# A variable-penalty alternating directions method for convex optimization *

Spyridon Kontogiorgis[†] Robert R. Meyer[‡]

**Abstract**

We study a generalized version of the method of alternating directions as applied to the minimization of the sum of two convex functions subject to linear constraints. The method consists of solving consecutively in each iteration two optimization problems which contain in the objective function both Lagrangian and proximal terms. The minimizers determine the new proximal terms and a simple update of the Lagrangian terms follows. We prove a convergence theorem which extends existing results by relaxing the assumption of uniqueness of minimizers. Another novelty is that we allow penalty matrices, and these may vary per iteration. This can be beneficial in applications, since it allows additional tuning of the method to the problem and can lead to faster convergence relative to fixed penalties. As an application, we derive a decomposition scheme for block angular optimization and present computational results on a class of dual block angular problems.

**Keywords:** parallel computing, alternating direction methods, decomposition, block angular programs.
**Abbreviated title:** variable-penalty ADI.

## 1 Introduction

We present an extended alternating directions method (ADI) for the general convex optimization problem

$$\begin{aligned} \min_{x,\, z} \quad & G_1(x) + G_2(z) \\ \text{subject to} \quad & Ax + b = Bz \end{aligned} \tag{1}$$

where $G_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $G_2 : \mathbb{R}^s \to \mathbb{R} \cup \{+\infty\}$ are extended-real valued, proper, closed, convex functions, $b \in \mathbb{R}^m$, and the matrices $A : \mathbb{R}^n \to \mathbb{R}^m$ and $B : \mathbb{R}^s \to \mathbb{R}^m$.

We make a basic assumption.

**Assumption 1.1** *Problem (1) admits a Lagrangian saddlepoint.*

*This material is based on research supported by the Air Force Office of Scientific Research Grant AFOSR-89-0410 and by NSF Grants CCR-8907671, CDA-9024618 and CCR-9306807.

†USAir Operations Research, 2345 Crystal Drive, Arlington, VA 22227.
‡Center for Parallel Optimization, Computer Sciences Department, University of Wisconsin–Madison, 1210 West Dayton Street, Madison, WI 53706.

The ADI method operates on the augmented Lagrangian associated with problem (1)

$$L_\lambda(x,z,p) := G_1(x) + G_2(z) + p^T(Ax+b-Bz) + \frac{\lambda}{2}\|Ax+b-Bz\|_2^2 \tag{2}$$

in which $p$ is a tentative dual multiplier and $\lambda$ is a positive scalar. From standard duality theory, if $(x^*, z^*, p^*)$ is a saddlepoint of the augmented Lagrangian, then $(x^*, z^*)$ is a primal solution of (1) and $p^*$ is an associated dual multiplier.

In the method of multipliers [26, 38, 39, 2] a saddlepoint is located by an iterative process, consisting of a minimization of the augmented Lagrangian, followed by a steepest-ascent update of the multipliers.

$$\begin{array}{rcl} (x^{t+1}, z^{t+1}) & \in & \underset{x,\,z}{\operatorname{argmin}} \;\; L_\lambda\ (x, z, p^t) \\ p^{t+1} & = & p^t + \lambda\ \nabla_p\ L_\lambda\ (x^{t+1}, z^{t+1}, p^t) \end{array}$$

A computational drawback of this algorithm is that the quadratic penalty term $\|Ax+b-Bz\|_2^2$ in $L_\lambda(x,z,p^t)$ is not separable with respect to $x$ and $z$. To overcome this, the primal minimization can be carried out in two steps in a block Gauss-Seidel fashion.

$$\begin{array}{rcl} x^{t+1} & \in & \underset{x}{\operatorname{argmin}} \;\; L_\lambda\ (x, z^t, p^t) \\ z^{t+1} & \in & \underset{z}{\operatorname{argmin}} \;\; L_\lambda\ (x^{t+1}, z, p^t) \\ p^{t+1} & = & p^t + \lambda\ \nabla_p\ L_\lambda\ (x^{t+1}, z^{t+1}, p^t) \end{array}$$

This is equivalent to the following scheme.

$$x^{t+1} \in \underset{x}{\operatorname{argmin}} \;\; G_1(x) + {p^t}^T \lambda A x + \frac{\lambda}{2}\left\|Ax+b-Bz^t\right\|_2^2 \tag{3}$$

$$z^{t+1} \in \underset{z}{\operatorname{argmin}} \;\; G_2(z) - {p^t}^T \lambda B z + \frac{\lambda}{2}\left\|Ax^{t+1}+b-Bz\right\|_2^2 \tag{4}$$

$$p^{t+1} = p^t + (Ax^{t+1} + b - Bz^{t+1}) \tag{5}$$

This is the ADI method; it consists of solving in each iteration two optimization problems, the objective function of which contains both Lagrangian and proximal terms. Then the Lagrangian terms are adjusted in proportion to the violation of the constraints. Arbitrary $p^0$ and $z^0$ can be chosen as starting point, and, in the basic case, the penalty $\lambda$ is constant.

The method has been studied extensively in the theoretical frameworks of both Lagrangian functions (Glowinski with Chan, Marrocco, Fortin and Le Tallec [16, 20]), and maximal monotone operators (Lions and Mercier [32], Gabay [18], Eckstein and Bertsekas [9, 12]). The connection to the proximal point algorithm is discussed in Rockafellar [40]. It has been shown [32] that the method is an instance of the Douglas-Rachford splitting [8] for finding a zero of a maximal monotone operator. Several other decomposition schemes, such as the algorithm of Han and Lou [24], Spingarn's method of partial inverses [42, 43], the progressive hedging algorithm of Rockafellar and Wets [41] and Golshtein's block method of convex programming [21, 22] are also instances of the Douglas-Rachford splitting (see [9] for a demonstration).

Eckstein and Bertsekas [12] and Cheng and Teboulle [5] have constructed ADI variants which allow for inexact minimization. The former also permits relaxation of the primal iterates; in the latter, quadratic proximal terms replace the augmented Lagrangian penalty terms. In the Peaceman-Rachford variant [18], based on [36], a multiplier update is interpolated between the two problems. This algorithm requires more stringent assumptions for convergence and is less robust numerically [16]. Fukushima [17] presents an ADI method for the dual problem of (1). ADI methods have been constructed for several other classes of problems, such as variational inequalities [19] and the monotone linear complementarity problem [13].

The existing literature on the basic ADI algorithm (3)-(5) treats the slightly simpler case where $B = I$ and $b = 0$. Convergence has been proved under assumption 1.1 and additional ones that guarantee that the original problem and/or the ADI problems are uniquely solvable. Such assumptions are: $A$ has full column rank and $G_2$ is the sum of a closed, proper, convex function and a strictly convex $C^1$ function [16, chapter 3]; $A$ has full column rank [9, 12], [3, chapter 3]. The dual algorithm in [17] requires that both primal and dual problems be feasible and that the solution set of the primal be bounded. We will show here that it is possible to dispense with the uniqueness assumption and still obtain a solution of (1) by ADI.

Previous theoretical work on (3)-(5) has focused on the case of a fixed penalty $\lambda$. It has been proven that, under additional assumptions of coercivity and Lipschitz continuity, the rate of convergence is linear. Then an optimal value of $\lambda$ exists and is related to the constants in these properties [32, section 1.3.3]. In the general case, a good value of $\lambda$ is determined empirically, after experimentation and examination of the characteristics of the problem. In network flow problems, accelerated convergence has been observed with heuristics that vary the penalty for finitely many iterations, or even use a separate penalty for each linear constraint [34, 10, 31].

In this article we present convergence results for a general variable penalty algorithm, in which a symmetric positive definite (spd) matrix $H^t$ is employed in the Lagrangian and proximal term, to allow for linear transformations of the constraints. This leads to the following extension of algorithm (3)-(5).

$$
\begin{aligned}
x^{t+1} &\in \operatorname*{argmin}_x \; G_1(x) + {p^t}^T H^t A x + \frac{1}{2}\left\| Ax + b - Bz^t \right\|^2_{H^t} && (6)\\
z^{t+1} &\in \operatorname*{argmin}_z \; G_2(z) - {p^t}^T H^t B z + \frac{1}{2}\left\| Ax^{t+1} + b - Bz \right\|^2_{H^t} && (7)\\
p^{t+1} &= p^t + (Ax^{t+1} + b - Bz^{t+1}) && (8)\\
\text{Update} \quad & H^t && (9)
\end{aligned}
$$

The next assumption guarantees that the algorithm is well-defined.

**Assumption 1.2** *Problems (6) and (7) are solvable.*

We impose a modest control on the growth of the penalty.

**Assumption 1.3** *The eigenvalues of the spd matrices $H^t$ are uniformly bounded from below away from zero, and, with finitely many exceptions, the eigenvalues of $H^t - H^{t+1}$ are nonnegative.*

This assumption implies that $\{H^t\}$ and its eigenvalues converge. (This can be shown by using lemma 2.9). This assumption allows us to vary the penalty in an arbitrary fashion for a finite number of iterations; thereafter, the variation must be such that $H^t - H^{t+1}$ is positive semidefinite. The added flexibility in the initial stages can be computationally beneficial.

This article is organized as follows: in section 2 we prove the main result, the convergence theorem for the algorithm (6)–(9), and present an implementable construction of a penalty $\{H^t\}$ satisfying assumption 1.3. A characteristic of the extended algorithm is that, although the primal iterates may not converge, they are feasible in the limit, and also the limit of the objective value is optimal. (This is in common with other algorithms, such as the subgradient method for nondifferentiable optimization in [37, section 5.3.2].) We present an example illustrating this characteristic in section 3. In section 4 we provide a sequence of corollaries that mainly address primal convergence and finite termination. As an application, in section 5 we derive a decomposition scheme for block angular optimization in which we also have convergence to the optimal objective value even in the absence of primal convergence. Finally, we examine computational performance on a class of dual block angular problems in section 6.

## 2 The convergence theorem

In this section we prove the following convergence theorem for the extended method.

