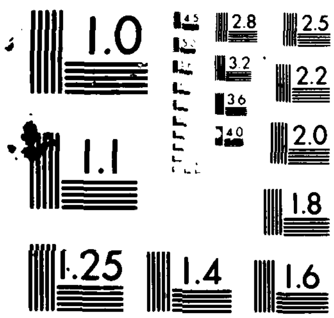
AD-A110 463 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
AN ALGORITHM FOR STRUCTURED, LARGE-SCALE QUADRATIC PROGRAMMING --ETC(U)
SEP 81 C D HA DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2276 NL

END
DATE
FILMED
3 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

MRC Technical Summary Report #2276

AN ALGORITHM FOR STRUCTURED, LARGE-SCALE QUADRATIC PROGRAMMING PROBLEMS

Cu Duong Ha

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

September 1981

Received August 12, 1981

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

AN ALGORITHM FOR STRUCTURED, LARGE-SCALE QUADRATIC PROGRAMMING PROBLEMS

Cu Duong Ha*

Technical Summary Report #2276
September 1981

ABSTRACT

An algorithm for structured, large-scale, convex quadratic programming problems is described. The structure of the constraint matrix is block diagonal with a small number of coupling constraints and variables. The algorithm utilizes twice a decomposition procedure that was developed earlier. The first time the decomposition procedure is used to break up the coupling constraints, and the second time it is used to break up the coupling variables. Preliminary computational results are also reported.

AMS(MOS) Subject Classifications: 90C20.

Key words: Quadratic Programs, Decomposition Method, Large-scale Problems, Computational Experience.

Work Unit Number 5 - Operations Research.

---

* Current Address: Mathematical Sciences Department
Virginia Commonwealth University
Richmond, Virginia 23284

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The block diagonal structure with a small number of coupling constraints and variables usually arises from the formulation of multitime period and multi-division production scheduling and distribution models in large corporations. Each block is concerned with the operation of one division during one time period considered in isolation. The coupling constraints arise from the use of common resources of all divisions or from combining the output of divisions to meet overall demands. The coupling variables represent activities that affect the operation of divisions in more than one time period.

In this paper we propose a decomposition method for convex quadratic programming problems having block diagonal structure with a small number of coupling constraints and variables. The basic idea of decomposition methods is to break up a large complex problem into a sequence of subproblems that are easier to solve.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

AN ALGORITHM FOR STRUCTURED, LARGE-SCALE QUADRATIC PROGRAMMING PROBLEMS

Cu Duong Ha*

1. Introduction

The problem that we shall be concerned with is

$$\begin{aligned}
\min \quad & \sum_{i=0}^{p} (\langle c_i, x_i \rangle + \tfrac{1}{2}\langle x_i, C_i x_i \rangle) \\
\text{subject to} \quad & B_1 x_0 + D_1 x_1 = d_1 \\
& B_2 x_0 + D_2 x_2 = d_2 \\
& \vdots \\
& B_p x_0 + D_p x_p = d_p \\
& A_0 x_0 + A_1 x_1 + A_2 x_2 + \ldots + A_p x_p = a \\
& x_i \geq 0 \quad i = 0,1,\ldots,p
\end{aligned} \tag{1.1}$$

where $x_i \in \mathbb{R}^{n_i}$, $c_i \in \mathbb{R}^{n_i}$, $d_i \in \mathbb{R}^{m_i}$, $a \in \mathbb{R}^{m_0}$, A_i, B_i, C_i, and D_i are matrices of dimensions $\ell \times n_i$, $m_i \times n_0$, $n_i \times n_i$ and $m_i \times n_i$, respectively; and $\langle \cdot, \cdot \rangle$ is the inner product.

We denote $x := (x_0, x_1, \ldots, x_p)$, $n := \sum_{i=0}^{p} n_i$.

We assume that the problem is feasible and C_i is symmetric and positive semidefinite for all i.

Note that we do not rule out the case $C_i \equiv 0$ for all i ; accordingly our algorithm can be used to solve block diagonal linear programming problems with coupling constraints and variables.

It is well-known that decomposition approaches for large-scale problems which uses dual methods have several disadvantages (see, for example, [Shapiro, 1979] or [Lasdon,1970]). To circumvent these drawbacks we developed a decomposition procedure by combining a dual

* Current Address: Matheamtical Sciences Department
Virginia Commonwealth University
Richmond, Virginia 23284

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

method and the proximal point algorithm and used it for solving block angular linear programming problems with coupling constraints and a large-scale, nonlinear problem of structural engineering ([Ha, 1981] and [Kaneko and Ha, 1980]). In this paper we apply the same decomposition procedure to the problem (1.1). The resulting problem is a quadratic programming problem having only coupling variables; and its dual problem has only coupling constraints. Then we modify our decomposition procedure to break up that dual problem completely into small subproblems.

The organization of this paper is as follows. Section 2 is the review of the decomposition procedure in [Ha, 1981]. Section 3 is the proposed solution method for the problem (1.1). Section 4 is the implementation and computational results of the algorithm.

