CENTER FOR
PARALLEL OPTIMIZATION

PARALLEL CONSTRAINT DISTRIBUTION

by

Michael C. Ferris and Olvi L. Mangasarian

Computer Sciences Technical Report #971

October 1990

PARALLEL CONSTRAINT DISTRIBUTION*

MICHAEL C. FERRIS† AND OLVI L. MANGASARIAN†

Abstract. Constraints of a mathematical program are distributed among parallel processors together with an appropriately constructed augmented Lagrangian for each processor which contains Lagrangian information on the constraints handled by the other processors. Lagrange multiplier information is then exchanged between processors. Convergence is established under suitable conditions for strongly convex quadratic programs and for general convex programs.

Key words. Parallel Optimization, Augmented Lagrangians, Quadratic Programs, Convex Programs

Abbreviated title. Parallel Constraint Distribution

1. Introduction.

We are concerned with the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \leq 0, \ldots, g_k(x) \leq 0 \end{array} \tag{1.1}$$

where $f, g_1, \ldots, g_k$ are differentiable convex functions from the n-dimensional real space $\mathbb{R}^n$ to $\mathbb{R}$, $\mathbb{R}^{m_1}, \ldots, \mathbb{R}^{m_k}$ respectively, with f being strongly convex on $\mathbb{R}^n$. Our principal aim is to distribute the k constraint blocks among k parallel processors together with an appropriately modified objective function. We then solve each of these k subproblems independently, share Lagrange multiplier information among the processors and repeat. Other recently proposed decomposition methods can be found in [20, 7, 5, 19]. The key to our approach lies in the precise form of the modified objective function to be optimized by each processor. Considerable experimentation with various Lagrangian terms [3] highlighted the difference between theoretical convergence and computational efficiency. We believe that we now have effective modified objectives for each processor that can best be described as augmented Lagrangian functions [17, 18, 1]. The modified objectives are made up of the original objective function plus augmented Lagrangian terms involving the constraints handled by the other processors. Computational experience on the Sequent Symmetry S-81 shared memory multiprocessor, with constraint distribution for quadratic programs derived from a least-norm solution of linear programs, has been encouraging. This is described in Section 4 of the paper. Section 2 is devoted to the quadratic programming case for which we obtain the strongest convergence results in Theorem 2.4. Thus under the assumption of a strongly convex quadratic objective and linear independence of each of the distributed constraint blocks, the parallel constraint distribution (PCD) algorithm converges from any starting point for a solvable problem. The key to the convergence proof is to show that in the dual space, the

* This material is based on research supported by the Air Force Office of Scientific Research Grant AFOSR-89-0410 and National Science Foundation Grants DCR-8521228 and CCR-8723091

† Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706

proposed parallel constraint distribution algorithm is equivalent to a subsequentially–convergent iterative method with step–size proposed in [10, Algorithm 2.1] for which full sequential convergence has just recently been established [8, 4, 16]. In Section 3 we establish a weaker convergence of the PCD algorithm (Theorem 3.2) for the general convex program (1.1) with strongly convex objective function. The method of proof in this section is entirely different from that of Section 2, and relies on the Lipschitz continuity of the solution variables of each subproblem in the fixed Lagrangian multipliers obtained from the other subproblems (Lemma 3.1). Unfortunately, to establish convergence, we need to assume that the distance between successive values of the multipliers approaches zero. We believe this assumption can be considerably relaxed and probably eliminated.

A word about our notation now. For a vector x in the n–dimensional real space $\mathbb{R}^n$, x_+ will denote the vector in $\mathbb{R}^n$ with components $(x_+)_i := \max\{x_i, 0\}$, $i = 1, \ldots, n$. The standard inner product of $\mathbb{R}^n$ will be denoted either by $\langle x, y\rangle$ or x^Ty. The Euclidean or 2–norm $(x^Tx)^{\frac{1}{2}}$, will be denoted by $\|\cdot\|$. For an $m \times n$ real matrix A, signified by $A \in \mathbb{R}^{m\times n}$, A^T will denote the transpose. The identity matrix of any order will be given by I. The nonnegative orthant in $\mathbb{R}^n$ will be denoted by $\mathbb{R}^n_+$. The term (s)psd will denote (symmetric) positive semidefinite, while (s)pd will denote (symmetric) positive definite.

2. Parallel constraint distribution for quadratic programs. For simplicity we consider a quadratic program with 3 blocks of inequality constraints. Routine extension to k blocks can be achieved by appropriate extension and permutation of subscripts. Equality constraints can also be incorporated in an straightforward manner. Consider then the problem

$$(2.1)\qquad \begin{array}{ll} \text{minimize} & c^Tx + \frac{1}{2}x^TQx \\ \text{subject to} & A_lx \le a_l,\ l = 1, 2, 3 \end{array}$$

where $c \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n\times n}$, $A_l \in \mathbb{R}^{m_l\times n}$, $a_l \in \mathbb{R}^{m_l}$ and Q is symmetric and positive definite. At iteration i we distribute the constraints of this problem among 3 parallel processors ($l = 1, 2, 3$) as follows

$$(2.2)\qquad \begin{array}{ll} \underset{x_l}{\text{minimize}} & c^Tx_l + \frac{1}{2}x_l^TQx_l + \frac{1}{2\gamma}\left[\sum_{\substack{j=1\\ j\ne l}}^{3} \left\|\left(\gamma(A_jx_l - a_j) + p_{jl}^i\right)_+\right\|^2\right] + r_l^ix_l \\ \text{subject to} & A_lx_l \le a_l \end{array}$$

where γ is a positive number and p_{jl}^i and r_l^i, $j, l = 1, 2, 3$ are defined below in (2.15) and (2.16). We note that the p_{jl}^i play the roles of multipliers and in fact converge to the optimal multipliers eventually, while r_l^i are substitution operators that replace estimates of the multipliers by their most recent values obtained from each of the other subproblems. Note also that even though the subproblems (2.2) are motivated by augmented Lagrangian ideas [17, 18, 1], they are rather different from being precisely an augmented

