Mathematical Programming 64 (1994) 81–101

A proximal-based decomposition method for convex minimization problems

Gong Chen, Marc Teboulle*

Department of Mathematics and Statistics, University of Maryland, Baltimore County Campus, Baltimore, MD 21228, USA

(Received 10 September 1992; Revised manuscript received 12 April 1993)

Abstract

This paper presents a decomposition method for solving convex minimization problems. At each iteration, the algorithm computes two proximal steps in the dual variables and one proximal step in the primal variables. We derive this algorithm from Rockafellar's proximal method of multipliers, which involves an augmented Lagrangian with an additional quadratic proximal term. The algorithm preserves the good features of the proximal method of multipliers, with the additional advantage that it leads to a decoupling of the constraints, and is thus suitable for parallel implementation. We allow for computing approximately the proximal minimization steps and we prove that under mild assumptions on the problem's data, the method is globally convergent and at a linear rate. The method is compared with alternating direction type methods and applied to the particular case of minimizing a convex function over a finite intersection of closed convex sets.

AMS Subject Classification: 90C25.

Key words: Convex programming; Proximal methods; Augmented Lagrangian; Decomposition-splitting methods

1. Introduction

This paper presents a decomposition algorithm for solving convex programming problems with separable structure. The method is based on the properties of the proximal point algorithm and its primal–dual application.

The proximal point algorithm [23] is an iterative method for finding a zero of a maximal monotone operator T, namely, a point $x^* \in \mathbb{R}^n$ such that $0 \in T(x^*)$. Starting at a point

*Corresponding author. Partially supported by Air Force Office of Scientific Research Grant 91-0008 and National Science Foundation Grant DMS-9201297.

SSDI 0025-5610(93)E0046-H

$x^0 \in \mathbb{R}^n$, the proximal point algorithm generates successively a sequence of points $x^{k+1} = (I + \lambda_k T)^{-1} x^k$, where λ_k is a positive scalar and the operator $(I + \lambda_k T)^{-1}$ is a single valued, nonexpansive operator on $\mathbb{R}^n$, [23]. For a survey we refer the reader to Lemaire [17], and for more recent convergence results on the proximal point algorithm, see Guler [12]. In [22], Rockafellar showed how the proximal point algorithm can be applied to convex programming problems in three different ways: the primal proximal minimization algorithm, the dual method of multipliers and the primal–dual proximal method of multipliers. Here, we are particularly interested in the *proximal method of multipliers* when applied to convex programs with separable structure of the form:

$$\text{(P)} \qquad \min\{f(x) + g(z) \colon Ax = z\} \tag{1.1}$$

where $f \colon \mathbb{R}^n \mapsto (-\infty, +\infty]$ and $g \colon \mathbb{R}^m \mapsto (-\infty, +\infty]$ are given closed proper convex functions, and A is a given $m \times n$ matrix. Note that many convex programs can be formulated in the generic form (1.1), see e.g. [10, 21]. The Lagrangian for problem (P) is defined by $L \colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \mapsto (-\infty, +\infty]$,

$$L(x, z, y) = f(x) + g(z) + \langle y, Ax - z \rangle \tag{1.2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and y is the Lagrangian multiplier associated with the constraint $Ax = z$.

The Lagrangian $L(x, z, y)$ is a closed convex–concave function. Therefore, the set-valued subdifferential mapping S on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ given by

$$S(x, z, y) = \partial_{x,z} L(x, z, y) \times \partial_y(-L(x, z, y)) \tag{1.3}$$

is maximal monotone [22]. A pair (x^*, z^*) is optimal for (P) and y^* is an optimal Lagrangian multiplier if and only if

$$L(x^*, z^*, y) \leqslant L(x^*, z^*, y^*) \leqslant L(x, z, y^*) \quad \forall (x, z) \in \mathbb{R}^{n \times m},\ \forall y \in \mathbb{R}^m, \tag{1.4}$$

that is, if and only if $0 \in S(x^*, z^*, y^*)$ (in the sequel, 0 denotes a zero element in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$).

In [22], Rockafellar applied the proximal point algorithm to S to produce the proximal method of multipliers for (P). At each iteration of this algorithm, given $\hat{x}, \hat{z}, \hat{y}$, one minimizes the augmented Lagrangian with respect to x and z,

$$L_\lambda(x, z, \hat{y}) = L(x, z, \hat{y}) + \tfrac{1}{2}\lambda \|Ax - z\|^2 + (1/(2\lambda))(\|x - \hat{x}\|^2 + \|z - \hat{z}\|^2) \tag{1.5}$$

to obtain the next iterates (x^+, z^+), and then updates the multiplier by the iteration

$$y^+ = \hat{y} + \lambda(Ax^+ - z^+) \tag{1.6}$$

where λ is a positive scalar.

Without the quadratic terms $\|x - \hat{x}\|^2 + \|z - \hat{z}\|^2$, the above method is the classical method of multipliers, see e.g. [2], which can be obtained by applying the proximal point algorithm to the dual of (P), [22]. As we shall see later, the additional quadratic terms will be of fundamental importance in the algorithm developed here.

The proximal method of multipliers is globally convergent under very mild assumptions on the problem's data, namely, convexity and existence of optimal solutions are enough to guarantee convergence. The principal disadvantage of this method when used in the context of separable problems like (P) is the presence of the expression $\|Ax-z\|^2$ in the augmented Lagrangian L_λ, which destroys the separability between x and z, since they are linked by the cross product term $z'Ax$, and thus prevent to minimize separately in x and z the augmented Lagrangian L_λ. This has been recognized as one of the major drawbacks of the augmented Lagrangian approach when applied to separable convex programs, and a number of strategies have been proposed and extensively studied in the literature for removing this difficulty, see e.g., [3, 5–8, 10, 13, 14, 16, 18, 19, 24–27].

In early 1958, Uzawa [28] suggested to simply minimize the Lagrangian function $L(x, z, y)$ with respect to x and z (with y fixed), thus preserving the separability in x, z, and then update the multiplier by the iteration (1.6). Uzawa's method is in fact a gradient method applied to the dual problem of (P), see e.g. [20]. The convergence of the method is guaranteed, if both f and g are strongly convex. This is a quite restrictive assumption, ruling out its potential applications for many interesting problems arising in application.

An important approach for eliminating the difficulties associated with the nonseparability of L_λ is a family of methods based on the *alternating direction method* of multipliers proposed by Gabay and Mercier [9], and Glowinski and Marrocco [11]. The idea of this approach is to alternate the minimization with respect to x and z, and is patterned after splitting methods used in numerical analysis, such as Douglas-Rachford, see e.g. [18]. At each iteration of this algorithm, the augmented Lagrangian

$$\mathscr{A}(x, z, y) = L(x, z, y) + \tfrac{1}{2}\lambda\|Ax-z\|^2$$

is first minimized with respect to x with z, y held fixed, then with respect to z and followed by an update of the multiplier y as in (1.6). These methods converge under some assumptions which will be discussed later in Section 5. A closely related method is the partial inverse of a monotone operator approach developed by Spingarn [24]. For further details and related works, we refer to [8, 10] and references therein, and [27].

Most recently, Eckstein and Bertsekas [4] developed a unified framework via monotone operator theory, allowing them to show that alternating type direction methods as well as a variety of other convex programming algorithms are in fact special cases of the proximal point algorithm, and thus demonstrating the power and versatility of proximal methods in the analysis and development of decomposition algorithms for convex optimization problems.