**Theorem 2.1** *Let assumptions 1.1, 1.2 and 1.3 hold. Then, for any sequence of iterates* $\{(x^t, z^t, p^t, H^t)\}$ *produced by the ADI algorithm (6)–(9),*
*(i)* $\{(Ax^t, Bz^t)\}$ *converges, and the limit satisfies the constraints of problem (1).*
*(ii)* $\{G_1(x^t)+G_2(z^t)\}$ *converges to the optimal value of the objective function for problem (1).*
*(iii)* $\{H^t p^t\}$ *converges to an optimal dual multiplier for problem (1).*
*(iv) Any minimizers of problems of the form (6) and (7) in which* $H^t p^t$, $Ax^{t+1}$ *and* $Bz^t$ *are fixed at their limit values are optimal for problem (1).*

Note that in (iii) we consider a dual multiplier independently of a primal solution. This is because of the following lemma, which describes a well-known property of saddlepoints: duals are associated with the problem, not with specific primal solutions.

**Lemma 2.2** *Let* $X^0$ *be a set in* $\mathbb{R}^n$*, and define the functions* $\theta : X^0 \to \mathbb{R}$ *and* $g : X^0 \to \mathbb{R}^s$*. Suppose that* $(x_1, u_1)$ *is a saddlepoint with respect to* $\theta$ *and* $g$*, i.e.*

$$\theta(x_1) + u^T g(x_1) \le \theta(x_1) + {u_1}^T g(x_1) \le \theta(x) + {u_1}^T g(x) \quad \forall u \in \mathbb{R}^s,\ \forall\ x \in X^0\ .$$

*If* $(x_2, u_2)$ *also satisfies these conditions, then so do* $(x_1, u_2)$ *and* $(x_2, u_1)$*.*

We will use the following minimum principle lemma, which specializes [4, theorem 2.3].

**Lemma 2.3** *Let* $X^0$ *be a convex subset of* $\mathbb{R}^n$*. Suppose* $J : X^0 \to \mathbb{R}$ *is a function of the form* $J = J_1 + J_2$*, where* $J_1$ *and* $J_2$ *are convex and* $J_2$ *is differentiable on* $X^0$*. If* $\bar{x}$ *minimizes* $J$*, then* $J_1(x) - J_1(\bar{x}) + \nabla J_2(\bar{x})^T(x - \bar{x}) \ge 0, \quad \forall x \in X^0$*.*

We will also use the following lemma, due to Cheng [6, lemma 2.1].

**Lemma 2.4** *Let* $\{a^t\}$ *and* $\{\epsilon^t\}$ *be two sequences of nonnegative numbers, with* $\sum_{t=0}^{\infty} \epsilon^t < \infty$ *and* $a^{t+1} \le a^t + \epsilon^t$*. Then* $\{a^t\}$ *converges.*

In order to prove theorem 2.1, we state and prove a collection of lemmas. We begin by showing, in lemma 2.5, that the iterates $\{(Ax^t, Bz^t, p^t)\}$ are bounded. Then we establish that $\{G_1(x^t) + G_2(z^t)\}$ converges to the optimal objective value for problem (1) and that, in the limit, $\{(Ax^t, Bz^t)\}$ satisfies the constraints of problem (1) (lemma 2.6). In lemma 2.7 we show that $\{Ax^t\}$, $\{Bz^t\}$ and $\{H^t p^t\}$ converge, and that the limit of $\{H^t p^t\}$ is an optimal dual for (1). Finally, lemma 2.8 shows how to obtain a primal solution for (1) by solving two minimization problems which employ the limits of $\{Ax^t\}$, $\{Bz^t\}$ and $\{H^t p^t\}$ as fixed terms in the objective.

The proof employs the saddlepoint approach of [3, chapter 3] and [20, chapter 3]. An important difference is that we do not assume minimizer uniqueness for problems (1), (6) or (7). Also, the use of variable penalty requires a more complex argument, involving lemma 2.4. Another novelty is the use of lemma 2.2, which allows us to consider primal and dual components of saddlepoints separately.

**Lemma 2.5** *Let the assumptions of theorem 2.1 hold. Then $\{Ax^t\}$, $\{Bz^t\}$ and $\{p^t\}$ are bounded.*

**Proof:** Applying lemma 2.3 to problem (6) with the correspondences

$$J_1(x) = G_1(x), \quad J_2(x) = {p^t}^T H^t Ax + \frac{1}{2}\left\|Ax + b - Bz^t\right\|^2_{H^t}$$

yields

$$[p^t + Ax^{t+1} + b - Bz^t]^T H^t A(x^{t+1} - x) \leq G_1(x) - G_1(x^{t+1}) \quad \forall x \in \mathbb{R}^n \tag{10}$$

which, after using (8), becomes

$$[p^{t+1} + B(z^{t+1} - z^t)]^T H^t A(x^{t+1} - x) \leq G_1(x) - G_1(x^{t+1}) \quad \forall x \in \mathbb{R}^n \tag{11}$$

Applying the lemma to problem (7) with the correspondences

$$J_1(z) = G_2(z), \quad J_2(z) = -{p^t}^T H^t Bz + \frac{1}{2}\left\|Ax^{t+1} + b - Bz\right\|^2_{H^t}$$

we get in a similar fashion

$$-{p^{t+1}}^T H^t B(z^{t+1} - z) \leq G_2(z) - G_2(z^{t+1}) \quad \forall z \in \mathbb{R}^s \tag{12}$$

Let $(x^*, z^*, p^*)$ be a saddlepoint for problem (1). Then

$$Ax^* + b = Bz^* \tag{13}$$

and also

$$G_1(x^*) + G_2(z^*) \leq G_1(x) + G_2(z) + {p^*}^T(Ax + b - Bz) \quad \forall x \in \mathbb{R}^n,\ \forall z \in \mathbb{R}^s$$

For the iterates $(x^t, z^t)$ this implies, after taking (8) into account,

$$G_1(x^*) + G_2(z^*) \leq G_1(x^{t+1}) + G_2(z^{t+1}) + {p^*}^T(p^{t+1} - p^t) \tag{14}$$

Substituting $x^*$ for $x$ in (11) we obtain

$$[p^{t+1} + B(z^{t+1} - z^t)]^T H^t A(x^{t+1} - x^*) \leq G_1(x^*) - G_1(x^{t+1}) \tag{15}$$

By (13) and (8)

$$A(x^{t+1} - x^*) = p^{t+1} - p^t + B(z^{t+1} - z^*) \tag{16}$$

and (15) can be rewritten as

$$[p^{t+1} + B(z^{t+1} - z^t)]^T H^t [p^{t+1} - p^t + B(z^{t+1} - z^*)] \le G_1(x^*) - G_1(x^{t+1}) \tag{17}$$

Substituting $z^*$ for $z$ in (12) we obtain

$$-\, {p^{t+1}}^T H^t B(z^{t+1} - z^*) \le G_2(z^*) - G_2(z^{t+1}) \tag{18}$$

Adding (14), (17) and (18) yields

$$\begin{aligned}(p^{t+1} - p^t)^T [H^t p^{t+1} - p^*] + [B(z^{t+1} - z^t)]^T H^t B(z^{t+1} - z^*) \\ + (p^{t+1} - p^t)^T H^t B(z^{t+1} - z^t) \quad \le 0\end{aligned} \tag{19}$$

The identities

$$2\,(p^{t+1} - p^t)^T [H^t p^{t+1} - p^*] = \|p^{t+1} - p^t\|_{H^t}^2 + \|p^{t+1} - (H^t)^{-1} p^*\|_{H^t}^2 - \|p^t - (H^t)^{-1} p^*\|_{H^t}^2$$

$$2\,[B(z^{t+1} - z^t)]^T H^t B(z^{t+1} - z^*) = \|B(z^{t+1} - z^t)\|_{H^t}^2 + \|B(z^{t+1} - z^*)\|_{H^t}^2 - \|B(z^t - z^*)\|_{H^t}^2$$

allow (19) to be written as

$$\begin{aligned}\|p^{t+1} - p^t\|_{H^t}^2 + \|B(z^{t+1} - z^t)\|_{H^t}^2 + 2(p^{t+1} - p^t)^T H^t B(z^{t+1} - z^t) \quad \le \\ \|p^t - (H^t)^{-1} p^*\|_{H^t}^2 + \|B(z^t - z^*)\|_{H^t}^2 - \|p^{t+1} - (H^t)^{-1} p^*\|_{H^t}^2 - \|B(z^{t+1} - z^*)\|_{H^t}^2\end{aligned} \tag{20}$$

The left hand side equals

$$\left\|(p^{t+1} - p^t) + B(z^{t+1} - z^t)\right\|_{H^t}^2 .$$

We can provide an upper bound for the right hand side. We define

$$\alpha^t := \left\|p^t - (H^t)^{-1} p^*\right\|_{H^t}^2 + \left\|B(z^t - z^*)\right\|_{H^t}^2 \tag{21}$$

By assumption 1.3, $H^t - H^{t+1}$ is ultimately positive semidefinite, and thus, for any vector $c$, ultimately $\|c\|_{H^t}^2 \ge \|c\|_{H^{t+1}}^2$. Then we have, for the right hand side

$$\begin{aligned}
&\|p^t - (H^t)^{-1} p^*\|_{H^t}^2 + \|B(z^t - z^*)\|_{H^t}^2 - \|p^{t+1} - (H^t)^{-1} p^*\|_{H^t}^2 - \|B(z^{t+1} - z^*)\|_{H^t}^2 && = \\
&\alpha^t - \|B(z^{t+1} - z^*)\|_{H^t}^2 - \|p^{t+1} - (H^t)^{-1} p^*\|_{H^t}^2 && \le \\
&\alpha^t - \|B(z^{t+1} - z^*)\|_{H^{t+1}}^2 - \|p^{t+1} - (H^t)^{-1} p^*\|_{H^t}^2 && = \\
&\alpha^t - \|B(z^{t+1} - z^*)\|_{H^{t+1}}^2 - \|p^{t+1}\|_{H^t}^2 - \|p^*\|_{(H^t)^{-1}}^2 + 2{p^{t+1}}^T p^* && \le \\
&\alpha^t - \|B(z^{t+1} - z^*)\|_{H^{t+1}}^2 - \|p^{t+1}\|_{H^{t+1}}^2 - \|p^*\|_{(H^t)^{-1}}^2 + 2{p^{t+1}}^T p^* && = \\
&\alpha^t - \alpha^{t+1} + \epsilon^t
\end{aligned}$$

for

$$\epsilon^t := \|p^*\|_{(H^{t+1})^{-1}}^2 - \|p^*\|_{(H^t)^{-1}}^2 \tag{22}$$

Thus (20) implies that ultimately

$$0 \le \left\|(p^{t+1} - p^t) + B(z^{t+1} - z^t)\right\|_{H^t}^2 \le \alpha^t - \alpha^{t+1} + \epsilon^t \tag{23}$$

This is the key inequality in establishing convergence. Since the matrix $H^t - H^{t+1}$ is ultimately positive semidefinite, so ultimately is $(H^{t+1})^{-1} - (H^t)^{-1}$ [27, corollary 7.7.4], and therefore $\epsilon^t \geq 0$, ultimately. A telescopic summation shows that

$$\sum_{t=0}^{k} \epsilon^t = \|p^*\|^2_{(H^{k+1})^{-1}} - \|p^*\|^2_{(H^0)^{-1}}, \quad \forall k \geq 0$$

Because of assumption 1.3, these partial sums are uniformly bounded from above. Then $\sum_{t=0}^{\infty} \epsilon^t$ converges and, by lemma 2.4, $\{\alpha^t\}$ converges. It follows that each of $\{p^t\}$ and $\{Bz^t\}$ is bounded. Then, by (16), $\{Ax^t\}$ is also bounded. ■

If the penalty were constant, $\epsilon^t$ in (23) would be identically zero. Then, for any saddlepoint $(x^*, z^*, p^*)$, $\{\alpha^t\}$ would be nonincreasing and *Fejér-monotone*, meaning that no step increases the distance to any solution point. This is also a property of the proximal point algorithm [42].