2. *A decompostion procedure ([Ha, 1981])*

In this section we consider the convex programming problem

$$\begin{aligned} &\min \quad f(x) \\ &\text{subject to} \qquad\qquad\qquad (2.1) \\ &\qquad x \in C \\ &\qquad Ax = a \end{aligned}$$

where $x \in \mathbb{R}^n$, $a \in \mathbb{R}^m$, A is an $m \times n$ matrix, f is a convex function defined on $\mathbb{R}^n$, and C is a nonempty closed convex set in $\mathbb{R}^n$. We assume that the problem (2.1) has optimal solutions.

Let $\{\lambda_k\}$ be a sequence of positive numbers bounded away from zero (i.e., $\lambda_k \geq \lambda > 0$ for all k). Let x^0 be an arbitrary point in $\mathbb{R}^n$. Suppose we already have $\{x^0, x^1, \ldots, x^k\}$, then x^{k+1} is the unique optimal solution of the problem

$$\min \quad f(x) + \frac{1}{2\lambda_k} \| x - x^k \|^2$$

$$\text{subject to} \qquad\qquad (2.2)$$

$$x \in C$$

$$Ax = a$$

Then by the proximal point method ([Rockafellar, 1976 a and b] and [Brézis and Lions, 1978]) the sequence $\{x^k\}$ will converge to an optimal solution of the problem (2.1).

The advantage of (2.2) over (2.1) is that the objective function of (2.2) is strongly convex although the function f may not be so. A function h defined on a convex set S is said to be strongly convex with modulus $\alpha > 0$ if

$$h((1-t)x + tz) \leqq (1-t)h(x) + t\,h(z) - \frac{1}{2}\alpha(1-t)t\,\| x - z \|^2$$

for all x and z in S and $0 < t < 1$.

In order to solve the problem (2.2) we use a dual method. The dual of (2.2) is

$$\max_{y \in \mathbb{R}^m} \quad g(y) \qquad (2.3)$$

where

$$g(y) := \min_{x \in C} \left(f(x) + \frac{1}{2\lambda_k} \|x - x^k\|^2 - \langle y , Ax - a \rangle\right) \quad (2.4)$$

Since $f(x) + \frac{1}{2\lambda_k} \| x - x^k\|^2$ is strongly convex. The problem (2.4) has a unique optimal solution for any y in $\mathbb{R}^m$. In other words, g(y) is finite everywhere. It can also be proved that the function g is Lipschitz continuously differentiable and the derivative g' of g at a point y is given by

$$g'(y) = Ax(y) - a$$

where x(y) is the optimal solution of (2.4) corresponding to the given y.

If y* is the optimal solution for (2.3) then x(y*) is the optimal solution for (2.2).

In summary, the general scheme for solving (2.1) is to choose points $x^0 \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and then solve the minimization problem (2.4) to obtain x(y), g(y) and g'(y). If y is not an optimal solution for (2.3), update y and solve (2.4) again until the optimal solution y* of (2.3) is obtained. If x(y*) is the same as x^0 then it is an optimal solution for the original problem (2.1); otherwise replace x^0 by x(y*) and repeat the procedure.

Note that in (2.4) we do not have the constraints Ax = a which can be the coupling constraints in the case that (2.1) is a structured large-scale problem. Because of computational experience with $\{\lambda_k\}$ in [Ha, 1981] we shall keep λ_k fixed; i.e., $\lambda_k = \lambda > 0$ for all k.

3. A method of solution for the problem (1.1)

We use the multilevel concept to explain our method of solution for the problem (1.1). It is a three level algorithm. The first level is the application of the decomposition procedure above to the problem (1.1). Starting from a given point $x^0 = (x_0^0, x_1^0, \ldots, x_p^0)$ We find the unique minimizer $\bar{x} = (\bar{x}_0, \bar{x}_1, \ldots \bar{x}_p)$ for the following problem.

$$\min \quad \sum_{i=0}^{p} (\langle c_i, x_i\rangle + \frac{1}{2}\langle x_i, C_i x_i\rangle + \frac{1}{2\lambda}\|x_i - x_i^0\|^2)$$

$$\text{subject to } \begin{array}{lllll} B_1x_0 & + D_1x_1 & & = & d_1 \\ B_2x_0 & + D_2x_2 & & = & d_2 \\ \vdots & & \ddots & & \vdots \\ B_px_0 & & + D_px_p & = & d_p \\ A_0x_0 & + A_1x_1 + & \ldots + A_px_p & = & a \end{array} \qquad (3.1)$$

$$x_i \geqq 0 \quad i = 0, 1, \ldots, p .$$

If $\bar{x}$ is close enough to x^0, $\bar{x}$ is approximately an optimal solution for the original problem (1.1) ; if not, we replace x^0 by $\bar{x}$ and solve (3.1) again. We solve (3.1) by a dual method. The dual problem of (3.1) is

$$\max g(y) \qquad (3.2)$$

where

$$g(y) := \min \quad \sum_{i=0}^{p} (\langle c_i, x_i \rangle + \frac{1}{2}\langle x_i, C_i x_i \rangle + \frac{1}{2\lambda} \|x_i - x_i^0\|^2) + \langle y, (\sum_{i=0}^{p} A_i x_i - a) \rangle$$

$$\begin{aligned} \text{subject to } & B_1 x_0 + D_1 x_1 = d_1 \\ & B_2 x_0 + D_2 x_2 = d_2 \\ & \vdots \\ & B_p x_0 + D_p x_p = d_p \\ & x_i \geq 0 \quad \text{for } i = 0,1,\ldots,p . \end{aligned} \tag{3.3}$$

