (84)$$

$$\begin{aligned} & = \frac{1}{c} (\langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle) \\ & \quad + \langle M(x^{k+1} - x^*), z^k - z^* \rangle + \langle M(x^{k+1} - x^*), z^{k+1} - z^* \rangle \\ & \quad + \frac{1}{c^2} \langle p^k - p^*, p^k - p^{k+1} \rangle \end{aligned} \quad (85)$$

$$\begin{aligned} & = \frac{1}{c} (\langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle) \\ & \quad + \frac{1}{2} \left( \|M(x^{k+1} - x^*)\|^2 + \|z^k - z^*\|^2 - \|Mx^{k+1} - z^k\|^2 \right) \end{aligned} \quad (86)$$

$$+ \frac{1}{2} \left( \|M(x^{k+1} - x^*)\|^2 + \|z^{k+1} - z^*\|^2 - \frac{1}{c^2} \|p^{k+1} - p^k\|^2 \right) \quad (87)$$

$$+ \frac{1}{2c^2} \left( \|p^k - p^*\|^2 + \|p^k - p^{k+1}\|^2 - \|p^{k+1} - p^*\|^2 \right), \quad (88)$$

where we use the multiplier update  $p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$  to obtain the term (85), we apply the identity  $\langle a, b \rangle = \frac{1}{2} (\|a\|^2 + \|b\|^2 - \|a - b\|^2)$  to obtain (86)-(88), and both (84) and the last term in (86) use that  $Mx^* = z^*$ . Multiplying the resulting inequality by  $2c^2$  and rearranging, we obtain

$$\begin{aligned} & c^2 \|z^{k+1} - z^*\|^2 + \|p^{k+1} - p^*\|^2 + c^2 \|Mx^{k+1} - z^k\|^2 \\ & \leq c^2 \|z^k - z^*\|^2 + \|p^k - p^*\|^2 + 2c (\langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle). \end{aligned} \quad (89)$$

We now bound the last term in (89), using the termination conditions for the algorithm's inner loops:

$$\begin{aligned} & \langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle \leq \|x^{k+1} - x^*\| \|y_1^{k+1}\| + \|z^{k+1} - z^*\| \|y_2^{k+1}\| \\ & \leq \|x^{k+1}\| \|y_1^{k+1}\| + \|x^*\| \|y_1^{k+1}\| + \|z^{k+1}\| \|y_2^{k+1}\| + \|z^*\| \|y_2^{k+1}\|. \end{aligned}$$

Whenever  $x^{k+1} \neq 0$  and  $z^{k+1} \neq 0$ , we then deduce that

$$\begin{aligned} & \langle x^{k+1} - x^*, y_1^{k+1} \rangle + \langle z^{k+1} - z^*, y_2^{k+1} \rangle \\ & \leq \|x^{k+1}\| \frac{\epsilon_{k+1}}{\|x^{k+1}\|} + \|x^*\| \frac{\epsilon_{k+1}}{\beta_1} + \|z^{k+1}\| \frac{\tau_{k+1}}{\|z^{k+1}\|} + \|z^*\| \frac{\tau_{k+1}}{\beta_2} \\ & \leq \left( 1 + \frac{\|x^*\|}{\beta_1} \right) \epsilon_{k+1} + \left( 1 + \frac{\|z^*\|}{\beta_2} \right) \tau_{k+1}, \end{aligned}$$

which may also easily be deduced to hold if  $x^{k+1} = 0$  or  $z^{k+1} = 0$ . Substituting this inequality into (89), we obtain

$$\begin{aligned} & c^2 \|z^{k+1} - z^*\|^2 + \|p^{k+1} - p^*\|^2 + c^2 \|Mx^{k+1} - z^k\|^2 \\ & \leq c^2 \|z^k - z^*\|^2 + \|p^k - p^*\|^2 + 2c \left[ \left(1 + \frac{\|x^*\|}{\beta_1}\right) \epsilon_{k+1} + \left(1 + \frac{\|z^*\|}{\beta_2}\right) \tau_{k+1} \right]. \end{aligned} \quad (90)$$

Adding this inequality for  $k = 0, \dots, K$  and canceling “telescoping” terms yields that for any  $K \geq 0$ ,

$$\begin{aligned} & c^2 \|z^{K+1} - z^*\|^2 + \|p^{K+1} - p^*\|^2 + c^2 \sum_{k=0}^K \|Mx^{k+1} - z^k\|^2 \\ & \leq c^2 \|z^0 - z^*\|^2 + \|p^0 - p^*\|^2 + 2c \left(1 + \frac{\|x^*\|}{\beta_1}\right) \sum_{k=0}^K \epsilon_{k+1} + 2c \left(1 + \frac{\|z^*\|}{\beta_2}\right) \sum_{k=0}^K \tau_{k+1}. \end{aligned} \quad (91)$$

By assumption, both  $\{\epsilon_k\}$  and  $\{\tau_k\}$  are summable, so the right-hand side of (91) converges to a finite limit as  $K \rightarrow \infty$ . From this boundedness, (91) allows us to draw the following conclusions:

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

From (90), since  $\|Mx^{k+1} - z^k\|^2$  is always nonnegative, we have for all  $k$  that

$$\begin{aligned} & c^2 \|z^{k+1} - z^*\|^2 + \|p^{k+1} - p^*\|^2 \\ & \leq c^2 \|z^k - z^*\|^2 + \|p^k - p^*\|^2 + 2c \left[ \left(1 + \frac{\|x^*\|}{\beta_1}\right) \epsilon_{k+1} + \left(1 + \frac{\|z^*\|}{\beta_2}\right) \tau_{k+1} \right]. \end{aligned} \quad (92)$$

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

$$\gamma_k = 2c \left[ \left(1 + \frac{\|x^*\|}{\beta_1}\right) \epsilon_{k+1} + \left(1 + \frac{\|z^*\|}{\beta_2}\right) \tau_{k+1} \right],$$

from which we can rewrite (92) as  $\alpha_{k+1} \leq \alpha_k + \gamma_k$ . Since  $\{\epsilon_k\}$  and  $\{\tau_k\}$  are summable, we know that  $\sum_{k=0}^{\infty} \gamma_k < \infty$ , so by Lemma 21, we conclude that

$$\{\alpha_k\} = \{c^2 \|z^k - z^*\|^2 + \|p^k - p^*\|^2\} = \{\|(cz^k, p^k) - (cz^*, p^*)\|^2\}$$

converges to a finite limit, and consequently so does  $\{\|(cz^k, p^k) - (cz^*, p^*)\|\}$ .  $\square$



**Proposition 25.** *If Assumption 1 holds, then the sequences generated by Algorithm 4.3.1 have the properties*

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|^2 < \infty \quad \sum_{k=0}^{\infty} \|Mx^{k+1} - z^{k+1}\|^2 < \infty \quad \sum_{k=0}^{\infty} \|p^{k+1} - p^k\|^2 < \infty.$$

*Proof.* By Lemma 17, we know that

$$\|Mx^{k+1} - z^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \leq \|Mx^{k+1} - z^k\|^2 + \frac{2}{c} \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle$$

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

$$\sum_{k=0}^K \|Mx^{k+1} - z^{k+1}\|^2 + \sum_{k=0}^K \|z^k - z^{k+1}\|^2 \leq \sum_{k=0}^K \|Mx^{k+1} - z^k\|^2 + \frac{2}{c} \sum_{k=0}^K \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle.$$

Now consider the limit as  $K \rightarrow \infty$ . Proposition 24 asserts that the first term on the right-hand side of the above inequality converges to a finite limit, while Lemma 23 guarantees that the same holds for the last term. Therefore we conclude that

$$\sum_{k=0}^{\infty} \|Mx^{k+1} - z^{k+1}\|^2 < \infty \quad \sum_{k=0}^{\infty} \|z^{k+1} - z^k\|^2 < \infty.$$

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

The following two propositions respectively summarize the dual and primal behavior of Algorithm 4.3.1.

**Proposition 26.** *Under Assumption 1, the sequences  $\{p^k\}$  and  $\{z^k\}$  generated by Algorithm 4.3.1 are bounded and all limit points of  $\{p^k\}$  are solutions to the dual problem (D). Furthermore, if  $\{x^k\}$  has at least one limit point, then  $\{p^k\}$  converges to a particular dual solution.*

*Proof.* We will use Proposition 18, whose assumptions are that  $\{Mx^k\}$  and  $\{p^k\}$  are bounded, along with

$$z^{k+1} - z^k \rightarrow 0 \quad Mx^k - z^k \rightarrow 0 \quad y_1^k \rightarrow 0 \quad y_2^k \rightarrow 0 \quad \langle y_1^k, x^k \rangle \rightarrow 0 \quad \langle y_2^k, z^k \rangle \rightarrow 0.$$

The boundedness of  $\{Mx^k\}$  and  $\{p^k\}$  was established in Proposition 24. The first two limit conditions above follow immediately from stronger results established in Proposition 25, while the remaining limit conditions similarly follow from stronger results already established in Lemma 22. Therefore, the hypotheses of Proposition 18 are satisfied, and we conclude that all limit points of  $\{p^k\}$  must be dual solutions.

Now suppose that  $\{x^k\}$  has at least one limit point  $x^\infty$ . Since Proposition 24 asserts that  $\{p^k\}$  and  $\{z^k\}$  are bounded, there exist  $z^\infty, p^\infty \in \mathbb{R}^m$  and an infinite set of indices  $\mathcal{K}$  for which  $(x^k, z^k, p^k) \rightarrow_{\mathcal{K}} (x^\infty, z^\infty, p^\infty)$ . We will use Lemma 14 to assert that  $(x^\infty, z^\infty, p^\infty)$  is KKT point. The hypotheses of Lemma 14 are

$$Mx^k - z^k \rightarrow 0 \quad Mx^{k+1} - z^k \rightarrow 0 \quad y_1^k \rightarrow 0 \quad y_2^k \rightarrow 0.$$

The condition  $Mx^k - z^k \rightarrow 0$  is a consequence of Proposition 25, while  $Mx^{k+1} - z^k \rightarrow 0$  follows from Proposition 24. The last two limit conditions above were established in Lemma 22. Therefore, Lemma 14 applies and  $(x^\infty, z^\infty, p^\infty)$  must be KKT point. Proposition 24 then asserts that  $\{\|(cz^k, p^k) - (cz^\infty, p^\infty)\|\}$  must converge to a finite limit, but since one of its limit points must be zero, the entire sequence converges and we deduce that  $p^k \rightarrow p^\infty$ .  $\square$

Note that since  $\{Mx^k\}$  must be bounded by Proposition 24, one possible sufficient condition for  $\{x^k\}$  to have a limit point is that  $M$  has full column rank.