**Lemma 2.6** *Let the assumptions of theorem 2.1 hold. Then,*
*(i)* $\{p^{t+1} - p^t\} \longrightarrow 0$ , $\{B(z^{t+1} - z^t)\} \longrightarrow 0$ , $\{Ax^t + b - Bz^t\} \longrightarrow 0$ .
*(ii)* $\{G_1(x^t) + G_2(z^t)\}$ *converges to the optimal objective value for problem (1).*

**Proof:** Substituting $z^t$ for $z$ in (12) yields

$$-\, {p^{t+1}}^T H^t B(z^{t+1} - z^t) \leq G_2(z^t) - G_2(z^{t+1}) \tag{24}$$

We write (12) for $t = t-1$ and let $z = z^{t+1}$. We obtain

$$-\, {p^t}^T H^{t-1} B(z^t - z^{t+1}) \leq G_2(z^{t+1}) - G_2(z^t) \tag{25}$$

Adding (24) and (25) yields, after rearranging,

$$(p^{t+1} - p^t)^T H^t B(z^{t+1} - z^t) \geq {p^t}^T (H^{t-1} - H^t) B(z^{t+1} - z^t) \tag{26}$$

We now combine (26) and (23) and get

$$\alpha^t - \alpha^{t+1} + \epsilon^t - 2\, {p^t}^T (H^{t-1} - H^t) B(z^{t+1} - z^t) \geq \left\| p^{t+1} - p^t \right\|^2_{H^t} + \left\| B(z^{t+1} - z^t) \right\|^2_{H^t} \geq 0 \tag{27}$$

Since $\{\alpha^t\}$ converges and $\{\epsilon^t\}$ converges to zero, $\{\alpha^t - \alpha^{t+1} + \epsilon^t\}$ converges to zero. Because of the boundedness of $\{p^t\}$ and $\{Bz^t\}$ and assumption 1.3, $\left\{{p^t}^T (H^{t-1} - H^t) B(z^{t+1} - z^t)\right\}$ converges to zero. Therefore, by dominance, both $\{p^{t+1} - p^t\}$ and $\{B(z^{t+1} - z^t)\}$ converge to zero. Then, because of (8), $\{Ax^t + b - Bz^t\}$ also converges to zero. This concludes the proof of part (i).

Let again $(x^*, z^*)$ be a solution of problem (1). We take limits in (14) and use the fact that $\{p^{t+1} - p^t\} \longrightarrow 0$. We get

$$G_1(x^*) + G_2(z^*) \leq \liminf_t \left\{ G_1(x^t) + G_2(z^t) \right\} \tag{28}$$

Adding (17) and (18) yields

$$\begin{aligned} G_1(x^*) + G_2(z^*) \;\geq\;& G_1(x^{t+1}) + G_2(z^{t+1}) + (p^{t+1} - p^t)^T H^t [p^t + B(z^{t+1} - z^t)] \\ &+ [B(z^{t+1} - z^t)]^T H^t B(z^{t+1} - z^*) \end{aligned} \tag{29}$$

We take limits in (29), using the boundedness of $\{p^t\}$ and $\{Bz^t\}$, assumption 1.3 and part (i). We obtain

$$\limsup_t \left\{G_1(x^t) + G_2(z^t)\right\} \leq G_1(x^*) + G_2(z^*) \tag{30}$$

Combining (28) and (30) we get

$$\limsup_t \left\{G_1(x^t) + G_2(z^t)\right\} \leq G_1(x^*) + G_2(z^*) \leq \liminf_t \left\{G_1(x^t) + G_2(z^t)\right\}$$

which implies that $\lim_t \{G_1(x^t) + G_2(z^t)\} = G_1(x^*) + G_2(z^*)$. This proves part (ii). ∎

**Lemma 2.7** *Let the assumptions of theorem 2.1 hold. Then,*
*(i)* $\{Ax^t\}$, $\{Bz^t\}$ *and* $\{H^tp^t\}$ *converge.*
*(ii) The limit of* $\{H^tp^t\}$ *is an optimal dual for problem (1).*

**Proof:** Let $(\beta, \varrho)$ be an accumulation point of the bounded sequence $\{(Bz^t, H^tp^t)\}$. We prove first that $\varrho$ is an optimal dual for problem (1), then we show that $(\beta, \varrho)$ is unique. The convergence of $\{Ax^t\}$ then follows from the convergence of $\{Bz^t\}$ and part (i) of lemma 2.6.

The sequence $\{(Ax^t, H^tp^t, Bz^t)\}$ is also bounded. Then there exists a subsequence $\{(Ax^j, H^jp^j, Bz^j)\}$, $j \in \mathcal{J}$, which converges to $(\alpha, \varrho, \beta)$, such that $\alpha + b = \beta$.

We add (11), (12) and (14) for iterates with indices in $\mathcal{J}$. We have

$$\begin{aligned} &G_1(x^*) + G_2(z^*) + {p^{j+1}}^T H^j(Ax^{j+1} + b - Bz^{j+1}) + [B(z^{j+1} - z^j)]^T H^j(Ax^{j+1} - Ax) \quad \leq \\ &G_1(x) + G_2(z) + p^{*T}(p^{j+1} - p^j) + {p^{j+1}}^T H^j(Ax + b - Bz) \\ &\qquad\qquad\qquad \forall j \in \mathcal{J}\ ,\ \forall x \in \mathbb{R}^n,\ \forall z \in \mathbb{R}^s \end{aligned} \tag{31}$$

Taking the limit, and using the facts that $\{Ax^j\}$ is bounded and that $\{Ax^j + b - Bz^j\} \longrightarrow 0$, $\{H^jp^{j+1}\} \longrightarrow \varrho$ and $\{B(z^j - z^{j+1})\} \longrightarrow 0$, we obtain

$$G_1(x^*) + G_2(z^*) \leq G_1(x) + G_2(z) + \varrho^T(Ax + b - Bz) \quad \forall x \in \mathbb{R}^n,\ \forall z \in \mathbb{R}^s \tag{32}$$

i.e. $\varrho$ is an optimal dual for problem (1).

We will now show that the point $(\beta, \varrho)$ is unique. Suppose $\{(Bz^t, H^tp^t)\}$ has another accumulation point $(\beta_1, \varrho_1)$. Then there exists a subsequence $\left\{(x^k, z^k, p^k)\right\}$, $k \in \mathcal{K}$, such that $\left\{(Ax^k + b, H^kp^k)\right\} \longrightarrow (\beta_1, \varrho_1)$, with $\varrho_1$ an optimal dual for problem (1), and also $\left\{G_1(x^k) + G_2(z^k)\right\} \longrightarrow G_1(x^*) + G_2(z^*)$. We will now retrace our analysis and show that $\varrho_1 = \varrho$ and $\beta_1 = \beta$.

Since $\varrho_1$ is an optimal dual, we have

$$G_1(x^*) + G_2(z^*) \leq G_1(x^{t+1}) + G_2(z^{t+1}) + {\varrho_1}^T(Ax^{t+1} + b - Bz^{t+1}) \tag{33}$$

We substitute $x^k$ for $x$ in (11), and get

$$G_1(x^{t+1}) + [p^{t+1} + B(z^{t+1} - z^t)]^T H^tA(x^{t+1} - x^k) \leq G_1(x^k) \quad \forall k \in \mathcal{K} \tag{34}$$

Substituting $z^k$ for $z$ in (12) yields

$$G_2(z^{t+1}) - {p^{t+1}}^T H^tB(z^{t+1} - z^k) \leq G_2(z^k) \quad \forall k \in \mathcal{K} \tag{35}$$

Adding (33), (34) and (35) yields

$$\begin{aligned} G_1(x^*) + G_2(z^*) + (H^t p^{t+1} - \varrho_1)^T (Ax^{t+1} + b - Bz^{t+1}) &\leq \\ G_1(x^k) + G_2(z^k) + p^{t+1^T} H^t (Ax^k + b - Bz^k) + [B(z^{t+1} - z^t)]^T H^t (Ax^k - Ax^{t+1}) & \end{aligned} \tag{36}$$

After taking the limit over $k \in \mathcal{K}$ and using the fact that $\left\{Ax^k\right\} \longrightarrow \beta_1 - b$ and $\left\{Bz^k\right\} \longrightarrow \beta_1$, the right hand side becomes

$$G_1(x^*) + G_2(z^*) + [B(z^{t+1} - z^t)]^T H^t (\beta_1 - b - Ax^{t+1})$$

and, with the use of (8), inequality (36) becomes

$$(p^{t+1} - p^t)^T [H^t p^{t+1} - \varrho_1] + [B(z^{t+1} - z^t)]^T H^t (Bz^{t+1} - \beta_1) + (p^{t+1} - p^t)^T H^t B(z^{t+1} - z^t) \leq 0$$

which is analogous to (19). Repeating the analysis from there on, we can show that the sequence

$$\left\{ \left\| p^t - (H^t)^{-1} \varrho_1 \right\|_{H^t}^2 + \left\| Bz^t - \beta_1 \right\|_{H^t}^2 \right\}$$

analogous to $\{\alpha^t\}$ of (21), is convergent. Then all its subsequences must have the same limit. For the subsequence with indices $k \in \mathcal{K}$, $\left\{(Bz^k, H^k p^k)\right\} \longrightarrow (\beta_1, \varrho_1)$, and thus the sequence has limit 0. Then, for the subsequence with $j \in \mathcal{J}$, we must have $\{(Bz^j, H^j p^j)\} \longrightarrow (\beta_1 \varrho_1)$. ∎

We now show how a primal solution for problem (1) can be obtained by solving two minimization problems which include in the objective the limit of $\{(Bz^t, H^t p^t)\}$. We call a function $\varphi : \mathbb{R}^m \to \mathbb{R}_+$ positive definite in case $\varphi(x) = 0$ if and only if $x = 0$. Then we have the following result.