Let $\bar{c}_i := c_i - \frac{1}{\lambda} x_i^0 + A_i^T y$

and $\bar{C}_i := C_i + \frac{1}{\lambda} I$

where I is the identity matrix with appropriate dimension; then $g(y)$ can be rewritten as

$$g(y) = -\langle a, y \rangle + \frac{1}{2\lambda} \|x_i^0\|^2 + g_1(y) ,$$

here, $g_1(y)$ is the optimal objective value of the problem:

$$\min \quad \sum_{i=0}^{p} (\langle \bar{c}_i, \bar{x}_i \rangle + \frac{1}{2} \langle x_i, \bar{C}_i x_i \rangle$$

$$\begin{aligned} \text{subject to } & B_1 x_0 + D_1 x_1 = d_1 \\ & B_2 x_0 + D_2 x_2 = d_2 \\ & \vdots \\ & B_p x_0 + D_p x_p = d_p \\ & x_i \geq 0 \quad \text{for } i = 0,1,\ldots,p . \end{aligned} \tag{3.4}$$

For a given y, the second level will give us x(y) the optimal solution of (3.4). Then we can compute g(y) and g'(y). If y is the optimal solution of (3.2) then x(y) is the variable $\bar{x}$ that we are looking for; otherwise, we can use any unconstrained optimization algorithm to update y and send the new value of y to the second level. The second level is concerned with the problem (3.4) for a given y receiving from the first level ($\bar{c}$ is a function of y). The ordinary dual of (3.4) can be written as ([Mangasarian, 1969]).

$$\min_{(x,v)} \quad \sum_{i=0}^{p} \left(\frac{1}{2}\langle x_i, \bar{C}_i x_i\rangle\right) + \sum_{i=1}^{p} \langle v_i, d_i\rangle \tag{3.5.1}$$

$$\text{subject to} \quad \left.\begin{array}{llll} \bar{C}_1 x_1 + D_1^T v_1 & & & \geq -\bar{c}_1 \\ & \bar{C}_2 x_2 + D_2^T v_2 & & \geq -\bar{c}_2 \\ & & \ddots & \vdots \\ & & \bar{C}_p x_p + D_p^T v_p & \geq -\bar{c}_p \end{array}\right\} \tag{3.5.2}$$

$$\bar{C}_0 x_0 + B_1^T v_1 + \quad . \quad . \quad . \quad + B_p^T v_p \geq -\bar{c}_0 \tag{3.5.3}$$

The problem above is a quadratic programming problem having block angular structure with (3.5.3) considered as the coupling constraints. It can be solved by the decomposition procedure outlined in Section 2. But in this special case, because the objective function (3.5.1) is already strongly convex with respect to x, we can simplify the procedure by adding only quadratic terms for the variable v. More specifically, let $\{\mu_k\}$ be a sequence of positive numbers bounded away from zero and v^0 be a starting point. Suppose we have generated from v^0 the sequence $\{(x^1,v^1),(x^2,v^2), \ldots, (x^k,v^k)\}$; let (x^{k+1}, v^{k+1}) be the optimal solution of the following problem

$$\min_{(x,v)} \quad \sum_{i=0}^{p} \left(\tfrac{1}{2}\langle x_i, \bar{C}_i x_i\rangle\right) + \sum_{i=1}^{p} \langle v_i, d_i\rangle + \frac{1}{2\mu_k}\|v-v^k\|^2$$

$$\begin{array}{llr} \text{subject to} & \bar{C}_1 x_1 + D_1^T v_1 & \geq -\bar{c}_1 \\ & \qquad \bar{C}_2 x_2 + D_2^T v_2 & \geq -\bar{c}_2 \\ & \qquad\qquad \ddots & \vdots \\ & \qquad\qquad\qquad \bar{C}_p x_p + D_p^T v_p & \geq -\bar{c}_p \\ & \bar{C}_0 x_0 + B_1^T v_1 + \quad \cdots \quad + B_p^T v_p & \geq -\bar{c}_0 \, . \end{array} \tag{3.6}$$

then the sequence $\{(x^k, v^k)\}$ will converge to an optimal solution for (3.5) (justifications will be given at the end of this section). Again, we solve (3.6) by a dual method. The dual of (3.6) is

$$\max_{u \geq 0} h(u) \tag{3.7}$$

where, for a fixed u,

$$h(u) = \min_{(x,v)} \quad \sum_{i=0}^{p} \left(\tfrac{1}{2}\langle x_i, \bar{C}_i x_i\rangle\right) + \sum_{i=1}^{p} \langle v_i, d_i\rangle + \frac{1}{2\mu_k}\|v-v^k\|^2 + \langle u, -\bar{c}_0 - \bar{C}_0 x_0 - \sum_{j=1}^{p} B_j^T v_j\rangle$$

$$\begin{array}{llr} \text{subject to} & \bar{C}_1 x_1 + D_1^T v_1 & \geq -\bar{c}_1 \\ & \qquad \bar{C}_2 x_2 + D_2^T v_2 & \geq -\bar{c}_2 \\ & \qquad\qquad \ddots & \\ & \qquad\qquad\qquad \bar{C}_p x_p + D_p^T v_p & \geq -\bar{c}_p \, . \end{array} \tag{3.8}$$