**Proposition 27.** *For the sequences  $\{x^k\}$  and  $\{z^k\}$  generated by Algorithm 4.3.1 under Assumption 1, we have*

$$Mx^k - z^k \rightarrow 0 \quad \lim_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} = f(x^*) + g(z^*).$$

*Proof.* The first (feasibility) claim is, as in the proof of the previous proposition, an immediate consequence of Proposition 25. In view of Lemma 22, Proposition 24, and Proposition 25, this second claim is a direct application of Proposition 19.  $\square$

#### 4.4 An approximate ADMM with relative error criteria

One drawback to Algorithm 4.3.1 above is that it is not inherently clear how to select the parameter sequences  $\{\epsilon_k\}$  and  $\{\tau_k\}$ . In the general development of proximal algorithms, it has become common starting with the work of Solodov and Svaiter in [36] to replace the use of 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 [14] with the relative-error non-alternating augmented Lagrangian techniques in [11], as opposed to the absolute-error techniques in [9]. As in [11], we introduce an auxiliary sequence into the iterative process, in this case  $\{(w_1^k, w_2^k)\}_{k=0}^\infty \subset \mathbb{R}^n \times \mathbb{R}^m$ . The other sequences generated by the algorithm,  $\{x^k\}, \{y_1^k\} \subset \mathbb{R}^n$  and  $\{z^k\}, \{y_2^k\}, \{p^k\} \subset \mathbb{R}^m$ , play similar roles similar to the corresponding sequences in Algorithm 4.3.1. To simplify the statement and analysis of the algorithm, we define the following notation for all  $k \geq 0$ :

$$\mathbf{w}^k = \begin{bmatrix} w_1^k \\ w_2^k \end{bmatrix} \quad \mathbf{y}^k = \begin{bmatrix} y_1^k \\ y_2^k \end{bmatrix} \quad \mathbf{x}^k = \begin{bmatrix} x^k \\ z^k \end{bmatrix}.$$

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  $\mathbf{w}^0 \in \mathbb{R}^n \times \mathbb{R}^m$  and develops sequences conforming to the following

recursive conditions for all  $k \geq 0$ :

$$y_1^{k+1} \in \partial_x \left[ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right]_{x=x^{k+1}} \quad (93)$$

$$y_2^{k+1} \in \partial_z \left[ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|z - Mx^{k+1}\|^2 \right]_{z=z^{k+1}} \quad (94)$$

$$\frac{2}{c} |\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle| + \|\mathbf{y}^{k+1}\|^2 \leq \sigma \|Mx^{k+1} - z^k\|^2 \quad (95)$$

$$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1}) \quad (96)$$

$$\mathbf{w}^{k+1} = \mathbf{w}^k - c\mathbf{y}^k. \quad (97)$$

Here, (93), (94), and (96) are essentially the same as in Algorithm 4.3.1, while (95) and (97) are essentially the same approximation criterion and update proposed in [11]. In fact, applying the algorithm of [11] to (P) with a constant penalty parameter would result in the same algorithm, except that the sequential conditions (93)-(94) would be replaced by the single coupled condition

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

We will show that the recursions (93)-(97) converge if (P) has a KKT point, essentially establishing that if one keeps the penalty parameter constant in the algorithm of [11] applied to (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.

We begin by establishing a Fejér monotonicity result:

**Proposition 28.** *Suppose that  $\{\mathbf{x}^k\}$ ,  $\{\mathbf{y}^k\}$ ,  $\{p^k\}$  and  $\{\mathbf{w}^k\}$  obey the recursion (93)-(97) for all  $k \geq 1$  and some fixed  $\sigma \in [0, 1)$ . Then if  $(x^*, z^*, p^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  is any KKT point of (P), The sequence  $\{\|(\mathbf{w}^k, cz^k, p^k) - (\mathbf{x}^*, cz^*, p^*)\|\}_{k=0}^\infty$  is nonincreasing and hence convergent. If the set of KKT points  $L^{-1}(0, 0, 0)$  is nonempty, then the following also hold:*

- The sequences  $\{p^k\}$ ,  $\{z^k\}$  and  $\{\mathbf{w}^k\}$  are bounded.
- $Mx^{k+1} - z^k \rightarrow 0$ ,  $\mathbf{y}^k \rightarrow 0$ , and  $\langle \mathbf{x}^k, \mathbf{y}^k \rangle \rightarrow 0$ .

*Proof.* As in the previous analysis, we let  $\mu^{k+1} = p^k + cMx^{k+1} - cz^k$  for all  $k$ . Let  $(x^*, z^*, p^*)$  be any KKT point of (P), which is necessarily a saddle point of  $L$  by Lemma 13. For any  $k \geq 0$ ,

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

where the last equality uses  $Mx^* = z^*$ . Thus,

$$\begin{aligned}
& \|p^{k+1} - p^*\|^2 + c^2 \|z^{k+1} - z^*\|^2 \\
&= \|p^k - p^*\|^2 + c^2 \|z^k - z^*\|^2 - \|p^k - p^{k+1}\|^2 - c^2 \|z^k - z^{k+1}\|^2 \\
&\quad - 2c \langle z^{k+1} - z^*, p^{k+1} - p^* \rangle - 2c \langle p^{k+1} - \mu^{k+1}, z^{k+1} - z^* \rangle \\
&\quad - 2c \langle -Mx^{k+1} + Mx^*, p^{k+1} - p^* \rangle. \tag{98}
\end{aligned}$$

Rewriting the last term in (98) as

$$\langle -Mx^{k+1} + Mx^*, p^{k+1} - p^* \rangle = \langle -Mx^{k+1} + Mx^*, p^{k+1} - \mu^{k+1} + \mu^{k+1} - p^* \rangle,$$

we next obtain, using  $Mx^* = z^*$  that

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

The quantity in brackets is simply  $\|Mx^{k+1} - z^k\|^2$ , so it follows that

$$\begin{aligned}
& \|p^{k+1} - p^*\|^2 + c^2 \|z^{k+1} - z^*\|^2 \\
&= \|p^k - p^*\|^2 + c^2 \|z^k - z^*\|^2 - c^2 \|Mx^{k+1} - z^k\|^2 \\
&\quad - 2c \langle z^{k+1} - z^*, p^{k+1} - p^* \rangle - 2c \langle -Mx^{k+1} - (-Mx^*), \mu^{k+1} - p^* \rangle. \tag{99}
\end{aligned}$$

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

$$\begin{aligned}
& \|\mathbf{w}^{k+1} - \mathbf{x}^*\|^2 = \|\mathbf{w}^k - c\mathbf{y}^{k+1} - \mathbf{x}^*\|^2 \\
&= \|\mathbf{w}^k - \mathbf{x}^*\|^2 - 2c \langle \mathbf{w}^k - \mathbf{x}^*, \mathbf{y}^{k+1} \rangle + c^2 \|\mathbf{y}^{k+1}\|^2 \\
&= \|\mathbf{w}^k - \mathbf{x}^*\|^2 - 2c \langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle - 2c \langle \mathbf{x}^{k+1} - \mathbf{x}^*, \mathbf{y}^{k+1} \rangle + c^2 \|\mathbf{y}^{k+1}\|^2. \tag{100}
\end{aligned}$$

Adding (99) to (100) and grouping terms yields

$$\begin{aligned}
& \|p^{k+1} - p^*\|^2 + c^2 \|z^{k+1} - z^*\|^2 + \|\mathbf{w}^{k+1} - \mathbf{x}^*\|^2 \\
&= \|p^k - p^*\|^2 + c^2 \|z^k - z^*\|^2 + \|\mathbf{w}^k - \mathbf{x}^*\|^2 \\
&\quad - 2c [\langle z^{k+1} - z^*, p^{k+1} - p^* \rangle + \langle z^{k+1} - z^*, y_2^{k+1} - 0 \rangle + \langle 0, x^{k+1} - x^* \rangle] \quad (A) \\
&\quad - 2c [\langle -Mx^{k+1} - (-Mx^*), \mu^{k+1} - p^* \rangle + \langle x^{k+1} - x^*, y_1^{k+1} - 0 \rangle + \langle 0, z^k - z^* \rangle] \quad (B) \\
&\quad - 2c \langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle \quad (C) \\
&\quad + c^2 \|\mathbf{y}^{k+1}\|^2 - c^2 \|Mx^{k+1} - z^k\|^2. \quad (101)
\end{aligned}$$

Since

$$(y_1^{k+1}, 0, -Mx^{k+1}) \in \partial L_1(x^{k+1}, z^k, \mu^{k+1}) \quad (0, 0, -Mx^*) \in \partial L_1(x^*, z^*, p^*),$$

it follows from the monotonicity of  $\partial L_1$  that

$$\langle -Mx^{k+1} - (-Mx^*), \mu^{k+1} - p^* \rangle + \langle x^{k+1} - x^*, y_1^{k+1} - 0 \rangle + \langle 0, z^k - z^* \rangle \geq 0. \quad (B')$$

Similarly, from

$$(0, y_2^{k+1}, z^{k+1}) \in \partial L_2(x^{k+1}, z^{k+1}, p^{k+1}) \quad (0, 0, z^*) \in \partial L_2(x^*, z^*, p^*),$$

the monotonicity of  $\partial L_2$  yields

$$\langle z^{k+1} - z^*, p^{k+1} - p^* \rangle + \langle z^{k+1} - z^*, y_2^{k+1} - 0 \rangle + \langle 0, x^{k+1} - x^* \rangle \geq 0. \quad (A')$$

For (C), it is always true that

$$-2c \langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle \leq 2c |\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle|. \quad (C')$$

Applying (B') (A') and (C') to (101), we obtain

$$\begin{aligned}
& \|p^{k+1} - p^*\|^2 + c^2 \|z^{k+1} - z^*\|^2 + \|\mathbf{w}^{k+1} - \mathbf{x}^*\|^2 \\
& \leq \|p^k - p^*\|^2 + c^2 \|z^k - z^*\|^2 + \|\mathbf{w}^k - \mathbf{x}^*\|^2 \\
& \quad + 2c |\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle| + c^2 \|\mathbf{y}^{k+1}\|^2 - c^2 \|Mx^{k+1} - z^k\|^2 \quad (102)
\end{aligned}$$

Multiplying the relative error condition (95) by  $c^2$  yields

$$2c |\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle| + c^2 \|\mathbf{y}^{k+1}\|^2 \leq c^2 \sigma \|Mx^{k+1} - z^k\|^2. \quad (103)$$

Combining (103) and (101), we obtain the inequality that is the key to the convergence analysis:

$$\begin{aligned}
& \|p^{k+1} - p^*\|^2 + c^2 \|z^{k+1} - z^*\|^2 + \|\mathbf{w}^{k+1} - \mathbf{x}^*\|^2 \\
& \leq \|p^k - p^*\|^2 + c^2 \|z^k - z^*\|^2 + \|\mathbf{w}^k - \mathbf{x}^*\|^2 - c^2 (1 - \sigma) \|Mx^{k+1} - z^k\|^2. \quad (104)
\end{aligned}$$

Since  $(1 - \sigma) > 0$ , we may immediately conclude that

$$\|(\mathbf{w}^{k+1}, cz^{k+1}, p^{k+1}) - (\mathbf{x}, cz^*, p^*)\| \leq \|(\mathbf{w}^k, cz^k, p^k) - (\mathbf{x}, cz^*, p^*)\|,$$

and so  $\{\|(\mathbf{w}^{k+1}, cz^{k+1}, p^{k+1}) - (\mathbf{x}, cz^*, p^*)\|\}$  is nonincreasing and convergent. It also follows immediately that the sequences  $\{p^k\}$ ,  $\{z^k\}$  and  $\{\mathbf{w}^k\}$  are bounded.

By summing (104) over  $k$ , we deduce that  $\{\|Mx^{k+1} - z^k\|^2\}$  is a summable sequence and hence that  $Mx^{k+1} - z^k \rightarrow 0$ . From the relative error condition (95), we then deduce that  $\{\|\mathbf{y}^k\|^2\}$  and  $\{|\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle|\}$  are both summable, thus that  $\{\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle\}$  is also summable and consequently that  $\mathbf{y}^k \rightarrow 0$  and  $\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle \rightarrow 0$ .

Finally, writing

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

we may reason as follows: since  $\{\mathbf{w}^k\}$  is bounded and  $\mathbf{y}^k \rightarrow 0$ , we have  $\langle \mathbf{w}^k, \mathbf{y}^{k+1} \rangle \rightarrow 0$ . Since we just established that  $\langle \mathbf{w}^k - \mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle \rightarrow 0$ , it thus follows that  $\langle \mathbf{x}^k, \mathbf{y}^k \rangle \rightarrow 0$ , and all the claims of the proposition are established.  $\square$

The role of the sequence  $\{\mathbf{w}^k\}$  is similar to that of the sequence  $\{w^k\}$  in [11]: it may be considered as accumulating the total subgradient error “drift” over the course of the algorithm.