**Lemma 2.8** *Let $\{(Bz^t, H^t p^t)\}$ converge to $(\beta, \varrho)$. Let $\tilde{x}$ solve*

$$\min_x \quad G_1(x) + \varrho^T Ax + \varphi_1(Ax + b - \beta)$$

*and let $\tilde{z}$ solve*

$$\min_z \quad G_1(z) - \varrho^T Bz + \varphi_2(\beta - Bz)$$

*in which $\varphi_1$ and $\varphi_2$ are continuous positive definite functions. Then $(\tilde{x}, \tilde{z})$ solves (1).*

**Proof :** Since $\tilde{x}$ is a minimizer, we have for all iterates $x^t$

$$G_1(\tilde{x}) + \varrho^T A\tilde{x} + \varphi_1(A\tilde{x} + b - \beta) \leq G_1(x^t) + \varrho^T Ax^t + \varphi_1(Ax^t + b - \beta) \tag{37}$$

Similarly we have for all iterates $z^t$

$$G_2(\tilde{z}) - \varrho^T B\tilde{z} + \varphi_2(\beta - B\tilde{z}) \leq G_2(z^t) - \varrho^T Bz^t + \varphi_2(\beta - Bz^t) \tag{38}$$

Combining (37) and (38) yields

$$\begin{aligned} G_1(\tilde{x}) + G_2(\tilde{z}) + \varrho^T (A\tilde{x} + b - B\tilde{z}) + \varphi_1(A\tilde{x} + b - \beta) + \varphi_2(\beta - B\tilde{z}) &\leq \\ G_1(x^t) + G_2(z^t) + \varrho^T (Ax^t + b - Bz^t) + \varphi_1(Ax^t + b - \beta) + \varphi_2(\beta - Bz^t) & \end{aligned}$$

Let now $(x^*, z^*)$ be a primal solution for (1). By lemma 2.6 and the continuity of $\varphi_1$ and $\varphi_2$, in the limit the above inequality becomes

$$G_1(\tilde{x}) + G_2(\tilde{z}) + \varrho^T(A\tilde{x} + b - B\tilde{z}) + \varphi_1(A\tilde{x} + b - \beta) + \varphi_2(\beta - B\tilde{z}) \leq G_1(x^*) + G_2(z^*) \quad (39)$$

Since $(x^*, z^*, \varrho)$ is a saddlepoint for (1), we have

$$G_1(x^*) + G_2(z^*) \leq G_1(\tilde{x}) + G_2(\tilde{z}) + \varrho^T(A\tilde{x} + b - B\tilde{z}) \quad (40)$$

Combining (39) and (40) shows that

$$\varphi_1(A\tilde{x} + b - \beta) + \varphi_2(\beta - B\tilde{z}) \leq 0$$

Since $\varphi_1$ and $\varphi_2$ are positive definite, this is possible only if $A\tilde{x} + b = \beta = B\tilde{z}$, i.e. if $(\tilde{x}, \tilde{z})$ is primal feasible for (1). Then from (39) and (40) it follows that

$$G_1(\tilde{x}) + G_2(\tilde{z}) = G_1(x^*) + G_2(z^*)$$

i.e. $(\tilde{x}, \tilde{z})$ is primal optimal for (1). ∎

A combination of the previous lemmas provides a proof for the master theorem.

**Proof of theorem 2.1:** Part (i) can be proven by combining part (i) of lemma 2.6 and part (i) of lemma 2.7. Part (ii) is proven in part (ii) of lemma 2.6 and part (iii) is proven in part (ii) of lemma 2.7. Finally, part (iv) is a special case of lemma 2.8, since a norm is a continuous positive definite function. ∎

We now display a sequence of spd matrices $\{H^t\}$ satisfying assumption 1.3. It is based on the following inequality, due to Weyl. We let $\lambda_k(A)$, $k = 1, \ldots, m$, denote the $k$-th largest eigenvalue of a real symmetric $m \times m$ matrix $A$, i.e. $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_m(A)$.

**Lemma 2.9** *[23, lemma 8.1.3] Let $A$ and $E$ be real symmetric $m \times m$ matrices. Then* $\lambda_k(A) + \lambda_m(E) \leq \lambda_k(A + E) \leq \lambda_k(A) + \lambda_1(E), \quad k = 1, \ldots, m.$

The following lemma describes the iterative construction.

**Lemma 2.10** *Given a scalar $L > 0$, construct $\{H^t\}$ as follows:*
Initialization. *Choose an spd matrix $H^0$ such that $L \leq \lambda_m(H^0)$.*
Iterative step. *Given $H^t$ such that $L \leq \lambda_m(H^t)$, pick a symmetric $m \times m$ matrix $E^t$ such that $0 \leq \lambda_k(E^t) \leq \lambda_m(H^t) - L$, $k = 1, \ldots, m$. Let $H^{t+1} = H^t - E^t$. Then,*
*(i) $H^t - H^{t+1}$ is positive semidefinite.*
*(ii) $H^{t+1}$ is spd, with eigenvalues satisfying $L \leq \lambda_k(H^{t+1})$, $k = 1, \ldots, m$.*

**Proof :** By construction, $H^t - H^{t+1} = E^t$ and $0 \leq \lambda_m(E^t)$, i.e. $H^t - H^{t+1}$ is positive semidefinite. This proves (i). For part (ii), lemma 2.9 shows that

$$\lambda_k(H^{t+1}) + \lambda_m(E^t) \leq \lambda_k(H^t) \leq \lambda_k(H^{t+1}) + \lambda_1(E^t), \quad k = 1, \ldots, m.$$

Since $\lambda_m(E^t)$ is nonnegative, $\lambda_k(H^{t+1}) \leq \lambda_k(H^t)$. On the other hand, since $\lambda_k(H^{t+1}) \geq \lambda_k(H^t) - \lambda_1(E^t)$, and, by construction, $\lambda_1(E^t) \leq \lambda_m(H^t) - L$, we have $\lambda_k(H^{t+1}) \geq \lambda_k(H^t) - \lambda_m(H^t) + L \geq L$. ∎

From Weyl's inequality we can infer that the condition $0 < L \leq \lambda_k(H^{t+1}) \leq \lambda_k(H^t)$, $k = 1, \ldots, m$, is ultimately necessary if assumption 1.3 is satisfied. If $\{H^t\}$ consists of diagonal matrices, this condition is also sufficient. In all cases, the condition implies that the eigenvalues of $H^t$ are uniformly bounded from above, and that for each $k = 1, \ldots, m$, the nonincreasing, bounded sequence $\{\lambda_k(H^t)\}$ converges to its infimum.

## 3 A simple example

We want to employ the ADI method to solve

$$\begin{array}{cc} \min\limits_{x_1, x_2, x_3, x_4 \geq 0} & x_1 - x_2 + x_3 + x_4 \\ \text{subject to} & x_1 - x_2 = 1 \\ & x_3 + x_4 = 1 \end{array} \tag{41}$$

This problem can be mapped to problem (1) in a variety of ways. Here we take

$$G_1(x_1, x_2, x_3) := \begin{cases} x_1 - x_2 + x_3 & \text{if } x_1 - x_2 = 1 \text{ and } x_1, x_2, x_3 \geq 0 \\ +\infty & \text{otherwise} \end{cases}$$

and

$$G_2(x_4) := \begin{cases} x_4 & \text{if } x_4 \geq 0 \\ +\infty & \text{otherwise} \end{cases}$$

and write problem (41) as

$$\begin{array}{cc} \min\limits_{x_1, x_2, x_3, x_4} & G_1(x_1, x_2, x_3) + G_2(x_4) \\ \text{subject to} & x_3 + x_4 = 1 \end{array}$$

This is in the form of problem (1), with $A = [0 \;\; 0 \;\; 1]$, $B = [-1]$ and $b = -1$. Observe that $A$ has not full column rank. Both $G_1$ and $G_2$ are convex, proper, closed functions. We take, for simplicity, $H^t = I$, $\forall t$. At each ADI iteration the two smaller problems are solved

$$\begin{array}{ccc} (x_1^{t+1}, x_2^{t+1}, x_3^{t+1}) \in & \operatorname*{argmin}\limits_{x_1, x_2, x_3 \geq 0} & x_1 - x_2 + x_3 + p^t x_3 + \frac{1}{2} \left\| x_3 + x_4^t - 1 \right\|_2^2 \\ & \text{subject to} & x_1 - x_2 = 1 \end{array}$$

$$x_4^{t+1} = \operatorname*{argmin}_{x_4 \geq 0} \;\; x_4 + p^t x_4 + \frac{1}{2} \left\| x_3^{t+1} + x_4 - 1 \right\|_2^2$$

and then the multipliers are updated

$$p^{t+1} = p^t + \left( x_3^{t+1} + x_4^{t+1} - 1 \right)$$

For initialization with arbitrary $(x_4^0, p^0) \geq 0$, the iterates are given in closed form by

$$\left( x_1^{t+1},\; x_2^{t+1},\; x_3^{t+1},\; x_4^{t+1},\; p^{t+1} \right) = \left( 1 + \alpha,\; \alpha,\; [-p^t]_+,\; 0,\; [p^t]_+ - 1 \right), \quad \text{for any } \alpha \geq 0$$

The sequence $\{(x_3^t, x_4^t, p^t)\}$ converges to $(1, 0, -1)$. For the choice $x_1^t = 1 + t$ and $x_2^t = t$, both sequences $\{x_1^t\}$ and $\{x_2^t\}$ are divergent. However, for any $t \geq 0$, the vector $(1 + t, t, 1, 0)$ is primal optimal for problem (41).