The third level is to solve the minimization problem (3.8) above for given u and v^k and then pass back to the second level its optimal solution (x(u) , v(u)). The problem (3.8) can be separated into (p+1) small subproblems. The subproblem corresponding to i=0 is

$$\min_{x_0} \frac{1}{2}\langle x_0,\bar{C}_0 x_0\rangle - \langle u,\bar{C}_0 x_0\rangle , \tag{3.9}$$

whose solution is $x_0 = u$; for $i \geq 1$, it is

$$\min_{(x_i,v_i)} \frac{1}{2}\langle x_i,\bar{C}_i x_i\rangle + \langle v_i d_i\rangle + \frac{1}{2\mu_k}\|v_i - v_i^k\|^2 - \langle u,B_i^T v_i\rangle \tag{3.10}$$

subject to $\bar{C}_i x_i + D_i^T v_i \geq -\bar{c}_i$

which can be solved by any algorithm for quadratic programming problems.

In the remainder of this section we shall prove a result to justify the second level. We consider the problem

$$\min_{(x,v)\in C} \theta(x) + \phi(v) \tag{3.11}$$

where C is a nonempty closed convex set, θ and ϕ are closed covex functions having real values.

We suppose the problem has optimal solutions.

Lemma 3.1

Let μ be a fixed positive number. Given any two points v^0 and v^1, we define

$$(\bar{x}^0,\bar{v}^0) := \underset{(x,v)\in C}{\operatorname{argmin}} \ \theta(x) + \phi(v) + \frac{1}{2\mu}\|v-v^0\|^2$$

and

$$(\bar{x}^1,\bar{v}^1) := \underset{(x,v)\in C}{\text{argmin}} \ \theta(x) + \phi(v) + \frac{1}{2\mu}\|v-v^1\|^2 ,$$

then

$$\|v^0-v^1\|^2 \geq \|\bar{v}^0-\bar{v}^1\|^2 + \|v^0-\bar{v}^0+\bar{v}^1-v^1\|^2 \tag{3.12}$$

and

$$\|v^0-v^1\| \geq \|\bar{v}^0-\bar{v}^1\| \tag{3.13}$$

Proof

It is clear that (3.13) follows from (3.12), so only (3.12) needs to be proven. We have

$$\|v^0-v^1\|^2 = \|v^0-\bar{v}^0+\bar{v}^0-\bar{v}^1+\bar{v}^1-v^1\|^2$$

$$= \|\bar{v}^0-\bar{v}^1\|^2 + \|v^0-\bar{v}^0+\bar{v}^1-v^1\|^2 + 2\langle \bar{v}^0-\bar{v}^1, v^0-\bar{v}^0+\bar{v}^1-v^1\rangle ,$$

so to prove (3.12) we shall show that

$$\langle \bar{v}^0-\bar{v}^1,\ v^0-\bar{v}^0+\bar{v}^1-v^1\rangle \geq 0 .$$

Let

$$F(x,v) := \theta(x) + \phi(v) + \psi_C(x,v)$$

and $$G(x,v) := (2\mu)^{-1}\ \|v-v^0\|^2 .$$

Since F and G are both proper convex and the effective domain of G is the entire (x,v)-space, we have ([Rockafellar, 1968])

$$\partial(F+G) = \partial F + \partial G \ .$$

It follows from the definition of the subdifferential that

$$(0,0) \in \partial(F+G)\,(\bar{x}^0,\bar{v}^0)$$

$$= \partial F(\bar{x}^0,\bar{v}^0) + \partial G(\bar{x}^0,\bar{v}^0)$$

$$= \partial F(\bar{x}^0,\bar{v}^0) + \{0\} \times \mu^{-1}(\bar{v}^0-v^0)$$

or

$$(0,\mu^{-1}(\bar{v}^0-v^0)) \in \partial F(\bar{x}^0,\bar{v}^0) \ .$$

Similarly, we have

$$(0,\mu^{-1}(\bar{v}^1-v^1)) \in \partial F(\bar{x}^1,\bar{v}^1) \ .$$

Therefore by the monotonicity of ∂F

$$\langle 0-0,\bar{x}^0-\bar{x}^1\rangle + \mu^{-1}\langle v^0-\bar{v}^0-v^1+\bar{v}^1,\bar{v}^0-\bar{v}^1\rangle \geq 0 \ ,$$

or

$$\langle \bar{v}^0-\bar{v}^1, v^0-\bar{v}^0+\bar{v}^1-v^1\rangle \geq 0 \ .$$

<u>Theorem 3.1</u>

In addition to the above assumptions, we assume that θ is strongly convex. Let $\{\mu_k\}$ be a sequence of positive numbers bounded away from zero and let v^0 be an arbitrary point. The sequence $\{(x^k,v^k)\}$ generated by

$$(x^{k+1},v^{k+1}) := \underset{(x,v)\in C}{\operatorname{argmin}} \ \theta(x) + \phi(v) + \frac{1}{2\mu_k}\|v-v^k\|^2$$

converges to a point (x^*,v^*) which is an optimal solution for (3.11).