Lagrangian. The motivation of this reformulation is that in each subproblem some constraints are treated explicitly as constraints while the remaining ones are treated as augmented Lagrangian terms in the objective function. The updating of the multipliers is done by solving the subproblems explicitly rather than the traditional, and often slow, gradient updating scheme in the dual space of the augmented Lagrangian approach [1]. The key to the convergence of our algorithm for the quadratic case is the choice of the parameters p^i_{jl} and r^i_l in such a way that the dual problems, associated with (2.2), result in a convergent iterative matrix-splitting method [10, 8, 13, 16] for a symmetric linear complementarity problem in the dual variables of the problem. This choice is by no means unique and we have experimented computationally with a number of choices for the p^i_{jl} and r^i_l which we report on in Section 4. We shall establish convergence of only one of our choices in this section of the paper, which may not necessarily be the best computationally. Further experimentation is needed to determine the best splitting. We now proceed to show how the parameters p^i_{jl} and r^i_l are chosen and to justify these choices from the point of view of a convergent matrix-splitting method.

Let $(\bar{x}^{i+1}_l, \bar{s}^{i+1}_l) \in \mathbb{R}^{n+m_l}$, $l = 1,2,3$, $i = 1, \dots$ satisfy the Karush–Kuhn–Tucker conditions [9] for subproblems (2.2). We shall signify this by

$$(\bar{x}^{i+1}_l, \bar{s}^{i+1}_l) \in \arg \mathrm{KKT}(2.2)$$

Hence $(\bar{x}^{i+1}_l, \bar{s}^{i+1}_l)$ satisfy the following Karush–Kuhn–Tucker conditions (where we have used the easily verified equivalence

$$b = d_+ \iff b - d \geq 0,\ b^T(b-d) = 0,\ b \geq 0$$

for any two vectors b and d in $\mathbb{R}^{m_l}$)

$$\text{(2.3)}\quad \begin{aligned} c + Q\bar{x}^{i+1}_l + \textstyle\sum_{\substack{j=1\\ j\neq l}}^3 A_j^T \left(\gamma(A_j\bar{x}^{i+1}_l - a_j) + p^i_{jl}\right)_+ + r^{i+1}_l + A_l^T\bar{s}^{i+1}_l &= 0 \\ \bar{s}^{i+1}_l = \left(\bar{s}^{i+1}_l + \gamma(A_l\bar{x}^{i+1}_l - a_l)\right)_+ & \end{aligned} \qquad l = 1,2,3$$

or equivalently

$$\text{(2.4)}\quad \begin{aligned} \bar{x}^{i+1}_l &= -Q^{-1}(c + \textstyle\sum_{\substack{j=1\\ j\neq l}}^3 A_j^T\bar{t}^{i+1}_{jl} + r^{i+1}_l + A_l^T\bar{s}^{i+1}_l) \\ \bar{s}^{i+1}_l &= \left(\bar{s}^{i+1}_l + \gamma(A_l\bar{x}^{i+1}_l - a_l)\right)_+ \\ \bar{t}^{i+1}_{jl} &= \left(\gamma(A_j\bar{x}^{i+1}_l - a_j) + p^i_{jl}\right)_+ \end{aligned} \qquad l = 1,2,3,\ j = 1,2,3,\ j \neq l$$

Elimination of $\bar{x}^{i+1}_l$ by using the first equation of (2.4) leads (after a bit of algebra) to the following symmetric linear complementarity problem (LCP) in the variable $\bar{z}^{i+1}$

$$\text{(2.5)}\qquad B\bar{z}^{i+1} + Cz^i + q \geq 0,\ \left\langle \bar{z}^{i+1}, B\bar{z}^{i+1} + Cz^i + q \right\rangle = 0,\ \bar{z}^{i+1} \geq 0$$

where

$$\text{(2.6)}\qquad \begin{aligned} \bar{z}^{i+1} &= (\bar{s}^{i+1}_1, \bar{s}^{i+1}_2, \bar{s}^{i+1}_3, \bar{t}^{i+1}_{12}, \bar{t}^{i+1}_{23}, \bar{t}^{i+1}_{31}, \bar{t}^{i+1}_{13}, \bar{t}^{i+1}_{21}, \bar{t}^{i+1}_{32}) \\ z^i &= (s^i_1, s^i_2, s^i_3, t^i_{12}, t^i_{23}, t^i_{31}, t^i_{13}, t^i_{21}, t^i_{32}) \end{aligned}$$

and

$$B = \begin{bmatrix} D & R_1^T & R_2^T \\ R_1 & I + D & R_1^T \\ R_2 & R_1 & I + D \end{bmatrix} \tag{2.7}$$

with

$$D = \gamma \begin{bmatrix} A_1 Q^{-1} A_1^T & 0 & 0 \\ 0 & A_2 Q^{-1} A_2^T & 0 \\ 0 & 0 & A_3 Q^{-1} A_3^T \end{bmatrix}$$

$$R_1 = \gamma \begin{bmatrix} 0 & A_1 Q^{-1} A_2^T & 0 \\ 0 & 0 & A_2 Q^{-1} A_3^T \\ A_3 Q^{-1} A_1^T & 0 & 0 \end{bmatrix} \tag{2.8}$$

$$R_2 = \gamma \begin{bmatrix} 0 & 0 & A_1 Q^{-1} A_3^T \\ A_2 Q^{-1} A_1^T & 0 & 0 \\ 0 & A_3 Q^{-1} A_2^T & 0 \end{bmatrix}$$