The algorithm developed in this paper is another manifestation of the proximal methodology for solving convex programs with separable structure. The decomposition method presented here will preserve the good features of the proximal method of multipliers; it globally converges to an optimal primal–dual solution under very mild assumptions, but will remove the difficulty of nonseparability associated with the augmented Lagrangian L_λ. Our method differs from the alternating direction methods. At each iteration of the algorithm, one replaces the minimization of the function (1.5) by the minimization of

$$L(x, z, \bar{w}) + (1/(2\lambda))(\|x-\bar{x}\|^2 + \|z-\bar{z}\|^2) \tag{1.7}$$

followed by the multiplier update (1.6). Here, $\bar{w}$ is an ''estimate'' of the multiplier obtained via a simple update rule similar to the one given by (1.6). Roughly speaking, our method consists of computing *two* proximal steps for the dual problem associated with (P), and one proximal step for the primal problem. The additional dual proximal step used to compute $\bar{w}$, allows for preserving the separability of the problem. A detailed description and motivation of the algorithm is given in Section 2. In Section 3, we prove global convergence of the method to an optimal solution of (P) and an optimal Lagrangian multiplier, under the same modest assumptions used in the proximal method of multipliers. The convergence is proved for an inexact version of the algorithm, where the minimization step is performed approximately according to a given stopping rule. The convergence rate analysis of the method is given in Section 4, where it is shown that the primal–dual sequence converges at a linear rate with a ratio given explicitly in terms of the problem's data. Our algorithm is then compared with alternating direction type methods in Section 5, and is applied to the minimization of a convex function over the intersection of a finite number of closed convex sets to obtain a new highly parallelizable algorithm for these problems. For notations or definitions of concepts not explicitly given in the paper, the reader is referred to Rockafellar's book [21].

2. A proximal-based decomposition method

To motivate the decomposition algorithm described below, let us first recall the proximal point algorithm when applied to find the minimizer of a closed proper convex function F on $\mathbb{R}^n$. In this case, starting from an arbitrary point $u^0 \in \mathbb{R}^n$, the iterative scheme of the proximal point algorithm is

$$u^{k+1} = (I + \lambda_k \partial F)^{-1}(u^k)$$

$$\Leftrightarrow \quad u^{k+1} = \arg\min\{F(u) + (1/(2\lambda_k))\|u-u^k\|^2\} \tag{2.1}$$

$$\Leftrightarrow \quad \frac{u^k - u^{k+1}}{\lambda_k} \in \partial F(u^{k+1}) \tag{2.2}$$

where, ∂F denotes the subdifferential of F. It is well known, see e.g. Lemaire [17], that the above iteration can be seen as an *implicit* discretization scheme for the differential inclusion

$$\begin{cases} -\mathrm{d}u/\mathrm{d}t \in \partial F(u), \quad t > 0, \\ u(0) = u^0. \end{cases} \tag{2.3}$$

Similarly, an *explicit* discretization scheme of (2.3) leads to the gradient method

$$(u^k - u^{k+1})/\lambda_k \in \partial F(u^k). \tag{2.4}$$

Let us now return to problem (P). As described in the introduction, in the proximal method of multipliers, one has to minimize the augmented Lagrangian L_λ given in (1.5) with respect to x and z, followed by the update rule for the multipliers (1.6). The minimization step is just the proximal point algorithm applied to the penalized Lagrangian $L(x, z, y^k) + \frac{1}{2}\lambda_k\|Ax-z\|^2$. Formally, using (2.2), one thus obtains the iterative scheme

$$(x^k-x^{k+1})/\lambda_k \in \partial f(x^{k+1}) + A^{\mathrm{T}}(y^k+\lambda_k(Ax^{k+1}-z^{k+1})), \tag{2.5}$$

$$(z^k-z^{k+1})/\lambda_k \in \partial g(z^{k+1}) - (y^k+\lambda_k(Ax^{k+1}-z^{k+1})). \tag{2.6}$$

The difficulty with the above scheme is that the iterates (x^{k+1}, z^{k+1}) are coupled by the implicit computation of the *constraints* given by the expression $(Ax^{k+1}-z^{k+1})$, and therefore the scheme cannot compute x^{k+1}, z^{k+1} separately. To remove this difficulty, we suggest to perform an explicit computation of the constraints expression in (2.5)–(2.6), which is thus replaced by the iterative scheme

$$(x^k-x^{k+1})/\lambda_k \in \partial f(x^{k+1}) + A^{\mathrm{T}}(y^k+\lambda_k(Ax^k-z^k)), \tag{2.7}$$

$$(z^k-z^{k+1})/\lambda_k \in \partial g(z^{k+1}) - (y^k+\lambda_k(Ax^k-z^k)). \tag{2.8}$$

The above scheme can thus be interpreted as a *partial implicit–explicit method*: implicit for the objectives and explicit for the constraints. Defining

$$p^{k+1} = y^k+\lambda_k(Ax^k-z^k) \tag{2.9}$$

and using (2.1), one can now rewrite (2.7)–(2.8) in the equivalent forms

$$x^{k+1} = \arg\min\{f(x) + \langle p^{k+1}, Ax\rangle + (1/(2\lambda_k))\|x-x^k\|^2\}, \tag{2.10}$$

$$z^{k+1} = \arg\min\{g(z) - \langle p^{k+1}, z\rangle + (1/(2\lambda_k))\|z-z^k\|^2\}. \tag{2.11}$$

Thus, the minimization of the augmented Lagrangian L_λ has now been decoupled in two separate minimization of *strongly convex functions* in the variables x and z. Finally, to complete one iteration of the proximal method of multipliers, the multipliers are updated according to the iteration

$$y^{k+1} = y^k+\lambda_k(Ax^{k+1}-z^{k+1}). \tag{2.12}$$

The iterations (2.9)–(2.12) are the decomposition algorithm we propose for solving problem (P). Note that the iteration (2.9) is exactly the same as the multiplier update rule (2.12), except that the constraints expression $Ax-z$ is now evaluated at the current iterate in (2.9).

As observed by Rockafellar [21], the Lagrangian dual multiplier update rule is nothing else but computing a proximal (maximization) iterate for the dual variable y, i.e.

$$y^{k+1} = \arg\max\{\langle y, Ax^{k+1}-z^{k+1}\rangle - (1/(2\lambda_k))\|y-y^k\|^2\}. \tag{2.13}$$

A similar computation will thus produce p^{k+1} by replacing $Ax^{k+1}-z^{k+1}$ with Ax^k-z^k in (2.13). Recalling the definition of the classical Lagrangian associated with (P), the four steps (2.9)–(2.12) of the algorithm can thus be written as

$$p^{k+1}=\arg\max\{L(x^k, z^k, y)-(1/(2\lambda_k))\|y-y^k\|^2\}, \tag{2.14}$$

$$x^{k+1}=\arg\min\{L(x, z^k, p^{k+1})+(1/(2\lambda_k))\|x-x^k\|^2\}, \tag{2.15}$$

$$z^{k+1}=\arg\min\{L(x^k, z, p^{k+1})+(1/(2\lambda_k))\|z-z^k\|^2\}, \tag{2.16}$$

$$y^{k+1}=\arg\max\{L(x^{k+1}, z^{k+1}, y)-(1/(2\lambda_k))\|y-y^k\|^2\}. \tag{2.17}$$

Therefore, our algorithm can be seen as performing two proximal steps in the dual variables, the predictor step p^{k+1} and the corrector step y^{k+1}, combined with a primal proximal step which is now separable in x and z. For ease of reference, we call this algorithm a *predictor corrector proximal multiplier* method, (PCPM).

It is interesting to note that the decoupling of the variables x and z can also be obtained by using a "kind" of linearization of the quadratic penalty term $\|Ax-z\|^2$; i.e. by linearizing the squared Euclidean norm $\frac{1}{2}\|\cdot\|^2$ at the current kth iterate. This strategy was proposed by Stephanopoulos and Westerberg [25] to construct separable classical augmented Lagrangian methods for nonconvex problems arising in engineering system optimization. A similar approach was further studied in [5]. None of these works considered adding the proximal quadratic terms. In [25] no convergence results are given, and none of our convergence analysis results, however, can be obtained from the results of [5].

As we shall see in the next sections, the interpretation of our algorithm via the proximal point method plays a central role in the convergence analysis of the PCPM method, and will demonstrate the benefits of the proximal framework.

3. Convergence of the PCPM

Consider the convex problem (P) defined in the introduction.

$$\text{(P)} \qquad \min\{f(x)+g(z): Ax=z\}. \tag{3.1}$$

The Lagrangian associated with (P) is defined by

$$L(x, z, y)=f(x)+g(z)+\langle y, Ax-z\rangle \tag{3.2}$$

and the dual of problem (P) is given by

$$\text{(D)} \qquad \max\left\{d(y):=\inf_{x,z} L(x, z, y)=-f^*(-A^{\mathrm{T}}y)-g^*(y)\right\} \tag{3.3}$$

where f^*, g^* are the conjugate functions of f, g respectively, see e.g. [21]. Throughout the rest of this paper, we make the following assumption.