Proof

Let $(\bar{x},\bar{v})$ be an optimal solution of (3.11). It is easy to see that $(\bar{x},\bar{v})$ is also the optimal solution of the following problem

$$\min_{(x,v)\in C} \ \theta(x) + \phi(v) + \frac{1}{2\mu_k}\|v-\bar{v}\|^2 .$$

Applying lemma 3.1 with $\mu = \mu_k$, $v^0 = \bar{v}^0 = \bar{v}$, $v^1 = v^k$ and $\bar{v}^1 = v^{k+1}$, we have

$$\|v^{k+1}-\bar{v}\| \leqq \|v^k-\bar{v}\| \tag{3.14}$$

and $$\|v^{k+1}-\bar{v}\|^2 + \|v^{k+1}-v^k\|^2 \leqq \|v^k-\bar{v}\|^2 .$$

The first inequality gives us the boundedness of the sequence $\{v^k\}$. The second shows that

$$\sum_{k=1}^{\infty} \|v^{k+1}-v^k\|^2 < \infty .$$

It implies that $\|v^{k+1}-v^k\| \to 0$ as $k \to \infty$. Now, we are going to show that the sequence $\{x^k\}$ is also bounded. First we have

$$\theta(x^{k+1}) + \phi(v^{k+1}) \leq \theta(x^{k+1}) + \phi(v^{k+1}) + \frac{1}{2\mu_k}\|v^{k+1}-v^k\|^2 \leq \theta(x^k) + \phi(v^k) ,$$

so $$\theta(x^{k+1}) \leq \theta(x^1) + \phi(v^1) - \phi(v^{k+1}) \quad \forall_k .$$

The boundedness of $\{v^k\}$ implies the boundedness of $\{\phi(v^k)\}$. Let M be a constant such that

$$\phi(v^1) - \phi(v^k) \leq M \quad \forall_k ,$$

then

$$x^k \in W := \{x \mid \theta(x) \leq \theta(x^1) + M\} \quad \forall_k .$$

The set W is compact by the strong convexity of ψ ; consequently $\{x^k\}$ is bounded.

Since the sequence $\{(x^k,v^k)\}$ is bounded, there is a subsequence $\{(x^{k_j},v^{k_j})\}$ that converges to a point $(x^*,v^*) \in C$. For any point $(x,v) \in C$, by lemma 3.1 we have

$$\theta(x) + \phi(v) \geq \theta(x^{k_j}) + \phi(v^{k_j}) + \frac{1}{\mu_{k_j}}\langle v^{k_j-1} - v^{k_j}, v - v^{k_j}\rangle .$$

Letting $k_j \to \infty$, we have

$$\theta(x^{k_j}) + \phi(v^{k_j}) \to \theta(x^*) + \phi(v^*)$$

$$v^{k_j-1} - v^{k_j} \to 0 .$$

Since $\{\mu_k\}$ is bounded away from zero

$$\frac{1}{\mu_{k_j}}\langle v^{k_j-1}-v^{k_j}, v-v^{k_j}\rangle \to 0 ,$$

so
$$\theta(x) + \phi(v) \geq \theta(x^*) + \phi(v^*) .$$

This is true for an arbitrary $(x,v) \in C$, so (x^*,v^*) solves the problem (3.11).

Using (3.14) with $\bar{v}$ replaced by v^* we have

$$\|v^{k+1}-v^*\| \leq \|v^k-v^*\| ,$$

so the sequence $\{v^k\}$ has to converge to v^*.

Suppose $\{x^k\}$ has two cluster points $\bar{x}$ and $\bar{\bar{x}}$, then both $(\bar{x},\bar{v})$ and $(\bar{\bar{x}},\bar{v})$ are optimal solutions for the problem

$$\min_{(x,v)\in C} \theta(x) + \phi(v) + \frac{1}{2\mu}\|v-\bar{v}\|^2$$

for any positive number μ. But the function $\theta(x)+\phi(v)+\frac{1}{2\mu}\|v-\bar{v}\|^2$ is strongly convex with respect to (x,v), so the problem above has a unique optimal solution. Hence $\bar{\bar{x}} = \bar{x}$ and the sequence $\{(x^k,v^k)\}$ converges to (x^*,v^*).

4. Implementation and computational results

At Level 1 the problem to be solved is (3.2) which is a nonlinear, unconstrained maximization problem, whose objective function $g(y)$ is differentiable and its derivative $g'(y)$ is available. To solve (3.2) we use the Broyden-Fletcher-Golfard-Shanno (BFGS) algorithm, which is available in the Harwell Subroutine Library and known as VA13A ([MACC, 1976]). For Level 2 we have to handle the nonlinear maximization problem (3.7) with nonnegativity constraints. In order to use an unconstrained algorithm we modify (3.5), (3.6) and (3.7) as follows.