and

$$Cz^i = \gamma \begin{bmatrix} A_1 Q^{-1} r_1^i \\ A_2 Q^{-1} r_2^i \\ A_3 Q^{-1} r_3^i \\ A_1 Q^{-1} r_2^i - p_{12}^i/\gamma \\ A_2 Q^{-1} r_3^i - p_{23}^i/\gamma \\ A_3 Q^{-1} r_1^i - p_{31}^i/\gamma \\ A_1 Q^{-1} r_3^i - p_{13}^i/\gamma \\ A_2 Q^{-1} r_1^i - p_{21}^i/\gamma \\ A_3 Q^{-1} r_2^i - p_{32}^i/\gamma \end{bmatrix}, \quad q = \gamma \begin{bmatrix} A_1 Q^{-1} c + a_1 \\ A_2 Q^{-1} c + a_2 \\ A_3 Q^{-1} c + a_3 \\ A_1 Q^{-1} c + a_1 \\ A_2 Q^{-1} c + a_2 \\ A_3 Q^{-1} c + a_3 \\ A_1 Q^{-1} c + a_1 \\ A_2 Q^{-1} c + a_2 \\ A_3 Q^{-1} c + a_3 \end{bmatrix} \tag{2.9}$$

The matrix C is determined by the choice of p_{jl}^i and r_l^i in (2.2) or equivalently in (2.9), and this is precisely where the power (and at the same time the difficulty) of the proposed method lies. The matrix C has to be chosen so that $B + C$ constitutes a "regular splitting" of some symmetric M, that is $M = B + C$, with M, B, and C satisfying certain properties such as:

$$M \text{ symmetric, } B - \tfrac{\lambda M}{2} \text{ pd for some } \lambda \in (0, 1] \tag{2.10}$$

$$\text{(2.11)} \qquad M \text{ symmetric psd, } B - \frac{\lambda M}{2} \text{ pd for some } \lambda \in (0,1]$$

This amounts to the key requirement that M be symmetric or symmetric psd, since B as defined by (2.7) is easily made pd and λ can be chosen sufficiently small to ensure that $B - \frac{\lambda M}{2}$ is pd. Under assumption (2.10), a solution of the LCP:

$$\text{(2.12)} \qquad Mz + q \geq 0, \ \langle z, Mz + q \rangle = 0, \ z \geq 0$$

is obtained [10] from each accumulation point of the sequence $\{z^i\}$ generated by solving the LCP (2.5) for $\bar{z}^{i+1}$ and then determining z^{i+1} by using the step-size λ, that is

$$\text{(2.13)} \qquad z^{i+1} = (1-\lambda)z^i + \lambda \bar{z}^{i+1}, \ \lambda \in (0,1]$$

Under assumption (2.11) the **whole** sequence $\{z^i\}$ generated by (2.5) and (2.13) converges to a solution of (2.12) provided the latter is solvable [8, 4, 16]. The simplest choice for p^i_{jl} and r^i_l we propose for the nonlinear case of Section 3 and for which we establish convergence under somewhat more stringent assumptions is the following

$$\text{(2.14)} \qquad p^i_{jl} = s^i_j, \ r^i_l = 0, \ l = 1,2,3, \ j = 1,2,3, \ j \neq l$$

Unfortunately this simple choice in the quadratic case leads to a **nonsymmetric** C and hence a nonsymmetric M in (2.12). The convergence conditions for splitting nonsymmetric LCP's are quite stringent [2, Chapter 5] and not useful for our proposed applications here. We have therefore settled on choices for the parameters p^i_{jl} and r^i_l which are more general than (2.14), and which generate a symmetric psd M. By choosing λ sufficiently small, it is easily seen that (2.11) is satisfied, because by (2.7), the matrix B is positive definite if we assume that each A_l, $l = 1,2,3$, has linearly independent rows. There are a number of choices of the p^i_{jl} and r^i_l that generate a symmetric psd M and hence a convergent scheme. Our preliminary computational experience does not provide a clear cut indication which is the best choice for p^i_{jl} and r^i_l among the convergent schemes. We believe this requires further theoretical and computational study. However, for concreteness, we wish to present at least one specific choice of C that results in the following choices of p^i_{jl} and r^i_l:

$$\text{(2.15)} \qquad \begin{aligned} r^i_1 &= A_2^T(s^i_2 - t^i_{21}) + A_3^T(s^i_3 - t^i_{31}) \\ r^i_2 &= A_1^T(s^i_1 - t^i_{12}) + A_3^T(s^i_3 - t^i_{32}) \\ r^i_3 &= A_1^T(s^i_1 - t^i_{13}) + A_2^T(s^i_2 - t^i_{23}) \end{aligned}$$

$$\text{(2.16)} \qquad \begin{aligned} p^i_{21} &= t^i_{21} + \gamma A_2 Q^{-1}(A_1^T(s^i_1 - t^i_{13}) + A_2^T(s^i_2 - t^i_{21}) + A_3^T(s^i_3 - t^i_{32})) \\ p^i_{31} &= t^i_{31} + \gamma A_3 Q^{-1}(A_1^T(s^i_1 - t^i_{12}) + A_2^T(s^i_2 - t^i_{23}) + A_3^T(s^i_3 - t^i_{31})) \\ p^i_{12} &= t^i_{12} + \gamma A_1 Q^{-1}(A_1^T(s^i_1 - t^i_{12}) + A_2^T(s^i_2 - t^i_{23}) + A_3^T(s^i_3 - t^i_{31})) \\ p^i_{32} &= t^i_{32} + \gamma A_3 Q^{-1}(A_1^T(s^i_1 - t^i_{13}) + A_2^T(s^i_2 - t^i_{21}) + A_3^T(s^i_3 - t^i_{32})) \\ p^i_{13} &= t^i_{13} + \gamma A_1 Q^{-1}(A_1^T(s^i_1 - t^i_{13}) + A_2^T(s^i_2 - t^i_{21}) + A_3^T(s^i_3 - t^i_{32})) \\ p^i_{23} &= t^i_{23} + \gamma A_2 Q^{-1}(A_1^T(s^i_1 - t^i_{12}) + A_2^T(s^i_2 - t^i_{23}) + A_3^T(s^i_3 - t^i_{31})) \end{aligned}$$