**Proposition 29.** *If the sequences  $\{x^k\}$ ,  $\{z^k\}$ ,  $\{p^k\}$ , and  $\{\mathbf{w}^k\}$  conform to the recursion (93)-(97), then*

$$Mx^{k+1} - z^{k+1} \rightarrow 0 \qquad z^{k+1} - z^k \rightarrow 0 \qquad p^{k+1} - p^k \rightarrow 0.$$

*Proof.* By Lemma 17, we have

$$\|Mx^{k+1} - z^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \leq \|Mx^{k+1} - z^k\|^2 + \frac{2}{c} \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle. \quad (105)$$

By Proposition 28,  $\{z^k\}$  is bounded,  $Mx^{k+1} - z^k \rightarrow 0$ , and  $\{y_2^k\} \rightarrow 0$ , so we may assert that

$$\|Mx^{k+1} - z^k\|^2 + \langle y_2^{k+1} - y_2^k, z^{k+1} - z^k \rangle \rightarrow 0.$$

It follows from (105) that  $\|Mx^{k+1} - z^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \rightarrow 0$ . The first two claims follow immediately, and the last claim is then a consequence of  $p^{k+1} - p^k = c(Mx^{k+1} - z^{k+1})$ .  $\square$

Following the pattern established in the previous subsection, next we show that all limit points of  $\{p^k\}$  are dual solutions. We also show that the primal sequences have desirable asymptotic behavior.

**Proposition 30.** *All accumulation points of the sequence  $\{p^k\}$  generated by the recursions (93)-(97) are solutions to the dual problem (D), all limit points of  $\{x^k\}$  are solutions to the primal problem (P), and  $\lim_{k \rightarrow \infty} \{Mx^k - z^k\} = 0$ . Furthermore, if  $(x^*, z^*)$  is an optimal solution to (P), we have*

$$\lim_{k \rightarrow \infty} \{f(x^k) + g(z^k)\} = f(x^*) + g(z^*).$$

*Proof.* By Propositions 28 and 29, we know that the sequences  $\{Mx^k\}$  and  $\{p^k\}$  generated by the recursions (93)-(97) are bounded and that

$$z^{k+1} - z^k \rightarrow 0 \quad Mx^{k+1} - z^{k+1} \rightarrow 0 \quad \langle y^{k+1}, x^{k+1} \rangle \rightarrow 0,$$

which establishes the claim regarding the sequence  $\{Mx^k - z^k\}$ . We may then apply Proposition 18 to obtain that every limit point of  $\{p^k\}$  is a dual solution.

We now consider the primal sequence  $\{x^k\}$ . From Lemma 14, we conclude that all its limit points are primal solutions. The final claim follows from Proposition 19.  $\square$

We close this section by showing that the sequence  $\{p^k\}$  defined by the recursions (93)-(97) converges to a dual solution. The key lemma we need here is by Alves and Svaiter [1].

**Lemma 31.** *[1, Lemma 2] Let  $W \times V$  be a nonempty set and let  $\{s_k = (\alpha_k, \beta_k)\}_{k \geq 0}$  be a sequence such that:*

1.  $\{\|s_k - s\|\}$  is nonincreasing for all  $s \in W \times V$ ;
2. Every limit point of  $\{\beta_k\}$  belongs to  $V$

*Then  $\{\beta_k\}$  converges to some element in  $V$ .*

**Proposition 32.** *If the set of KKT points for (P) is nonempty, then the sequence  $\{p^k\}$  defined by the recursions (93)-(97) converges to a solution of (D).*

*Proof.* Let  $(x, cz^*, p^*)$  be an arbitrary KKT point, and thus a saddle point of  $L$ . From Proposition 28, we know that  $\{\|(\mathbf{w}^k, cz^k, p^k) - (x, cz^*, p^*)\|\}$  is nonincreasing. For all  $k \geq 0$  let  $\alpha_k = (\mathbf{w}^{k+1}, cz^{k+1})$  and  $\beta_k = p^k$ , and define  $V$  to be the set all solutions of (D). By Proposition 30, we know that every limit point of  $\{p^k\}$  belongs to  $V$ . Therefore, Lemma 31 implies that  $\{p^k\}$  converges to a solution of (D).  $\square$

Propositions 32 and 30 together summarize the convergence properties of the recursions (93)-(97). We next consider to the implementation of concrete algorithms that conform to these recursions.

## 4.5 A partially inexact ADMM with a relative error criterion

In many applications of the ADMM, subproblem (3) may be solved exactly. For instance, in  $L_1$ -regularized data fitting problems, one typically has  $g(z) = \nu \|z\|_1$  for some  $\nu > 0$ . In this case, the exact solution to the  $g$  subproblem (3) can be easily computed exactly by the *soft thresholding operator* given by

$$z_i^{k+1} = \text{sgn} \left( M_i^\top x^{k+1} + \frac{1}{c} p_i^k \right) \max \left\{ 0, \left| M_i^\top x^{k+1} + \frac{1}{c} p_i^k \right| - \frac{\nu}{c} \right\} \quad i = 1, \dots, m, \quad (106)$$

where the vector  $M_i$  denotes the  $i^{\text{th}}$  row of the matrix  $M$ . We now give a concrete algorithm that applies (93)-(97) to such situations:

---

**Algorithm 4.5.1** Partially inexact ADMM with a relative error criterion

---

**initialization:** Pick scalar parameters  $c > 0$ , and  $\sigma \in [0, 1)$ , along with initial points  $x^0, w^0 \in \mathbb{R}^n$  and  $p^0, z^0 \in \mathbb{R}^m$

**repeat** {**for**  $k = 0, 1, 2, \dots$ }

**repeat** {**for**  $l = 1, 2, \dots$ }

        Improve the solution to  $x^{k+1} \approx \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\}$  by taking  $(x^{k,l}, y^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$

**until**  $\frac{2}{c} |\langle w^k - x^{k,l}, y^{k,l} \rangle| + \|y^{k,l}\|^2 \leq \sigma \|Mx^{k,l} - z^k\|^2$

$x^{k+1} = x^{k,l}, y^{k+1} = y^{k,l}$

$z^{k+1} = \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\}$

$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$

$w^{k+1} = w^k - cy^k$

**until** Overall convergence

---

**Proposition 33.** *Suppose Assumption 1 holds. If the inner loop of (over  $l$ ) of Algorithm 4.5.1 always terminates in a finite number of iterations, all limit points of the sequence  $\{(x^k, z^k, p^k)\}$  generated by the algorithm are the KKT points of (P),  $\{p^k\}$  converges to an optimal solution of the dual problem (D), and  $\lim_{k \rightarrow \infty} f(x^k) + g(z^k) = f(x^*) + g(z^*)$ , where  $(x^*, z^*)$  is any optimal solution of (P). If the inner loop cycles infinitely over  $l$  for some iteration  $k = \bar{k}$ , then for all limit points  $x^\infty$  of  $\{x^{\bar{k},l}\}_{l=1}^\infty$  one has that  $(x^\infty, z^{\bar{k}}, p^{\bar{k}})$  is a KKT point of (P).*

*Proof.* Consider the case that the inner loop always terminates finitely. In this case, it is easily verified that Algorithm 4.5.1 generates sequences conforming to the recursions (93)-(97) with  $y_2^k = 0$  for all  $k$ . The relevant conclusions then follow from Propositions 30 and 32.

The remaining case is that the first loop executes an infinite number of times at outer iteration  $k = \bar{k}$ , which cannot happen unless

$$\lim_{l \rightarrow \infty} \{Mx^{\bar{k},l} - z^{\bar{k}}\} = 0 \qquad \lim_{l \rightarrow \infty} y^{\bar{k},l} = 0.$$

Then by Corollary 16 all limit points of  $\{(x^{\bar{k},l}, z^{\bar{k}}, p^{\bar{k}})\}$  are the KKT points.  $\square$

Regarding the existence of limit points of  $\{x^{\bar{k},l}\}_{l=1}^\infty$  in the case of the inner loop running indefinitely, we may appeal to Lemma 5: specifically, if the solution set of the  $x$  subproblem is bounded, then  $\{x^{\bar{k},l}\}_{l=1}^\infty$  is bounded and hence must have limit points. If the solution of the  $x$  subproblem is unique, then  $\{x^{\bar{k},l}\}_{l=1}^\infty$  must converge to it, so the unique limit  $x^\infty$  of  $\{x^{\bar{k},l}\}_{l=1}^\infty$  is a solution of (P) by Proposition 33.

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

While the partially inexact Algorithm 4.5.1 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 4.6.1** 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  $Mx^1 \neq z^0$

**repeat** {**for**  $k = 0, 1, 2, \dots$ }

$l \leftarrow 0$

**repeat** {**for**  $t = 0, 1, \dots$ }

**repeat** {with increasing  $l$ }

$l \leftarrow l + 1$

$(z^{k,l}, y_2^{k,l}) = \mathcal{G}(p^k, x^{k+1}, c, z^k, l)$

**until**  $\frac{2}{c} \left| \langle w_2^k - z^{k,l}, y_2^{k,l} \rangle \right| + \|y_2^{k,l}\|^2 \leq \alpha^t (1 - \tau) \sigma \|Mx^{k+1} - z^k\|^2$

$p^{k,l} = p^k + c(Mx^{k+1} - z^{k,l})$

$w_2^{k,l} = w_2^k - cy_2^{k,l}$

$\bar{x}^{k,0} = x^{k+1}$

**repeat** {**for**  $l' = 0, 1, \dots$ }

$(x^{k,t,l'}, y_1^{k,t,l'}) = \mathcal{F}(p^{k,l}, z^{k,l}, c, \bar{x}^{k,t-1}, l')$

**accept**  $\leftarrow \left( \frac{2}{c} \left| \langle w_1^k - x^{k,t,l'}, y_1^{k,t,l'} \rangle \right| + \|y_1^{k,t,l'}\|^2 < \tau \sigma \|Mx^{k,t,l'} - z^{k,l}\|^2 \right)$

**until accept or**  $\left( \|y_1^{k,t,l'}\|^2 \leq \beta \|y_2^{k,l}\|^2 \text{ and } \|Mx^{k,t,l'} - z^{k,l}\|^2 \leq \gamma \|y_2^{k,l}\|^2 \right)$

$\bar{x}^{k,t} = x^{k,t,l'}; \bar{z}^{k,t} = z^{k,l}; \bar{p}^{k,t} = p^{k,l}; \bar{y}_1^{k,t} = y_1^{k,t,l'}; \bar{y}_2^{k,t} = y_2^{k,l}$

**until accept**

$x^{k+2} = x^{k,t,l'}; z^{k+1} = z^{k,l}; p^{k+1} = p^{k,l}; y_1^{k+1} = y_1^{k,t,l'}; y_2^{k+1} = y_2^{k,l}; w_2^{k+1} = w_2^{k,l}$

$w_1^{k+2} = w_1^{k+1} - cy_1^{k+1}$

**until** Overall convergence

---

Compared to Algorithm 4.3.1, this algorithm has a somewhat complicated structure. The reason is that if one simply modifies Algorithm 4.5.1 to include a second approximation loop in the manner of Algorithm 4.3.1, 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 4.5.1 to include an approximate  $z$  minimization, one might at first consider the following algorithm:

**initialization:** Pick  $c > 0$ ,  $\sigma \in [0, 1)$ ,  $x^0, w^0 \in \mathbb{R}^n$ , and  $p^0, z^0 \in \mathbb{R}^m$

**repeat** {**for**  $k = 0, 1, 2, \dots$ }

**repeat** {**for**  $l = 1, 2, \dots$ }

$(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$

**until**  $\frac{2}{c} \left| \langle w^k - x^{k,l}, y_1^{k,l} \rangle \right| + \|y_1^{k,l}\|^2 \leq \sigma \|Mx^{k,l} - z^k\|^2$

