



PB93-228740

NTIS[®]
Information is our business.

ALTERNATING DIRECTION MULTIPLIER DECOMPOSITION OF CONVEX PROBLEMS

DISSEMINATION STATEMENT 2
Approved for public release;
Distribution Unlimited

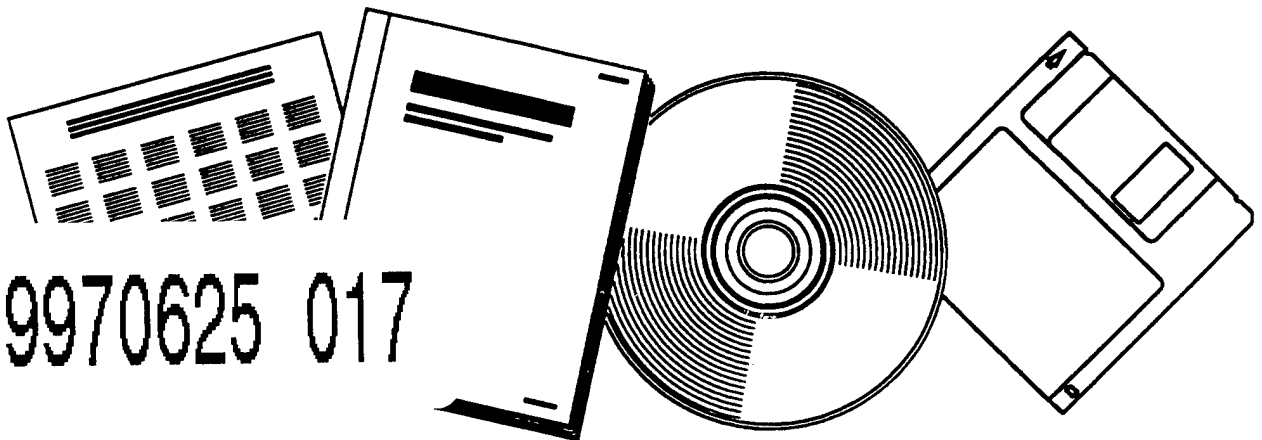
THINKING MACHINES CORP.
CAMBRIDGE, MA

OCT 91

DTIC QUALITY INSPECTED 3

19970625 017

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service





PB93-228740

Alternating Direction Multiplier Decomposition of Convex Problems

J. Eckstein

DTIC QUALITY INSURANCE

Thinking Machines Corporation

Technical Report Series

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

Tmc-211

REPORT DOCUMENTATION PAGE



PB93-228740

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE Oct. 1991	3. REPORT TYPE AND DATES COVERED Technical	
4. TITLE AND SUBTITLE Alternating direction multiplier decomposition of convex programs			5. FUNDING NUMBERS	
6. AUTHOR(S) Jonathan Eckstein			NSF ECS-8519058 Army research office DAAL03-86-K-0171	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Thinking Machines Corp. 245 First Street Cambridge, MA 02142-1264			8. PERFORMING ORGANIZATION REPORT NUMBER TMC 211	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) NSF 1800 G street NW Washington DC 20550 Dept of Army The Pentagon, Washington, DC 20310			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This paper focuses on two kinds of problems, monotropic programs and block-seperable problems. For block-seperable problems, the algorithm obtained bears some resemblance to an earlier method due to Spingarn, But solves a larger number of simpler subproblems at each iteration. Its fundamental operation is projection onto the epigraph of a convex function. For monotropic programs, one obtains a compact method that has some interesting properties when specialized to lenear programming, and, for quadratic problems, has been shown to be competitive in the massively parallel environment.				
14. SUBJECT TERMS Optimization			15. NUMBER OF PAGES 23	
17. SECURITY CLASSIFICATION OF REPORT unclassified			16. PRICE CODE	
18. SECURITY CLASSIFICATION OF THIS PAGE SAR		19. SECURITY CLASSIFICATION OF ABSTRACT SAR		20. LIMITATION OF ABSTRACT SAR

Alternating Direction Multiplier Decomposition of Convex Programs

Jonathan Eckstein

*Mathematical Sciences Research
Thinking Machines Corporation
245 First Street
Cambridge, MA 02142*

October, 1991

Abstract

The alternating direction method of multipliers, and extensions thereof, can be used to derive potentially parallel decomposition algorithms for convex programming. This paper focuses on two kinds of problems, monotropic programs (including linear programs) and block-separable problems. For block-separable problems, the algorithm obtained bears some resemblance to an earlier method due to Spingarn, but solves a larger number of simpler subproblems at each iteration. Its fundamental operation is projection onto the epigraph of a convex function. For monotropic programs, one obtains a compact method that has some interesting properties when specialized to linear programming, and, for quadratic problems, has been shown to be competitive in practice in a massively parallel environment.

Some of the research described here was performed at MIT and supported by the Army Research Office under grant number DAAL03-86-K-0171, and by the National Science Foundation under grant number ECS-8519058. It was supervised by Dimitri P. Bertsekas, for whose support I am grateful.

1. Introduction

Consider the convex programming problem

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) + g(\mathbf{z}) \\ & \text{such that} && \mathbf{M}\mathbf{x} = \mathbf{z} \quad , \end{aligned} \quad (1)$$

where $f: \mathcal{R}^n \rightarrow (-\infty, +\infty]$ and $g: \mathcal{R}^s \rightarrow (-\infty, +\infty]$ are closed proper convex, and \mathbf{M} is some $s \times n$ matrix. The *alternating direction method of multipliers* for (1) is the given by the recursions

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \frac{\lambda}{2} \|\mathbf{M}\mathbf{x} - \mathbf{z}^k\|^2 \right\} \\ \mathbf{z}^{k+1} &= \arg \min_{\mathbf{z}} \left\{ g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z} \rangle + \frac{\lambda}{2} \|\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}\|^2 \right\} \\ \mathbf{p}^{k+1} &= \mathbf{p}^k + \lambda (\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \quad , \end{aligned} \quad (2)$$

where λ is a given positive scalar. This method resembles the conventional Hestenes-Powell method of multipliers for (1) (see for example [1]), except that it minimizes the augmented Lagrangian function

$$L_{\lambda}(\mathbf{x}, \mathbf{z}, \mathbf{p}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{p}, \mathbf{M}\mathbf{x} - \mathbf{z} \rangle + \frac{\lambda}{2} \|\mathbf{M}\mathbf{x} - \mathbf{z}\|^2 \quad (3)$$

first with respect to \mathbf{x} , and then with respect to \mathbf{z} , rather than with respect to both \mathbf{x} and \mathbf{z} simultaneously. Thus, \mathbf{x} and \mathbf{z} are effectively decoupled, and the method avoids difficulties arising from the non-separable cross term $2\mathbf{z}^T \mathbf{M}\mathbf{x}$ arising from the quadratic penalty in (3). However, the method still retains many of the theoretical convergence advantages of the method of multipliers over algorithms using the simple Lagrangian $L_0(\mathbf{x}, \mathbf{z}, \mathbf{p}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{p}, \mathbf{M}\mathbf{x} - \mathbf{z} \rangle$.

The alternating direction method of multipliers was first introduced by Glowinski and Marocco [11], and by Gabay and Mercier [9]. Fortin and Glowinski [7] furthered the theory of the method, and Gabay [8] demonstrated its relationship to the Douglas-Rachford splitting procedure for monotone operators [13]. Eckstein and Bertsekas [4] have combined this relationship with some proximal point analysis to obtain a generalization of the algorithm. A convergence theorem for this generalized method is restated in Section 2 of this paper, and forms the basis of the following

analysis. Alternative analyses of the alternating direction method of multipliers may be found in [10] and [2, pp. 253-261].

This paper presents two ways of transforming convex programs into the form (1), and then studies the algorithms resulting from applying (2) and its generalizations. In both cases, significant decomposition of the original problem and extensive opportunities for parallelism result.