Assumption A. There exist (x^*, z^*) and y^* optimal solutions for problems (P) and (D) respectively, i.e. there exists (x^*, z^*, y^*) which is a saddle point of L,

$$L(x^*, z^*, y)\leqslant L(x^*, z^*, y^*)\leqslant L(x, z, y^*) \quad \forall (x, z)\in\mathbb{R}^{n\times m},\ \forall y\in\mathbb{R}^m. \tag{3.4}$$

Note that since (P) is a convex program, the existence of an optimal dual Lagrange multiplier y^* associated with the constraint $Ax = z$ is guaranteed if a constraint qualification holds for problem (P) (see [21]), namely if,

$$\text{there exist } x^* \in \text{ri}(\text{dom} f) \text{ and } z^* \in \text{ri}(\text{dom } g) \text{ satisfying } Ax^* = z^*. \tag{3.5}$$

The PCPM algorithm that we suggest for solving (P) allows for approximate minimization with respect to x and z. In the sequel, we use the notation,

$$\varepsilon\text{-arg min } F(u) = \{v\colon F(v) \leqslant \inf F + \varepsilon\}$$

where F is a given function and $\varepsilon \geqslant 0$.

Algorithm I (*inexact*). Starting with an initial arbitrary triple $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, a sequence $(x^k, z^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, $k \geqslant 0$, is successively generated by the following steps:

Step 1. Compute

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

Step 2. Solve

$$x^{k+1} = \alpha_k\text{-arg min}\{f(x) + \langle p^{k+1}, Ax\rangle + (1/(2\lambda_k)) \|x - x^k\|^2\},$$

$$z^{k+1} = \beta_k\text{-arg min}\{g(z) - \langle p^{k+1}, z\rangle + (1/(2\lambda_k)) \|z - z^k\|^2\}.$$

Step 3. Compute

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

Here, $\{\alpha_k\}$ and $\{\beta_k\}$ are sequences such that

$$\forall k, \quad \alpha_k, \beta_k \geqslant 0, \qquad \sum_{k=0}^{\infty} \sqrt{\alpha_k} < \infty, \qquad \sum_{k=0}^{\infty} \sqrt{\beta_k} < \infty,$$

and $\{\lambda_k\}$ is a sequence of positive scalars, to be specified later.

When $\alpha_k = \beta_k = 0$, we obtain the algorithm with exact minimization. We will denote the exact minimizers by $\bar{x}^{k+1}$, $\bar{z}^{k+1}$.

The convergence of the PCPM algorithm strongly relies on its proximal nature. The following result gives two basic estimates which will be used in the convergence analysis. These estimates have been shown in Auslender [1]. For completeness, we include here a short proof. We would like to stress that the estimate given below in Lemma 3.1(i), while simple, is powerful in its implications on the convergence analysis of proximal methods, see e.g., the elegant new convergence results obtained by Guler [12].

Let $F\colon \mathbb{R}^n \mapsto (-\infty, +\infty]$ be a closed proper convex function and define

$$\bar{u}^{k+1} = \text{arg min}\{F(u) + (1/(2\lambda_k)) \|u - u^k\|^2\},$$

$$u^{k+1}=\varepsilon_k\text{-arg min}\{F(u)+(1/(2\lambda_k))\|u-u^k\|^2\}. \tag{3.6}$$

Lemma 3.1. *For any* $k\geqslant 0$,

$$\text{(i)}\quad 2\lambda_k[F(\bar{u}^{k+1})-F(u)]\leqslant(\|u^k-u\|^2-\|\bar{u}^{k+1}-u\|^2-\|\bar{u}^{k+1}-u^k\|^2)\quad \forall u\in\mathbb{R}^n;$$

$$\text{(ii)}\quad \|\bar{u}^{k+1}-u^{k+1}\|\leqslant\sqrt{2\varepsilon_k\lambda_k}. \tag{3.7}$$

Proof. Let $\psi_k(u):=F(u)+(1/(2\lambda_k))\|u-u^k\|^2$. By the definition of $\bar{u}^{k+1}$ we have $0\in\partial\psi_k(\bar{u}^{k+1})$, since ψ_k is strongly convex with modulus $1/\lambda_k$ (see, Rockafellar [23, Proposition 6]), it follows that

$$2\lambda_k[\psi_k(u)-\psi_k(\bar{u}^{k+1})]\geqslant\|\bar{u}^{k+1}-u\|^2\quad\forall u$$

and (i) is proved. By the definition of u^{k+1} we have $\psi_k(v)-\psi_k(u^{k+1})\geqslant-\varepsilon_k\ \forall v$. Setting $u=u^{k+1}$ and $v=\bar{u}^{k+1}$ in the above two inequalities respectively, and adding, we obtain (ii). □

Remark 3.1. As noticed by Auslender [1], with $\varepsilon'_k:=\sqrt{2\varepsilon_k\lambda_k}$ the scheme (3.6) implies Lemma 3.1 (ii) and then one recover the approximate criterion A suggested by Rockafellar [22, p. 100]. Note that here as in Rockafellar, we assume that $\sum_{k=0}^{\infty}\varepsilon'_k<\infty$ (cf. the sequences $\{\alpha_k\}$ and $\{\beta_k\}$ defined above).

We define

$$\begin{aligned}\bar{y}^{k+1}&=y^k+\lambda_k(A\bar{x}^{k+1}-\bar{z}^{k+1})\\&=\arg\min\{-L(\bar{x}^{k+1},\bar{z}^{k+1},y)+(1/(2\lambda_k))\|y-y^k\|^2\}\end{aligned} \tag{3.8}$$

(see (2.13)). In the next result, we establish two fundamental estimates relating the exact and inexact iterates from an optimal solution.

Proposition 3.1. *For all* $k\geqslant 0$,

$$\begin{aligned}\text{(i)}\quad \|\bar{x}^{k+1}-x^*\|^2+\|\bar{z}^{k+1}-z^*\|^2\leqslant{}&\|x^k-x^*\|^2+\|z^k-z^*\|^2\\&-2\lambda_k\langle p^{k+1}-y^*,A\bar{x}^{k+1}-\bar{z}^{k+1}\rangle\\&-\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2\};\end{aligned}$$

$$\begin{aligned}\text{(ii)}\quad \|\bar{y}^{k+1}-y^*\|^2\leqslant{}&\|y^k-y^*\|^2-\{\|p^{k+1}-\bar{y}^{k+1}\|^2+\|p^{k+1}-y^k\|^2\}\\&-2\lambda_k\{\langle y^*-\bar{y}^{k+1},A\bar{x}^{k+1}-\bar{z}^{k+1}\rangle+\langle\bar{y}^{k+1}-p^{k+1},Ax^k-z^k\rangle\}.\end{aligned}$$

Proof. From Step 2 of Algorithm I, the sequences $\{\bar{x}^k\}$, $\{\bar{z}^k\}$ are obtained by applying the exact proximal point algorithm (i.e. $\alpha_k=\beta_k=0$) to the separable Lagrangian in (x,z),

$$(x,z)\mapsto L(x,z,p^{k+1})=f(x)+g(z)+\langle p^{k+1},Ax-z\rangle. \tag{3.9}$$

Applying Lemma 3.1 (i) with the choice $F := L$ at the optimal point $x = x^*$ and $z = z^*$, we obtain

$$2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})]$$
$$\leqslant \|x^k - x^*\|^2 + \|z^k - z^*\|^2 - \{\|\bar{x}^{k+1} - x^*\|^2 + \|\bar{z}^{k+1} - z^*\|^2\}$$
$$- \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\}.$$

Since (x^*, z^*, y^*) is a saddle point for $L(x, z, y)$, we also have

$$2\lambda_k[L(x^*, z^*, p^{k+1}) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*)] \leqslant 0.$$

Adding the above two inequalities, we obtain (i) after rearranging terms. Similarly, since the sequences $\{p^{k+1}\}$, $\{\bar{y}^{k+1}\}$ are obtained by applying the proximal point algorithm to $F(y) := -L(x^k, z^k, y)$ and $F(y) := -L(\bar{x}^{k+1}, \bar{z}^{k+1}, y)$ respectively, applying once again Lemma 3.1 (i) at $y = \bar{y}^{k+1}$, $y = y^*$ respectively, we obtain

$$2\lambda_k[L(x^k, z^k, \bar{y}^{k+1}) - L(x^k, z^k, p^{k+1})]$$
$$\leqslant \|y^k - \bar{y}^{k+1}\|^2 - \|p^{k+1} - \bar{y}^{k+1}\|^2 - \|p^{k+1} - y^k\|^2,$$
$$2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})]$$
$$\leqslant \|y^k - y^*\|^2 - \|\bar{y}^{k+1} - y^*\|^2 - \|\bar{y}^{k+1} - y^k\|^2.$$

Adding the above two inequalities, and rearranging terms we obtain (ii). □

We can now state and prove our convergence result.