$x^{k+1} = x^{k,l}, y_1^{k+1} = y_1^{k,l}$

**repeat** {**for**  $l = 1, 2, \dots$ }

$(z^{k,l}, y_2^{k,l}) = \mathcal{G}(p^k, x^{k+1}, c, z^k, l)$

**until**  $\frac{2}{c} \left| \langle w_2^k - z^{k,l}, y_2^{k,l} \rangle \right| + \|y_2^{k,l}\|^2 \leq (1 - \tau) \sigma \|Mx^{k+1} - z^{k,l}\|^2$

$z^{k+1} = z^{k,l}, y_2^{k+1} = y_2^{k,l}$

$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$

$w^{k+1} = w^k - cy^{k+1}$

**until** Overall convergence

Unfortunately, it is conceivable for the second loop over  $l$  in this algorithm to become “trapped” in an infinite loop if both sides of the termination condition

$$\frac{2}{c} \left| \langle w_2^k - z^{k,l}, y_2^{k,l} \rangle \right| + \|y_2^{k,l}\|^2 \leq (1 - \tau)\sigma \|Mx^{k+1} - z^{k,l}\|^2$$

converge to zero. If the prior  $x$  minimization were exact, this infinite inner loop would converge to a solution, but unfortunately the prior minimization over  $x$  need not be exact. Essentially, 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 some desirable form of convergence in all cases in the analysis below. However, to be able to express the algorithm using conventional block structure, we reorder its components so that it appears to start with the  $z$ -minimization step, resulting in Algorithm 4.6.1. We now establish the convergence properties of Algorithm 4.6.1:

**Lemma 34.** *If the inner loops of Algorithm 4.6.1 always terminate finitely, the algorithm produces sequences conforming to the recursions (93)-(97).*

*Proof.* Consider the loop termination conditions

$$\frac{2}{c} \left| \langle w_1^k - x^{k+1}, y_1^{k+1} \rangle \right| + \|y_1^{k+1}\|^2 \leq \tau\sigma \|Mx^{k+1} - z^k\|^2 \quad (107)$$

$$\frac{2}{c} \left| \langle w_2^k - z^{k+1}, y_2^{k+1} \rangle \right| + \|y_2^{k+1}\|^2 \leq \alpha^t(1 - \tau)\sigma \|Mx^{k+1} - z^k\|^2. \quad (108)$$

Using that  $\tau \in [0, 1]$ ,  $\alpha \in (0, 1)$ , and  $t$  is a nonnegative integer, (108) implies that

$$\frac{2}{c} \left| \langle w_2^k - z^{k+1}, y_2^{k+1} \rangle \right| + \|y_2^{k+1}\|^2 \leq (1 - \tau)\sigma \|Mx^{k+1} - z^k\|^2 \quad (109)$$

Adding (107) and (109), we obtain the relative error condition (95). 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  $w^{k+1}$  near the end of the outer loop, it may be verified that the algorithm produces sequences conforming to (93)-(97).  $\square$

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 such cases it would be much simpler to use Algorithm 5.1.2 instead.

In addition to Assumption 2, we make the following mild assumption, as in Lemma 20:

**Assumption 35.** *For any  $p, z \in \mathbb{R}^m \times \mathbb{R}^m$ ,  $c > 0$ ,  $x \in \mathbb{R}^n$ , the sequence  $\{\mathcal{F}_1(p^\tau, z^\tau, c, x, l)\}_{l=1}^\infty$  is bounded.*

Note that all sequences  $\{z^l\}$  generated by the  $g$  approximation scheme  $\mathcal{G}$  must be bounded, because they must converge to the exact solution of the subproblem, as shown in Lemma 6.

**Lemma 36.** *In Algorithm 4.6.1, the labeled “with increasing  $l$ ” always terminates finitely.*

*Proof.* By the initialization step of the algorithm,  $Mx^{k+1} - z^k$  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  $Mx^{k+1} - z^k$  in this case. Therefore, we have  $\|Mx^{k+1} - z^k\| > 0$  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.  $\square$

While the  $l$  loop must always terminate finitely, it is possible that the loop over  $l'$  may run indefinitely. In this case, we are in essentially the same situation as in Algorithm 4.5.1, and we can show convergence of the inner loop to a KKT point

**Lemma 37.** *If the loop over  $l'$  in Algorithm 4.6.1 does not terminate, then  $(x^{k,t,\infty}, p^{k,l}, z^{k,l})$  is a KKT point of (P) for any limit point  $x^{k,t,\infty}$  of  $\{x^{k,t,l'}\}_{l'=0}^\infty$ .*

*Proof.* We begin by claiming that the  $l'$  loop only can run indefinitely if  $y_2^{k,l} = 0$  and  $\lim_{l' \rightarrow \infty} Mx^{k,t,l'} = z^{k,l}$ . To establish this claim, first suppose that  $Mx^{k,t,l'} \not\rightarrow z^{k,l}$ . Then the left-hand side of **accept** converges to zero, while the condition's right-hand side is non-negative and does not converge to zero, so **accept** will eventually become true and the  $l'$  loop must terminate. Now instead suppose that  $Mx^{k,t,l'} \rightarrow z^{k,l}$  but  $y_2^{k,l} \neq 0$ . In this case, the alternative conditions  $\|y_1^{k,t,l'}\|^2 \leq \beta \|y_2^{k,l}\|^2$  and  $\|Mx^{k,t,l'} - z^{k,l}\|^2 \leq \gamma \|y_2^{k,l}\|^2$  must eventually hold for sufficiently large  $l'$ , and again the  $l'$  loop must terminate. This establishes the claim.

So, since we are assuming that the  $l'$  loop runs indefinitely, we know that  $y_2^{k,l} = 0$  and  $\lim_{l' \rightarrow \infty} Mx^{k,t,l'} = z^{k,l}$ . Since  $y_2^{k,l} = 0$ , we know that  $z^{k,l}$  must be an exact minimizer of previous  $g$  subproblem. Furthermore, the properties of the  $\mathcal{F}$ -procedure from Assumption 2 guarantee that  $\lim_{l' \rightarrow \infty} y_1^{k,t,l'} = 0$ . For any limit point  $x^{k,t,\infty}$  of  $\{x^{k,t,l'}\}_{l'=0}^\infty$ , Corollary 16 then implies that  $(x^{k,t,\infty}, p^{k,l}, z^{k,l})$  is a KKT point.  $\square$

We may refer to Lemma 5 for conditions under which  $\{x^{k,t,l'}\}_{l'=0}^\infty$  must have limit points or must converge.

Next, we consider the scenario that both of the loops over  $l$  and  $l'$  terminate finitely, but the loop over  $t$  runs indefinitely. In this case, we again obtain convergence to a KKT point.

**Lemma 38.** *In Algorithm 4.6.1, if the loop over  $t$  runs indefinitely for some  $k$ , then any limit point of  $\{\bar{x}^{k,t}, \bar{z}^{k,t}, \bar{p}^{k,t}\}_{t=0}^\infty$  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  $l'$  must always terminate finitely, indicating that the alternative  $l'$  termination conditions  $\|y_1^{k,t,l'}\|^2 \leq \beta \|y_2^{k,l}\|^2$  and  $\|M\bar{x}^{k,t,l'} - \bar{z}^{k,l}\|^2 \leq \gamma \|y_2^{k,l}\|^2$  hold. By construction, we have  $\alpha \in (0, 1)$  and thus  $\lim_{t \rightarrow \infty} \alpha^t = 0$ , so we know that  $\lim_{t \rightarrow \infty} \bar{y}_2^{k,t} = 0$ . Consequently, the alternative termination condition in the loop over  $l'$  ensures that  $\lim_{t \rightarrow \infty} \bar{y}_1^{k,t} = 0$  and  $\lim_{t \rightarrow \infty} M\bar{x}^{k,t} - \bar{z}^{k,t} = 0$ . Furthermore,  $\{\bar{z}^{k,t}\}_{t=0}^\infty$  must be convergent since it is a subsequence of the convergent sequence  $\{z^{k,l}\}_{l=1}^\infty$ . The convergence of  $\{z^{k,l}\}_{l=1}^\infty$  implies the convergence of  $\{p^{k,l}\}_{l=1}^\infty$ . The conclusion then holds by Lemma 15.  $\square$

Finally we are ready to show the convergence of Algorithm 4.6.1:

**Proposition 39.** *Under Assumptions 1 and 35, Algorithm 4.6.1 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  $\{(x^k, z^k, p^k)\}$  generated by the algorithm is a KKT points of (P),  $\{p^k\}$  converges to an optimal solution of the dual problem (D), and  $\lim_{k \rightarrow \infty} f(x^k) + g(z^k) = f(x^*) + g(z^*)$ , where  $(x^*, z^*)$  is any optimal solution of (P).*
2. *The loop over  $l'$  runs indefinitely for some  $k$ ,  $t$ , and  $l$ , in which case  $(x^{k,t,\infty}, p^{k,l}, z^{k,l})$  is a KKT point of (P) for any limit point  $x^{k,t,\infty}$  of  $\{x^{k,t,l'}\}_{l'=0}^\infty$ .*
3. *The loop over  $t$  runs indefinitely for some  $k$ , in which case every limit point of the sequence  $\{\bar{x}^{k,t}, \bar{z}^{k,t}, \bar{p}^{k,t}\}_{t=0}^\infty$  is a KKT point of (P).*

*Proof.* By Lemma 36, the  $l$  loop always terminates finitely, so the only possibilities are infinite loops over  $k$ ,  $l'$ , or  $t$ . Lemma 37 guarantees the claimed convergence in the case of an infinite  $l'$  loop. Similarly, Lemma 38 guarantees the claimed convergence if the  $t$  loop does not terminate finitely. In the case of an infinite  $k$  loop, Lemma 34 asserts that the sequences generated by the algorithm obey the recursions (93)-(97) for all  $k \geq 1$  (although not necessarily  $k = 0$ , but that is of no consequence), and therefore we may apply Propositions 30 and 32 to obtain the claimed result.  $\square$

**Remark:** In Algorithm 4.6.1, the expression  $\alpha^t$  may be replaced  $\zeta(t)$ , where  $\zeta$  is any function  $\mathbb{N} \rightarrow (0, 1]$  such that  $\lim_{t \rightarrow \infty} \zeta(t) = 0$ , and the convergence results given above will still hold. Only a minor modification to the proof of Lemma 38 is required.

## 5 Numerical Tests

### 5.1 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 Section 3 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 4.3. 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  $\|x^{k,l}\| \leq \beta_1$  and  $\|z^{k,l}\| \leq \beta_2$  throughout this algorithm, it coincides with the simpler Algorithm 3.1.1 of Section 3.1. 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 4.4 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 5.1.1** Exact version of ADMM: `admm_exact`

---

**initialization:** Pick  $c > 0$  and initial points  $p^0, z^0 \in \mathbb{R}^m$   
**repeat** {**for**  $k = 0, 1, 2, \dots$ }  
    **repeat** {**for**  $l = 1, 2, \dots$ }  
        Improve the solution to  $x^{k+1} \approx \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\}$  by taking  $(x^{k,l}, y^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$   
    **until** Inner loop convergence  
     $x^{k+1} = x^{k,l}$   
     $y^{k+1} = y^{k,l}$   
     $z^{k+1} = \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\}$   
     $p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$   
**until** Overall convergence

---

For `admm_primDR`, we simply use Algorithm 3.4.1. In the case of `admm_abssum`, we specialize Algorithm 3.1.1 to the case that the  $g$  minimization is exact, resulting in the following algorithm:

---

**Algorithm 5.1.2** 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  $\{\epsilon_k\}$   
**repeat** {**for**  $k = 0, 1, 2, \dots$ }  
    **repeat** {**for**  $l = 1, 2, \dots$ }  
        Improve the solution to  $x^{k+1} \approx \arg \min_x \left\{ f(x) + \langle p^k, Mx \rangle + \frac{c}{2} \|Mx - z^k\|^2 \right\}$  by taking  $(x^{k,l}, y_1^{k,l}) = \mathcal{F}(p^k, z^k, c, x^k, l)$   
    **until**  $\|y^{k,l}\| \leq \epsilon_{k+1}$   
     $x^{k+1} = x^{k,l}$   
     $y^{k+1} = y^{k,l}$   
     $z^{k+1} = \arg \min_z \left\{ g(z) - \langle p^k, z \rangle + \frac{c}{2} \|Mx^{k+1} - z\|^2 \right\}$   
     $p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1})$   
**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 4.5.1, since we are only testing the case in which the  $g$  minimization is exact.