The first transformation, analyzed in Section 3 of this paper, is applicable to *monotropic programs* [16], that is, separable convex programs having only linear constraints. It results in an algorithm sometimes called the "alternating step method." This algorithm has an interesting interpretation when applied to linear programming, and has been shown to be computationally competitive for certain kinds of quadratic-cost problems on massively parallel computer architectures [6]. No complete derivation of it has been published, although a preliminary analysis (based on the current author's early research) appears in [2, p. 254].

The second transformation, presented in Section 4, is for block-separable problems of the sort studied by Spingarn [18]. The approach given here yields a method that solves a much larger number of simpler subproblems at each iteration, each with far fewer points of nondifferentiability. This algorithm is called an *epigraphic projection method* because the basic calculation in each subproblem is projection onto the epigraph [14, p. 23] of a convex function. A preliminary, unpublished analysis of this epigraphic projection method appears in [3].

Section 5 of this paper gives some closing observations.

2. A Generalized Alternating Direction Method of Multipliers

In this paper, $\| \cdot \|$ denotes the usual euclidean norm in \mathfrak{R}^n , and $\mathbf{a} \approx_{\epsilon} \mathbf{b}$ will be a shorthand for $\| \mathbf{a} - \mathbf{b} \| \leq \epsilon$. Convex analysis notation will be adopted from [14].

The convex program (1) can also be expressed as

$$\underset{\mathbf{x} \in \mathfrak{R}^n}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) \quad , \quad (4)$$

and, attaching a vector of multipliers $\mathbf{p} \in \mathfrak{R}^s$ to the constraints in (1), one obtains a symmetrical dual problem

$$\underset{\mathbf{p} \in \mathfrak{R}^s}{\text{minimize}} \quad f^*(-\mathbf{M}^T \mathbf{p}) + g^*(\mathbf{p}) \quad , \quad (5)$$

where the asterisk denotes the convex conjugacy operation [14]. A pair $(\mathbf{x}^*, \mathbf{p}^*) \in \mathfrak{R}^n \times \mathfrak{R}^s$ is said to be a *Kuhn-Tucker pair* for (1) or (4) if $-\mathbf{M}^T \mathbf{p}^* \in \partial f(\mathbf{x}^*)$ and $\mathbf{p}^* \in \partial g(\mathbf{M}\mathbf{x}^*)$, where “ ∂ ” denotes the subgradient mapping. It is a basic exercise in convex analysis to show that $(\mathbf{x}^*, \mathbf{p}^*)$ is a Kuhn-Tucker pair if and only if \mathbf{x}^* is optimal for (1) and (4), and \mathbf{p}^* is optimal for (5).

We now present a convergence theorem for a generalization of the alternating direction method of multipliers (2); a proof may be found in [4].

Theorem 1. Consider a convex program in the form (1) or (4), where \mathbf{M} has full column rank. Let $\mathbf{p}^0, \mathbf{z}^0 \in \mathfrak{R}^s$, and suppose we are given some scalar $\lambda > 0$, nonnegative summable sequences $\{\mu_k\}$ and $\{\nu_k\}$, and

$$\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2) \quad , \quad 0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 \quad .$$

Suppose $\{\mathbf{x}^k\}$, $\{\mathbf{z}^k\}$, and $\{\mathbf{p}^k\}$ conform, for all $k \geq 0$, to

$$\mathbf{x}^{k+1} \underset{\mu_k}{\approx} \underset{\mathbf{x}}{\text{arg min}} \left\{ f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \frac{\lambda}{2} \|\mathbf{M}\mathbf{x} - \mathbf{z}^k\|^2 \right\} \quad (6)$$

$$\mathbf{z}^{k+1} \underset{\nu_k}{\approx} \underset{\mathbf{z}}{\text{arg min}} \left\{ g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z} \rangle + \frac{\lambda}{2} \|\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{z}^k - \mathbf{z}\|^2 \right\} \quad (7)$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{z}^k - \mathbf{z}^{k+1}) \quad . \quad (8)$$

Then if (1) has a Kuhn-Tucker pair, $\{\mathbf{x}^k\}$ converges to a solution of (1) and $\{\mathbf{p}^k\}$ converges to a solution of the dual problem (5). Furthermore, $\{\mathbf{z}^k\}$ converges to $\mathbf{M} \left(\lim_{k \rightarrow \infty} \mathbf{x}^k \right)$. If (5) has no opti-

mal solution, then at least one of the sequences $\{p^k\}$ or $\{z^k\}$ is unbounded.

The algorithm (6)-(8) reduces to the alternating direction method of multipliers in the case that $\rho_k = 1$ for all k , and all minimizations are performed exactly.

3. Monotropic Programming

A *monotropic program* [16] is a convex programming problem taking the canonical form

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^n h_j(x_j) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b} \\ & && \mathbf{x} \in \mathfrak{R}^n, \end{aligned} \tag{9}$$

where the $h_j: \mathfrak{R} \rightarrow (-\infty, +\infty]$, $j=1, \dots, n$, are closed proper convex, \mathbf{A} is a real $m \times n$ matrix, and $\mathbf{b} \in \mathfrak{R}^m$.

Since the h_j can take on the value $+\infty$, they may impose implicit interval constraints on the x_j . Thus, by using slack variables and manipulating of the h_j , any convex optimization problem with a separable lower semicontinuous objective and a finite number of linear equality and inequality constraints can be converted to the form (9).

Given some monotropic program (9), define $d(i)$, the *degree of constraint i* , to be the number of nonzero elements in row i of \mathbf{A} . If \mathbf{A} is the node-arc incidence matrix of a network or graph, then this definition agrees with the usual notion of the degree of node i in the corresponding graph. Let $\mathbf{A}_{.j}$ denote column j of \mathbf{A} , and let \mathbf{A}_i denote row i of \mathbf{A} . The *surplus* or *residual* $r_i(\mathbf{x})$ of constraint i of the system $\mathbf{Ax} = \mathbf{b}$, with respect to the primal variables \mathbf{x} , is $b_i - \mathbf{A}_i \cdot \mathbf{x}$.

Now consider the conversion of (9) to the form (1). Here, f will be defined on \mathfrak{R}^n and g on \mathfrak{R}^{mn} . Index the components of vectors $\mathbf{z} \in \mathfrak{R}^{mn}$ as z_{ij} , where $1 \leq i \leq m$ and $1 \leq j \leq n$. Then let

$$f(\mathbf{x}) = \sum_{j=1}^n h_j(x_j) \tag{10}$$

$$C = \left\{ \mathbf{z} \in \mathfrak{R}^{mn} \mid \sum_{j=1}^n z_{ij} = b_i \quad \forall i = 1, \dots, m, \quad a_{ij} = 0 \Rightarrow z_{ij} = 0 \right\} \quad (11)$$

$$g(\mathbf{z}) = \begin{cases} 0, & \mathbf{z} \in C \\ +\infty, & \mathbf{z} \notin C \end{cases} \quad (12)$$

$$\mathbf{M} = \begin{bmatrix} \left[\text{diag}(\mathbf{A}_1) \right] \\ \left[\text{diag}(\mathbf{A}_2) \right] \\ \vdots \\ \left[\text{diag}(\mathbf{A}_m) \right] \end{bmatrix}. \quad (13)$$

where, for any vector $\mathbf{v} \in \mathfrak{R}^n$, $\text{diag}(\mathbf{v})$ denotes the $n \times n$ matrix \mathbf{D} with entries $d_{jj} = v_j$ along the diagonal, and zeroes elsewhere. It is easily confirmed that, under (10)-(13), problems (1) and (9) are equivalent. Furthermore, unless \mathbf{A} has a column that consists entirely of zeroes, the matrix \mathbf{M} of (13) has full column rank.