# 4 Corollaries

## 4.1 An interchange variant

There is enough flexibility in the variable penalty algorithm to allow for the rearrangement of the minimization problems per iteration.

**Corollary 4.1** *Theorem 2.1 is valid for the algorithm (6)–(9) with steps (6) and (7) interchanged.*

This is due to the symmetry of problem (1) in $x$ and $z$, in both the objective and the constraints. From a computational perspective, we would choose to solve first the problem for which it is easier to generate a good starting point, or which contains more data from the original problem.

## 4.2 Primal convergence

In certain cases the convergence result can be strengthened. In particular, when the matrices $A$ and $B$ have linearly independent columns, we can guarantee primal convergence, as well.

**Corollary 4.2** *Let the assumptions of theorem 2.1 hold, and let $A$ and $B$ have full column rank. Then $\{(x^t, z^t)\}$ is uniquely defined and converges to a primal solution of problem (1).*

**Proof:** If $A$ and $B$ have full column rank, then the objective function in problems (6) and (7) is strongly convex, and thus the minimizers are unique. By part (i) of theorem 2.1, $\{(Ax^t, Bz^t)\}$ converges, and the limit satisfies the constraints of problem (1). Since $A$ and $B$ have full column rank, $\{x^t\}$ and $\{z^t\}$ converge to $\bar{x}$ and $\bar{z}$, respectively, so that $A\bar{x} + b = B\bar{z}$, i.e. $(\bar{x}, \bar{z})$ is primal feasible for problem (1). Then,

$$G_1(\bar{x}) + G_2(\bar{z}) \geq \lim_t \left\{ G_1(x^t) + G_2(z^t) \right\}$$

where the right hand side is the optimal value, by part (ii) of theorem 2.1. $G_1$ and $G_2$ are both closed, convex functions, thus lower semicontinuous, and so we must have

$$G_1(\bar{x}) \leq \liminf_t \left\{ G_1(x^t) \right\} \quad \text{and} \quad G_2(\bar{z}) \leq \liminf_t \left\{ G_2(z^t) \right\}$$

Combining the three inequalities yields

$$G_1(\bar{x}) + G_2(\bar{z}) \leq \liminf_t \left\{ G_1(x^t) \right\} + \liminf_t \left\{ G_2(z^t) \right\} \leq \lim_t \left\{ G_1(x^t) + G_2(z^t) \right\} \leq G_1(\bar{x}) + G_2(\bar{z})$$

which shows that the value $G_1(\bar{x}) + G_2(\bar{z})$ is optimal. ∎

A weaker result holds if we just assume that $\{(x^t, z^t)\}$ has accumulation points. These are guaranteed to exist if the effective domains of $G_1$ and $G_2$ or the nonempty level sets of $G_1 + G_2$ are compact.

**Corollary 4.3** *Let the assumptions of theorem 2.1 hold. Then any accumulation point $(\bar{x}, \bar{z})$ of $\{(x^t, z^t)\}$ is a primal solution of problem (1).*

**Proof:** After going to a subsequence, if necessary, $\{(x^t, z^t)\}$ converges to $\{(\bar{x}, \bar{z})\}$. In view of part (i) of theorem 2.1, $A\bar{x} + b = B\bar{z}$, i.e. $(\bar{x}, \bar{z})$ is primal feasible. By arguments similar to those in the proof of the previous corollary, the value $G_1(\bar{x}) + G_2(\bar{z})$ is optimal. ∎

### 4.3 Finite termination

We can provide a sufficient condition for finite termination at an optimal point. Note that this corollary does not require assumption 1.3.

**Corollary 4.4** *Let assumptions 1.1 and 1.2 hold. If iterates $(x^t, z^t, p^t)$ and $(x^{t+1}, z^{t+1}, p^{t+1})$ are such that $(p^t, Bz^t) = (p^{t+1}, Bz^{t+1})$, then $(x^{t+1}, z^{t+1}, H^t p^{t+1})$ is a saddlepoint for problem (1).*

**Proof:** By (8), $p^t = p^{t+1}$ implies

$$Ax^{t+1} + b = Bz^{t+1} \tag{42}$$

i.e. $(x^{t+1}, z^{t+1})$ is feasible for (1). Let $(x^*, z^*)$ be optimal for (1). Then

$$G_1(x^*) + G_2(z^*) \le G_1(x^{t+1}) + G_2(z^{t+1}) \tag{43}$$

Using the hypothesis in (29) yields

$$G_1(x^{t+1}) + G_2(z^{t+1}) \le G_1(x^*) + G_2(z^*) \tag{44}$$

Combining (43) and (44) we get

$$G_1(x^{t+1}) + G_2(z^{t+1}) = G_1(x^*) + G_2(z^*) \tag{45}$$

i.e. the value of the objective function at $(x^{t+1}, z^{t+1})$ is optimal. We now add (11), (12) and (45) and use the hypothesis and (42). We obtain

$$G_1(x^*) + G_2(z^*) \le G_1(x) + G_2(z) + {p^{t+1}}^T H^t(Ax + b - Bz) \quad \forall x \in \mathbb{R}^n,\ \forall z \in \mathbb{R}^s \tag{46}$$

This shows that $H^t p^{t+1}$ is an optimal dual for (1). ∎

If this condition is used as a stopping rule in the example in section 3, the algorithm (for nonnegative start) terminates finitely at an optimal point.

## 5 An ADI decomposition scheme for block angular problems

Several classes of models in applied optimization, including multicommodity network flow [1] and stochastic scenario analysis [41], require solving convex block angular problems (**CBA**) of the following form:

$$\begin{array}{llcccccccc}
\min\limits_{x_{[1]}, x_{[2]}, \ldots, x_{[K]}} & f_{[1]}(x_{[1]}) & + & f_{[2]}(x_{[2]}) & + & \ldots & + & f_{[K]}(x_{[K]}) & & \\
\text{subject to} & A_{[1]}x_{[1]} & & & & & & & = & b_{[1]} \\
 & & & A_{[2]}x_{[2]} & & & & & = & b_{[2]} \\
 & & & & & \ddots & & & & \\
 & & & & & & & A_{[K]}x_{[K]} & = & b_{[K]} \\
 & D_{[1]}x_{[1]} & + & D_{[2]}x_{[2]} & + & \ldots & + & D_{[K]}x_{[K]} & \le & \mathtt{d} \\
 & \multicolumn{9}{c}{0 \le x_{[i]} \le u_{[i]}, \quad i = 1, \ldots, K.}
\end{array} \tag{47}$$

Each function $f_{[i]}$ is finite-valued, convex and continuous, and, in many applications, quadratic in the vector $x_{[i]}$, i.e. $f_{[i]}(x_{[i]}) = c_{[i]}{}^T x_{[i]} + x_{[i]}{}^T Q_{[i]} x_{[i]}$, in which $Q_{[i]}$ is a real symmetric positive semidefinite matrix. This includes the case of linear objective ($Q_{[i]} = 0$). We denote by $\psi(\cdot \mid \mathcal{B}_{[i]})$ the indicator function of the feasible set $\mathcal{B}_{[i]}$ for the block constraints

$$\mathcal{B}_{[i]} := \{x_{[i]} \mid A_{[i]} x_{[i]} = b_{[i]} \text{ and } 0 \le x_{[i]} \le u_{[i]}\}$$

The variables interact only in the coupling constraints, defined by the matrix $\left[D_{[1]} \dots D_{[K]}\right]$ and the shared resource vector $\mathtt{d}$.

For this class of problems we will derive a decomposition scheme based on the extended ADI method. In mapping **CBA** onto problem (1), we choose to incorporate the block constraints in the definition of $G_1$ and represent the coupling constraints in the definition of $G_2$ and as explicit linear equality constraints. (Other mappings are discussed in [30].) We define

$$G_1\left(x_{[1]}, \dots, x_{[K]}\right) := \sum_{i=1}^{K} f_{[i]}(x_{[i]}) + \psi(x_{[i]} \mid \mathcal{B}_{[i]})$$

and

$$G_2\left(d_{[1]}, \dots, d_{[K]}\right) := \begin{cases} 0 & \text{if } \sum_{i=1}^{K} d_{[i]} \le \mathtt{d} \\ +\infty & \text{otherwise} \end{cases}$$

Problem **CBA** can be written as

$$\begin{array}{ll} \min\limits_{x_{[i]},\, d_{[i]}} & G_1\left(x_{[1]}, \dots, x_{[K]}\right) + G_2\left(d_{[1]}, \dots, d_{[K]}\right) \\ \text{subject to} & D_{[i]} x_{[i]} = d_{[i]}, \quad i = 1, \dots, K \end{array}$$

which is in the form of problem (1), with the correspondences $b \leftarrow 0$, $A \leftarrow \mathcal{D} := \operatorname{diag}\left(D_{[1]}, \dots, D_{[K]}\right)$, and $B \leftarrow I$. A multiplier vector $p_{[i]}$ is paired with each block of constraints $D_{[i]} x_{[i]} = d_{[i]}$, $i = 1, \dots, K$. For reasons to be explained soon, we employ a diagonal positive penalty matrix $\Lambda^t$, common to all blocks. We let the diagonal matrix $\Lambda_K^t$ consist of $K$ copies of $\Lambda^t$ placed along the diagonal. We write in shorthand $x$ for the concatenation of the vectors $x_{[1]}, x_{[2]}, \dots, x_{[K]}$, and similarly for $d$ and $p$. At each iteration we solve two problems

$$\begin{array}{lll} x^{t+1} \in & \operatorname*{argmin}\limits_{x} & f(x) + {p^t}^T \Lambda_K^t \mathcal{D} x + \frac{1}{2}\left\|\mathcal{D}x - d^t\right\|^2_{\Lambda_K^t} \\ & \text{subject to} & \left.\begin{array}{c} A_{[i]} x_{[i]} = b_{[i]} \\ 0 \le x_{[i]} \le u_{[i]} \end{array}\right\} \; i = 1, \dots, K \end{array} \tag{48}$$

$$\begin{array}{lll} d^{t+1} = & \operatorname*{argmin}\limits_{d} & -{p^t}^T \Lambda_K^t d + \frac{1}{2}\left\|\mathcal{D}x^{t+1} - d\right\|^2_{\Lambda_K^t} \\ & \text{subject to} & \sum_{i=1}^{K} d_{[i]} \le \mathtt{d} \end{array} \tag{49}$$

Then we update the multipliers

$$p^{t+1} = p^t + \mathcal{D}x^{t+1} - d^{t+1} \tag{50}$$

and the penalty matrix $\Lambda^t$.