## 5.2 Termination criteria and algorithm parameters

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

$$\text{dist}_\infty(0, \partial_x [f(x) + g(Mx)]_{x=x^k}) \leq \epsilon, \quad (110)$$

where  $\text{dist}_\infty(t, S) = \inf\{\|t - s\|_\infty \mid s \in S\}$ , 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

$$\|y^{k,l}\| \leq \frac{\epsilon}{10}, \quad (111)$$

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  $\frac{2}{c} |\langle w^k - x^{k,l}, y^{k,l} \rangle| + \|y^{k,l}\|^2 \leq \sigma \|Mx^{k,l} - z^k\|^2$  to be significantly more restrictive than (111). 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 (110). 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 (111) holds.

For `admm_abssum`, we need to select a positive absolutely summable sequence  $\{\epsilon_k\}$ . 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.

## 5.3 LASSO regression

A simple and common problem class that fits readily into the form (1) is the “LASSO” [38] or “compressed sensing” problem. This problem is an  $L_1$ -regularized version of linear regression, taking the form

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \nu \|x\|_1, \quad (112)$$

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  $Ax = b$ , with a

preference for making the solution  $x \in \mathbb{R}^n$  sparse. Letting  $f(x) = \frac{1}{2} \|Ax - b\|^2$ ,  $g(z) = \nu \|z\|_1$ , and  $M = I$  the LASSO problem (112) may be written as

$$\begin{aligned} \min \quad & f(x) + g(z) \\ \text{s. t.} \quad & x = z. \end{aligned} \quad (113)$$

Applying the ADMM to (113), the  $x$ -minimization subproblem (2) reduces to solving a system of linear equations involving the matrix  $A^\top A + cI$ :

$$x^{k+1} = (A^\top A + cI)^{-1} (A^\top b + cz^k - p^k). \quad (114)$$

Note that the  $A^\top A + cI$  is always invertible since  $c > 0$ . We use the conjugate gradient method as given in [27, Algorithm 5.2] to solve the linear system (114).

The  $z$ -minimization subproblem (3) reduces to the soft thresholding operator (106) with  $M = I$ :

$$z_i^{k+1} = \text{sgn}(x_i^{k+1} + \frac{1}{c}p_i^k) \max \left\{ 0, \left| x_i^{k+1} + \frac{1}{c}p_i^k \right| - \frac{\nu}{c} \right\}, \quad i = 1, \dots, n. \quad (115)$$

This calculation is straightforward and requires a constant amount of time per element  $z_i^{k+1}$ . The multiplier update  $p^{k+1} = p^k + c(x^{k+1} - z^{k+1})$  has a similar property, so the  $x$  minimization dominates run time of this application if the ADMM if (114) is solved exactly.

As in [3, Section 11.1], we scaled  $b$  and the columns of  $A$  to have unit  $\ell_2$  norm and set the regularization parameter  $\nu$  to  $0.1 \|A^\top b\|_\infty$ . For `admm_exact`, the termination condition for the inner loop (111) becomes

$$\|(A^\top A + cI)x^{k+1} - (A^\top b + cz^k - p^k)\| \leq 10^{-7} \quad (116)$$

We performed our LASSO tests on four categories of datasets:

*Gene expression*: Six standard cancer DNA microarray datasets from [5]. These instances have dense, wide, and relatively small matrices  $A$ , with the number of rows  $m \in [42, 102]$ , and the number of columns  $n \in [2000, 6033]$ .

*Single-Pixel Camera*: Four dense compressed image sensing problems from [7] with  $m \in [410, 4770]$  and  $n \in [1024, 16384]$ .

*Finance*: A single large dense financial dataset [19] with  $m = 30465$  and  $n = 216842$ .

*PEMS*: A single large, wide, and dense dataset from the California Department of Transportation [23] with  $m = 267$  and  $n = 138672$ .

Figures 1, 2, and 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 4, 5, and 6 depict the cumulative total number of inner iterations for 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.

Figure 1: LASSO — number of outer iterations for the cancer datasets.

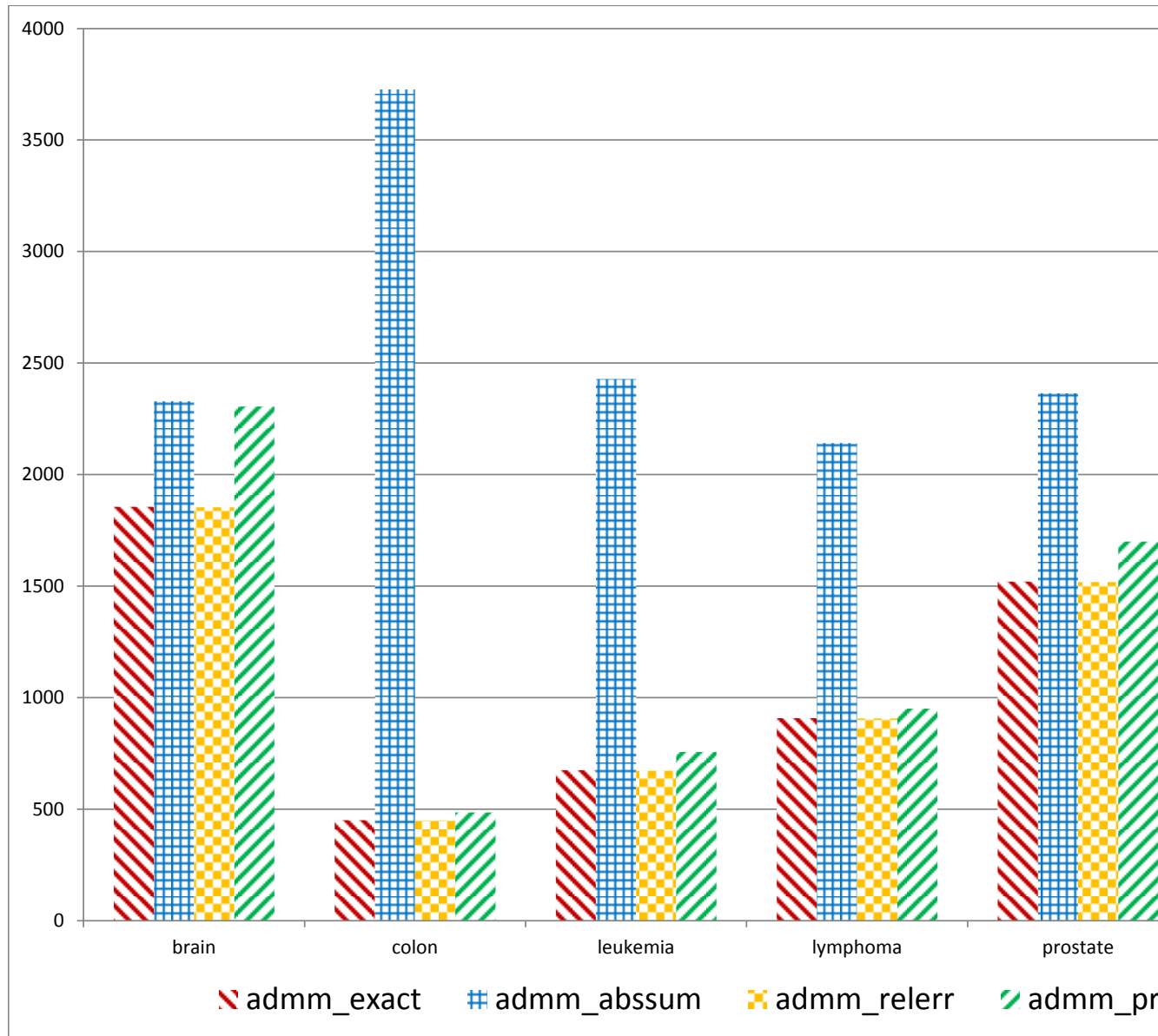




Figure 2: LASSO — number of outer iterations for the image datasets.

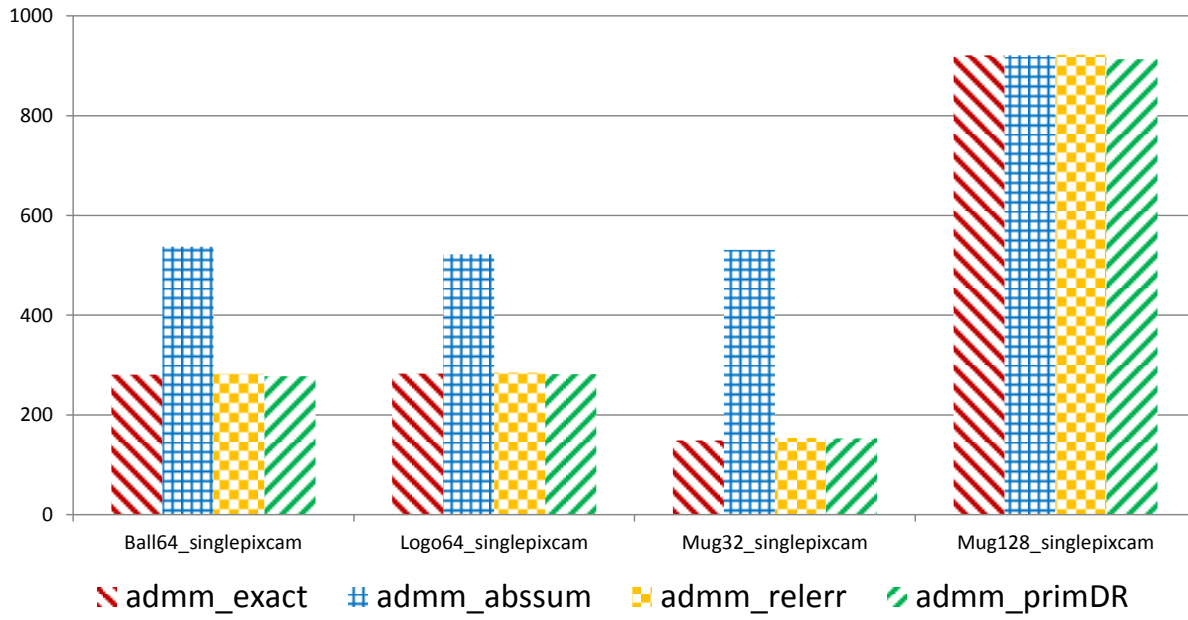


Figure 3: LASSO — number of outer iterations for PEMS and finance1000.