We note that the r_l^i play the role of a substitution operator in the sense that they substitute the latest Lagrange multiplier value s_l^i obtained from each subproblem solution for t_{jl}^i, both of which eventually converge to an optimal Lagrange multiplier value. The p_{jl}^i terms are essentially multiplier value estimates given by t_{jl}^i plus additional terms that converge to zero. The additional terms are added in order to produce a symmetric C and hence a symmetric M. These above choices of p_{jl}^i and r_l^i lead to the following matrix C defined through the relations (2.9) and (2.6),

$$C = \begin{bmatrix} R_1 + R_2 & -R_1^T & -R_2^T \\ -R_1 & -I + R_1 + R_2 & -R_1^T \\ -R_2 & -R_1 & -I + R_1 + R_2 \end{bmatrix} \tag{2.17}$$

with R_1 and R_2 defined in (2.8). Addition of the matrices B and C gives the symmetric block–diagonal matrix M

$$M = \begin{bmatrix} H & 0 & 0 \\ 0 & H & 0 \\ 0 & 0 & H \end{bmatrix} \tag{2.18}$$

where

$$H := D + R_1 + R_2 = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} Q^{-1} \begin{bmatrix} A_1^T & A_2^T & A_3^T \end{bmatrix} \tag{2.19}$$

Note that if our original quadratic program (2.1) is feasible, then it it solvable. Hence its Wolfe dual is solvable, which is equivalent to the solvability of the LCP (2.12) with M as defined in (2.18) and q as in (2.9). In fact the LCP (2.12) constitutes a replication of the Wolfe dual 3 times.

We are now ready to define the PCD algorithm for the quadratic program (2.1).

2.1. PCD algorithm for quadratic programming.

1. Start with any s_l^0, t_{jl}^0, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$.
2. Compute r_l^0, p_{jl}^0, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$ from (2.15) and (2.16).
3. Having s_l^i, t_{jl}^i, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$ compute:
   (a) $(\bar{x}_l^{i+1}, \bar{s}_l^{i+1}) \in \arg \mathrm{KKT}(2.2)$, $l = 1,2,3$
   (b) $\bar{t}_{jl}^{i+1} = \left(\gamma(A_j \bar{x}_l^{i+1} - a_j) + p_{jl}^i\right)_+$, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$
   (c) $(s_l^{i+1}, t_{jl}^{i+1}) = (1-\lambda)(s_l^i, t_{jl}^i) + \lambda(\bar{s}_l^{i+1}, \bar{t}_{jl}^{i+1})$, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$ with $\lambda \in (0,1]$ satisfying (2.25) below.

2.2. Remark. We note that the subproblems (2.2) of the PCD Algorithm 2.1 split the constraints of the original quadratic program (2.1) between them in the form of split explicit constraints as well as augmented Lagrangian terms involving the other constraints. The principal objective that has been achieved is that the **explicit** constraints of each of the subproblems are a subset of the constraints of the original problem.

2.3. Remark: Monotone LCP as dual of nonsmooth convex quadratic program. It is interesting to note that the PCD algorithm is a matrix-splitting iterative method for a monotone LCP that can be associated with a dual formulation of a nonsmooth convex program. Thus consider such a nonsmooth program:

$$\begin{array}{lll} (2.20) & \underset{x}{\text{minimize}} & c^T x + \frac{1}{2} x^T Q x + \frac{1}{2} \left\| (Hx - h)_+ \right\|^2 \\ & \text{subject to} & Bx \le b \end{array}$$

where Q is spsd. The necessary and sufficient Karush–Kuhn–Tucker conditions for this problem are

$$\begin{array}{ll} (2.21) & \begin{array}{c} c + Qx + H^T (Hx - h)_+ + B^T s = 0 \\ s = (s + Bx - b)_+ \end{array} \end{array}$$

Defining a new variable t as

$$(2.22) \qquad t = (Hx - h)_+$$

and solving the first Karush–Kuhn–Tucker condition for x gives

$$(2.23) \qquad x = -Q^{-1}(H^T t + B^t s + c)$$

Substituting for x in the second equation of (2.21) and in (2.22) gives the following monotone linear complementarity problem in the variables (s, t)

$$(2.24) \qquad \begin{array}{c} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} BQ^{-1}B^T & BQ^{-1}H^T \\ HQ^{-1}B^T & I + HQ^{-1}H^T \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} + \begin{bmatrix} BQ^{-1}c + b \\ HQ^{-1}c + h \end{bmatrix} \ge 0 \\ (s^T, t^T) \begin{bmatrix} v \\ w \end{bmatrix} = 0, \quad \begin{bmatrix} s \\ t \end{bmatrix} \ge 0 \end{array}$$

We then have the following duality relation between the nonsmooth convex program (2.20) and the monotone LCP (2.24). For each solution (s, t) of (2.24), x defined by (2.23) is the unique solution of (2.20). Conversely, for each Karush–Kuhn–Tucker point (x, s) of (2.20), the point (s, t), with t defined by (2.22), solves (2.24). Note that the symmetric LCP (2.24) is equivalent to the following quadratic program in (s, t):

$$\underset{(s,t)\ge 0}{\text{minimize}} \quad \frac{1}{2}(s^T, t^T) \begin{bmatrix} BQ^{-1}B^T & BQ^{-1}H^T \\ HQ^{-1}B^T & I + HQ^{-1}H^T \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} + (s^T, t^T) \begin{bmatrix} BQ^{-1}c + b \\ HQ^{-1}c + h \end{bmatrix}$$