Problem (48) decomposes to the following $K$ block problems.

$$\begin{array}{rcc} x_{[i]}^{t+1} \in & \underset{x_{[i]}}{\operatorname{argmin}} & f_{[i]}(x_{[i]}) + {p_{[i]}^t}^T \Lambda^t D_{[i]} x_{[i]} + \frac{1}{2} \left\| D_{[i]} x_{[i]} - d_{[i]}^t \right\|_{\Lambda^t}^2 \\ & \text{subject to} & A_{[i]} x_{[i]} = b_{[i]}, \quad 0 \leq x_{[i]} \leq u_{[i]} \end{array} \tag{51}$$

Since $D_{[i]}$ may not have full column rank, $x_{[i]}^{t+1}$ may not be unique. This nonuniqueness can be dealt with by the convergence theory we developed in section 2. Problem (49) has a strongly convex objective and therefore it is uniquely solvable. Because of our choice of diagonal penalty, the solution can be expressed in closed form.

$$d_{[i]}^{t+1} = D_{[i]} x_{[i]}^{t+1} + p_{[i]}^t - \frac{1}{K} \left[ \sum_{i=1}^{K} \left( p_{[i]}^t + D_{[i]} x_{[i]}^{t+1} \right) - \mathtt{d} \right]_+ \tag{52}$$

Substitution in (50) yields

$$p_{[i]}^{t+1} = \frac{1}{K} \left[ \sum_{i=1}^{K} \left( p_{[i]}^t + D_{[i]} x_{[i]}^{t+1} \right) - \mathtt{d} \right]_+ \tag{53}$$

which shows that, for $t \geq 1$, the multipliers are equal across all blocks and nonnegative.

This is a *resource proximization* (**RP**) splitting [31], in which the activities $x_{[i]}^t$ always satisfy the block constraints and the target resource allocations $d_{[i]}^t$ always satisfy the coupling constraints. In the objective of problem (51) the vector $D_{[i]} x_{[i]}$, which reflects the consumption of the shared resource $\mathtt{d}$, is penalized by both price and proximal terms; the iterative adjustments (52) and (53) are such that, in the limit, consumption matches an optimal allocation. This is shown in the following theorem, which specializes the general theorem 2.1.

**Theorem 5.1** *Assume that* **CBA** *admits a Lagrangian saddlepoint, and that each function $f_{[i]}$ is either quadratic or has bounded level sets over the feasible set for the corresponding block constraints. Let $\lambda_k(\Lambda^t)$, the eigenvalues of the diagonal positive matrices $\{\Lambda^t\}$, ultimately satisfy $L \leq \lambda_k(\Lambda^{t+1}) \leq \lambda_k(\Lambda^t)$, for $L > 0$ given. Then any sequence $\{x^t\}$ produced by the algorithm (48)–(50) for arbitrary start $(p^0, d^0, \Lambda^0)$ is such that*

*(i)* $\sum_{i=1}^{K} \lim_t \left\{ D_{[i]} x_{[i]}^t \right\} \leq \mathtt{d}$.

*(ii)* $\sum_{i=1}^{K} \left\{ f_{[i]}(x_{[i]}^t) \right\}$ *converges to the optimal value for* **CBA**.

*(iii)* $\left\{ x_{[i]}^t \right\}$ *converges for all $i \in \{1, \ldots, K\}$ such that $D_{[i]}$ has full column rank.*

**Proof:** The functions $G_1$ and $G_2$ are convex and closed, by construction, and also proper, since **CBA** is solvable, by hypothesis. The assumptions on the solvability of **CBA** and on the $f_{[i]}$'s are sufficient to guarantee the solvability of problem (48). The objective in problem (49) is strongly convex and continuous, and therefore has compact level sets. The feasible region is a nonempty polyhedral set, thus closed, and its intersection with a nonempty level set is a compact set. By the Bolzano-Weierstass theorem, the infimum of the continuous objective over this intersection is attained; hence problem (49) is solvable. Also, by construction, the penalty matrices $\{\Lambda^t\}$ satisfy assumption 1.3. Thus all assumptions of theorem 2.1 are met.

By lemma 2.6 (i), $\left\{d^t_{[i]}\right\}$ converges, say to $d^*_{[i]}$. By part (i) of theorem 2.1, $\left\{D_{[i]}x^t_{[i]} - d^t_{[i]}\right\}$ converges to zero and therefore $\left\{D_{[i]}x^t_{[i]}\right\}$ converges to $d^*_{[i]}$. Since $\sum_{i=1}^K d^{t+1}_{[i]} \leq \mathtt{d}$, we must have $\sum_{i=1}^K d^*_{[i]} \leq \mathtt{d}$. This proves part (i). Part (ii) follows from part (ii) of theorem 2.1. Part (iii) follows from part (i). If all matrices $D_{[i]}$ have full column rank, $\{x^t\}$ converges to a primal solution of **CBA**, by corollary 4.2. ∎

Our definitions of $G_1$ and $G_2$ have resulted in a coarse grain decomposition algorithm for **CBA**: the first ADI problem decomposes into independent block problems which can even be solved in parallel, while the second ADI problem has a simple closed form solution. Appropriate definitions of $G_1$ and $G_2$ can lead to a fine grain (activity-level) decomposition scheme [14]. The choice of granularity depends on the architecture of the target computing environment: a coarse grain method may perform better in a cluster of workstations, while a fine grain one may be better suited to a massively parallel system.

In [31] we present computational results for an ultimately-fixed-penalty variant of the coarse grain **RP** decomposition on the Connection Machine 5 parallel supercomputer. The CM-5 can be viewed as a cluster of powerful processors linked by fast networks. On this system, the **RP** algorithm solved large-scale multicommodity network flow problems one to two orders of magnitude faster than the serial optimizer MINOS 5.4 on a DEC 5000 workstation.

# 6 Computational experiments

In the basic ADI method (3)-(5) a single penalty is used and is held fixed over all iterations. In this case the computational performance depends strongly on the value of the penalty. Experience on a variety of applications [16, chapter 5], [9, chapter 7], [17] has shown that if the penalty is chosen too small or too large the solution time can significantly increase. For certain simple problems an optimal value of $\lambda$ can be found by spectral techniques [16, chapter 1]. In the general case the choice of a good value of $\lambda$ is a question of considerable experimentation and of familiarity with the characteristics of the problem. In such cases an appropriate variable penalty heuristic can result in computational savings.

To illustrate this we apply both the fixed- and the variable-penalty methods to a problem from the ADI literature and compare results. We consider the Fermat-Weber problem

$$\min_{z \in \mathbb{R}^n} \sum_{i=1}^{K} a_i \left\| z - b_{[i]} \right\|_2 \tag{54}$$

in which the vectors $b_{[i]}$ and the weights $a_i > 0$ are given. For $n = 2$ the problem has a single-facility location interpretation: $b_{[i]}$ are shipment centers, represented as points in the plane; the sought minimizer is the location of the facility to be built, such that the sum of the transportation costs between the centers and the facility is minimized, where each cost is proportional to the euclidean distance.

Specialized algorithms for this problem are reviewed in [35]. To cast it in a format suitable for ADI, we introduce auxiliary vectors of variables $x_{[1]}, \ldots, x_{[K]}$, the combination of which plays the role of the $x$ variables, and rewrite it in a dual block angular form, as in [16, section 3.7.3]

$$\begin{array}{cc} \min\limits_{x_{[1]},\ldots,x_{[K]},z\in \mathbb{R}^n} & \sum_{i=1}^{K} a_i \left\| x_{[i]} \right\|_2 \\ \text{subject to} & x_{[i]} = z - b_{[i]}, \quad i = 1,\ldots,K \end{array}$$

We pair a multiplier vector $p_{[i]}$ with each block of constraints and use a separate penalty value $\lambda_i$ for each block. At each iteration we solve two strongly convex problems

$$x^{t+1} = \operatorname*{argmin}_{x} \sum_{i=1}^{K} \left[ a_i \left\| x_{[i]} \right\|_2 - x_{[i]}{}^T \lambda_i^t p_{[i]}^t + \frac{\lambda_i^t}{2} \left\| z^t - b_{[i]} - x_{[i]} \right\|_2^2 \right] \tag{55}$$

$$z^{t+1} = \operatorname*{argmin}_{z} \sum_{i=1}^{K} \left[ z^T \lambda_i^t p_{[i]}^t + \frac{\lambda_i^t}{2} \left\| z - b_{[i]} - x_{[i]}^{t+1} \right\|_2^2 \right] \tag{56}$$

Then we update the multipliers

$$p_{[i]}^{t+1} = p_{[i]}^t + (z^{t+1} - b_{[i]} - x_{[i]}^{t+1}) \tag{57}$$

and the penalties $\lambda_i$. Because of the proximal terms, both problems have unique, closed form solutions.

$$x_{[i]}^{t+1} = \left[ 1 - \frac{a_i}{\lambda_i^t \left\| \beta_{[i]}^t \right\|_2} \right]_+ \beta_{[i]}^t, \text{ for } \beta_{[i]}^t := z^t - b_{[i]} + p_{[i]}^t, \quad i = 1,\ldots,K \tag{58}$$

$$z^{t+1} = (\sum_{i=1}^{K} \lambda_i^t)^{-1} \sum_{i=1}^{K} \lambda_i^t \left[ b_{[i]} + x_{[i]}^{t+1} - p_{[i]}^t \right] \tag{59}$$

For a single penalty $\lambda$, the updates agree with (7.40)-(7.43) in [16, section 3.7.3]. In this case, by (59) and (57) we have that $\sum_{i=1}^{K} p_{[i]}^{t+1} = 0$, for $t \geq 0$, and thus (59) simplifies to

$$z^{t+1} = \frac{1}{K} \sum_{i=1}^{K} \left[ b_{[i]} + x_{[i]}^{t+1} \right], \qquad t \geq 1 \tag{60}$$

We note that the earlier theory in [16, chapter 3] cannot characterize the convergence of this iterative scheme, because the objective function in (54) is not strictly convex. In contrast, our corollary 4.2 guarantees that $\{z^t\}$ converges to a primal solution of the problem.