We now apply the generalized alternating direction method (6)-(8) to this reformulation of the monotropic program. By the separability of f and the form of \mathbf{M} , (6) can now be decomposed into n independent computations of the form

$$x_j^{k+1} \underset{\varepsilon_k}{\approx} \underset{x_j}{\arg \min} \left\{ h_j(x_j) + \left(\sum_{i=1}^m p_{ij}^k a_{ij} \right) x_j + \frac{\lambda}{2} \sum_{i: a_{ij} \neq 0} (a_{ij} x_j - z_{ij}^k)^2 \right\},$$

where $\varepsilon_k = \mu_k / \sqrt{n}$. Now consider the computation of \mathbf{z}^{k+1} in (7). For any $\mathbf{z} \in \mathfrak{R}^{mn}$, and $1 \leq i \leq m$, let $\mathbf{z}_i \in \mathfrak{R}^n$ be the subvector of \mathbf{z} with components $z_{ij}, j=1, \dots, n$. Let

$$C_i = \left\{ \mathbf{z}_i \in \mathfrak{R}^n \mid \sum_{j=1}^n z_{ij} = b_i, a_{ij} = 0 \Rightarrow z_{ij} = 0 \right\},$$

so that $C = C_1 \times \dots \times C_m$. Finding \mathbf{z}^{k+1} can be reduced to m independent calculations

$$\mathbf{z}_i^{k+1} \underset{\frac{v_k}{\sqrt{m}}}{\approx} \underset{\mathbf{z}_i \in C_i}{\operatorname{arg\,min}} \left\{ -\langle \mathbf{p}_i^k, \mathbf{z}_i \rangle + \frac{\lambda}{2} \sum_{j=1}^n (\rho_k a_{ij} x_j^{k+1} + (1 - \rho_k) z_{ij}^k - z_{ij})^2 \right\}.$$

To solve this problem exactly for each i , we attach a single Lagrange multiplier π_i^{k+1} to the constraint $\sum_{j=1}^n z_{ij} = b_i$ defining C_i . Some algebraic manipulations of the Karush/Kuhn-Tucker conditions yield

$$\pi_i^{k+1} = \frac{1}{d(i)} \left(\lambda \rho_k r_i(\mathbf{x}^{k+1}) - \sum_{j: a_{ij} \neq 0} p_{ij}^k \right) \quad (14)$$

$$z_{ij}^{k+1} = \begin{cases} \rho_k a_{ij} x_j^{k+1} + (1 - \rho_k) z_{ij}^k - \frac{p_{ij}^k + \pi_i^{k+1}}{\lambda}, & a_{ij} \neq 0 \\ 0, & a_{ij} = 0. \end{cases} \quad (15)$$

Finally, consider the multiplier update (8). For (i, j) such that $a_{ij} = 0$, one has simply $p_{ij}^{k+1} = p_{ij}^k$. For $a_{ij} \neq 0$,

$$\begin{aligned} p_{ij}^{k+1} &= p_{ij}^k + \lambda (\rho_k a_{ij} x_j^{k+1} + (1 - \rho_k) z_{ij}^k - z_{ij}^{k+1}) \\ &= p_{ij}^k + \lambda \left(-\frac{1}{\lambda} [p_{ij}^k + \pi_i^{k+1}] \right) \\ &= -\pi_i^{k+1}. \end{aligned}$$

It is therefore possible to eliminate $\{\mathbf{p}^k\}$ from the algorithm in favor of the lower-dimensional sequence $\{\boldsymbol{\pi}^k\} \subseteq \mathfrak{R}^n$, obtaining

$$z_{ij}^{k+1} = \begin{cases} \rho_k a_{ij} x_j^{k+1} + (1 - \rho_k) z_{ij}^k - \frac{r_i(\mathbf{x}^{k+1})}{d(i)}, & a_{ij} \neq 0 \\ 0, & a_{ij} = 0 \end{cases}$$

from (15). It now turns out that $\{z^k\}$ can also be replaced by a lower-dimensional sequence $\{y^k\} \subseteq \mathfrak{R}^n$, as follows:

Lemma 1. Assume that z^0 is of the form

$$z_{ij}^0 = \begin{cases} a_{ij}y_j^0 - \frac{r_i(y^0)}{d(i)}, & a_{ij} \neq 0 \\ 0, & a_{ij} = 0 \end{cases}$$

for some $y^0 \in \mathfrak{R}^n$. Define the sequence $\{y^k\}$ via the recursion

$$y^{k+1} = (1 - \rho_k)y^k + \rho_k x^{k+1} \quad \forall k \geq 0.$$

Then for all $k \geq 0$,

$$z_{ij}^k = \begin{cases} a_{ij}y_j^k - \frac{r_i(y^k)}{d(i)}, & a_{ij} \neq 0 \\ 0, & a_{ij} = 0. \end{cases} \quad (16)$$

Proof. Assuming (16) holds, we have, when $a_{ij} \neq 0$,

$$\begin{aligned} & a_{ij}y_j^{k+1} - \frac{1}{d(i)} r_i(y^{k+1}) \\ &= a_{ij}((1 - \rho_k)y_j^k + \rho_k x_j^{k+1}) - \frac{1}{d(i)} (b_i - \langle A_i, y^{k+1} \rangle) \\ &= a(1 - \rho_k)a_{ij}y_j^k + \rho_k a_{ij}x_j^{k+1} - \frac{1}{d(i)} ((1 - \rho_k)(b_i - \langle A_i, y^k \rangle) + \rho_k(b_i - \langle A_i, x^{k+1} \rangle)) \\ &= a(1 - \rho_k)[a_{ij}y_j^k - \frac{1}{d(i)} r_i(y^k)] + \rho_k a_{ij}x_j^{k+1} - \frac{\rho_k}{d(i)} r_i(x^{k+1}) \\ &= a(1 - \rho_k)z_{ij}^k + \rho_k a_{ij}x_j^{k+1} - \frac{\rho_k}{d(i)} r_i(x^{k+1}) \\ &= az_{ij}^{k+1}. \end{aligned}$$

So, (16) holds with k replaced by $k+1$, and the claim follows by induction. ■

In view of Lemma 1, the method (6)-(8) reduces to

$$\begin{aligned}
x_j^{k+1} &\underset{\varepsilon_k}{\underset{x_j}{\operatorname{arg\,min}}} \left\{ h_j(x_j) - \langle \mathbf{A}_{\cdot j}, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda}{2} \sum_{i: a_{ij} \neq 0} \left(a_{ij} x_j - \left(a_{ij} y_j^k + \frac{r_i(\mathbf{y}^k)}{d(i)} \right) \right) \right\} \\
\pi_i^{k+1} &= \pi_i^k + \frac{\lambda \rho_k}{d(i)} r_i(\mathbf{x}^{k+1}) \\
\mathbf{y}^{k+1} &= \rho_k \mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{y}^k .
\end{aligned} \tag{17}$$

Collecting terms in the squared expressions and assuming \mathbf{A} has no all-zero columns, one obtains

$$\begin{aligned}
x_j^{k+1} &\underset{\varepsilon_k}{\underset{x_j}{\operatorname{arg\,min}}} \left\{ h_j(x_j) - \langle \mathbf{A}_{\cdot j}, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda}{2} \|\mathbf{A}_{\cdot j}\|^2 \left[x_j - \left(y_j^k + \|\mathbf{A}_{\cdot j}\|^{-2} \sum_{i=1}^m \frac{a_{ij} r_i(\mathbf{y}^k)}{d(i)} \right) \right] \right\} \\
\pi_i^{k+1} &= \pi_i^k + \frac{\lambda \rho_k}{d(i)} r_i(\mathbf{x}^{k+1}) \\
\mathbf{y}^{k+1} &= \rho_k \mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{y}^k ,
\end{aligned} \tag{18}$$

where $\mathbf{y}^0 \in \mathfrak{R}^n$ and $\boldsymbol{\pi}^0 \in \mathfrak{R}^m$ are completely arbitrary. One may call this algorithm an *alternating step method* because of the alternating updates to the primal variables \mathbf{x}^k and \mathbf{y}^k and dual variables $\boldsymbol{\pi}^k$. The updates of the primal variables x_j and y_j are completely independent over j , and the updates of the dual variables π_i are likewise independent over i . Thus, the method has the potential for massive parallelism.