We are now ready to establish convergence of the PCD Algorithm 2.1

THEOREM 2.4 (PCD CONVERGENCE FOR QUADRATIC PROGRAMS). *Let (2.1) be feasible and let Q be spd and let each of A_l, $l = 1, 2, 3$, have linearly independent* *rows. Then the sequence* $\left\{s_l^i, t_{jl}^i\right\}$, $l = 1,2,3,\ j = 1,2,3,\ j \neq l,\ i = 0,1,\ldots,$ *generated by the PCD Algorithm 2.1 converges to* $(\bar{s}_l, \bar{t}_{jl})$, $l = 1,2,3,\ j = 1,2,3,\ j \neq l$ *and each of the sequences* $\{x_l^i\}$, $l = 1,2,3$, *converges to the unique solution* $\bar{x}$ *of (2.1). Furthermore, each* $(\bar{x}, \bar{s}_l)$ *and each* $(\bar{x}, \bar{t}_{jl})$, $l = 1,2,3,\ j = 1,2,3,\ j \neq l$ *is a Karush–Kuhn–Tucker point of (2.1), and* $\bar{p}_{jl} = \bar{t}_{jl}$, $l = 1,2,3,\ j = 1,2,3,\ j \neq l$

Proof. Let $\bar{x}_l^{i+1}$, $l = 1,2,3$, be the unique solution of the subproblems (2.2). Hence $\bar{x}_l^{i+1}$ and some $\bar{s}_l^{i+1} \in \mathbb{R}^{m_l}$ satisfy the Karush–Kuhn–Tucker conditions (2.3), or equivalently, $(\bar{x}_l^{i+1}, \bar{s}_l^{i+1})$ and some $\bar{t}_{jl}^{i+1}$, $l = 1,2,3,\ j = 1,2,3,\ j \neq l$ satisfy (2.4). This in turn is equivalent to $\bar{z}^{i+1}$, as defined by (2.6), satisfying the LCP (2.5). By the choice of r_l^i, p_{jl}^i, $l = 1,2,3,\ j = 1,2,3,\ j \neq l$ of (2.15) and (2.16) we have that the matrix C, as defined in (2.17), and hence also the matrix $M = B + C$, given by (2.18) and (2.19), are symmetric and psd with B pd by the linear independence assumption. Thus if λ is chosen sufficiently small, and specifically such that

$$(2.25) \qquad 0 < \lambda \leq \min\{1, 2(\text{min eigenvalue(B) / max eigenvalue(M)})\}$$

it follows that (2.10) above (which is condition (6) of [10]) and (2.11) above (which is condition (4.1) of [8]) are satisfied. Hence, since the LCP (2.12) is solvable, the sequence $\{z^i\}$ converges [8, Theorem 2 and Example 3] to a solution of the LCP (2.12), and by $z^{i+1} = (1-\lambda)z^i + \lambda\bar{z}^{i+1}$, so does the sequence $\{\bar{z}^i\}$. It follows by (2.3), (2.4), (2.15) and (2.16) that in the limit we have

$$\left.\begin{aligned} c + Q\bar{x}_l + \textstyle\sum_{\substack{j=1\\ j\neq l}}^3 A_j^T\bar{t}_{jl} + \bar{r}_l + A_l^T\bar{s}_l &= 0 \\ \bar{s}_l &= (\bar{s}_l + \gamma(A_l\bar{x}_l - a_l))_+ \\ \bar{t}_{jl} &= (\gamma(A_j\bar{x}_l - a_j) + \bar{p}_{jl})_+ \end{aligned}\right. \qquad l = 1,2,3, j = 1,2,3, j \neq l$$

However, from (2.15) it follows that

$$\bar{r}_l = \textstyle\sum_{\substack{j=1\\ j\neq l}}^3 A_j^T(\bar{s}_j - \bar{t}_{jl}) \quad l = 1,2,3$$

and hence that

$$(2.26) \qquad \left.\begin{aligned} c + Q\bar{x}_l + \textstyle\sum_{j=1}^3 A_j^T\bar{s}_j &= 0 \\ \bar{s}_l &= (\bar{s}_l + \gamma(A_l\bar{x}_l - a_l))_+ \end{aligned}\right. \quad l = 1,2,3$$

It is now clear form the nonsingularity of Q that

$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 =: \bar{x}$$

Conditions (2.26) become then the necessary and sufficient conditions for $\bar{x}$ to be the unique solution of (2.1) with multipliers as indicated in the statement of the theorem. Furthermore, since $\bar{z} = (\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{t}_{12}, \bar{t}_{23}, \bar{t}_{31}, \bar{t}_{13}, \bar{t}_{21}, \bar{t}_{32})$ solves the 3–block LCP (2.12) with identical M and q sub–blocks as defined by (2.18) and (2.9) respectively, it follows that each of $(\bar{s}_1, \bar{s}_2, \bar{s}_3)$, $(\bar{t}_{12}, \bar{t}_{23}, \bar{t}_{31})$ and $(\bar{t}_{13}, \bar{t}_{21}, \bar{t}_{32})$ solve any one of the 3 sub–blocks of LCP (2.12) and hence [12, Corollary 2] their differences lie in the nullspace of H. Thus

$$(2.27)\qquad H\left(\begin{bmatrix}\bar{s}_1\\ \bar{s}_2\\ \bar{s}_3\end{bmatrix}-\begin{bmatrix}\bar{t}_{12}\\ \bar{t}_{23}\\ \bar{t}_{31}\end{bmatrix}\right)=0 \text{ and } H\left(\begin{bmatrix}\bar{s}_1\\ \bar{s}_2\\ \bar{s}_3\end{bmatrix}-\begin{bmatrix}\bar{t}_{13}\\ \bar{t}_{21}\\ \bar{t}_{32}\end{bmatrix}\right)=0$$

Relations (2.27), and relations (2.16) in the limit, imply that $\bar{p}_{jl} = \bar{t}_{jl}$, $l = 1,2,3$, $j = 1,2,3$, $j \neq l$. □

3. Parallel constraint distribution for convex programs. We extend our ideas now to general convex programs with strongly convex objective functions. For simplicity of notation we consider the 2–block problem

$$(3.1)\qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \leq 0,\ g_2(x) \leq 0 \end{array}$$

where $f: \mathbb{R}^n \to \mathbb{R}$, $g_1: \mathbb{R}^n \to \mathbb{R}^{m_1}$, $g_2: \mathbb{R}^n \to \mathbb{R}^{m_2}$ are differentiable convex functions on $\mathbb{R}^n$, with f strongly convex with modulus k, and g_1, g_2 Lipschitz continuous with constant K on $\mathbb{R}^n$. We begin with the following straightforward Lipschitz continuity result.