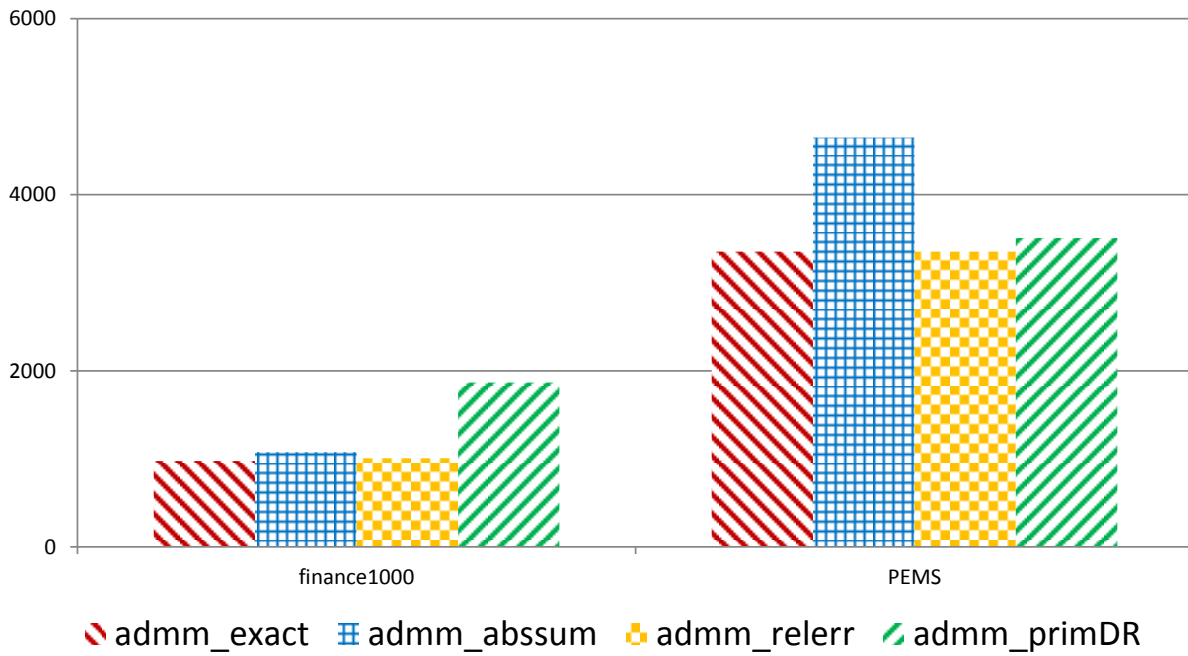


Figure 4: LASSO — total number of inner iterations for the cancer datasets.

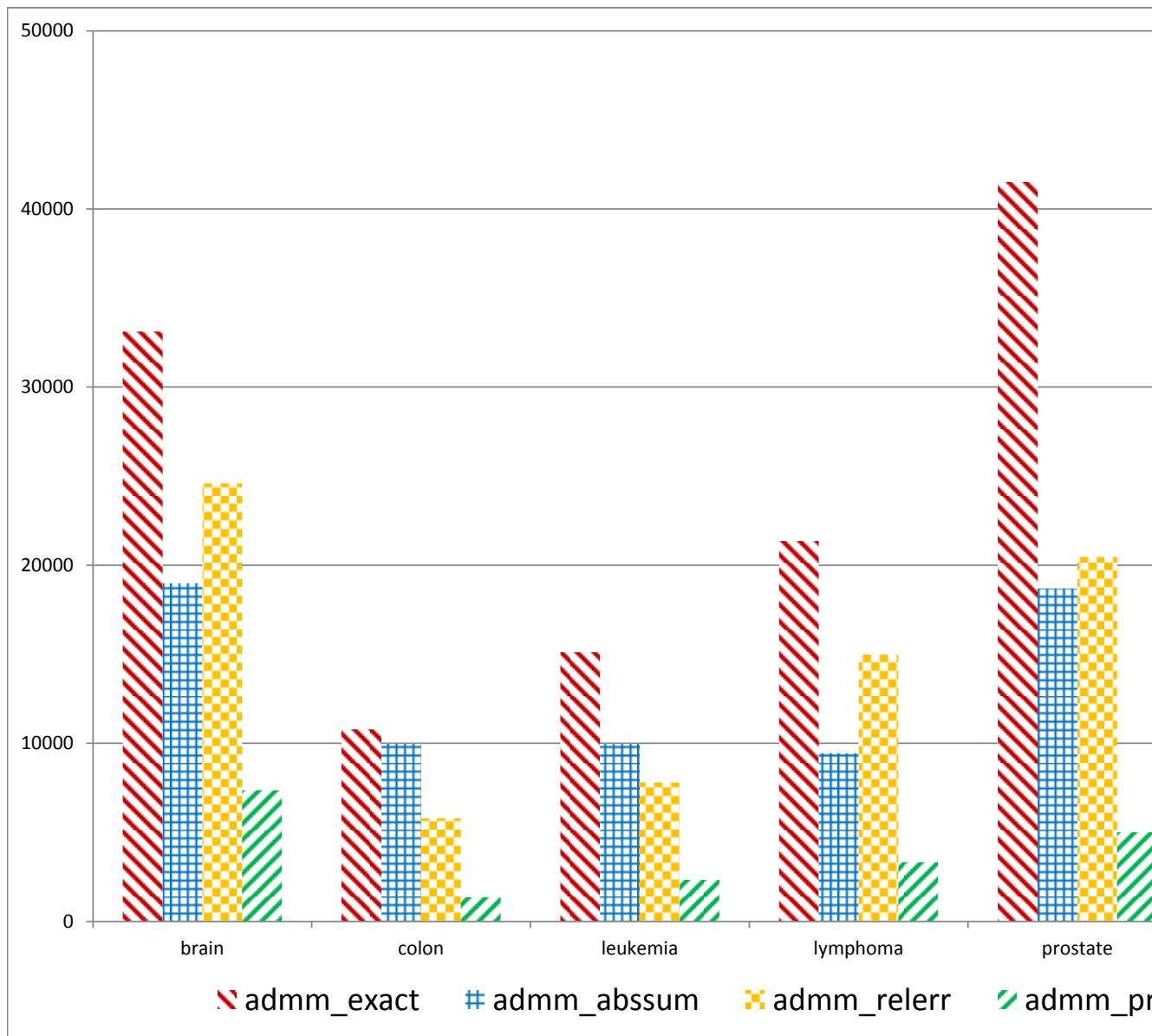


Figure 5: LASSO — total number of inner iterations for the image datasets.

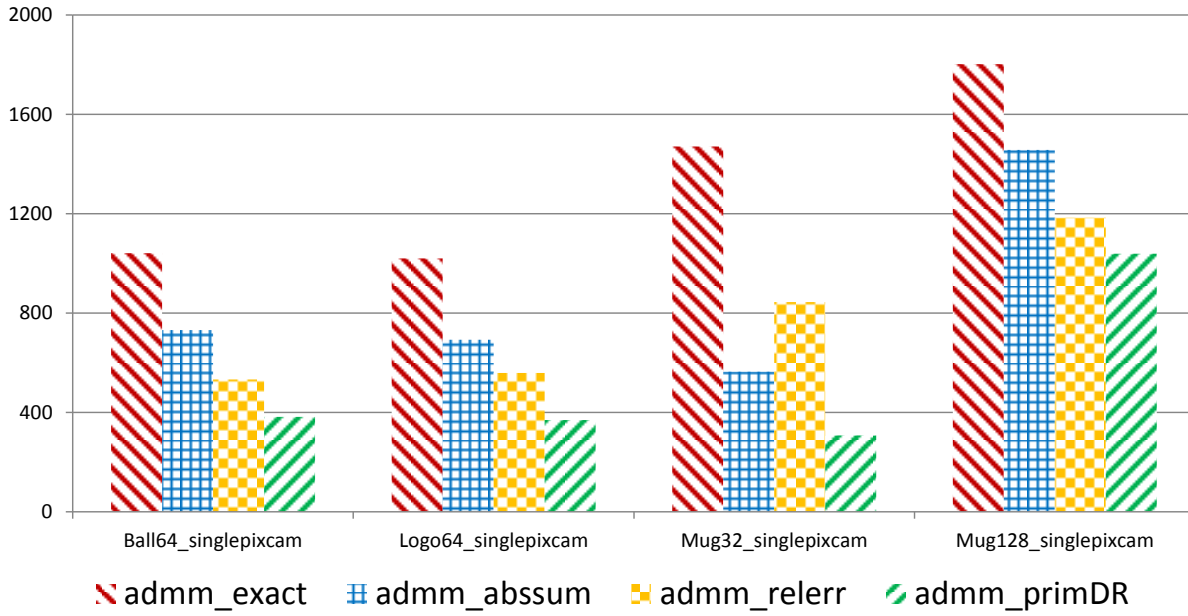
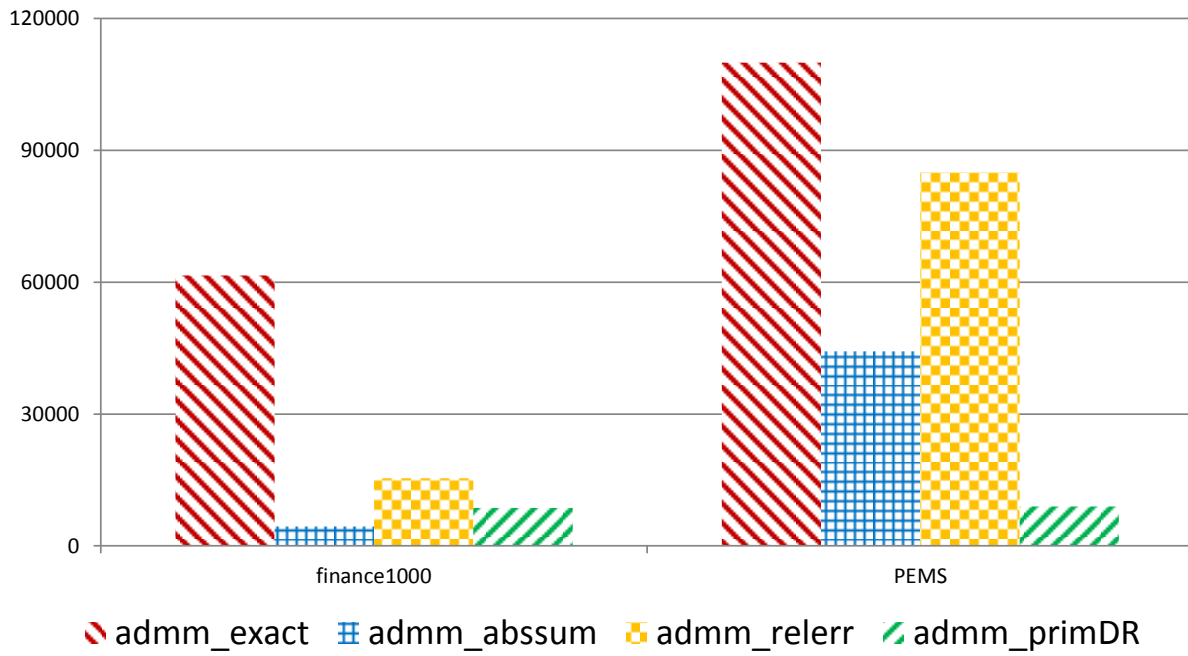


Figure 6: LASSO — total number of inner iterations for PEMS and finance1000.