Now consider the convergence of (18). Recall that the dual of the monotropic program (9) may be written [16, Chapter 11]

$$\underset{\boldsymbol{\pi} \in \mathfrak{R}^m}{\operatorname{minimize}} \sum_{j=1}^n h_j^* \left(\langle \mathbf{A}_{\cdot j}, \boldsymbol{\pi} \rangle \right) - \mathbf{b}^\top \boldsymbol{\pi} , \tag{19}$$

where h_j^* denotes the convex conjugate of h_j .

Theorem 2. Let $\{\varepsilon_k\} \subseteq [0, \infty)$ be summable and suppose $0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2$. If \mathbf{A} has no identically zero columns and both the monotropic program (9) and its dual (19) have optimal solutions with finite objective value, then any sequences $\{\mathbf{x}^k\}$ and $\{\boldsymbol{\pi}^k\}$ conforming the recursion (18) re-

spectively converge to such optimal solutions. If (19) has no optimal solution, then, given any two sequences $\{\mathbf{y}^k\}$ and $\{\boldsymbol{\pi}^k\}$ conforming to (18), one of them must be unbounded.

Proof. Suppose $p_{ij} = -\pi_i$ whenever $a_{ij} \neq 0$. Then

$$-\mathbf{M}^T \mathbf{p} = \left[\begin{array}{c} - \sum_{\substack{i=1, \dots, m \\ a_{ij} \neq 0}} a_{ij} (-\pi_i) \\ \vdots \\ \vdots \end{array} \right]_{j=1}^n = \langle \mathbf{A}, \mathbf{j}, \boldsymbol{\pi} \rangle. \quad (20)$$

Applying basic convex analysis to g , we find that

$$\partial g(\mathbf{z}) = \begin{cases} \left\{ \mathbf{p} \in \mathfrak{R}^{mn} \mid a_{ij} \neq 0, a_{il} \neq 0 \Rightarrow p_{ij} = p_{il} \right\}, & \mathbf{z} \in C \\ \emptyset, & \text{otherwise} \end{cases}$$

$$\partial g^*(\mathbf{p}) = \begin{cases} C, & a_{ij} \neq 0, a_{il} \neq 0 \Rightarrow p_{ij} = p_{il} \\ \emptyset, & \text{otherwise} \end{cases}$$

$$g^*(\mathbf{p}) = \begin{cases} -\mathbf{b}^T \boldsymbol{\pi}, & \exists \boldsymbol{\pi} \in \mathfrak{R}^m: a_{ij} \neq 0 \Rightarrow p_{ij} = -\pi_i \\ +\infty, & \text{otherwise} \end{cases}$$

In view of (20), it follows that the dual problems (5) and (19) are equivalent in the sense that

$$f^*(-\mathbf{M}^T \mathbf{p}) + g^*(\mathbf{p}) = h_j^*(\langle \mathbf{A}, \mathbf{j}, \boldsymbol{\pi} \rangle) - \mathbf{b}^T \boldsymbol{\pi}$$

whenever $p_{ij} = -\pi_i$ for all i , and that if $p_{ij} \neq p_{il}$ for any i, j , and l such that a_{ij} and a_{il} are nonzero, then $f^*(-\mathbf{M}^T \mathbf{p}) + g^*(\mathbf{p}) = \infty$. Thus, if \mathbf{x}^* and $\boldsymbol{\pi}^*$ are respectively optimal for the monotropic program and its dual, then \mathbf{x}^* and $\mathbf{z}^* = \mathbf{M}\mathbf{x}^*$ are optimal for (1) and \mathbf{p}^* defined by $p_{ij}^* = -\pi_i^*$ is optimal for (5), so $(\mathbf{x}^*, \mathbf{p}^*)$ form a Kuhn-Tucker pair in the sense of Theorem 1. Since \mathbf{A} has no zero columns, \mathbf{M} has full column rank, and Theorem 1 with $\mu_k = \sqrt{n}\varepsilon_k$ and $\nu_k = 0$ for all k gives convergence of $\{\mathbf{x}^k\}$ and $\{\mathbf{p}^k\}$ in algorithm (6)-(8). By the derivation above and the equivalence of the primal and dual problems, this implies the convergence of $\{\mathbf{x}^k\}$ and $\{\boldsymbol{\pi}^k\}$ to their respective optima (note that nothing is said in this case about convergence of $\{\mathbf{y}^k\}$). When the monotropic

dual (19) does not have a finite optimum, then the dual equivalence already shown implies (5) has none either, and one can again invoke Theorem 1. ■

The method (17) is based on an underlying alternating direction iteration (6)-(8) in which some variables, namely $\{\mathbf{p}^k\}$ and $\{\mathbf{z}^k\}$, have dimension mn . However, it reduces to a form in which all the variables have dimension either m or n . A convergence proof involving only the lower-dimensional sequences would be very appealing, but so far none has been found.

Consider now the (completely general) linear programming problem

$$\begin{aligned} & \text{minimize} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \\ & && \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{l} \in [-\infty, \infty]^n$, $\mathbf{u} \in (-\infty, \infty]^n$, and \mathbf{A} is an $m \times n$ real matrix. This can easily be converted to the form (9) by combining the constraints $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$ with the objective function. It then turns out that the variables of (19) are the usual linear programming dual variables. The minimization step of (18) may then be done exactly in closed form, and the method reduces after some algebra to

$$\begin{aligned} x_j^{k+1} &= \max \left\{ l_j, \min \left\{ u_j, y_j^k + \frac{1}{\|\mathbf{A}_{.j}\|^2} \left[\sum_{i=1}^m \frac{a_{ij} r_i(\mathbf{y}^k)}{d(i)} \right] - \frac{c_j - \mathbf{A}_{.j}^\top \boldsymbol{\pi}^k}{\lambda} \right\} \right\} \\ \pi_i^{k+1} &= \pi_i^k + \frac{\lambda \rho_k}{d(i)} r_i(\mathbf{x}^{k+1}) \\ \mathbf{y}^{k+1} &= \rho_k \mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{y}^k \end{aligned} \tag{21}$$

Setting $\rho_k = 1$ for all k yields the even simpler method

$$\begin{aligned} x_j^{k+1} &= \max \left\{ l_j, \min \left\{ u_j, x_j^k + \frac{1}{\|\mathbf{A}_{.j}\|^2} \left[\sum_{i=1}^m \frac{a_{ij} r_i(\mathbf{x}^k)}{d(i)} \right] - \frac{c_j - \mathbf{A}_{.j}^\top \boldsymbol{\pi}^k}{\lambda} \right\} \right\} \\ \pi_i^{k+1} &= \pi_i^k + \frac{\lambda \rho_k}{d(i)} r_i(\mathbf{x}^{k+1}) \end{aligned} \tag{22}$$

Here, each primal variable x_j is first adjusted by terms proportional to its reduced cost $c_j - A_{.j}^T \pi^k$, and the amount of violation of constraints in which it is involved. Then, it is projected onto its valid range $[l_j, u_j] \cap \mathcal{R}$. After this, each dual variable is adjusted proportionally to any remaining constraint violation, and the process repeats. Primal feasibility, dual feasibility, and complementary slackness are not maintained; instead, all are gradually satisfied as the algorithm progresses towards a fixed point of the recursions (21) or (22). The algorithm is highly parallel; indeed, the only need for communication between processors responsible for the various primal and dual variables in the computation and dissemination of the surpluses $r_i(\mathbf{x}^k)$.

Linear programming algorithms of this type are also known to have linear convergence rate — see [5] and [3]. However, they were tested in [3] and found to converge very slowly in practice on the majority of NETGEN-generated [12] minimum cost flow problems. However, (18) may also be done in closed form when the cost function is quadratic, and for NETGEN-like quadratic problems, computational tests have been quite encouraging; see [6] for details of both the specialization of the algorithm and its performance.