LEMMA 3.1. *Let f, g_1, g_2 be differentiable convex functions on $\mathbb{R}^n$ with f strongly convex with modulus k and let g_1 be Lipschitz continuous with constant K on $\mathbb{R}^n$. Let g_2 satisfy a constraint qualification on the nonempty set $\{x \mid g_2(x) \leq 0\}$. Then*

$$x(u_1) := \arg\min \left\{ f(x) + \frac{1}{2\gamma}\left\{ \left\| (\gamma g_1(x) + u_1)_+ \right\|^2 \right\} \mid g_2(x) \leq 0 \right\}$$

is Lipschitz continuous on $\mathbb{R}^{m_1}_+$ with Lipschitz constant $\frac{K}{2k}(1+\sqrt{1+4k/\gamma K^2})$.

Proof. Let u_1, $\bar{u}_1 \in \mathbb{R}^{m_1}_+$ and $x := x(u_1)$ and $\bar{x} := x(\bar{u}_1)$. By the Karush–Kuhn–Tucker conditions, there exist v_2, $\bar{v}_2 \in \mathbb{R}^{m_2}$ such that

$$\nabla f(x) + (\gamma g_1(x) + u_1)_+^T \nabla g_1(x) + v_2^T \nabla g_2(x) = 0$$
$$g_2(x) \leq 0,\ \langle v_2, g_2(x)\rangle = 0,\ v_2 \geq 0$$

and

$$\nabla f(\bar{x}) + (\gamma g_1(\bar{x}) + \bar{u}_1)_+^T \nabla g_1(\bar{x}) + \bar{v}_2^T \nabla g_2(\bar{x}) = 0$$
$$g_2(\bar{x}) \leq 0,\ \langle \bar{v}_2, g_2(\bar{x})\rangle = 0,\ \bar{v}_2 \geq 0$$

By the strong convexity of f we have that

$$k\left\|\bar{x} - x\right\|^2 \leq (\nabla f(\bar{x}) - \nabla f(x))(\bar{x} - x)$$

This together with the Karush–Kuhn–Tucker conditions gives

$$
\begin{aligned}
&k\|\bar{x}-x\|^2 \\
&\leq \left((\gamma g_1(\bar{x})+\bar{u}_1)_+^T\nabla g_1(\bar{x})+\bar{v}_2^T\nabla g_2(\bar{x})-(\gamma g_1(x)+u_1)_+^T\nabla g_1(x)-v_2^T\nabla g_2(x)\right)(x-\bar{x}) \\
&\leq \left\langle(\gamma g_1(x)+u_1)_+-(\gamma g_1(\bar{x})+\bar{u}_1)_+, g_1(\bar{x})-g_1(x)\right\rangle+\langle v_2-\bar{v}_2, g_2(\bar{x})-g_2(x)\rangle
\end{aligned}
$$

where the last inequality above follows from the following inequality

$$\langle w-\bar{w}, h(x)-h(\bar{x})\rangle \leq (w^T\nabla h(x)-\bar{w}^T\nabla h(\bar{x}))(x-\bar{x})$$

for a convex differentiable $h: \mathbb{R}^n \to \mathbb{R}^k$ and w, $\bar{w} \in \mathbb{R}^k_+$. The Karush–Kuhn–Tucker conditions allow us to drop nonpositive term $\langle v_2-\bar{v}_2, g_2(\bar{x})-g_2(x)\rangle$ thus giving us

$$k\|\bar{x}-x\|^2 \leq \left\langle(\gamma g_1(x)+u_1)_+-(\gamma g_1(\bar{x})+\bar{u}_1)_+, g_1(\bar{x})-g_1(x)\right\rangle$$

From the fundamental properties of the projection operator $()_+$, we have for y, $z \in \mathbb{R}^m$

$$\left\langle y-z, (y)_+-(z)_+\right\rangle \geq 0$$

so that

$$
\begin{aligned}
k\|\bar{x}-x\|^2 &\leq \frac{1}{\gamma}\left\langle(\gamma g_1(x)+u_1)_+-(\gamma g_1(\bar{x})+\bar{u}_1)_+, u_1-\bar{u}_1\right\rangle \\
&\leq \frac{1}{\gamma}\left\|(\gamma g_1(x)+u_1)_+-(\gamma g_1(\bar{x})+\bar{u}_1)_+\right\|\|u_1-\bar{u}_1\| \\
&\leq \frac{1}{\gamma}\|\gamma g_1(x)+u_1-\gamma g_1(\bar{x})-\bar{u}_1\|\|u_1-\bar{u}_1\| \\
&\leq K\|x-\bar{x}\|\|u_1-\bar{u}_1\|+\frac{1}{\gamma}\|u_1-\bar{u}_1\|^2
\end{aligned}
$$

Defining $d := \|\bar{x}-x\|$ and $e := \|u_1-\bar{u}_1\|$ we obtain the quadratic inequality in d

$$kd^2-Ked-\frac{1}{\gamma}e^2 \leq 0$$

and hence d must lie between the roots

$$d=\frac{Ke \pm \sqrt{K^2e^2+4ke^2/\gamma}}{2k}$$

Thus

$$d \leq \frac{K}{2k}[1+\sqrt{1+4k/\gamma K^2}]e$$

which gives the required Lipschitz continuity. □

We are now able to state a parallel constraint distribution algorithm for the convex program (3.1) and establish its convergence.