## 5.4 $L_1$ -regularized logistic regression

Logistic regression with  $L_1$  regularization has been proposed as a promising method for feature selection in classification problems [15, 26]. Given a training dataset consisting of  $m$  pairs  $(a_i, b_i)$ , 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

$$\min_{\substack{w \in \mathbb{R}^n \\ v \in \mathbb{R}}} \sum_{i=1}^m \log(1 + \exp(-b_i(a_i^\top w + v))) + \nu \|w\|_1. \quad (117)$$

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  $A = [a_1, \dots, a_m]^\top$ . We set the problem parameter  $\nu$  in the same manner as in [3, Section 11.2]. To apply the ADMM to (117), we may formulate it as

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

where  $f((v, w)) = \sum_{i=1}^m \log(1 + \exp(-b_i(a_i^\top w + v)))$  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  $a'_i = b_i a_i$  for  $i = 1, \dots, m$ , yielding

$$f((v, w)) = \sum_{i=1}^m \log(1 + \exp(-a_i'^\top w - b_i v)) \quad (118)$$

We assemble the vectors  $a'_i$  into a matrix  $A' = [a'_1, \dots, a'_m]^\top$  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 (118), we obtain

$$\nabla_v f((v, w)) = - \sum_{i=1}^m \frac{\exp(-a_i'^\top w - b_i v)}{1 + \exp(-a_i'^\top w - b_i v)} b_i = -b^\top \left( \frac{\exp(-A'w - bv)}{1 + \exp(-A'w - bv)} \right), \quad (119)$$

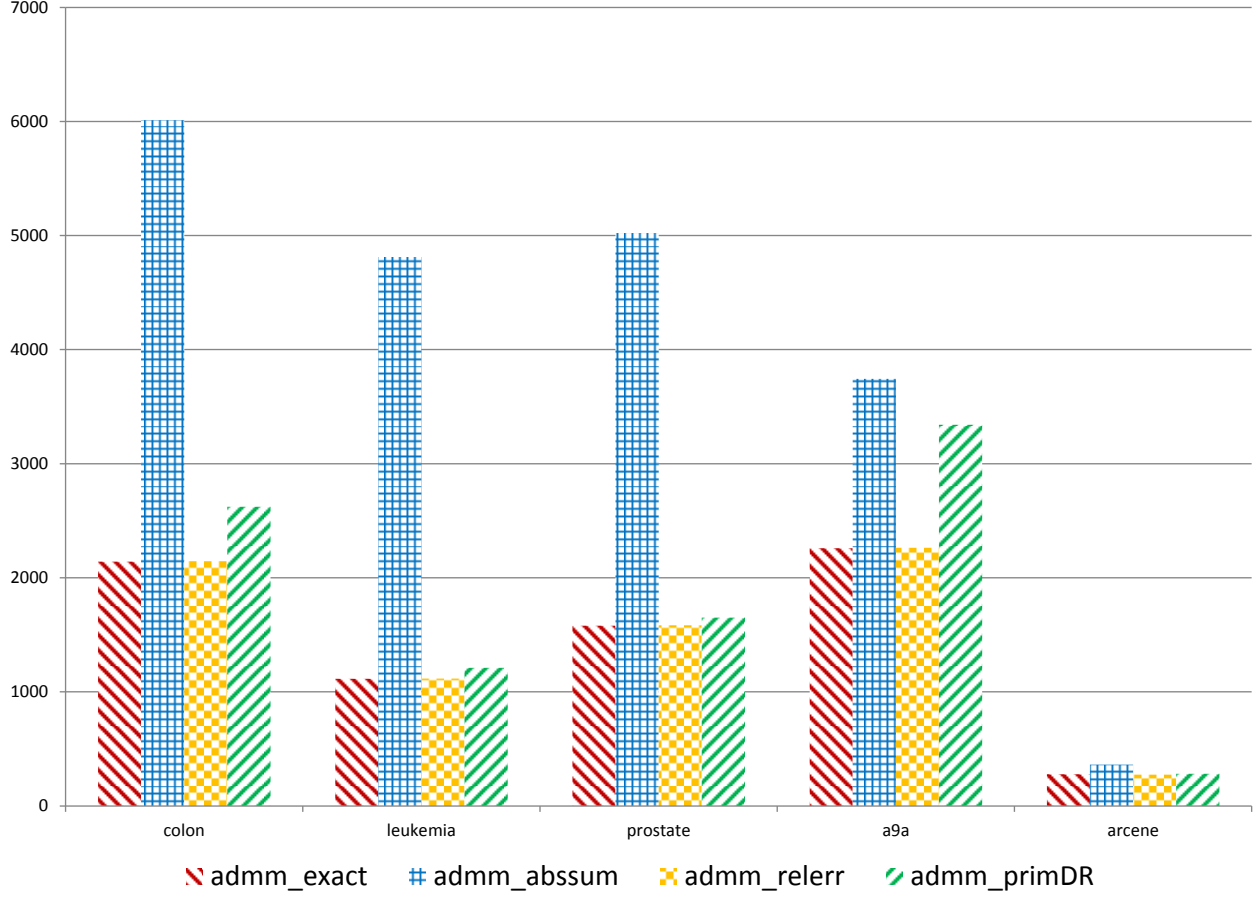
where the exponentiation, addition, and division operations in the final parenthesized expression are interpreted as being applied componentwise. Using a similar notation, we also find that

$$\nabla_w f((v, w)) = - \sum_{i=1}^m \frac{\exp(-a_i'^\top w - b_i v)}{1 + \exp(-a_i'^\top w - b_i v)} a'_i = -A'^\top \left( \frac{\exp(-A'w - bv)}{1 + \exp(-A'w - bv)} \right). \quad (120)$$

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

$$y = \nabla f(x) + p + c(x - z) = \left( \nabla f_v((v, w)), \nabla f_w((v, w)) \right) + p + c(x - z). \quad (121)$$

Figure 7: Logistic regression — number of outer iterations.

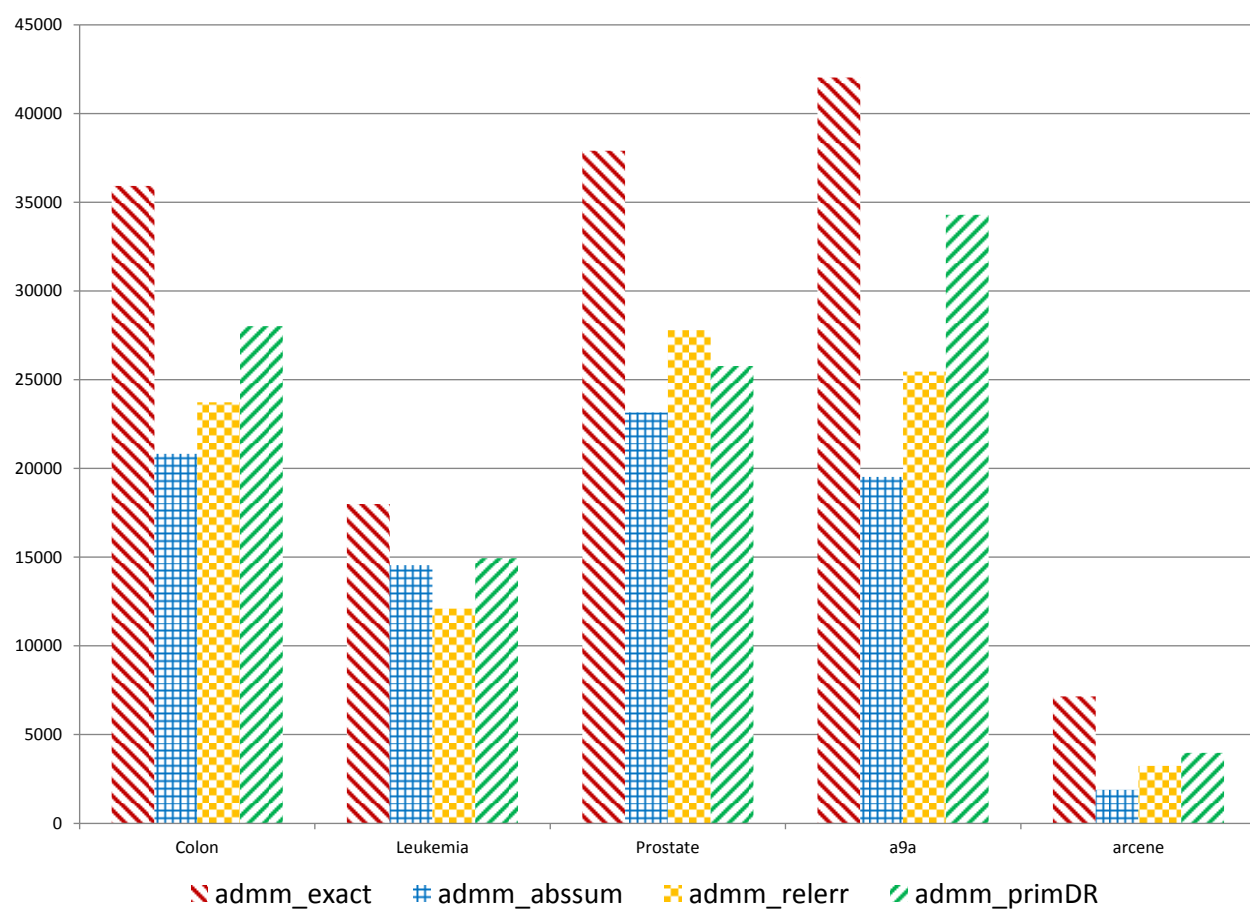


To approximately solve the  $f$  subproblem, we employ the limited-memory BFGS (L-BFGS) method [25] for unconstrained nonlinear optimization. We solve the second ( $g$ ) subproblem exactly using essentially the same soft thresholding operator (106) (applied only to the  $w$  component) used in the LASSO problem.

For test data, we selected the cancer datasets from [5] that have  $b_i \in \{-1, 1\}$  for all  $i$ . In addition, we also used the *a9a* [13] and *Arcene* [17] datasets, which are both available from the UCI Machine Learning Repository [23] (where *a9a* is called *adult*). Both of these datasets are sparse, and *a9a* has  $m = 32,561$  and  $n = 123$ , while *Arcene* has  $m = 900$  and  $n = 10,000$ .

Figure 7 shows the number of outer iterations for each algorithm, revealing a pattern similar to that for the LASSO problem. Figure 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.

Figure 8: Logistic regression — total number of inner iterations.



## 5.5 Sparse inverse covariance selection

The covariance selection problem was first introduced in [4], 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* [12, 20]. 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 dependencies between their expression profiles are indicated by graph edges.

In this problem, we are given a dataset of vectors  $a_1, \dots, a_m \in \mathbb{R}^n$  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*

$$\min_{X \in \mathbb{S}_{++}^n} \mathbf{Tr}(SX) - \log \det X + \nu \|X\|_1, \quad (122)$$

where  $\|X\|_1 = \sum_{i=1}^n \sum_{j=1}^n |x_{ij}|$ . The model's goal is to minimize the negative log-likelihood of the sample, combined with an  $\ell_1$  regularization term promoting sparsity of the solution [2]. Note that  $-\log \det X = \log \frac{1}{\det X} = \log \det (X^{-1})$ .

If one lets  $f(X) = \mathbf{Tr}(SX) - \log \det X + \nu \|X\|_1$  and  $g(X) = \nu \|X\|_1$ , then (treating the unknown  $X$  as a vector) problem (122) fits the standard ADMM problem form with  $M = I$ , that is,  $\min_X \{f(X) + g(X)\}$ . Applying the ADMM, we obtain the recursions

$$X^{k+1} = \arg \min_{X \in \mathbb{S}_{++}^n} \left\{ \mathbf{Tr}(SX) - \log \det X + \frac{c}{2} \|X - Z^k + \frac{1}{c} P^k\|_F^2 \right\} \quad (123)$$

$$Z_{ij}^{k+1} = \text{sgn} \left( X_{ij}^{k+1} + \frac{1}{c} P_{ij}^k \right) \max \left\{ 0, \left| X_{ij}^{k+1} + \frac{1}{c} P_{ij}^k \right| - \frac{\nu}{c} \right\} \quad i, j = 1, 2, \dots, n. \quad (124)$$

$$P^{k+1} = P^k + c \left( X^{k+1} - Z^{k+1} \right), \quad (125)$$

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 [22]. For some examples of applying ADMM to SICS, see [34] and [42], the latter showing that ADMM outperforms other methods for this problem.