4. Block-Separable Problems: an Epigraphic Projection Method

Let us now turn to a more general class of convex programs, as studied in [18, Section 4]. Let $\{1, \dots, n\}$ be partitioned into $d \geq 2$ nonempty subsets $N_j, j=1, \dots, d$, and let $n_j = |N_j|$ for all j . For any $\mathbf{x} \in \mathcal{R}^n$, let $\mathbf{x}_j \in \mathcal{R}^{n_j}$ denote the subvector of \mathbf{x} with components $x_q, q \in N_j$. A convex program is called *block-separable* if for some such partition, it takes the form

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^d h_{0j}(\mathbf{x}_j) \\ & \text{subject to} && \sum_{j=1}^d h_{ij}(\mathbf{x}_j) \leq 0 \quad i = 1, \dots, m, \end{aligned} \tag{23}$$

where the h_{ij} are convex functions on \mathcal{R}^{n_j} for $i=0, \dots, m$ and $j=1, \dots, d$. To simplify the analysis, assume that the h_{ij} are finite on \mathcal{R}^{n_j} for $i \geq 1$. For the h_{0j} , assume, slightly more generally than in [18], that

$$h_{0j}(x_j) = \begin{cases} \bar{h}_{0j}(x_j) & x_j \in C_j \\ +\infty & x_j \notin C_j \end{cases}, \quad (24)$$

where the \bar{h}_{0j} are finite and convex throughout \mathfrak{R}^{n_j} , and the $C_j \subseteq \mathfrak{R}^{n_j}$ are closed convex sets. Let $C = C_1 \times \dots \times C_d \subseteq \mathfrak{R}^n$.

We start by expressing (23) as a minimum problem for a sum of *three* convex functions. First, some notation: let the components of vectors $\mathbf{u} \in \mathfrak{R}^{md}$ be written u_{ij} , for $i=1, \dots, m$ and $j=1, \dots, d$. Similarly, denote the components of vectors $\mathbf{z} \in \mathfrak{R}^{mn}$ by z_{iq} , for $i=1, \dots, m$ and $q=1, \dots, n$, and let \mathbf{H} be the $mn \times n$ matrix taking vectors $\mathbf{x} \in \mathfrak{R}^n$ to $\mathbf{z} \in \mathfrak{R}^{mn}$, where $z_{iq} = x_q$ for all i and q . That is,

$$\mathbf{H} = \left. \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \right\} m \text{ times.}$$

For any $\mathbf{z} \in \mathfrak{R}^{mn}$, let \mathbf{z}_{ij} denote the subvector of \mathbf{z} consisting of the z_{iq} , $q \in N_j$. Now define functions

$$F_1: \mathbf{x} \in \mathfrak{R}^n \mapsto \sum_{j=1}^d h_{0j}(x_j)$$

$$F_2: \mathbf{u} \in \mathfrak{R}^{md} \mapsto \begin{cases} 0, & \sum_{j=1}^d u_{ij} = 0 \quad \forall i = 1, \dots, m \\ +\infty, & \text{otherwise} \end{cases}$$

$$F_3: (\mathbf{z}, \mathbf{u}) \in \mathfrak{R}^{mn} \times \mathfrak{R}^{md} \mapsto \begin{cases} 0, & h_{ij}(\mathbf{z}_{ij}) \leq u_{ij} \quad \forall i = 1, \dots, m, j = 1, \dots, d \\ +\infty, & \text{otherwise.} \end{cases}$$

F_1, F_2 , and F_3 are closed proper convex. Furthermore, the problem

$$\begin{aligned} & \text{minimize } F_1(\mathbf{x}) + F_2(\mathbf{u}) + F_3(\mathbf{H}\mathbf{x}, \mathbf{u}) \\ & \mathbf{x} \in \mathfrak{R}^n \\ & \mathbf{u} \in \mathfrak{R}^{md} \end{aligned} \quad (25)$$

is equivalent to (23) in that \mathbf{x} is feasible with finite objective value in (23) if and only if there exists $\mathbf{u} \in \mathfrak{R}^{md}$ such that $F_1(\mathbf{x}) + F_2(\mathbf{u}) + F_3(\mathbf{H}\mathbf{x}, \mathbf{u})$ is finite, in which case

$$F_1(\mathbf{x}) + F_2(\mathbf{u}) + F_3(\mathbf{H}\mathbf{x}, \mathbf{u}) = \sum_{j=1}^d h_{0j}(\mathbf{x}_j) .$$

Now consider how to convert (25) into the canonical form (4). One approach would be to set

$$\begin{aligned} f(\mathbf{x}, \mathbf{u}) &= F_1(\mathbf{x}) + F_3(\mathbf{H}\mathbf{x}, \mathbf{u}) \\ g(\mathbf{x}, \mathbf{u}) &= F_2(\mathbf{u}) \end{aligned}$$

and let \mathbf{M} bet the $(n+md)$ -dimensional identity matrix. Applying algorithm (6)-(8) and setting $\mu_k = \nu_k = 0$ for simplicity then yields (after considerable algebra) the recursions

$$\mathbf{x}_j^{k+1} = \arg \min_{\mathbf{x}_j \in \mathfrak{R}^{n_j}} \left\{ h_{0j}(\mathbf{x}_j) + \frac{\lambda}{2} \|\mathbf{x}_j - \mathbf{y}_j^k\|^2 + \frac{\lambda}{2} \sum_{i=1}^m \max^2 \left\{ 0, h_{ij}(\mathbf{x}_j) - \nu_{ij}^k + \frac{1}{\lambda} \pi_i^k \right\} \right\} \quad (26)$$

$$u_{ij}^{k+1} = \max \left\{ \nu_{ij}^k + \frac{1}{\lambda} \pi_i^k, h_{ij}(\mathbf{x}_j^{k+1}) \right\} \quad (27)$$

$$\mathbf{y}^{k+1} = \rho_k \mathbf{x}^{k+1} + (1 - \rho_k) \mathbf{y}^k \quad (28)$$

$$\nu_{ij}^{k+1} = \rho_k u_{ij}^{k+1} + (1 - \rho_k) \nu_{ij}^k - \frac{\rho_k}{d} \sum_{l=1}^d u_{il}^{k+1} \quad (29)$$

$$\pi_i^{k+1} = \pi_i^k + \frac{\lambda \rho_k}{d} \sum_{l=1}^d u_{il}^{k+1} . \quad (30)$$

Here, $\{\mathbf{x}^k\}$ and $\{\mathbf{y}^k\}$ are in \mathfrak{R}^n , $\{\mathbf{u}^k\}$ and $\{\mathbf{v}^k\}$ are in \mathfrak{R}^{md} with $\sum_{l=1}^d \nu_{il}^0 = 0$ for all i , and $\{\pi^k\}$ is in \mathfrak{R}^m . Further setting $\rho_k = 1$ for all k and eliminating $\{\mathbf{y}^k\}$ yields the decomposition method proposed by Spingarn [18, Section 4]. Spingarn's derivation is different, being based on the notion of the *partial inverse* of a monotone operator [17,18], but the fundamental principle is the same as that behind the alternating direction method of multipliers, as pointed out in [4].

One problem with the method (26)-(30) is that the minimand in (26) may possess a very compli-

cated pattern of nondifferentiable points, mitigating against the use of standard unconstrained methods to obtain \mathbf{x}_j^{k+1} . Fundamentally, this difficulty arises because (23) has not been decomposed "vertically," that is, with respect to i , even though it has been decomposed "horizontally," that is, with respect to j .