We convert the inequality coupling constraints (3.5.3) into equality constraints by adding slack varibles w; (3.5) then becomes

$$\min_{(x,v),w} \sum_{i=0}^{p} \left(\frac{1}{2}\langle x_i, \bar{C}_i x_i\rangle\right) + \sum_{i=1}^{p} \langle v_i, d_i\rangle$$

$$\begin{aligned}
\text{subject to } & \bar{C}_1 x_1 + D_1^T v_1 && \geq -\bar{c}_1 \\
& \bar{C}_2 x_2 + D_2^T v_2 && \geq -\bar{c}_2 \\
& \quad \ddots && \quad \vdots \\
& \bar{C}_p x_p + D_p^T v_p && \geq -\bar{c}_p \\
& \bar{C}_0 x_0 + B_1^T v_1 + \cdots + B_p^T v_p - w && = -\bar{c}_0 \\
& w \geq 0 .
\end{aligned} \tag{3.5'}$$

Instead of (3.6) and (3.7) we have

$$\min_{(x,v),w} \sum_{i=0}^{p} \left(\frac{1}{2}\langle x_i, \bar{C}_i x_i\rangle + \sum_{i=1}^{p}\langle v_i, d_i\rangle + \frac{1}{2\mu_k}\|v-v^k\|^2 + \frac{1}{2\mu_k}\|w-w^k\|^2\right.$$

$$\begin{array}{llllll} \text{subject to} & \bar{C}_1 x_1 + D_1^T v_1 & & & & \geq -\bar{c}_1 \\ & & \bar{C}_2 x_2 + D_2^T v_2 & & & \geq -\bar{c}_2 \\ & & & \ddots & & \vdots \\ & & & & \bar{C}_p x_p + D_p^T v_p & \geq -\bar{c}_p \\ & \bar{C}_0 x_0 + B_1^T v_1 + & \cdot & \cdot & \cdot \; + B_p^T v_p - w & = -\bar{c}_0 \\ & & & w \geq 0 & & \end{array} \qquad (3.6')$$

and

$$\max_{u \in \mathbb{R}^{n_0}} h(u) \qquad (3.7')$$

where, for a fixed u ,

$$h(u) = \min_{(x,v),w} \sum_{i=0}^{p} \left(\frac{1}{2}\langle x_i, C_i, x_i\rangle\right) + \sum_{i=1}^{p}\langle v_i, d_i\rangle + \frac{1}{2\mu_k}\|v-v^k\|^2 + \frac{1}{2\mu_k}\|w-w^k\|^2 +$$

$$\langle u, -\bar{c}_0 - \bar{C}_0 x_0 - \sum_{i=1}^{p} B_i^T v_i + w\rangle$$

$$\begin{array}{lllll} \text{subject to} & \bar{C}_1 x_1 + D_1^T v_1 & & & \geq -\bar{c}_1 \\ & & \bar{C}_2 x_2 + D_2^T v_2 & & \geq -\bar{c}_2 \\ & & & \ddots & \vdots \\ & & & \bar{C}_p x_p + D_p^T v_p & \geq -\bar{c}_p \\ & & w \geq 0 \; . & & \end{array} \qquad (3.8')$$

Note that (3.7') is an unconstrained maximization problem. The function $h(u)$ is Lipschitz continuously differentiable and its derivatives can be computed easily (Theorem 2.2 in [Ha, 1981]). We also use the BFGS algorithm to solve it.

At Level 3, in addition to (p+1) subproblems (3.9) and (3.10) there is one more subproblem, namely

$$\min_{w \geq 0} \frac{1}{2\mu_k} \|w-w^k\|^2 + \langle u,w \rangle \quad . \tag{4.1}$$

It optimal solution $w(u)$ is given by

$$w(u)_j = \begin{cases} (w^k-\mu_k u)_j & \text{if } (w^k-\mu_k u)_j \geq 0 \\ 0 & \text{otherwise .} \end{cases}$$

Now, we are going to convert the subproblem (3.10) into a quadratic programming problem with only nonnegativity constraints and with the dimension of the variables being n_i (compared to $n_i + m_i$ of (3.10)). First, we define $\bar{d}_i$ to be

$$\bar{d}_i := d_i - B_i u - \frac{1}{\mu_k} v_i^k ,$$

so that the problem (3.10) can be rewritten (except for the constant term $\frac{1}{2\mu_k} \|v_i^k\|^2$) as

$$\min_{(x_i,v_i)} \quad \frac{1}{2}\langle x_i,\bar{C}_i x_i \rangle + \frac{1}{2\mu_k}\langle v_i,v_i \rangle + \langle v_i,\bar{d}_i \rangle$$

$$\text{subject to} \quad -\bar{C}_i x_i - D_i^T v_i \leq \bar{c}_i \quad .$$

Its ordinary dual problem is

$$\max_{(x_i,v_i),z} \; -\frac{1}{2}\langle x_i,\bar{C}_i x_i\rangle - \frac{1}{2\mu_k}\langle v_i,v_i\rangle - \langle \bar{c}_i,z\rangle$$

$$\text{subject to} \quad \bar{C}_i x_i - \bar{C}_i^T z = 0 \qquad (4.2)$$

$$\frac{1}{\mu_k} v_i - D_i z = -\bar{d}_i$$

$$z \geqq 0$$

From the constraints we have

$$\bar{C}_i x_i = \bar{C}_i^T z \ .$$

Since $\bar{C}_i$ is symmetric and positive definite, this implies

$$x_i = z \ .$$

We also have

$$v_i = \mu_k(D_i z - \bar{d}_i)$$

$$= \mu_k(D_i x_i - \bar{d}_i)$$

Replacing z by x_i and v_i by the expression above in (4.2), (4.2) is equivalent to

$$\max_{x_i \geqq 0} \; -\frac{1}{2}\langle x_i,\bar{C}_i x_i\rangle - \frac{1}{2\mu_k}\langle \mu_k(D_i x_i-\bar{d}_i),\ \mu_k(D_i x_i-\bar{d}_i)\rangle - \langle \bar{c}_i,x_i\rangle$$

or

$$\min_{x_i \geq 0} \frac{1}{2}\langle x_i, (\bar{C}_i + \mu_k D_i^T D_i) x_i \rangle + \langle x_i, \bar{c}_i - \mu_k D_i^T \bar{d}_i \rangle \quad . \tag{4.3}$$