It is possible to develop an analytical solution to subproblem (123). First, we know that

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

Using the first order optimality condition for the convex problem (123), whose solution must lie in the interior of the open cone  $\mathbb{S}_{++}^n$ , we obtain

$$0 \in \partial \left\{ \mathbf{Tr}(SX) - \log \det X + \frac{c}{2} \|X - Z^k + \frac{1}{c} P^k\|_F^2 \right\} \quad (126)$$

$$\Leftrightarrow 0 = S - X^{-1} + P^k + c \left( X - Z^k \right). \quad (127)$$

Rearranging (127), we obtain

$$cX - X^{-1} = cZ^k - P^k - S. \quad (128)$$

Thus,  $X^{k+1}$  can be obtained by solving (128) for  $X$ . Next we take the orthogonal eigenvalue decomposition of the symmetric matrix on the left-hand side of (128), obtaining some diagonal matrix  $\Lambda^{k+1} = \mathbf{diag}(\lambda_1^{k+1}, \dots, \lambda_n^{k+1})$  and orthogonal matrix  $Q^{k+1}$  such that

$$cX^{k+1} - (X^{k+1})^{-1} = Q^{k+1} \Lambda^{k+1} (Q^{k+1})^\top. \quad (129)$$

Multiplying by  $(Q^{k+1})^\top$  from the left and by  $Q^{k+1}$  from the right on both sides of (128), we obtain

$$c\tilde{X}^{k+1} - (\tilde{X}^{k+1})^{-1} = \Lambda^{k+1}, \quad (130)$$

where  $\tilde{X}^{k+1} = (Q^{k+1})^\top X^{k+1} Q^{k+1}$ . We now can construct a diagonal solution  $\tilde{X}^{k+1}$  of (130). To find each entry  $\tilde{X}_{ii}$  on the diagonal of  $\tilde{X}^{k+1}$ , we need to solve  $c\tilde{X}_{ii}^{k+1} - 1/\tilde{X}_{ii}^{k+1} = \lambda_i^{k+1}$ , and because all the  $\tilde{X}_{ii}$  must be nonnegative, we have

$$\tilde{X}_{ii}^{k+1} = \frac{\lambda_i^{k+1} + \sqrt{(\lambda_i^{k+1})^2 + 4c}}{2c}. \quad (131)$$

At this point, one can see that the solution of equation (128),  $X^{k+1}$ , is

$$X^{k+1} = Q^{k+1} \tilde{X}^{k+1} (Q^{k+1})^\top. \quad (132)$$

The  $g$  minimization (124) is the same soft thresholding operation as in (115), applied throughout a symmetric matrix. Clearly, the most time-consuming part of (123)-(125) is the eigenvalue decomposition (129) required by the  $x$  minimization (123). In our numerical experiments, we used the Jacobi iterative method from [33] 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 (129):

$$c\tilde{X}^{k,l} - (\tilde{X}^{k,l})^{-1} \approx Q^{k,l} \Lambda^{k,l} (Q^{k,l})^\top. \quad (133)$$

This equation may be solved using appropriate modifications of (131) and (132), namely

$$\tilde{X}_{ii}^{k,l} = \frac{\lambda_i^{k,l} + \sqrt{(\lambda_i^{k,l})^2 + 4c}}{2c} \quad X^{k,l} = Q^{k,l} \tilde{X}^{k,l} (Q^{k,l})^\top, \quad (134)$$



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 (132). 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 (123).

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 [22]:

*Lymph node status* (**Lymph**): This dataset is derived from [6], 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 [6]. 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 [39].  $S$  has dimension  $n = 843$  and its rank is 117.

*Leukemia* (**Leukemia**): A gene expression data set from [41] that has dimension  $n = 1255$  and rank 71.

*Hereditary breast cancer* (**Hereditarybc**): This data set is from [18]. 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 [22] 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 9 and 10. Figure 9 depicts the number outer iterations, which show very little variation between the various ADMM variants. Figure 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.

## References

- [1] Maicon M. Alves and Benar 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] 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.
- [3] 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.

Figure 9: SICS — number of outer iterations.

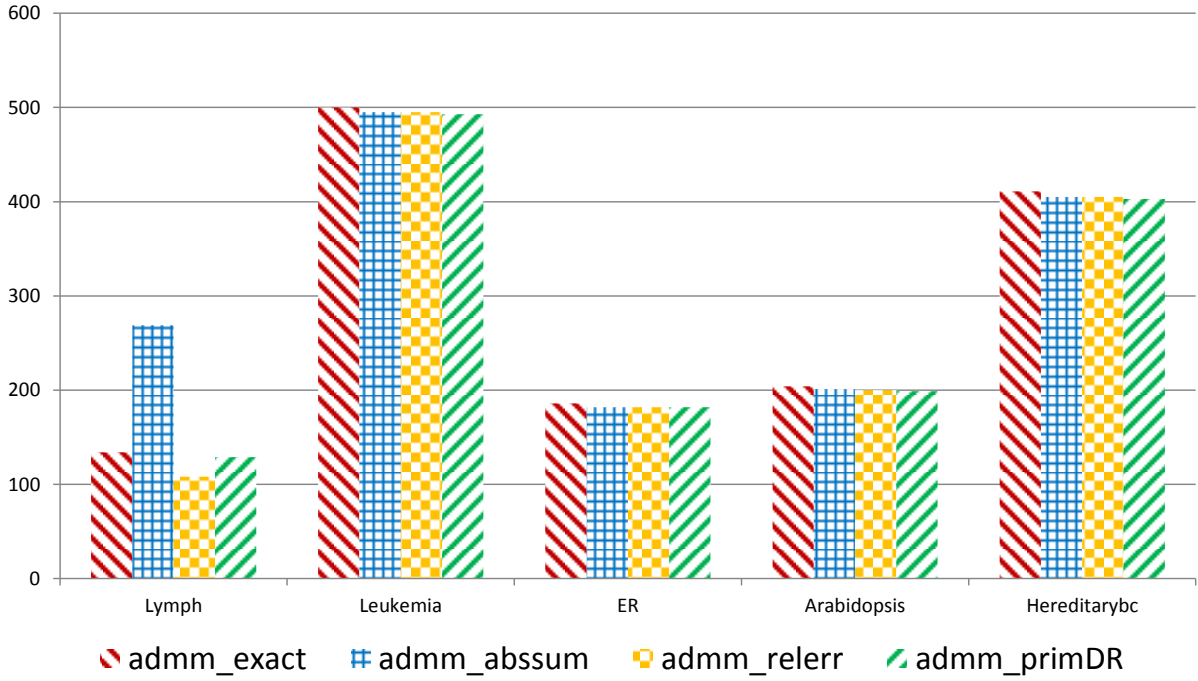
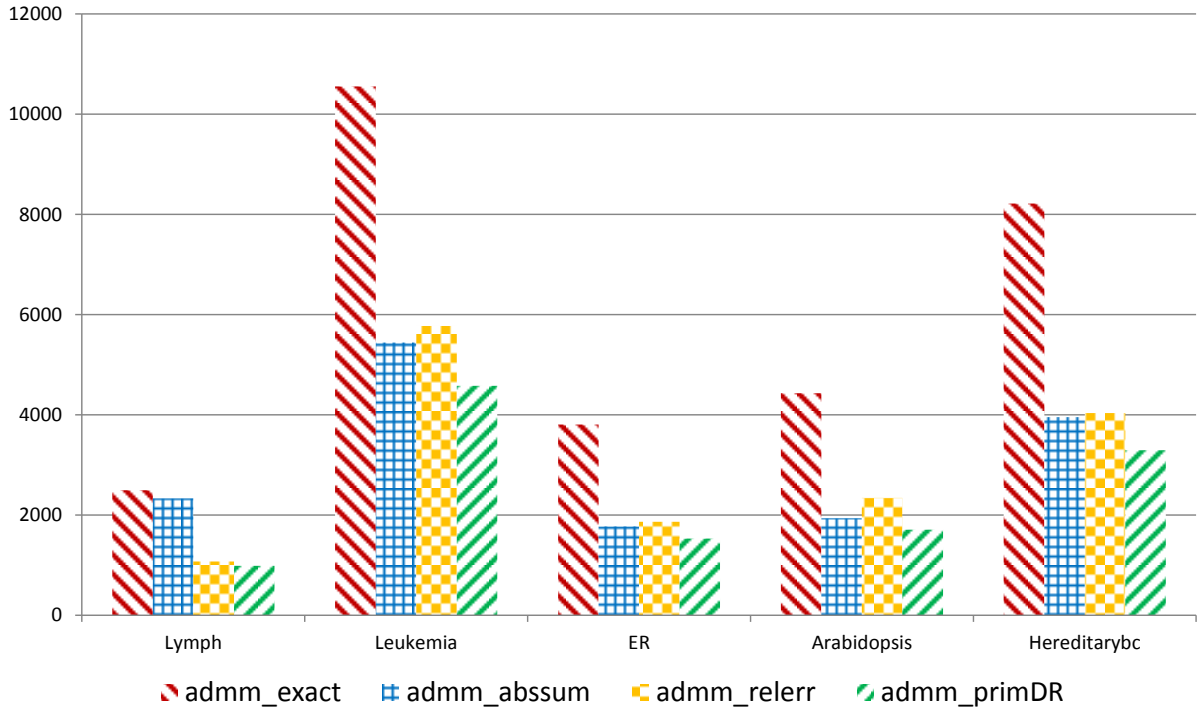


Figure 10: SICS — total number of inner iterations.



- [4] Arthur P. Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972.
- [5] Marcel Dettling and Peter Bühlmann. Finding predictive gene groups from microarray data. *J. Multivariate Anal.*, 90(1):106–131, 2004.
- [6] Adrian Dobra. Variable selection and dependency networks for genomewide data. *Bio-statistics*, 10(4):621–639, 2009.
- [7] 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.
- [8] Jonathan Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. PhD thesis, Massachusetts Institute of Technology, 1989.
- [9] Jonathan Eckstein. A practical general approximation criterion for methods of multipliers based on Bregman distances. *Math. Program.*, 96(1):61–86, 2003.
- [10] 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.
- [11] Jonathan Eckstein and Paulo J. S. Silva. A practical relative error criterion for augmented Lagrangians. *Math. Program.*, 141(1-2):319–348, 2013.
- [12] David Edwards. *Introduction to Graphical Modeling*. Springer-Verlag, New York, second edition, 2000.
- [13] 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.
- [14] 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 boundary-value problems*, volume 15 of *Studies in Mathematics and its Applications*, pages 97–146. North-Holland, Amsterdam, 1983.
- [15] James Franklin. The elements of statistical learning: data mining, inference and prediction. *Math. Intelligencer*, 27(2):83–85, 2005.
- [16] 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, Amsterdam, 1983.

- [17] 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, Cambridge, MA, USA, 2005.
- [18] 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.
- [19] Shimon Kogan, Dmitry 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.
- [20] Steffen L. Lauritzen. *Graphical Models*, volume 17. The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.
- [21] Jim Lawrence and Jonathan E. Spingarn. On fixed points of nonexpansive piecewise isometric mappings. *Proc. London Math. Soc.*, 55(3):605–624, 1987.
- [22] Lu Li and Kim-Chuan Toh. An inexact interior point method for  $L_1$ -regularized sparse covariance selection. *Math. Program. Comput.*, 2(3-4):291–315, 2010.
- [23] Moshe Lichman. UCI machine learning repository, 2013.
- [24] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two non-linear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.
- [25] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Program.*, 45(3):503–528, 1989.
- [26] 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.
- [27] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, New York, second edition, 2006.
- [28] Boris T. Polyak. *Introduction to Optimization*. Optimization Software, Inc., New York, 1987.
- [29] R. Tyrrell Rockafellar. Local boundedness of nonlinear, monotone operators. *Michigan Math. J.*, 16:397–407, 1969.
- [30] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [31] R. Tyrrell Rockafellar. *Conjugate Duality and Optimization*. SIAM, Philadelphia, 1974.

- [32] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.
- [33] Heinz Rutishauser. The Jacobi method for real symmetric matrices. *Numer. Math.*, 9(1):1–10, 1966.
- [34] 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 2101–2109, 2010. Proceedings of “Neural Information Processing Systems 2010”.
- [35] Mikhail V. Solodov and Benar F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.
- [36] Mikhail V. Solodov and Benar F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.
- [37] Mikhail V. Solodov and Benar 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.
- [38] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.
- [39] 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.
- [40] Ming Yan and Wotao Yin. Self equivalence of the alternating direction method of multipliers. In Roland Glowinski, Stanley J. Osher, and Wotao Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, Cham, Switzerland, 2016.
- [41] 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.
- [42] Xiaoming Yuan. Alternating direction method for covariance selection models. *J. Sci. Comput.*, 51(2):261–273, 2012.