An alternative approach to converting (23) to the form (4) is to combine F_1 and F_2 , as opposed to F_1 and F_3 . Because F_1 depends only on \mathbf{x} , and F_2 depends only on \mathbf{u} , one thus preserves the structure of the three-way splitting embodied in (25), and the resulting algorithm achieves a more thorough decomposition of the problem. We let

$$f(\mathbf{x}, \mathbf{u}) = F_1(\mathbf{x}) + F_2(\mathbf{u}) = \begin{cases} \sum_{j=1}^d h_{0j}(\mathbf{x}_j) , & \sum_{j=1}^d u_{ij} = 0 \quad \forall i = 1, \dots, m \\ +\infty , & \text{otherwise} \end{cases} \quad (31)$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{H} & \\ & \mathbf{I} \end{bmatrix} \quad (32)$$

$$g(\mathbf{z}, \mathbf{u}) = F_3(\mathbf{z}, \mathbf{u}) \quad (33)$$

In (32), \mathbf{M} is of dimension $(n+md) \times (mn+md)$, and has full column rank from the definition of \mathbf{H} . Under (31)-(33), problem (4) is equivalent to (25), and hence to (23).

We now apply the alternating direction method (6)-(8). The role of the " \mathbf{x}^k " variables will be played by pairs $(\mathbf{x}^k, \mathbf{u}^k) \in \mathfrak{R}^n \times \mathfrak{R}^{md}$, the place of the " \mathbf{z}^k " variables will be taken by pairs $(\mathbf{y}^k, \mathbf{v}^k) \in \mathfrak{R}^{mn} \times \mathfrak{R}^{md}$, and the role of the multipliers " \mathbf{p}^k " will be assumed by pairs $(\mathbf{p}^k, \mathbf{q}^k) \in \mathfrak{R}^{mn} \times \mathfrak{R}^{md}$. By the separability of \mathbf{x} and \mathbf{u} in (31), (6) decomposes immediately into

$$\mathbf{x}^{k+1} \underset{\varepsilon_k}{\approx} \arg \min_{\mathbf{x}} \left\{ F_1(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{H}\mathbf{x} \rangle + \frac{\lambda}{2} \|\mathbf{H}\mathbf{x} - \mathbf{z}^k\|^2 \right\} \quad (34)$$

$$\mathbf{u}^{k+1} \underset{\delta_k}{\approx} \arg \min_{\mathbf{u}} \left\{ F_2(\mathbf{u}) + \langle \mathbf{q}^k, \mathbf{u} \rangle + \frac{\lambda}{2} \|\mathbf{u} - \mathbf{v}^k\|^2 \right\} , \quad (35)$$

where $\sqrt{\varepsilon_k^2 + \delta_k^2} = \mu_k$. In view of the structure of F_1 and F_2 , (34) and (35) may be satisfied by

setting

$$\mathbf{x}_j^{k+1} \underset{\frac{\mu_k}{\sqrt{d}}}{\approx} \arg \min_{\mathbf{x}_j \in \mathfrak{R}^{n_j}} \left\{ h_{0j}(\mathbf{x}_j) + \left\langle \sum_{i=1}^m \mathbf{p}_{ij}^k, \mathbf{x}_j \right\rangle + \frac{m\lambda}{2} \left\| \mathbf{x}_j - \frac{1}{m} \sum_{i=1}^m \mathbf{z}_{ij}^k \right\|^2 \right\} \quad (36)$$

$$\mathbf{u}_{ij}^{k+1} = \left(v_{ij}^k - \sum_{l=1}^d v_{il}^k \right) - \frac{1}{\lambda} \left(q_{ij}^k - \sum_{l=1}^d q_{il}^k \right). \quad (37)$$

Next, one must apply (7), which now decomposes into md calculations of the form

$$\begin{pmatrix} \mathbf{z}_{ij}^{k+1} \\ v_{ij}^{k+1} \end{pmatrix} \underset{\frac{\mu_k}{\sqrt{md}}}{\approx} \arg \min_{\substack{\mathbf{z}_{ij} \in \mathfrak{R}^{n_j} \\ v_{ij} \in \mathfrak{R}}} \left\{ g_{ij}(\mathbf{z}_{ij}, v_{ij}) - \left\langle \mathbf{p}_{ij}^k, \mathbf{z}_{ij} \right\rangle - q_{ij}^k v_{ij} + \frac{\lambda}{2} \left\| \mathbf{z}_{ij} - (\rho_k \mathbf{x}_j^{k+1} + (1-\rho_k) \mathbf{z}_{ij}^k) \right\|^2 + \frac{\lambda}{2} \left(v_{ij} - (\rho_k u_{ij}^{k+1} + (1-\rho_k) v_{ij}^k) \right)^2 \right\}.$$

Completing the square and rewriting,

$$\begin{pmatrix} \mathbf{z}_{ij}^{k+1} \\ v_{ij}^{k+1} \end{pmatrix} \underset{\frac{\mu_k}{\sqrt{md}}}{\approx} \arg \min_{h_{ij}(\mathbf{z}_{ij}) \leq v_{ij}} \left\{ \frac{\lambda}{2} \left\| \mathbf{z}_{ij} - (\rho_k \mathbf{x}_j^{k+1} + (1-\rho_k) \mathbf{z}_{ij}^k + \frac{1}{\lambda} \mathbf{p}_{ij}^k) \right\|^2 + \frac{\lambda}{2} \left(v_{ij} - (\rho_k u_{ij}^{k+1} + (1-\rho_k) v_{ij}^k + \frac{1}{\lambda} q_{ij}^k) \right)^2 \right\}, \quad (38)$$

that is, one must project the point

$$\begin{pmatrix} \rho_k \mathbf{x}_j^{k+1} + (1-\rho_k) \mathbf{z}_{ij}^k + \frac{1}{\lambda} \mathbf{p}_{ij}^k \\ \rho_k u_{ij}^{k+1} + (1-\rho_k) v_{ij}^k + \frac{1}{\lambda} q_{ij}^k \end{pmatrix} = \rho_k \begin{pmatrix} \mathbf{x}_j^{k+1} \\ u_{ij}^{k+1} \end{pmatrix} + (1-\rho_k) \begin{pmatrix} \mathbf{z}_{ij}^k \\ v_{ij}^k \end{pmatrix} + \frac{1}{\lambda} \begin{pmatrix} \mathbf{p}_{ij}^k \\ q_{ij}^k \end{pmatrix}$$

onto the epigraph

$$\text{epi } h_{ij} = \left\{ (\mathbf{x}_j, v) \in \mathfrak{R}^{n_j+1} \mid h_{ij}(\mathbf{x}_j) \leq v \right\}.$$

Lemma 2. Suppose that $h: \mathfrak{R}^l \rightarrow \mathfrak{R}$ is convex and everywhere finite. Then the projection (\mathbf{y}^*, v^*) of any $(\mathbf{x}, u) \in \mathfrak{R}^l \times \mathfrak{R}$ onto $\text{epi } h$ can be accomplished by the following procedure:

if $h(\mathbf{x}) \leq u$
 then $(\mathbf{y}^*, v^*) := (\mathbf{x}, u)$;
 else begin;
 $\mathbf{y}^* := \arg \min_{\mathbf{y}} \{ \|\mathbf{y} - \mathbf{x}\|^2 + (h(\mathbf{y}) - u)^2 \}$;
 $v^* := h(\mathbf{y}^*)$;
 end;

Proof. If $h(\mathbf{x}) \leq u$, then $(\mathbf{x}, u) \in \text{epi}(h)$, so $(\mathbf{y}^*, v^*) := (\mathbf{x}, u)$. Otherwise, $(\mathbf{x}, u) \notin \text{epi}(h)$. Being convex and finite everywhere, h is a continuous function, therefore $\eta(\mathbf{y}, v) = h(\mathbf{y}) - v$ is continuous. Now, $\text{epi } h = \{(\mathbf{y}, v) \mid \eta(\mathbf{y}, v) \leq 0\}$, so $\text{int}(\text{epi } h) = \{(\mathbf{y}, v) \mid \eta(\mathbf{y}, v) < 0\}$ and the boundary of $\text{epi } h$, $\text{bd}(\text{epi } h)$, is $\{(\mathbf{y}, v) \mid \eta(\mathbf{y}, v) = 0\}$. The projection of (\mathbf{x}, u) onto $\text{epi}(h)$ must lie in $\text{bd}(\text{epi } h)$, and therefore must be of the form $(\mathbf{y}, h(\mathbf{y}))$. The suggested procedure minimizes the distance between (\mathbf{x}, u) and points of this form. ■