Remark

If at Level 2 we also add the quadratic term $\frac{1}{2\mu_k}\|x-x^k\|^2$ into the objective function (using the proximal point algorithm); the subproblem (3.10) of level 3 would be:

$$\min_{(x_i, v_i)} \quad \frac{1}{2}\langle x_i, \bar{C}_i x_i \rangle + \langle v_i, d_i \rangle + \frac{1}{2\mu_k}\|v_i - v_i^k\|^2 + \frac{1}{2\mu_k}\|x_i - x_i^k\|^2 - \langle u, B_i^T v_i \rangle$$

subject to $\bar{C}_i x_i + D_i^T v_i \geq -\bar{c}_i$.

Following the same arguments as above one still could convert this problem into a problem whose only variables are x_i . But instead of (4.3) one would have

$$\min_{x_i} \quad \frac{1}{2}\langle x_i, (\bar{C}_i + \mu_k^{-1} I + (I + \mu_k^{-1}\bar{C}_i^{-1}) D_i^T D_i (I + \mu_k^{-1}\bar{C}_i^{-1})) x_i \rangle$$
$$+ \mu_k \langle x_i, (I + \mu_k^{-1}\bar{C}_i^{-1}) D_i^T (-\mu_k^{-1} D_i \bar{C}_i^{-1} x_i^k - \bar{d}_i) \rangle$$

subject to $(\mu_k I + \bar{C}_i^{-1}) x_i \geq \bar{C}_i^{-1} x_i k$

which is much more complicated.

There are several algorithms for solving the quadratic programming problems (4.3). With pivoting-type algorithms we cannot use the optimal solution of the previous iteration as the starting point for the current iteration. The Best-Ritter algorithm ([Best-Ritter, 1976]) which is a conjugate direction method allows us to do that, so we use it to solve (4.3).

As in the case of λ we keep μ_k fixed; i.e., $\mu_k = \mu > 0$ for all k. In that case the matrices

$$\bar{C}_i + \mu D_i^T D_i \qquad (4.4)$$

remain unchanged throughout the process, so they need to be computed only once.

We wrote a computer program to test the algorithm, using FORTRAN V on the UNIVAC 1110 of the Madison Academic Computing Center of the University of Wisconsin-Madison. In the program we set the stopping criteria as follows.

$\epsilon_1 = 10^{-4}$. If

$$\|x^{k+1} - x^k\| \leq \epsilon_1$$

or

$$\|x^{k+1} - x^k\| \leq \epsilon_1 \|x^k\|$$

then x^{k+1} is considered to be an optimal solution for the problem (1.1).

$\epsilon_2 = 10^{-5}$. This is the tolerance of the BFGS algorithm (problem 2.2). That means a solution y is accepted as optimal if a relative change of size ϵ_2 in the components of y does not reduce the objective value.

We have the same tolerances for Level 2.

For Level 3 we set $\epsilon_3 = 10^{-5}$.

We generated test problems by predetermining the size of the problem, the size of blocks, and the number of coupling constraints and variables; then generating randomly data of the problem. The matrices A_i,

B_i and D_i are 90% dense. Each entry of those matrices is a pseudo-random number in the range [-50, 50], obtained by the random number routines of the Madison Academic Computing Center [MACC, 1978]. Each positive semi-definite matrix C_i is of the form $E_i^T E_i$, where E_i is a randomly generated matrix. The cost coefficients c_i are pseudo-random numbers in the range [-10, 10]. To be sure that the problem is feasible we randomly generated a sequence of integers in between 0 and 5 (considered as a feasible point) and then multiplied them in a proper way with the matrices A_i,B_i and D_i to obtain the right-hand side coefficients. The structures of the test problems are given in Table 4.1 below.

Problem	Problem Size	Number Of		
		Blocks	Coupling Variables	Coupling Constraints
I	50 × 100	10	5	5
II	59 × 125	16	2	3
III	70 × 150	20	2	3

Table 4.1 Structures of Test Problems

We compared the computational results with the results obtained by using the LCPL package available at the Madison Academic Computing Center of the University of Wisconsin-Madison. LCPL is a program for solving linear complementarity problems and quadratic programming problems. It was developed by Tomlin [1976] at the Systems Optimization Laboratory of Stanford University. It uses a variant of Lemke's method such that only the nonzero coefficents are stored in packed form and the basis inverse is maintained in standard product form. Computational results are shown in Table 4.2.

Problem	LCPL	Our Algorithm		
		$\lambda = 10,\ \mu = 50$	$\lambda = 10,\ \mu = 100$	$\lambda = 10,\ \mu = 200$
I	10.	40.	18.	19.
II	14.	33.	36.	(1)
III	(2)	158.	34.	29.

Table 4.2 CPU Time (secs) of Test Problems

(1) Numerical errors.