Theorem 3.1. *Consider problem* (P) *for which Assumption* A *holds, that is there exists a primal–dual optimal solution* (x^*, z^*, y^*). *Let* $\{x^k, z^k, y^k\}$ *be the sequence generated by Algorithm* I. *If* $\{\lambda_k\}$ *satisfies*

$$\varepsilon \leqslant \lambda_k \leqslant \min\left(\frac{1-\varepsilon}{2}, \frac{1-\varepsilon}{2\|A\|}\right) \quad \forall k \geqslant 0 \tag{3.10}$$

for some $0 < \varepsilon \leqslant \min(\frac{1}{3}, 1/(2\|A\| + 1))$, *then* $\{x^k\}$ *converges to* x^*, $\{z^k\}$ *converges to* Ax^* *and* $\{y^k\}$ *converges to* y^*.

Proof. We denote $w = (x, y, z)$ with the associated norm $\|w\|^2 = \|x\|^2 + \|z\|^2 + \|y\|^2$. By adding the inequalities (i)–(ii) derived in Proposition 3.1, we obtain

$$\|\bar{w}^{k+1} - w^*\|^2 \leqslant \|w^k - w^*\|^2 - \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\}$$
$$- \{\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} + \rho \tag{3.11}$$

where $\rho := 2\lambda_k\langle \bar{y}^{k+1} - p^{k+1}, A(\bar{x}^{k+1} - x^k) - (\bar{z}^{k+1} - z^k)\rangle$.

Using (3.8) and Step 1 of Algorithm I, we have

$$\rho = 2\lambda_k^2 \|A(\bar{x}^{k+1}-x^k)-(\bar{z}^{k+1}-z^k)\|^2$$

$$\leqslant 4\lambda_k^2\|A\|^2\|\bar{x}^{k+1}-x^k\|^2+4\lambda_k^2\|\bar{z}^{k+1}-z^k\|^2. \tag{3.12}$$

Combining (3.11) with inequality (3.12), it follows that

$$\|\bar{w}^{k+1}-w^*\|^2$$
$$\leqslant \|w^k-w^*\|^2-(1-4\lambda_k^2\|A\|^2)\|\bar{x}^{k+1}-x^k\|^2-(1-4\lambda_k^2)\|\bar{z}^{k+1}-z^k\|^2$$
$$-\{\|p^{k+1}-\bar{y}^{k+1}\|^2+\|p^{k+1}-y^k\|^2\}. \tag{3.13}$$

Since by (3.10) we assumed that $\varepsilon\leqslant\lambda_k\leqslant\min(\frac{1}{2}(1-\varepsilon),\ (1-\varepsilon)/(2\|A\|))$, we obtain from (3.13),

$$\|\bar{w}^{k+1}-w^*\|^2\leqslant\|w^k-w^*\|^2-\varepsilon\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2$$
$$+\|p^{k+1}-\bar{y}^{k+1}\|^2+\|p^{k+1}-y^k\|^2\}. \tag{3.14}$$

Now we need to find an estimate for $\|w^{k+1}-w^*\|$. By the triangle inequality and (3.14), we obtain

$$\|w^{k+1}-w^*\|\leqslant\|\bar{w}^{k+1}-w^{k+1}\|+\|\bar{w}^{k+1}-w^*\|\leqslant\|\bar{w}^{k+1}-w^{k+1}\|+\|w^k-w^*\|. \tag{3.15}$$

We also have,

$$\|\bar{w}^{k+1}-w^{k+1}\|^2=\|\bar{x}^{k+1}-x^{k+1}\|^2+\|\bar{z}^{k+1}-z^{k+1}\|^2+\|\bar{y}^{k+1}-y^{k+1}\|^2$$
$$\leqslant(1+2\lambda_k^2\|A\|^2)\|\bar{x}^{k+1}-x^{k+1}\|^2+(1+2\lambda_k^2)\|\bar{z}^{k+1}-z^{k+1}\|^2 \tag{3.16}$$

by using the definitions of y^{k+1}, $\bar{y}^{k+1}$ given respectively in Step 3 of Algorithm I and (3.8).

By Lemma 3.1 (ii), we have

$$\|\bar{x}^{k+1}-x^{k+1}\|^2\leqslant2\lambda_k\alpha_k,\qquad \|\bar{z}^{k+1}-z^{k+1}\|^2\leqslant2\lambda_k\beta_k. \tag{3.17}$$

Combining (3.17) and (3.16), we then obtain

$$\|\bar{w}^{k+1}-w^{k+1}\|\leqslant\sqrt{2\lambda_k\alpha_k(1+2\lambda_k^2\|A\|^2)+2\lambda_k\beta_k(1+2\lambda_k^2)}$$
$$\leqslant c_1\sqrt{\alpha_k}+c_2\sqrt{\beta_k} \tag{3.18}$$

where $0\leqslant c_1, c_2<\infty$, since by the assumption (3.10), $\{\lambda_k\}$ is bounded.

Using (3.18) in the most left-hand part of inequality (3.15), we then have

$$\|w^{k+1}-w^*\|\leqslant\|w^k-w^*\|+c_1\sqrt{\alpha_k}+c_2\sqrt{\beta_k} \tag{3.19}$$

which because of $\sum_{k=0}^{\infty}\sqrt{\alpha_k}<\infty$ and $\sum_{k=0}^{\infty}\sqrt{\beta_k}<\infty$ implies that $\{w^k\}$ is bounded, and the existence of

$$\lim_{k\to\infty}\|w^k-w^*\|=\mu<\infty. \tag{3.20}$$

From (3.18), we also have under our assumptions for $\{\alpha_k\}$ and $\{\beta_k\}$ that $\|\bar{w}^{k+1}-w^{k+1}\|\to 0$ and therefore using (3.15), we obtain

$$\lim_{k\to\infty} \|\bar{w}^{k+1}-w^*\| = \mu < \infty.$$

Therefore, by taking the limits on both sides of (3.14), we obtain (since $\varepsilon > 0$)

$$\begin{aligned} &\|\bar{x}^{k+1}-x^k\|\to 0, \quad \|\bar{z}^{k+1}-z^k\|\to 0, \\ &\|p^{k+1}-\bar{y}^{k+1}\|\to 0, \quad \|p^{k+1}-y^k\|\to 0. \end{aligned} \tag{3.21}$$

Since $\{w^k\}$ is bounded, there exists a limit point w^∞ i.e. there exists a subsequence $\{w^{k_j}=(x^{k_j}, z^{k_j}, y^{k_j})\}\to w^\infty = (x^\infty, z^\infty, y^\infty)$. We now show that w^∞ is a saddle point of $L(x, z, y)$.

Applying Lemma 3.1(i) to Step 2 of Algorithm I at any fixed x and z, we have

$$\begin{aligned} &2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, p^{k+1}) - L(x, z, p^{k+1})] \\ &\quad \leqslant \|x^k-x\|^2 + \|z^k-z\|^2 - \{\|\bar{x}^{k+1}-x\|^2 + \|\bar{z}^{k+1}-z\|^2\} \\ &\quad\quad - \{\|\bar{x}^{k+1}-x^k\|^2 + \|\bar{z}^{k+1}-z^k\|^2\}. \end{aligned} \tag{3.22}$$

Taking the limit over the appropriate subsequences on both sides of (3.22) and using (3.21), we have

$$L(x^\infty, z^\infty, y^\infty) - L(x, z, y^\infty) \leqslant 0 \tag{3.23}$$

since λ_k is bounded below by $\varepsilon > 0$.