In general, let the notation $\text{epp}(h, (\mathbf{x}, u))$ stand for the projection of (\mathbf{x}, u) onto $\text{epi } h$. Note that for general convex functions taking the value $+\infty$, the procedure of Lemma 2 should be replaced by

$$\begin{aligned}
 \mathbf{y}^* &:= \arg \min_{\mathbf{y}} \left\{ \|\mathbf{y} - \mathbf{x}\|^2 + \max^2\{0, h(\mathbf{y}) - u\} \right\} \\
 v^* &:= \max\{u, h(\mathbf{y}^*)\}.
 \end{aligned}$$

Here, such a situation is precluded by the assumptions on (23). In view of (36), (37), and (38), (6)-(8) now reduce to the following recursions:

$$\mathbf{x}_j^{k+1} \underset{\phi_k}{\approx} \arg \min_{\mathbf{x}_j \in \mathcal{R}^{n_j}} \left\{ h_{0j}(\mathbf{x}_j) + \left\langle \sum_{i=1}^m \mathbf{p}_{ij}^k, \mathbf{x}_j \right\rangle + \frac{m\lambda}{2} \left\| \mathbf{x}_j - \frac{1}{m} \sum_{i=1}^m \mathbf{z}_{ij}^k \right\|^2 \right\} \quad (39)$$

$$u_{ij}^{k+1} = \left(v_{ij}^k - \sum_{l=1}^d v_{il}^k \right) - \frac{1}{\lambda} \left(q_{ij}^k - \sum_{l=1}^d q_{il}^k \right) \quad (40)$$

$$\begin{pmatrix} \mathbf{z}_{ij}^{k+1} \\ v_{ij}^{k+1} \end{pmatrix} \underset{\theta_k}{\approx} \text{epp} \left(h_{ij}, \begin{pmatrix} \rho_k \mathbf{x}_j^{k+1} + (1 - \rho_k) \mathbf{z}_{ij}^k + \frac{1}{\lambda} \mathbf{p}_{ij}^k \\ \rho_k u_{ij}^{k+1} + (1 - \rho_k) v_{ij}^k + \frac{1}{\lambda} q_{ij}^k \end{pmatrix} \right) \quad (41)$$

$$\mathbf{p}_{ij}^{k+1} = \mathbf{p}_{ij}^k + \lambda \left(\rho_k \mathbf{x}_{ij}^{k+1} + (1 - \rho_k) \mathbf{z}_{ij}^k - \mathbf{z}_{ij}^{k+1} \right) \quad (42)$$

$$q_{ij}^{k+1} = q_{ij}^k + \lambda \left(\rho_k u_{ij}^{k+1} + (1 - \rho_k) v_{ij}^k - v_{ij}^{k+1} \right) \quad , \quad (43)$$

where $\{\phi_k\}$ and $\{\theta_k\}$ are nonnegative, summable sequences. This method will be called an *epigraphic projection method* because of step (41). The problem has been more thoroughly decomposed than in the method (26)-(30) in that no subproblem involves more than one function h_{ij} . Step (41), in particular, decomposes into md , as opposed to d , separate tasks, affording a large potential for parallelism. Furthermore, if all the h_{ij} are smooth, then all subproblem minimands, including those needed to implement (41), are smooth. By comparison, the minimand in (26) will in general be nonsmooth, even if all the h_{ij} are smooth.

In some cases, it may be possible to carry out the epigraphic projections (41) exactly; in other cases, they might have to be performed approximately to some accuracy $\phi_k \geq 0$. In the subroutine outlined in Lemma 2, this involves finding some $\tilde{\mathbf{y}} \approx_{\alpha_k} \mathbf{y}^*$ such that $h(\tilde{\mathbf{y}}) \approx_{\beta_k} h(\mathbf{y}^*)$, where $\sqrt{\alpha_k^2 + \beta_k^2} \leq \phi_k$. Guaranteeing such conditions is potentially complicated in the general case, so we will not discuss this issue any further here; however, working out similar approximation conditions for (26)-(27) seems much more complicated, due to coupling between h_{ij} . See [15] for some exemplary approximate solution criteria.

Consider now the convergence of (39)-(43). The proof requires a Slater condition similar to that of [18], but modified to accommodate the more general h_{0j} of (24).

Theorem 3. Consider the problem (23), where the h_{ij} are everywhere finite-valued convex for $i = 1, \dots, m$, and the h_{0j} are as in (24). Suppose $\{\phi_k\}$ and $\{\theta_k\}$ are nonnegative summable sequences, $0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2$, and the following Slater condition is met:

$$\exists \bar{\mathbf{x}} \in \text{ri}(\text{dom } h_{01}) \times \dots \times \text{ri}(\text{dom } h_{0d}) : \forall i = 1, \dots, m, \sum_{j=1}^d h_{ij}(\bar{\mathbf{x}}_j) < 0 \quad . \quad (44)$$

Then, if (23) has an optimal solution with finite objective value, the sequence $\{x^k\}$ produced by the epigraphic projection method (39)-(43), with arbitrary initial conditions, will converge to a solution of (23).

Proof. First consider, under (31)-(33), the problem

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}, \mathbf{u}) + g(\mathbf{M}(\mathbf{x}, \mathbf{u})) , \\ & (\mathbf{x}, \mathbf{u}) \in \mathcal{R}^n \times \mathcal{R}^{md} \end{aligned} \tag{45}$$

which is of form (4). If this problem has a Kuhn-Tucker point, then, in view of the above derivation, setting $\mu_k = \sqrt{d}\phi_k$ and $\nu_k = \sqrt{md}\theta_k$ in Theorem 1 gives the desired result. So, it suffices to show (45) has a Kuhn-Tucker point.

Suppose x^* is optimal for (23). Then there must exist a $u^* \in \mathcal{R}^{md}$ such that (x^*, u^*) is optimal for (25) and hence for the equivalent problem (45), that is

$$(0, 0) \in \partial[f + g \circ \mathbf{M}](x^*, u^*) .$$

Suppose it were true that

$$\partial[f + g \circ \mathbf{M}](\mathbf{x}, \mathbf{u}) = \partial f(\mathbf{x}, \mathbf{u}) + \mathbf{M}^T \partial g(\mathbf{M}(\mathbf{x}, \mathbf{u})) \quad \forall (\mathbf{x}, \mathbf{u}) \in \mathcal{R}^n \times \mathcal{R}^{md} . \tag{46}$$

Then there would have to exist $(p^*, q^*) \in \partial g(\mathbf{M}(x^*, u^*))$ such that $-\mathbf{M}^T(p^*, q^*) \in \partial f(x^*, u^*)$, and $((x^*, u^*), (p^*, q^*))$ would be a Kuhn-Tucker pair for (45). So, it suffices to prove (46).