THEOREM 3.2 (PCD ALGORITHM AND CONVERGENCE FOR CONVEX PROGRAMS). *Let $f: \mathbb{R}^n \to \mathbb{R}$, $g_1: \mathbb{R}^n \to \mathbb{R}^{m_1}$, $g_2: \mathbb{R}^n \to \mathbb{R}^{m_2}$ be continuously differentiable convex functions on $\mathbb{R}^n$ with f strongly convex and g_1, g_2 Lipschitz continuous on $\mathbb{R}^n$. Let $g(x) := \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix}$ and let g_1 and g_2 satisfy some constraint qualification on the nonempty sets $\{x \mid g_1(x) \le 0\}$ and $\{x \mid g_2(x) \le 0\}$ respectively. Define $s^i := \begin{bmatrix} s_1^i \\ s_2^i \end{bmatrix} \in \mathbb{R}^{m_1+m_2}$ and start with $s_1^0 = 0$, $s_2^0 = 0$. Given s^i determine s^{i+1} as follows:*

$$(x_1^{i+1}, s_1^{i+1}) \in \arg\mathrm{KKT}(\min\left\{ \mathrm{f(x)} + \frac{1}{2\gamma}\left\|\left(\gamma \mathrm{g_2(x)} + \mathrm{s_2^i}\right)_+\right\|^2 \mid \mathrm{g_1(x)} \le 0\right\})$$

$$(x_2^{i+1}, s_2^{i+1}) \in \arg\mathrm{KKT}(\min\left\{ \mathrm{f(x)} + \frac{1}{2\gamma}\left\|\left(\gamma \mathrm{g_1(x)} + \mathrm{s_1^i}\right)_+\right\|^2 \mid \mathrm{g_2(x)} \le 0\right\})$$

Assume that $\{s^{i+1} - s^i\} \to 0$, then for each accumulation point $\bar{s}$ such that $\{s^{i_j}\} \to \bar{s}$, $\left\{x_1^{i_j}\right\}$ and $\left\{x_2^{i_j}\right\}$ converge to $\arg\min\{f(x) \mid g(x) \le 0\}$.

Proof. By Lemma 3.1, $x_1^{i+1} := x_1(s^i)$ $x_2^{i+1} := x_2(s^i)$ are continuous. Let $\{s^{i_j}\} \to \bar{s}$. Hence $\{s^{i_j+1}\} \to \bar{s}$, $\left\{x_1^{i_j}\right\} \to \bar{x}_1$ and $\left\{x_2^{i_j}\right\} \to \bar{x}_2$. Invoking the continuity of the Karush–Kuhn–Tucker conditions we have at these limits

$$\begin{gathered}\nabla f(\bar{x}_1) + (\gamma g_2(\bar{x}_1) + \bar{s}_2)_+^T \nabla g_2(\bar{x}_1) + \bar{s}_1^T \nabla g_1(\bar{x}_1) = 0 \\ \bar{s}_1 = (\gamma g_1(\bar{x}_1) + \bar{s}_1)_+\end{gathered}$$

and

$$\begin{gathered}\nabla f(\bar{x}_2) + (\gamma g_1(\bar{x}_2) + \bar{s}_1)_+^T \nabla g_1(\bar{x}_2) + \bar{s}_2^T \nabla g_2(\bar{x}_2) = 0 \\ \bar{s}_2 = (\gamma g_2(\bar{x}_2) + \bar{s}_2)_+\end{gathered}$$

Hence

$$\nabla f(\bar{x}_1) + (\gamma g_2(\bar{x}_1) + \bar{s}_2)_+^T \nabla g_2(\bar{x}_1) + (\gamma g_1(\bar{x}_1) + \bar{s}_1)_+^T \nabla g_1(\bar{x}_1) = 0$$

and

$$\nabla f(\bar{x}_2) + (\gamma g_1(\bar{x}_2) + \bar{s}_1)_+^T \nabla g_1(\bar{x}_2) + (\gamma g_2(\bar{x}_2) + \bar{s}_2)_+^T \nabla g_2(\bar{x}_2) = 0$$

Thus

$$\bar{x}_1 = \bar{x}_2 = \arg\min\{f(x) + \frac{1}{2\gamma}\left\|(\gamma g(x) + \bar{s})_+\right\|^2\}$$

because the objective of the last minimization problem is strongly convex. Hence $(\bar{x}_1, \bar{s})$ and $(\bar{x}_2, \bar{s})$ satisfy the Karush–Kuhn–Tucker conditions of $\min\{f(x) \mid g(x) \le 0\}$ and thus $\bar{x}_1 = \bar{x}_2 = \arg\min\{f(x) \mid g(x) \le 0\}$. □

4. Computational experience. We have tested out the algorithms of the previous sections on some linear programming problems. The standard form linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

has the dual problem

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \leq c \end{array} \tag{4.1}$$

and these problems are in precisely the form of our preceding discussion except the objective is not strongly convex. In order to strongly convexify the objective we have used the least two–norm formulation [14, 11], where for $\epsilon \in (0, \bar{\epsilon}]$ for some $\bar{\epsilon} > 0$, the solution of

$$\begin{array}{ll} \text{minimize} & -b^T y + \frac{\epsilon}{2} y^T y \\ \text{subject to} & A^T y \leq c \end{array} \tag{4.2}$$

is the least two–norm solution of (4.1). For the purpose of our computation, a value of $\epsilon = 10^{-6}$ was used.

We have split up the problems as follows: firstly the user has specified the number of processors available and the problem has been split into that many blocks. If the number of constraints in each block is not the same we have added combinations of constraints form other blocks to make the number of constraints in each block equal with the aim of balancing the load in each processor.