Similarly, by applying Lemma 3.1(i) to (3.8) at any fixed y, we obtain

$$\begin{aligned} &2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, y) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})] \\ &\quad \leqslant \|y^k-y\|^2 - \|\bar{y}^{k+1}-y\|^2 - \|\bar{y}^{k+1}-y^k\|^2 \\ &\quad \leqslant \|y^k-y\|^2 - \|\bar{y}^{k+1}-y\|^2. \end{aligned} \tag{3.24}$$

Taking the limit over the appropriate subsequences on both sides of (3.24), we then have

$$L(x^\infty, z^\infty, y) - L(x^\infty, z^\infty, y^\infty) \leqslant 0. \tag{3.25}$$

Combining (3.23) and (3.25), which hold for any (x, z, y), we have thus shown that $(x^\infty, z^\infty, y^\infty)$ is a saddle point of $L(x, z, y)$.

To complete the proof, it remains to show that $\{w_k\}$ has a unique limit point. Let w_1^∞ and w_2^∞ be any two limit points of $\{w_k\}$. As shown above, both are saddle points of $L(x, z, y)$, and hence

$$\lim_{k\to\infty} \|w^k - w_i^\infty\| = \mu_i < \infty, \quad i=1, 2.$$

Now we can use the same argument as given in [23, p. 885]. Writing

$$\|w^k - w_1^\infty\|^2 - \|w^k - w_2^\infty\|^2 = -2\langle w^k, w_1^\infty - w_2^\infty\rangle + \|w_1^\infty\|^2 - \|w_2^\infty\|^2$$

and passing to the limit we obtain for each limit point w_1^∞, w_2^∞ of $\{w_k\}$,

$$\begin{aligned}\mu_1^2 - \mu_2^2 &= -2\langle w_1^\infty, w_1^\infty - w_2^\infty\rangle + \|w_1^\infty\|^2 - \|w_2^\infty\|^2 = -\|w_1^\infty - w_2^\infty\|^2\\ &= -2\langle w_2^\infty, w_1^\infty - w_2^\infty\rangle + \|w_1^\infty\|^2 - \|w_2^\infty\|^2 = \|w_1^\infty - w_2^\infty\|^2.\end{aligned}$$

Therefore, we must have $\|w_1^\infty - w_2^\infty\| = 0$ and hence w^∞ is unique. □

4. Rate of convergence

In this section we consider the rate at which the sequence $\{x^k, z^k, y^k\}$ generated by Algorithm I converges to a saddle point of the Lagrangian $L(x, z, y)$. To prove this result, we need some further assumptions on the problem's data. Let T be a set valued maximal monotone operator on $\mathbb{R}^n$. Following Rockafellar [23], we say that the mapping T^{-1} is Lipschitz continuous at the origin with modulus $a \geqslant 0$, if there exists a unique solution $\bar{u}$ such that $0 \in T(\bar{u})$ and for some $\tau > 0$, we have $\|u - \bar{u}\| \leqslant a\|v\|$, whenever $v \in T(u)$ and $\|v\| \leqslant \tau$.

Recall that (cf. (1.3))

$$S(x, z, y) = \partial_{x,z} L(x, z, y) \times \partial_y(-L(x, z, y)) \tag{4.1}$$

is a maximal monotone operator on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and that solving problem (P) is equivalent to finding a zero of S.

Assumption B. S^{-1} is Lipschitz continuous at the origin with modulus $a \geqslant 0$.

Note that for problem (P) we have

$$S^{-1}(v_1, v_2, v_3) = \arg\min_{x,z}\max_{y}\{L(x, z, y) - \langle x, v_1\rangle - \langle z, v_2\rangle + \langle y, v_3\rangle\}$$

and therefore Assumption B can be interpreted in terms of the problem's data as: there exists a unique saddle point w^* such that for some $\tau > 0$, we have $\|w - w^*\| \leqslant a\|v\|$, whenever $\|v\| \leqslant \tau$ and $w := (x, y, z) \in S^{-1}(v_1, v_2, v_3)$.

Assumption C. The sequence $\{x^k, z^k, y^k\}$ is generated by Algorithm I under the approximate criterion

$$\|x^{k+1} - \bar{x}^{k+1}\| \leqslant \eta_k \|x^{k+1} - x^k\|, \qquad \|z^{k+1} - \bar{z}^{k+1}\| \leqslant \eta_k \|z^{k+1} - z^k\|, \tag{4.2}$$

where $\eta_k \geqslant 0$ and $\sum_0^\infty \eta_k < \infty$.

Note that the use of the same η_k for the approximation criterion (4.2) is just to simplify notation in the analysis below. In fact if one chooses different sequences $\eta_k^i \geqslant 0$, $\sum_0^\infty \eta_k^i < \infty$, $i = 1, 2$, then one should simply define $\eta_k = \max(\eta_k^1, \eta_k^2)$ in (4.2).

The assumptions B and C were suggested by Rockafellar [22, p. 100], to derive the rate of convergence of the proximal method of multipliers, and will also be used here to derive the rate of convergence for Algorithm I. Note however that the additional proximal dual step in Algorithm I precludes a direct application of the results developed in [22] and required a specific analysis (see also Remark 4.1 below).

Before proving our convergent rate result, we need the following lemma.

Lemma 4.1. *If the Assumption* C *holds, then*

$$\|w^{k+1}-\bar{w}^{k+1}\| \leqslant \delta_k \|w^{k+1}-w^k\| \tag{4.3}$$

where $\delta_k = \eta_k \max(\sqrt{1+2\lambda_k^2 \|A\|^2}, \sqrt{1+2\lambda_k^2})$.

Proof. By (2.12), (3.8) and (4.2), we have

$$\begin{aligned} \|y^{k+1}-\bar{y}^{k+1}\| &= \lambda_k \|A(x^{k+1}-\bar{x}^{k+1}) - (z^{k+1}-\bar{z}^{k+1})\| \\ &\leqslant \eta_k\lambda_k(\|A\|\,\|x^{k+1}-x^k\| + \|z^{k+1}-z^k\|). \end{aligned} \tag{4.4}$$

Using the inequality $(r+q)^2 \leqslant 2(r^2+q^2)$, in (4.4) we obtain

$$\|y^{k+1}-\bar{y}^{k+1}\|^2 \leqslant 2(\eta_k\lambda_k)^2(\|A\|^2\|x^{k+1}-x^k\|^2 + \|z^{k+1}-z^k\|^2). \tag{4.5}$$

Therefore from (4.2), (4.5) and the definition of δ_k given in the lemma,

$$\begin{aligned} \|w^{k+1}-\bar{w}^{k+1}\|^2 &= \|x^{k+1}-\bar{x}^{k+1}\|^2 + \|z^{k+1}-\bar{z}^{k+1}\|^2 + \|y^{k+1}-\bar{y}^{k+1}\|^2 \\ &\leqslant \eta_k^2(\|x^{k+1}-x^k\|^2 + \|z^{k+1}-z^k\|^2) \\ &\quad + 2(\eta_k\lambda_k)^2(\|A\|^2\|x^{k+1}-x^k\|^2 + \|z^{k+1}-z^k\|^2) \\ &\leqslant \delta_k^2(\|x^{k+1}-x^k\|^2 + \|z^{k+1}-z^k\|^2) \\ &\leqslant \delta_k^2\|w^{k+1}-w^k\|^2. \qquad \square \end{aligned}$$

We can now state and prove our convergent rate result.

Theorem 4.1. *Consider problem* (P) *for which Assumptions* A, B *and* C *hold. Let* $\{x^k, z^k, y^k\}$ *be a bounded sequence generated by Algorithm* I, *and let* $\{\lambda_k\}$ *satisfies* (3.10). *Then,* $w^k = \{x^k, z^k, y^k\}$ *converges linearly to the unique optimal solution* $w^* = (x^*, z^*, y^*)$, *that is, there exists an integer* $\bar{k}$ *such that, for all* $k \geqslant \bar{k}$,

$$\|w^{k+1}-w^*\| \leqslant \theta_k \|w^k - w^*\|, \tag{4.6}$$

where $1 > (\sqrt{4a^2+\varepsilon^3}+2a)/(2\sqrt{4a^2+\varepsilon^3}) \geqslant \theta_k$, *for all* $k \geqslant \bar{k}$.