Define the linear space

$$U = \left\{ \mathbf{u} \in \mathcal{R}^{md} \mid \sum_{j=1}^d u_{ij} = 0 \quad \forall i = 1, \dots, m \right\} ,$$

so that

$$\text{ri}(\text{dom } f) = \text{ri}(\text{dom } F_1) \times \text{ri}(\text{dom } F_2) = (\text{ri}(\text{dom } h_{01}) \times \dots \times \text{ri}(\text{dom } h_{0d})) \times U .$$

For $i = 1, \dots, m$, let

$$r_i = -\sum_{j=1}^d h_{ij}(\bar{x}_j) > 0 ,$$

and define $\bar{\mathbf{u}} \in \mathfrak{R}^{md}$ via $\bar{u}_{ij} = h_{ij}(\bar{x}_j) + \frac{r_i}{d}$. Then, $\bar{\mathbf{u}} \in U$, and $(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in \text{ri}(\text{dom } f)$. Now,

$$\text{dom}(g \circ \mathbf{M}) = \{(\mathbf{x}, \mathbf{u}) \mid h_{ij}(\mathbf{x}_j) - u_{ij} \leq 0, i = 1, \dots, m, j = 1, \dots, d\} .$$

The h_{ij} , $i = 1, \dots, m$, are finite and convex, hence continuous, so the functions

$$\eta_{ij}(\mathbf{x}_j, u_{ij}) = h_{ij}(\mathbf{x}_j) - u_{ij}$$

are also continuous. For all i and j ,

$$\eta_{ij}(\bar{x}_j, \bar{u}_{ij}) = h_{ij}(\bar{x}_j) - \left(h_{ij}(\bar{x}_j) + \frac{r_i}{d} \right) = -\frac{r_i}{d} < 0 .$$

It follows that there is an open neighborhood of $(\bar{\mathbf{x}}, \bar{\mathbf{u}})$ that is contained in $\text{dom}(g \circ \mathbf{M})$, and hence that

$$(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in \text{int}(\text{dom}(g \circ \mathbf{M})) = \text{ri}(\text{dom}(g \circ \mathbf{M})) .$$

Therefore, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in \text{ri}(\text{dom } f) \cap \text{ri}(\text{dom}(g \circ \mathbf{M}))$, and

$$\partial[f + g \circ \mathbf{M}](\mathbf{x}, \mathbf{u}) = \partial f(\mathbf{x}, \mathbf{u}) + \partial[g \circ \mathbf{M}](\mathbf{x}, \mathbf{u}) \quad \forall (\mathbf{x}, \mathbf{u}) \in \mathfrak{R}^n \times \mathfrak{R}^{md}$$

by [14, Theorem 23.8]. Similarly,

$$\mathbf{M}(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in \text{int}(\text{dom } g) = \text{ri}(\text{dom } g) ,$$

so

$$\partial[g \circ \mathbf{M}](\mathbf{x}, \mathbf{u}) = \mathbf{M}^\top \partial g(\mathbf{M}(\mathbf{x}, \mathbf{u})) \quad \forall (\mathbf{x}, \mathbf{u}) \in \mathfrak{R}^n \times \mathfrak{R}^{md}$$

by [14, Theorem 23.9]; (46) is thus established. ■

Following Theorems 23.8 and 23.9 of [14], alternatives to the condition (44) are possible when some or all of the h_{ij} are polyhedral.

5. Conclusion

This paper has attempted to demonstrate by example the power and generality of alternating direction multiplier algorithms for decomposing convex programs, and thus parallelizing their solution. The approach given here subsumes Spingarn's partial inverse technique [17,18]; for an explanation, see [4] or [3]. The key step in the alternating direction multiplier approach to decomposition is careful conversion of the problem of interest into the form (1) or (4). The way this conversion is done can strongly influence the degree of decomposition attained, as evidenced by the comparison of method (26)-(30) to the epigraphic projection procedure (39)-(43).

So far, it is not easy to tell when methods derived in this manner will perform well in practice. For example, the methods of Section 3 are impractical for linear-cost problems, but competitive for quadratic ones. Perhaps future research will resolve such questions.

References

- [1] Bertsekas, D. P. (1982), *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic Press).
- [2] Bertsekas, D. P., Tsitsiklis, J. (1989), *Parallel and Distributed Computation: Numerical Methods* (Englewood Cliffs: Prentice-Hall).
- [3] Eckstein, J. (1989), *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*. Doctoral dissertation, department of civil engineering, Massachusetts Institute of Technology. Available as report LIDS-TH-1877, Laboratory for Information and Decision Sciences, MIT, or report CICS-TH-140, Brown/Harvard/MIT Center for Intelligent Control.
- [4] Eckstein, J., Bertsekas, D. P. (1989). *On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators*. Report CICS-P-167, Brown/Harvard/MIT Center for Intelligent Control Systems, or Harvard Business School working paper 90-033 (to appear in *Mathematical Programming*, 1992).
- [5] Eckstein, J. and Bertsekas, D. P. (1990). *An Alternating Direction Method for Linear Programming*. working paper 90-063, Harvard Business School, Boston, MA.
- [6] Eckstein, J. (1990). *Implementing and Running the Alternating Step Method on the Connection Machine CM-2*. Working Paper 91-005, Harvard Business School, Boston, MA. (to appear in the *ORSA Journal on Computing*, 1992).
- [7] Fortin, M., Glowinski, R. (1983), "On Decomposition-Coordination Methods Using an Augmented Lagrangian". In M. Fortin, R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-

Holland).

- [8] Gabay, D. (1983), "Applications of the Method of Multipliers to Variational Inequalities". In M. Fortin, R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-Holland).
- [9] Gabay, D., Mercier B. (1976), "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations". *Computers and Mathematics with Applications* 2:17-40.
- [10] Glowinski, R., Le Tallec, P. (1987), *Augmented Lagrangian Methods for the Solution of Variational Problems*. MRC Technical Summary Report #2965, Mathematics Research Center, University of Wisconsin – Madison.
- [11] Glowinski, R., Marroco, A. (1975), "Sur L'Approximation, par Elements Finis d'Ordre Un, et la Resolution, par Penalisation-Dualité, d'une Classe de Problemes de Dirichlet non Lineares". *Revue Française d'Automatique, Informatique et Recherche Opérationnelle* 9(R-2):41-76.
- [12] Klingman, D., Napier, A., Stutz, J., (1974), "NETGEN — A Program for Generating Large-Scale (Un)capacitated Assignment, Transportation, and Minimum Cost Network Problems". *Management Science* 20:814-822.
- [13] Lions, P.-L., Mercier, B. (1979), "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis* 16(6):964-979.
- [14] Rockafellar, R. T. (1970), *Convex Analysis* (Princeton: Princeton University Press).
- [15] Rockafellar, R. T. (1976), "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research* 1(2):97-116.
- [16] Rockafellar, R. T. (1984), *Network Flows and Monotropic Optimization* (New York: John Wiley).
- [17] Spingarn, J. E. (1983), "Partial inverse of a monotone operator," *Applied Mathematics and Optimization* 10:247-265.
- [18] Spingarn, J. E. (1985), "Application of the Method of Partial Inverses to Convex Programming: Decomposition". *Mathematical Programming* 32:199-233.

NTIS does not permit return of items for credit or refund. A replacement will be provided if an error is made in filling your order, if the item was received in damaged condition, or if the item is defective.

Reproduced by NTIS

National Technical Information Service
Springfield, VA 22161

*This report was printed specifically for your order
from nearly 3 million titles available in our collection.*

For economy and efficiency, NTIS does not maintain stock of its vast collection of technical reports. Rather, most documents are printed for each order. Documents that are not in electronic format are reproduced from master archival copies and are the best possible reproductions available. If you have any questions concerning this document or any order you have placed with NTIS, please call our Customer Service Department at (703) 487-4660.

About NTIS

NTIS collects scientific, technical, engineering, and business related information — then organizes, maintains, and disseminates that information in a variety of formats — from microfiche to online services. The NTIS collection of nearly 3 million titles includes reports describing research conducted or sponsored by federal agencies and their contractors; statistical and business information; U.S. military publications; audiovisual products; computer software and electronic databases developed by federal agencies; training tools; and technical reports prepared by research organizations worldwide. Approximately 100,000 *new* titles are added and indexed into the NTIS collection annually.

For more information about NTIS products and services, call NTIS at (703) 487-4650 and request the free *NTIS Catalog of Products and Services*, PR-827LPG, or visit the NTIS Web site
<http://www.ntis.gov>.

NTIS

***Your indispensable resource for government-sponsored
information—U.S. and worldwide***