The PCD Algorithm 3.2 of Section 3 was implemented on the Sequent Symmetry S–81 shared memory multiprocessor. The subproblems were solved on each processor using MINOS 5.3 a more recent version of [15]. The explicit constraints in each subproblem remained fixed throughout the computation but the blocks were not chosen to satisfy the linear independence assumption.

We have used the following scheme to update the augmented Lagrangian parameter, γ. Initially it is set at 10 and is increased by a factor of 4 only when the norm of the violation of the constraints increases.

The step–length λ in the method (which is needed in the convergence proof) was chosen by several techniques. One technique was to choose a fixed positive step–length $\lambda < 1$. With a step–length of 1 we found that the algorithm did fail to converge in several instances as the theory would suggest (see Table 2). Another heuristic technique was to choose the step–length between 0.4 and 1.0 to minimize the augmented Lagrangian. This has proven to be robust and results in a good saving in iterations.

The algorithm was terminated whenever the difference in the primal objective value of (4.1) and its dual objective value normalized by their sum differed by less than 10^{-5}. The constraint violation was also required to be less than this tolerance.

Problem	Variables	Constraints	Blocks			
			3	6	9	18
Ex6	3	5	10	10		
Ex9	5	11	10	12		
Ex10	6	14	11	12	13	
AFIRO	27	51	16	16	16	16
ADLittle	56	138	14	27	19	27

TABLE 1
Numerical results with fixed $\lambda = 0.7$

Problem	Variables	Constraints	Blocks			
			3	6	9	18
Ex6	3	5	2	2		
Ex9	5	11	4	*		
Ex10	6	14	4	4	4	
AFIRO	27	51	20	14	*	*
ADLittle	56	138	*	*	*	*

TABLE 2
Numerical results with fixed $\lambda = 1.0$

Problem	Variables	Constraints	Blocks			
			3	6	9	18
Ex6	3	5	2	2		
Ex9	5	11	4	5		
Ex10	6	14	4	4	4	
AFIRO	27	51	13	15	15	14
ADLittle	56	138	12	14	14	15

TABLE 3
Numerical results with variable λ

Tables 1, 2 and 3 summarize preliminary numerical results for the PCD Algorithm 3.2 on the Sequent Symmetry S–81 for 5 small linear programs reformulated as in (4.2). The first three are homemade test problems, while the last two, AFIRO and ADLittle, are from the NETLIB collection [6]. In the tables, an empty column entry signifies that we did not perform the computation. The character * signifies that the algorithm did not terminate. Note that for the algorithm does fail when a full step is taken (see Table 2) as may be expected from Theorem 2.4 where the step–size λ must satisfy (2.25). The heuristic step–size outlined above performs the best (see Table 3).

The key observation to make is that the total number of iterations required for accurate solutions (tolerance $< 10^-5$) can be achieved with a small number of iterations (2–13 iterations for 3 blocks and 14–15 iterations for 18 blocks). The fact that the number of iterations remains essentially constant for increasing number of blocks is encouraging and leads us to believe that the PCD is worthy of additional theoretical and computational study.

REFERENCES

[1] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[2] R.W. Cottle, J.-S. Pang, and R.E. Stone. *The Linear Complementarity Problem.* Academic Press, New York, 1991.

[3] R. De Leone and O.L. Mangasarian. Parallel proximal point decomposition of linear programming constraints. SIAM National Meeting, Chicago, Illinois, July 16–20, 1990.

[4] A.R. de Pierro and A.N. Iusem. Convergence properties of iterative methods for symmetric positive semidefinite linear complementarity problems. Technical report, Instituto de Matematica, Elasticita e Ciencia da Computacao, Universidade Estadual de Campinas, CP 6065, Campinas, 13081, SP, Brazil, 1990.

[5] J. Eckstein. The Lions–Mercier splitting algorithm and the alternating direction method as instances of the proximal point algorithm. Technical Report LIDS-P-1769, MIT, Cambridge, Massachusetts, 1988.

[6] D.M. Gay. Electronic mail distribution of linear programming test problems. *COAL Newsletter*, 13:10–12, 1985.

[7] S.-P. Han. A decomposition method and its application to convex programming. *Mathematics of Operations Research*, 14:237–248, 1989.

[8] Z.-Q. Luo and P. Tseng. On the convergence of a matrix splitting algorithm for the symmetric monotone linear complementarity problem. Technical Report LIDS-P-1884, MIT, Cambridge, Massachusetts, 1989. To appear, SIAM Journal on Control and Optimization.

[9] O.L. Mangasarian. *Nonlinear Programming.* McGraw–Hill, New York, 1969.

[10] O.L. Mangasarian. Solution of symmetric linear complementarity problems by iterative methods. *Journal of Optimization Theory and Applications*, 22(4):465–485, August 1977.

[11] O.L. Mangasarian. Normal solutions of linear programs. *Mathematical Programming Study*, 22:206–216, 1984.

[12] O.L. Mangasarian. A simple characterization of solution sets of convex programs. *Operations Research Letters*, 7(1):21–26, 1988.

[13] O.L. Mangasarian. On the convergence of iterates of an inexact matrix splitting algorithm for the symmetric monotone linear complementarity problem. Technical Report 917, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, March 1990. To appear in SIAM Journal on Optimization, 1 1991.

[14] O.L. Mangasarian and R.R. Meyer. Nonlinear perturbation of linear programs. *SIAM Journal*

on Control and Optimization, 17(6):745–752, November 1979.

[15] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, December 1983.

[16] J.-S. Pang. Convergence of splitting and Newton methods for complementarity problems: An application of some sensitivity results. Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, September 1990.

[17] R.T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12:268–285, 1974.

[18] R.T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976.

[19] R.T. Rockafellar and R.J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. Working paper, University of California–Davis, California, 1989. To appear in Mathematics of Operations Research.

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