Proof. Under our assumptions, w^k is bounded and hence from (3.17) with the choice

$$\alpha_k = \eta_k^2 \frac{\|x^{k+1}-x^k\|^2}{2\lambda_k}, \qquad \beta_k = \eta_k^2 \frac{\|z^{k+1}-z^k\|^2}{2\lambda_k}$$

(recall that $\lambda_k > \varepsilon > 0$), Theorem 3.1 holds and w^k converges to w^* by Theorem 3.1. We now establish the rate of convergence. From Step 2 of Algorithm I, the sequences $\{\bar{x}^k\}$, $\{\bar{z}^k\}$ are obtained by applying the exact proximal point algorithm (i.e. with $\alpha_k = \beta_k = 0$), to the Lagrangian given in (3.9) and thus we obtain

$$0 \in \partial f(\bar{x}^{k+1}) + A^{\mathrm{T}} p^{k+1} + (\bar{x}^{k+1} - x^k)/\lambda_k$$
$$= \partial f(\bar{x}^{k+1}) + A^{\mathrm{T}} \bar{y}^{k+1} - A^{\mathrm{T}}(\bar{y}^{k+1} - p^{k+1}) + (\bar{x}^{k+1} - x^k)/\lambda_k, \tag{4.7}$$
$$0 \in \partial g(\bar{z}^{k+1}) - p^{k+1} + (\bar{z}^{k+1} - z^k)/\lambda_k$$
$$= \partial g(\bar{z}^{k+1}) - \bar{y}^{k+1} + \bar{y}^{k+1} - p^{k+1} + (\bar{z}^{k+1} - z^k)/\lambda_k. \tag{4.8}$$

From (3.8), we have $\bar{y}^{k+1} = y^k + \lambda_k(A\bar{x}^{k+1} - \bar{z}^{k+1})$ which we rewrite as

$$-(\bar{y}^{k+1} - y^k)/\lambda_k = \bar{z}^{k+1} - A\bar{x}^{k+1}. \tag{4.9}$$

Since

$$\partial_x L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) = \partial f(\bar{x}^{k+1}) + A^{\mathrm{T}} \bar{y}^{k+1},$$
$$\partial_z L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) = \partial g(\bar{z}^{k+1}) - \bar{y}^{k+1},$$
$$\partial_y(-L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})) = \{\bar{z}^{k+1} - A\bar{x}^{k+1}\},$$

from (4.7)–(4.9) and the definition of S (cf. (4.1)), by rearranging terms we obtain

$$(\pi_k, \mu_k, \gamma_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) \tag{4.10}$$

where

$$\pi_k := A^{\mathrm{T}}(\bar{y}^{k+1} - p^{k+1}) - (\bar{x}^{k+1} - x^k)/\lambda_k, \tag{4.11}$$
$$\mu_k := -(\bar{y}^{k+1} - p^{k+1}) - (\bar{z}^{k+1} - z^k)/\lambda_k, \tag{4.12}$$
$$\gamma_k := -(\bar{y}^{k+1} - y^k)/\lambda_k. \tag{4.13}$$

Recall that from (2.9), $p^{k+1} = y^k + \lambda_k(Ax^k - z^k)$, then by subtracting the latter from (3.8), we have

$$\bar{y}^{k+1} - p^{k+1} = \lambda_k[A(\bar{x}^{k+1} - x^k) - (\bar{z}^{k+1} - z^k)]. \tag{4.14}$$

Substituting (4.14) in (4.11)–(4.13), we obtain

$$\pi_k = \lambda_k A^{\mathrm{T}}[A(\bar{x}^{k+1} - x^k) - (\bar{z}^{k+1} - z^k)] - (\bar{x}^{k+1} - x^k)/\lambda_k$$
$$= (\lambda_k^2 A^{\mathrm{T}} A - I)(\bar{x}^{k+1} - x^k)/\lambda_k - \lambda_k A^{\mathrm{T}}(\bar{z}^{k+1} - z^k),$$
$$\mu_k = -\lambda_k[A(\bar{x}^{k+1} - x^k) - (\bar{z}^{k+1} - z^k)] - (\bar{z}^{k+1} - z^k)/\lambda_k$$
$$= (\lambda_k^2 - 1)(\bar{z}^{k+1} - z^k)/\lambda_k - \lambda_k A(\bar{x}^{k+1} - x^k),$$
$$\gamma_k = -(\bar{y}^{k+1} - y^k)/\lambda_k.$$

Since

$$\varepsilon \leqslant \lambda_k \leqslant \min\left(\frac{1-\varepsilon}{2}, \frac{1-\varepsilon}{2\|A\|}\right),$$

from (3.21), we have $(\pi_k, \mu_k, \gamma_k) \to (0, 0, 0)$. Choose $\hat{k}$ so that $\|(\pi_k, \mu_k, \gamma_k)\| < \tau$ for all $k \geqslant \hat{k}$. Then using Assumption B and the fact $0 \in S(x^*, z^*, y^*)$ with the choice

$$u = (\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}), \qquad \bar{u} = (x^*, z^*, y^*), \qquad v = (\pi_k, \mu_k, \gamma_k)$$

we have

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\| &= \|(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) - (x^*, z^*, y^*)\| \\ &\leqslant a\|(\pi_k, \mu_k, \gamma_k)\| \quad \forall k \geqslant \hat{k}. \end{aligned} \tag{4.15}$$

We now estimate the right-hand side of (4.15). Using the definition of (π_k, μ_k, γ_k) and the inequality $(r+q)^2 \leqslant 2(r^2+q^2)$, we obtain

$$\|\pi_k\|^2 \leqslant \frac{2}{\lambda_k^2}\|\lambda_k^2 A^{\mathrm{T}}A - I\|^2 \|\bar{x}^{k+1} - x^k\|^2 + 2\lambda_k^2\|A\|^2\|\bar{z}^{k+1} - z^k\|^2,$$

$$\|\mu_k\|^2 \leqslant \frac{2}{\lambda_k^2}(\lambda_k^2 - 1)^2\|\bar{z}^{k+1} - z^k\|^2 + 2\lambda_k^2\|A\|^2\|\bar{x}^{k+1} - x^k\|^2,$$

$$\begin{aligned} \|\gamma_k\|^2 &= \frac{1}{\lambda_k^2}\|\bar{y}^{k+1} - y^k\|^2 = \frac{1}{\lambda_k^2}\|(\bar{y}^{k+1} - p^{k+1}) + (p^{k+1} - y^k)\|^2 \\ &\leqslant \frac{2}{\lambda_k^2}(\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2). \end{aligned}$$

Therefore

$$\begin{aligned} \|(\pi_k, \mu_k, \gamma_k)\|^2 &= \|\pi_k\|^2 + \|\mu_k\|^2 + \|\gamma_k\|^2 \\ &\leqslant D\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 \\ &\qquad + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\}, \end{aligned} \tag{4.16}$$

where

$$D := 2\max\left\{\frac{1}{\lambda_k^2}\|\lambda_k^2 A^{\mathrm{T}}A - I\|^2 + \lambda_k^2\|A\|^2, \frac{1}{\lambda_k^2}(\lambda_k^2 - 1)^2 + \lambda_k^2\|A\|^2, \frac{1}{\lambda_k^2}\right\}.$$

By simple algebra, using (3.10) (which implies $\lambda_k \leqslant \frac{1}{2}$ and $\lambda_k \leqslant 1/(2\|A\|)$), one can verify that $D \leqslant 4/\lambda_k^2$. Hence, combining (4.15) and (4.16) we obtain for all $k \geqslant \hat{k}$,

$$\begin{aligned} \lambda_k^2\|\bar{w}^{k+1} - w^*\|^2 &\leqslant 4a^2\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 \\ &\qquad + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\}, \end{aligned} \tag{4.17}$$

which together with (3.14) implies that,

$$4a^2\|\bar{w}^{k+1} - w^*\|^2 + \varepsilon\lambda_k^2\|\bar{w}^{k+1} - w^*\|^2 \leqslant 4a^2\|w^k - w^*\|^2 \quad \forall k \geqslant \hat{k}.$$