To assess the impact of the penalty value on performance, we generated 20 classes of data, with the number of points $K$ in $\{10, 15, 25, 50, 75\}$ and the dimension $n$ in $\{2, 4, 8, 16\}$. For each class we generated 49 random problems. The weights $a_i$ were uniformly distributed in $[1, 10]$, while the components of $b$ were uniformly distributed in $[10, 100]$.

In all runs we chose initial values $z^0 = p^0 = 0$. We terminated a run when all components of two successive $(z, p)$ iterates agreed to at least $D$ significant digits, for $D = 6$ and $D = 8$. All runs were done on an IBM RS-6000/590 workstation using double precision arithmetic. We also solved the problems with the special-purpose Weiszfeld algorithm as emended in [35], and compared results. The objective function values at termination agreed to $6 - 7$ digits.

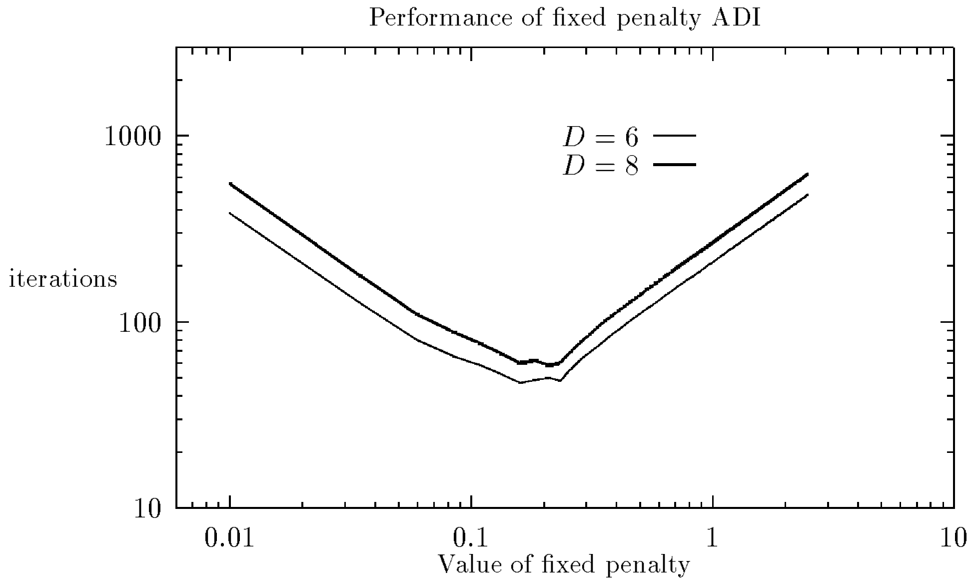


FIGURE 1: *Number of ADI iterations as a function of the fixed penalty value on a Fermat-Weber problem with $K = 15$, $n = 4$, for termination accuracy of $D = 6$ and $D = 8$ significant digits. Only a narrow range of penalty values results in low iteration count (below 100).*

In Figure 1 we display the number of iterations to termination of the fixed penalty algorithm on an example problem with $K = 15$ and $n = 4$, as the penalty ranges from 0.01 to 2.5. We observe that only a very narrow range of penalty values offers good performance: the algorithm terminates in at most a hundred iterations only if the penalty is in [0.06, 0.45] for $D = 6$ and in [0.08, 0.36] for $D = 8$. Performance deteriorates dramatically for values outside this interval. This can be attributed to the following: If $\lambda >> 1$, then $x_{[i]}^{t+1} \simeq \beta_{[i]}^{t}$ and therefore $z^{t+1} \simeq z^{t}$, i.e. the new estimate for $z$ is very close to the previous one. On the other hand, if $0 < \lambda << 1$, then $x_{[i]}^{t+1} \simeq 0$ and $z^{t+1} \simeq \frac{1}{K}\sum_{i=1}^{K} b_{[i]}$, i.e. $z^{t+1}$ defaults to the average of the observations, a poor estimate, in general, because it ignores the weights $a_i$. Small computational progress is made in both cases.

In this example the best penalty is 0.16 for 6-digit accuracy, resulting in 47 iterations, and 0.21 for 8-digit accuracy, resulting in 58 iterations. For penalty values in [0.01, 2.5], the median number of iterations was 263 for 6-digit accuracy, and 337 for 8-digit accuracy. Interchanging the order in which problems (55) and (56) are solved (with $x^0 = p^0 = 0$) yielded similar results: for 6-digit accuracy, the best count was 45, for $\lambda = 0.16$, and the median was 249; for 8-digit accuracy, the best count was 57, for $\lambda = 0.185$, and the median was 323.

Using the variable penalty heuristic we describe next, on the algorithm (55)–(57) without interchange, we solved this problem in 35 and 45 iterations, for $D = 6$ and $D = 8$, respectively.

We chose penalties $\lambda_i^0$ to make $z^1$ a weighted combination of the vectors $b_{[i]}$: specifically, $\lambda_i^0 = 2\, a_i / \left\| b_{[i]} \right\|_2$, which yielded $z^1 = 0.5\, (\sum_{i=1}^{K} a_i b_{[i]} / \left\| b_{[i]} \right\|_2) \;/\; (\sum_{i=1}^{K} a_i / \left\| b_{[i]} \right\|_2)$. In choosing the limiting value $L$ for the variable penalty, we considered the fact that, in the objective of problem (55), the original terms, with weights $a_i$, compete with the penalty terms, with weights $\lambda_i$. Thus, for balancing purposes, we let $L$ be a multiple of the average weight, $L = \theta\, \frac{1}{K}\sum_{i=1}^{K} a_i$,

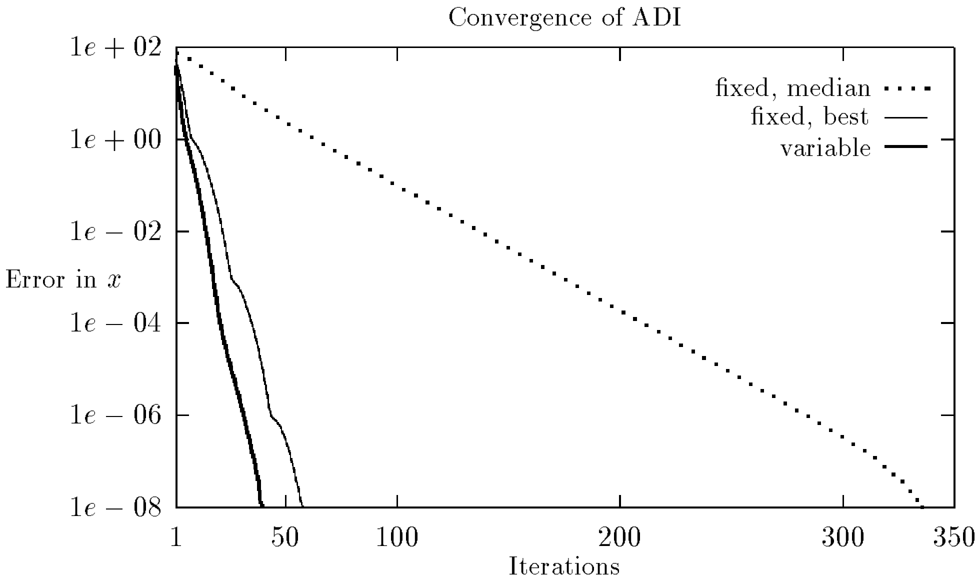


FIGURE 2: *$l_2$-error in the $z$ iterates for the Fermat-Weber problem with $K = 15$, $n = 4$, for the variable- and fixed-penalty ADI methods. The variable-penalty method converges faster than the best case of the fixed-penalty method.*

with $\theta$ chosen such that the initial penalties $\lambda_i^0$ lie on both sides of $L$. After experimentation we set $\theta = 0.075/n$. For the example problem this yielded $L = 0.092$. The initial penalties were in $[0.019,\ 0.234]$; nine were above $L$ and six were below. The penalties were updated every $T = 10$ iterations. If a penalty were initially below $L$, it was increased by a factor of 1.05 until it exceeded $L$. Otherwise, it was reduced by a factor of 0.98 down to $L$.

$$\lambda_i^t = \begin{cases} 1.05\,\lambda_i^{t-T}, & \text{if } \lambda_i^{t-T} < L \\ \max\{0.98\,\lambda_i^{t-T},\ L\}, & \text{otherwise} \end{cases}$$

The increasing update could thus be selected only a finite number of times. Thus assumption 1.3 was met.

In Figure 2 we display the iterative decrease in the error magnitude for the $z$ iterates in the example problem with $D = 8$. The thick line corresponds to the variable penalty algorithm, while the thin line corresponds to the fixed penalty algorithm with the best value $\lambda = 0.21$. The dotted line corresponds to the fixed penalty algorithm for $\lambda = 1.285$, which results in the median number of iterations among all values in the interval $[0.01,\ 2.5]$. The rate of convergence was almost linear for all three cases. The reduced count of iterations resulted from a larger decrease of the error per iteration: the average decrease in the error was 6% for the median, 32% for the best of fixed penalty and 42% for the variable penalty.

For this example we also ran the algorithm with all penalties fixed at the limit $L$. The performance was markedly worse than that of the variable case, since it took 63 iterations to termination for $D = 6$ and 84 iterations for $D = 8$.