(2) The run stopped after 255 iterations because of the insufficiency of space to store the inverse of the basis in product form (exceeded the limit of 8000 words of the eta file).

From Table 4.2, we see that the computation time of the decomposition procedure is not yet as good as that of LCPL. Several points are relevant to this comparison. First, our computer program was written for the purpose of solving a limited number of test problems and investigating some of the computational aspects of the method; consequently, it was not written with a great a programming efficiency as was LCPL. Secondly, the decomposition approaches are intended to handle large problems; however our test problems are by no means large.

In terms of memory storage requirements, our method is much better than LCPL. The requirements of our method are as follows (we do not count the storage for problem data since it is approximately the same for both methods):

Storage requirements (neglecting small items)

Level 1: We need to store x^0 , $\bar{x}$ and five other vectors of the same dimensions n, so the total number of words is 7n. The requirements for a BFGS routine with the dimension of the variables being ℓ are $3\ell + \ell(\ell + 13)/2$.

Level 2: Storage for the vectors $(v^k, w^k), (\bar{v}, \bar{w})$ and (v^{k+1}, w^{k+1}) is $3((\sum_{i=0}^{p} m_i) + n_0)$; working space needed is $2((\sum_{i=1}^{p} m_i) + n_0)$. There is another BFGS routine with the dimension of the variables being n_0. Its storage requirements are $3n_0 + n_0(n_0 + 13)/2$.

Level 3: For a quadratic programming problem with the dimension of the variables being N and with only nonnegativity constraints, the storage requirements of the Best-Ritter algorithm are $2N^2 + 10N$. To accomodate problems of sizes n_i, $i = 1, 2, \ldots, p$, N should be

$$N = \max \{n_i \mid i = i, 2, \ldots, p\} .$$

In order to use the optimal solutions of the previous iteration as the starting point for the current iterations we need to keep information about the previous iteration; that requires $n + \sum_{i=0}^{p} (n_i^2)$ words of memory. We also need storage for the matrices

(4.4). They are symmetric, so the requirements are

$$\sum_{i=0}^{p} \frac{n_i(n_i+1)}{2}$$

Storage requirements for our test problems are shown in Table 4.3:

Level	Problem I	Problem II	Problem III
1	700	875	1050
2	60	33	33
3	250	290	345
4	60	21	21
5	2055	2390	2684
Total	2495	3609	4133

Table 4.3 Storage Requirements (words) for Test Problems

From Table 4.3, we see that, for problem III, our method required only 4133 words of memory, while LCPL used up the limit of 8000 words for eta file alone and still has not reached the optimal solution of the problem (see Table 4.2). With larger problems, the decomposition procedure would lead to even greater savings in memory.

ACKNOWLEDGEMENTS

This research is a part of the author's doctoral thesis submitted to the Department of Industrial Engineering at the University of Wisconsin-Madison under the supervision of Professor Stephen M. Robinson. I would like to express my sincere gratitude to Professor Stephen M. Robinson for his advice and support.

References

Best, M. J. and Ritter, K. (1976), "An effective algorithm for quadratic minimization problems", Math. Research Center, University Wisconsin-Madison,Tech. Rep. #1691.

Brézis , H. and Lions, P. L. (1978), "Produits infinit de Résolvantes," Israel J. Math., Vol. 12, pp. 329-345.

Ha, C. D. (1981), "A decomposition method and its application to block angular linear programs," Math. Research Center, Univ. Wisconsin-Madison, Tech. Rep. #2174.

Kaneko, I. and Ha, C. D. (1980), "A decomposition procedure for large scale optimal plastic design problem," Math. Research Center, Univ. Wisconsin-Madison, Tech. Rep. #2075.

Lasdon, L. S. (1970), Optimization Theory for Large Systems, Macmillan.

Madison Academic Computing Center (MACC, 1976), Nonlinear programming routines, reference manual.

Madison Academic Computing Center (MACC, 1978), Random number routines, reference manual.

Mangasarian, O. L. (1969), Nonlinear Programming, McGraw-Hill.

Rockafellar, R. T. (1968), "Convex function, monotone operators and variational inequalities," in Theory and Applications of Manotone Operators, Ghizetti, (ed.).

Rockafellar, R. T. (1976b), "Augmented lagrangians and applications of the proximal point algorithm in convex programming," Math. O.R., Vol. 1, No. 2.

Shapiro, J. F. (1979), Mathematical Programming, Wiley-Interscience.

Tomlin, J. A. (1976), "Robust implementation of Lemke's method for the linear complementarity problem," Tech. Rep. SOL 76-24, Systems Optim. Lab., Stanford University.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2276	2. GOVT ACCESSION NO. AD-A110 463	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN ALGORITHM FOR STRUCTURED, LARGE-SCALE QUADRATIC PROGRAMMING PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Cu Duong Ha		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 5 - Operations Research
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE September 1981
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quadratic programs, Decomposition method, Large-scale problems, Computational experience

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An algorithm for structured, large-scale, convex quadratic programming problems is described. The structure of the constraint matrix is block diagonal with a small number of coupling constraints and variables. The algorithm utilizes twice a decomposition procedure that was developed earlier. The first time the decomposition procedure is used to break up the coupling constraints, and the second time it is used to break up the coupling variables. Preliminary computational results are also reported.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DAT
ILM