Defining $\nu_k := 2a/\sqrt{4a^2+\varepsilon\lambda_k^2}$, the latter inequality reduces to

$$\|\bar{w}^{k+1}-w^*\| \leqslant \nu_k \|w^k-w^*\|, \quad k \geqslant \hat{k}. \tag{4.18}$$

But

$$\|w^{k+1}-w^*\| \leqslant \|w^{k+1}-\bar{w}^{k+1}\| + \|\bar{w}^{k+1}-w^*\|, \tag{4.19}$$

and invoking Lemma 4.1, we have

$$\|w^{k+1}-\bar{w}^{k+1}\| \leqslant \delta_k \|w^{k+1}-w^k\| \leqslant \delta_k \|w^{k+1}-w^*\| + \delta_k \|w^k-w^*\|. \tag{4.20}$$

Therefore using (4.20) and (4.18) in (4.19) we obtain

$$\|w^{k+1}-w^*\| \leqslant \delta_k \|w^{k+1}-w^*\| + \delta_k \|w^k-w^*\| + \nu_k \|w^k-w^*\|, \quad k \geqslant \hat{k},$$

which proved (4.6) with $\theta_k = (\nu_k+\delta_k)/(1-\delta_k)$. Since $\delta_k \to 0$, $\lambda_k > \varepsilon$, and

$$\frac{\sqrt{4a^2+\varepsilon^3}+2a}{2\sqrt{4a^2+\varepsilon^3}} > \frac{2a}{\sqrt{4a^2+\varepsilon^3}} \geqslant \frac{2a}{\sqrt{4a^2+\varepsilon\lambda_k^2}} = \nu_k,$$

for some $\bar{k} \geqslant \hat{k}$, we have

$$1 > \frac{\sqrt{4a^2+\varepsilon^3}+2a}{2\sqrt{4a^2+\varepsilon^3}} \geqslant \theta_k. \quad \square$$

Remark 4.1. It is interesting to notice that even though Algorithm I is quite different from the proximal method of multipliers, one is still able to obtain a linear rate of convergence result. Once again, Proposition 3.1, which allows us to get the estimate (3.14), is the key in the above analysis. Note however that a superlinear rate of convergence derived for the proximal method of multipliers, when $\lambda_k \to \infty$, is not applicable for the PCPM, since here $\{\lambda_k\}$ must stay bounded above.

5. Comparison with other methods and potential applications

5.1. Comparison with alternating direction methods

It is interesting to compare our algorithm with various types of well known splitting methods used for decomposition in convex programming. We will focus on alternating direction methods of multipliers. These methods are taking roots from their similarity with some methods for solving differential equations, such as the Douglas-Rachford scheme see e.g. [18]. For further details and references, we refer the reader to the recent books of Bertsekas and Tsitsiklis [3] and Glowinski and Le Tallec [10].

The basic idea underlying alternating direction methods is a relaxation approach, whereby given an augmented Lagrangian associated with (P), one first minimizes it with respect to x and then with respect to z, and the multipliers y are updated via the usual augmented

Lagrangian update rule. This approach removes the difficulty of the joint minimization in x and z, and thus preserves separability. As we shall see below, algorithms produced from this approach bear similarity with our algorithm, but are different both in the computational steps, and in the assumptions involved in the problem's data.

We will concentrate on comparing our algorithm with the alternating direction method of multipliers proposed by Gabay and Mercier [9] and Glowinski and Marrocco [11], and with some of its variant and related methods.

Consider the following augmented Lagrangian associated with (P):

$$\mathscr{A}(x, z, y) = f(x) + g(z) + \langle y, Ax - z\rangle + \tfrac{1}{2}\lambda \|Ax - z\|^2. \tag{5.1}$$

As explained above, by minimising $\mathscr{A}$ first over x then over z, the alternating direction method of multipliers takes the basic form:

Algorithm A [9–11].

$$x^{k+1} \in \arg\min\{f(x) + \langle y^k, Ax\rangle + \tfrac{1}{2}\lambda \|Ax - z^k\|^2\},$$
$$z^{k+1} = \arg\min\{g(z) - \langle y^k, z\rangle + \tfrac{1}{2}\lambda \|Ax^{k+1} - z\|^2\},$$
$$y^{k+1} = y^k + \lambda(Ax^{k+1} - z^{k+1}),$$

where λ is a fixed positive scalar.

More recently, Tseng [27] proposed another variant of Algorithm A, that he called the alternating minimization algorithm. The key difference with Algorithm A is that the usual Lagrangian function $L(x, z, y)$ replaces the augmented Lagrangian function $\mathscr{A}(x, z, y)$ when the minimization is taken with respect to x. The alternating minimization algorithm takes the form:

Algorithm B [27].

$$x^{k+1} = \arg\min\{f(x) + \langle y^k, Ax\rangle\},$$
$$z^{k+1} = \arg\min\{g(z) - \langle y^k, z\rangle + \tfrac{1}{2}\lambda_k \|Ax^{k+1} - z\|^2\},$$
$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} - z^{k+1}),$$

where $\{\lambda_k\}$ satisfies some conditions (see (B3) below).

For a detailed comparison between Algorithms A and B, we refer the reader to Tseng [27]. For ease of comparison between Algorithms A and B, and our algorithm, we use the exact version of Algorithm I (i.e. set $\alpha_k = \beta_k = 0$ in Step 2 of Algorithm I), which will be called in the sequel, Algorithm E. First we observe that in all three algorithms, the multiplier update rules are the same. The main differences between Algorithm E and Algorithms A and B are in the minimization steps with respect to x and z (note that for Algorithms A and B the minimization steps with respect to z are identical); and by the fact that our Algorithm E requires the additional predictor multiplier step

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

However, this additional step is a simple update similar to the multiplier update rule, and allows for obtaining the convergence of our algorithm under the only assumption that there exists a primal–dual optimal solution. On the other hand, with this assumption at hand, the following additional assumptions are needed to establish the convergence of Algorithms A and B:

For Algorithm A:

(A1) The matrix A^TA is positive definite i.e., the matrix A has full column rank. This is needed to have a well defined x^k, when performing the minimization with respect to x.

Note, however that the parameter $\lambda>0$, needs not to be chosen from a restricted range or changed at each iteration k. But see also Glowinski and Le Tallec [10, p. 85] for a choice of λ which is not fixed at each iteration.

For Algorithm B:

(B1) The function f is strongly convex with modulus $\alpha>0$.

(B2) $\{\lambda_k\}$ satisfies $\varepsilon\leqslant\lambda_k\leqslant 4\alpha/\|A\|^2-\varepsilon$ for all k, for some $\varepsilon\in(0, 2\alpha/\|A\|^2)$.

Note that Algorithm E requires an assumption similar to (B2) for $\{\lambda_k\}$ (cf. (3.10)), except that we need not to know α, since we emphasize, that in Algorithm E neither f nor g need to be strongly convex.

We note that only recently, Eckstein and Bertsekas [4] were the first to prove the convergence of the alternating direction method for solving problem (P) allowing approximate minimization in x and z as in Algorithm I. We are unaware of any convergence rate results for the alternating direction method. For the alternating minimization algorithm, Tseng [27] proved that at least linear rate of convergence can be achieved, under the additional assumption that one of the operators, $A\partial f^*A^T$ or ∂g^*, is strongly monotone.

Finally, we point out that in both Algorithms A and B, one needs to know the vector x^{k+1} in order to update z^{k+1}. On the other hand, in Algorithm E, x^{k+1} and z^{k+1} are computed separately from each other, and thus this offers a further degree of parallelization. This will be exemplified below.

5.2. Potential applications

In [13], Han and Lou proposed a parallel decomposition algorithm for minimizing a strongly convex and differentiable function over the intersection of closed convex sets. This algorithm has been further studied, under different assumptions in [15], [19] and [27]. It was shown in Iusem and De Pierro [15] and Tseng [27], that Han-Lou algorithm is in fact a special case of the alternating minimization algorithm.