In table 1 we compare the performance of this variable penalty method against fixed penalty, for an accuracy of at least 6 and 8 significant digits. The table reports the median number

| Performance of ADI | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of iterations | | | | | | | |
| $n$ | $K$ | variable | fixed penalty | | percentile | variable | fixed penalty | | percentile |
| | | penalty | best | median | fix $\leq$ var | penalty | best | median | fix $\leq$ var |
| 2 | 10 | 49 | 49 | 210 | 7.55 | 65 | 64 | 274 | 7.31 |
| | 15 | 55 | 47 | 211 | 6.90 | 72 | 62 | 283 | 6.90 |
| | 25 | 74 | 55 | 225 | 12.18 | 95 | 70 | 300 | 12.33 |
| | 50 | 87 | 60 | 222 | 21.71 | 118 | 79 | 293 | 22.49 |
| | 75 | 102 | 65 | 231 | 24.16 | 132 | 82 | 302 | 24.84 |
| 4 | 10 | 39 | 38 | 258 | 4.39 | 52 | 48 | 349 | 4.35 |
| | 15 | 42 | 39 | 249 | 4.98 | 56 | 50 | 334 | 5.24 |
| | 25 | 44 | 39 | 252 | 3.06 | 57 | 50 | 333 | 3.27 |
| | 50 | 48 | 43 | 259 | 3.73 | 62 | 53 | 332 | 4.12 |
| | 75 | 52 | 45 | 262 | 6.31 | 68 | 56 | 338 | 7.10 |
| 8 | 10 | 34 | 35 | 338 | 0.96 | 44 | 45 | 446 | 1.20 |
| | 15 | 36 | 37 | 340 | 0.73 | 46 | 46 | 446 | 0.90 |
| | 25 | 37 | 37 | 330 | 1.33 | 48 | 45 | 426 | 1.51 |
| | 50 | 38 | 39 | 342 | 1.14 | 49 | 48 | 439 | 1.43 |
| | 75 | 40 | 41 | 345 | 0.86 | 52 | 50 | 434 | 1.29 |
| 16 | 10 | 33 | 35 | 464 | 0.41 | 42 | 43 | 598 | 0.59 |
| | 15 | 33 | 35 | 431 | 0.37 | 43 | 44 | 563 | 0.65 |
| | 25 | 34 | 37 | 437 | 0.29 | 43 | 46 | 554 | 0.41 |
| | 50 | 36 | 38 | 465 | 0.20 | 45 | 47 | 584 | 0.33 |
| | 75 | 36 | 39 | 476 | 0.31 | 46 | 48 | 598 | 0.39 |
| | | 6 accurate digits | | | | 8 accurate digits | | | |

TABLE 1: *Number of iterations for the ADI method with fixed and variable penalty on the random Fermat-Weber problems.*

of iterations for the 49 problems in each class. We ran the fixed penalty algorithm with 100 penalty values equally spaced in the interval [0.01, 2.5]. In the column labeled 'best' we list the fewest number of iterations to termination; they correspond to the best choice of fixed penalty. The column labeled 'median' lists the median number of iterations over all fixed penalty choices in [0.01, 2.5]. Under 'percentile' we list the percentage of fixed penalty values which result in an iteration count no worse than that of the variable penalty, aggregated over all 49 problems in the corresponding class.

The table indicates that, as the dimension $n$ of the problem increases, the benefit of maintaining multiple varying penalties becomes more pronounced, as the fixed penalty percentile decreases at an almost quadratic rate for many cases. For sufficiently large problems, the percentile ranking of the variable penalty method is better than the 99th percentile.

# 7 Conclusions and future directions

In recent years there has been a renewed interest in both the theoretical and computational properties of the alternating directions method for optimization, especially in the framework of parallel computing. The basic method exhibits many desirable characteristics, such as convergence under mild assumptions, stability (due to the proximal terms) and flexibility in the implementation: the ADI problems may be solved inexactly, their order may be interchanged, the primal iterates may be relaxed and the starting point may be arbitrary. Another attractive feature, given today's diverse parallel computing systems, is the capability of the method to lead to both fine- and coarse-level decomposition algorithms for large scale problems, such as block angular ones.

In this article we have extended the ADI method along two directions: we characterized convergence in the absence of uniqueness of minimizers (absence of strong convexity, essentially) and in the presence of variable positive definite penalty. The first extension allowed us to derive a new decomposition scheme for the block angular problem. The second one can lead to the design of efficient heuristics for the acceleration of convergence.

In the future we plan to investigate further the computational benefits of variable penalty. For instance, in the examples we presented we employed only diagonal penalties; the theory allows general spd matrices $H^t$. A possible strategy is to choose $H^t$ such that the quadratic proximal terms are approximately diagonalized. Techniques for the local acceleration of linear convergence, such as Aitken's $\Delta^2$-method [25, section 5.9], may also be beneficial.

An open problem is convergence under partial updates in (8), i.e. when, in the computation of the new multipliers, the old value at iteration $t$ is used for some components of $(x, z)$ and the new value at iteration $t+1$ is used for the rest. These incomplete updates may be computationally attractive in a distributed environment where communication is expensive or the solution times for the subproblems in a decomposition vary widely. Another attractive option is the modification of the multiplier updates to include second order (Hessian) information. Such updates, although computationally expensive, may yield faster convergence.

When applied to linear block angular problems, the coarse grain ADI decomposition schemes require solving quadratic problems at each iteration. To overcome this computational drawback we may iteratively linearize the quadratic term, as done in [44, 33], [15, section 2.5.1] for the method of multipliers, or replace it with a piecewise linear local approximation, as done in the convex optimization methods in [29, 28].

On the theoretical level, an open issue is whether the convergence properties are preserved if the quadratic penalty is replaced by other penalty functions, such as Bregman's [45, 11], or the class of strongly convex functions employed in the auxiliary problem method [7].

# References


[1] A.A. Assad. Multicommodity network flows: A survey. *Networks*, 8:37–91, 1978.

[2] D.P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12:133–145, 1976.

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

[4] J. Céa. *Lectures on Optimization – Theory and Algorithms*. Lectures on Mathematics and Physics, vol. 53. Tata Institute, Bombay, 1978. Distributed by Springer Verlag, Berlin.

[5] G. Cheng and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming, Series A*, 64(1):81–110, 1994.

[6] Y.C. Cheng. On the gradient-projection method for solving the nonsymmetric linear complementarity problem. *Journal of Optimization Theory and Applications*, 43:527–541, 1984.

[7] G. Cohen and D.L. Zhu. Decomposition coordination methods in large scale optimization problems: The nondifferentiable case and the case of augmented lagrangians. In J.B. Cruz, editor, *Advances in Large Scale Systems Theory and Applications*, pages 203–266. JAI Press, Greenwich, CT, 1983.

[8] J. Douglas and H.H. Rachford Jr. On the numerical solution of heat conduction problems in two- and three-space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

[9] J. Eckstein. *Splitting Methods for Monotone Operators with Applications to Parallel Optimization.* PhD thesis, Massachusetts Institute of Technology, Department of Civil Engineering, 1989. Available as Technical Report LIDS-TH-1877, MIT 1989.

[10] J. Eckstein. The alternating step method for monotropic programming on the Connection Machine CM-2. *ORSA Journal on Computing*, 5(1):293–318, 1993.

[11] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18:202–226, 1993.

[12] J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point method for maximal monotone operators. *Mathematical Programming, Series A*, 55(3):293–318, 1992.

[13] J. Eckstein and M. Ferris. Operator splitting methods for monotone linear complementarity problems. Technical Report TMC-239, Thinking Machines Corporation, Cambridge, Massachusetts, 1992.

[14] J. Eckstein and M. Fukushima. Some reformulations and applications of the alternating direction method of multipliers. In W.W. Hager, D.W. Hearn, and P.M. Pardalos, editors, *Large Scale Optimization: State of the Art*, pages 115–134. Kluwer Academic, Dordrecht, The Netherlands, 1994.

[15] W. Findeisen, F.N. Bailey, M. Brdyś, K. Malinowski, P. Tatjewski, and A. Woźniak. *Control and Coordination in Hierarchical Systems.* IIASA series. John Wiley & Sons, 1980.

[16] M. Fortin and R. Glowinski, editors. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems.* North–Holland, Amsterdam, 1983.

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

[18] D. 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-Valued Problems*, pages 299–331. North–Holland, Amsterdam, 1983.

[19] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2:17–40, 1976.

[20] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* Society for Industrial and Applied Mathematics, 1989.

[21] Ye. G. Golshtein. The block method of convex programming. *Soviet Mathematics Doklady*, 33:584–587, 1986.

[22] Ye. G. Golshtein. A general approach to decomposition of optimization systems. *Soviet Journal of Computer and Systems Sciences*, 25(3):105–114, 1987.

[23] G.H. Golub and C.F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, second edition, 1989.

[24] S.P. Han and G. Lou. A parallel algorithm for a class of convex problems. *SIAM Journal on Control and Optimization*, 26:345–355, 1988.

[25] P. Henrici. *Elements of Numerical Analysis.* John Wiley, New York, 1966.

[26] M. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[27] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1993.

[28] P.V. Kamesam and R.R. Meyer. Multipoint methods for separable nonlinear networks. *Mathematical Programming Study*, 22:185–205, 1984.

[29] C.Y. Kao and R.R. Meyer. Secant approximation methods for convex optimization. *Mathematical Programming Study*, 14:143–162, 1981.

[30] S. Kontogiorgis. *Alternating Directions Methods for the Parallel Solution of Large-Scale Block-Structured Optimization Problems.* PhD thesis, University of Wisconsin–Madison, Department of Computer Sciences, 1994. Available as Report MP 94-13.

[31] S. Kontogiorgis, R. De Leone, and R.R. Meyer. Alternating direction splittings for block angular parallel optimization. *Journal of Optimization Theory and Applications*, 90(1), 1996. To appear.

[32] P.L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[33] J.M. Mulvey and A. Ruszczyński. A diagonal quadratic approximation method for large scale linear programs. *Operations Research Letters*, 12:205–215, 1992.

[34] J.M. Mulvey and H. Vladimirou. Solving multistage stochastic networks: An application of scenario aggregation. *Networks*, 21(6):619–643, 1990.

[35] L.M. Ostresh Jr. On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research*, 26:597–609, 1978.

[36] D.W. Peaceman and H.H. Rachford Jr. The numerical solution of parabolic and elliptic differential equations. *SIAM Journal*, 3:28–42, 1955.

[37] B.T. Polyak. *Introduction to Optimization.* Optimization Software, Inc., Publications Division, 1987.

[38] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, 1969.

[39] R.T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12:555–562, 1973.

[40] R.T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[41] R.T. Rockafellar and R. J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.

[42] J.E. Spingarn. Partial inverse of a monotone operator. *Applied Mathematics and Optimization*, 10:247–265, 1983.

[43] J.E. Spingarn. Applications of the method of partial inverses to convex programming: Decomposition. *Mathematical Programming*, 32:199–223, 1985.

[44] 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*, 15:285–309, 1975.

[45] P. Tseng and D.P. Bertsekas. On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming*, 60:1–19, 1993.