We conclude this section by applying Algorithm E to that problem, providing another new highly parallelizable algorithm, which does not require the strong convexity (or differentiability) assumption imposed in [13], [19] and [27].

Consider the following convex programming problem:

$$\min\left\{h(x): x\in \bigcap_{i=1}^{l} C_i\right\} \tag{5.2}$$

where $h:\mathbb{R}^n\mapsto(-\infty, +\infty]$ is a closed proper convex function, and each C_i is a closed convex set in $\mathbb{R}^n$. Introducing the artificial variables $z_i\in\mathbb{R}^n$, $i=1, \ldots, l$, we can rewrite (5.2) in the following equivalent form:

$$\min\left\{h(x)+\sum_{i=1}^{l}\delta(z_i\,|\,C_i): x=z_i,\ i=1, \ldots, l\right\} \tag{5.3}$$

where $\delta(\cdot\,|\,C_i)$ is the indicator function for C_i. Identifying, $h(x)=f(x)$, $g:\mathbb{R}^{nl}\mapsto(-\infty, +\infty]$ with $g(z)=\Sigma_{i=1}^{l}\delta(z_i|C_i)$ where $z=(z_1, z_2, z_3, \ldots, z_l)\in\mathbb{R}^{nl}$, and $A=[I, I, \ldots, I]'\in\mathbb{R}^{n\times nl}$, problem (5.3) is clearly a special case of problem (P). Applying Algorithm E to problem (5.3) produces the following iterations.

Algorithm C. Starting with an arbitrary point (x^0, z^0, y^0), for $k\geqslant 0$,

$$p_i^{k+1}=y_i^k+\lambda_k(x^k-z_i^k),\quad i=1, \ldots, l,$$

$$x^{k+1}=\arg\min\left\{h(x)+\sum\langle p_i^{k+1}, x\rangle+(1/(2\lambda_k))\,\|x-x^k\|^2\right\}$$

$$z_i^{k+1}=\arg\min_{z_i\in C_i}\{-\langle p_i^{k+1}, z_i\rangle+(1/(2\lambda_k))\,\|z_i-z_1^k\|^2\},\quad i=1, \ldots, l,$$

$$y_i^{k+1}=y_i^k+\lambda_k(x^{k+1}-z_i^{k+1}),\quad i=1, \ldots, l.$$

Here, $y_i\in\mathbb{R}^n$, $i=1, \ldots, l$, is the Lagrangian multipliers associated with the constraints $x=z_i$. Note that the above iterations are highly parallelizable and that the minimization with respect to z is equivalent to compute a projection on each of the set C_i, i.e. $z_i^{k+1}=P_{C_i}(z_i^k+\lambda_k p_i^{k+1})$, compare with [13, 27].

To prove convergence of Algorithm C, we assume that Assumption A holds. In this particular case, the latter will be satisfied if

$$\mathrm{ri}(\mathrm{dom}\,h)\cap\mathrm{ri}(C_1)\cap\cdots\cap\mathrm{ri}(C_l)\neq\emptyset. \tag{5.4}$$

Applying Theorem 3.1, we then obtain:

Theorem 5.1. *Let $\{x^k, z^k, y^k\}$ be the sequence generated by Algorithm* C *and assume that* (5.4) *holds. If $\{\lambda_k\}$ satisfies*

$$\varepsilon\leqslant\lambda_k\leqslant(1-\varepsilon)/(2\sqrt{l})\quad\forall k\geqslant 0 \tag{5.5}$$

for some $0 \leqslant \varepsilon \leqslant 1/(2\sqrt{l}+1)$, *then* $\{x^k\}$ *converges to* x^*, $\{z_i^k\}$ *converges to* x^* *for all* $i=1, \ldots, l$ *and* $\{y_i^k\}$ *converges to* y_i^*, $i=1, \ldots, l$. □

We finally mention that one can also apply Theorem 4.1 to problem (5.2) to derive a rate of convergence result for Algorithm C. Furthermore, if the function $h(x)$ is given separable, then the minimization step for the variable x can be performed componentwise, and since we need not to assume the strong convexity of the objective function, our algorithm is in particular also applicable to linear programming problems.

References

[1] A. Auslender, ''Numerical methods for nondifferentiable convex optimization,'' *Mathematical Programming Study* 30 (1987) 102–126.

[2] D.P. Bertsekas, *Constrained Optimization and Lagrangian Multipliers* (Academic Press, New York, 1982).

[3] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[4] J. Eckstein and D.P. Bertsekas, ''On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators,'' *Mathematical Programming* 55 (1992) 293–318.

[5] W. Findeisen, F.N. Bailey, M. Brdys, K. Malinowski, P. Tatjewski and A. Wozniak, *Control and Coordination in Hierarchical Systems* (Wiley, New York, 1980).

[6] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems* (North-Holland, Amsterdam, 1983).

[7] M. Fukushima, ''Application of the alternating direction method of multipliers to separable convex programming problems,'' *Computational Optimization and Applications* 1 (1992) 93–111.

[8] D. Gabay, ''Applications of the method of multipliers to variational inequalities,'' in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems* (North-Holland, Amsterdam, 1983) pp. 299–331.

[9] D. Gabay and B. Mercier, ''A dual algorithm for the solution of nonlinear variational problems via finite-element approximations,'' *Computors and Mathematics with Applications* 2 (1976) 17–40.

[10] R. Glowinski and P. Le Tallec, ''Augmented lagrangian and operator-splitting methods in nonlinear mechanics,'' in: *SIAM Studies in Applied Mathematics* (SIAM, Philadelphia, PA, 1989).

[11] R. Glowinski and A. Marrocco, ''Sur l'approximation par éléments finis d'ordre un, et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet nonlinéaires,'' *Revue Française d'Automatique, Informatique et Recherche Opérationelle* 2 (1975) 41–76.

[12] O. Guler, ''On the convergence of the proximal point algorithm for convex minimization,'' *SIAM Journal on Control and Optimization* 29 (1991) 403–419.

[13] S.P. Han and G. Lou, ''A parallel algorithm for a class of convex programs,'' *SIAM Journal on Control and Optimization* 26 (1988) 345–355.

[14] S. Ibaraki, M. Fukushima and T. Ibaraki, ''Primal–dual proximal point algorithm for linearly constrained convex programming problems,'' Technical Report, No. 91016, Kyoto University (Kyoto, Japan, 1991).

[15] A.N. Iusem and A.R. De Pierro, ''On the convergence of Han's method for convex programming with quadratic objective,'' *Mathematical Programming* 52 (1991) 265–284.

[16] L.S. Lasdon, *Optimization Theory for Large Systems* (Macmillan, New York, 1970).

[17] B. Lemaire, ''The proximal algorithm,'' *International Series of Numerical Mathematics* 87 (1989) 73–87.

[18] P.L. Lions and B. Mercier, ''Splitting algorithms for the sum of two nonlinear operators,'' *SIAM Journal on Numerical Analysis* 16 (1979) 964–979.

[19] K. Mouallif, V.H. Nguyen and J.J. Strodiot, ''A perturbed parallel decomposition method for a class of nonsmooth convex minimization problems,'' *SIAM Journal on Control and Optimization* 29 (1991) 829–847.

[20] B.T. Polyak, *Introduction to Optimization* (Optimization Software, New York, 1987).

[21] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
[22] R.T. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research* 1 (1976) 97–116.
[23] R.T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization* 14 (1976) 877–898.
[24] J.E. Spingarn, "Applications of the method of partial inverses to convex programming: decomposition," *Mathematical Programming* 32 (1985) 199–223.
[25] G. Stephanopoulos and A.W. Westerberg, "The use of Hestenes' method of multipliers to resolve dual gaps in engineering system optimization," *Journal of Optimization Theory and Applications* 15 (1975) 285–309.
[26] A. Tanikawa and H. Mukai, "A new technique for nonconvex primal–dual decomposition," *IEEE Transactions on Automatic Control* AC-30 (1985) 133–143.
[27] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization* 29 (1991) 119–138.
[28] H. Uzawa, "Iterative methods for concave programming," in: K.J. Arrow, L. Hurwicz and H. Uzawa, eds., *Studies in Linear and Nonlinear Programming* (Stanford University Press, Stanford, CA, 1958) pp. 154–165.