BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 17, Number 2, June 1989

A PARTIALLY SIMULTANEOUS EXTENSION OF HILDRETH'S ITERATIVE ROW-ACTION METHOD FOR QUADRATIC PROGRAMMING

BY

HON-JANG YANG (楊宏章)

Abstract. An *extended partially simultaneous* version of Hildreth's iterative algorithm for norm minimization over linear inequalities is presented. Proofs are given showing that the algorithm converges to the solution in the feasible case.

1. Introduction. Linearly constrained quadratic optimization problems derived from computing the least element of polyhedra appear in various fields of application and it is not rare to encounter large-scale or even huge-scale problems. Usually the matrix describing the constraints will be sparse, but all too often no special structure pattern is detectable in it. In such cases *row-action* methods are frequently used. A row-action method is an iterative procedure which requires in each step only the current point and one row of the matrix, and performs no transformation on the matrix elements, see Censor [2]. Image reconstruction from projections is an important application of these type of methods, see, e. g., Herman and Lent [4], Herman, Lent and Lutz [6]. Hildreth's quadratic programming procedure [7] is a row-action method whose capabilities for solving large scale problems were numerically demostrated by Herman and Lent [5].

Iusem and De Pierro [9] mentioned that Kaczmarz [10] and Ciminno [3] proposed methods for finding a feasible solution for a system of linear equations. In such a case, they can be described as follows: Kaczmarz's approach, as presented by Agmon [1] for

Received by the editors March 26, 1988.
AMS Subject Classification: 52A20, 65F10, 90C20.
Key words: Quadratic programming, least square, projection method.

the inequality case, starts with an arbitrary point and generates the iterates by using the inequalities in a cyclic way and taking as next iterate the orthogonal projection of the current iterate onto the half space defined by the current inequality (projecting a point onto a halfspace means projecting it onto the associated hyperplane if the point lies outside the half space and leaving it unmodified otherwise). In Cimmino's method, on the other hand, the current iterate is projected onto all the half spaces and the new iterate is a convex combination of such projections.

Hildreth's method can be seen as a modification of Kaczmarz-Agmon's procedure for the case when the required solution is not just a feasible point but a norm minimizer one [9]. The minimum norm solution may be justified by the minimum variance property in the application to image reconstruction [6]. Lent and Censor [12] cast Hildreth's algorithm into a compact form and introduce in it a sequence $\{r^{(k)}\}$ of *relaxation parameters.* The option of using relaxation parameters has been shown to be a very important tool in practical implementations of other row-action methods (see [4] and [6]). Another feature they introduced into Hildreth's algorithm is the *almost cyclic control,* meaning that the rows of the matrix may be taken up, as the iterations proceed, in a manner which is less restrictive than cyclic control. The introduction of almost cyclic control lays the foundations for the method of quadratic optimization over pairs of inequalities (i.e., interval constraints) developed in Herman and Lent [5].

Iusem and De Pierro [9] present a Cimmino-like *simultaneous* version of Hildreth's iterative algorithm casted by Lent and Censor [12]. Their implementation has two features which make it different from Lent and Censor's version. First, it is appropriate for parallel processing computers, since all the projections of the current iterate can be calculated simultaneously (that is why they call their method "simultaneous"). Second, both in the infeasible case and feasible case, their primal sequence converges to a norm minimizing least square solution.

In this paper we propose an algorithm (described in the next section) which include Lent-Censor's Kaczmarz-like algorithm and Iusem-De Pierro's Cimmino-like algorithm as special cases, and prove convergence in the feasible case.

2. A partially simultaneous iterative row-action (PSIRA) algorithm. Hildreth's quadratic programming procedure [7] is an iterative method for finding an approximating solution of the problem:

$$\min \frac{1}{2}\langle By, y\rangle + \langle y, d\rangle \qquad (2.1)$$
$$\text{such that } Gy \leq h,$$

where B is a positive-definite $n \times n$ matrix, G an $m \times n$ matrix, $y \in R^n$, $d \in R^n$ and $h \in R^m$; $\langle \cdot, \cdot \rangle$ stands for the inner product in the n-dimensional Euclidean space R^n.

Using the Choleski decomposition with $B = D^t D$ and letting $y = D^{-1}x - B^{-1}d$, the problem (2.1) is transformed into the following:

Standard problem

$$\min \frac{1}{2}\|x\|^2 \qquad (2.2)$$
$$\text{such tha } Ax \leq b,$$

where $A = GD^{-1}$ is an $m \times n$ matrix and $b = h + GB^{-1}d$; $\|\cdot\|$ stands for the Euclidean norm in R^n.

The application of PSIRA algorithm to a given quadratic optimization problem is governed by a sequence of subsets which specifies the rows of the matrix A to be taken up at a given iterative step. This sequence will be called the *control* of the algoritm.

Let $\Pi \equiv \{I_1, I_2, \cdots, I_s\}$ be a class of nonempty subsets of $I \equiv \{1, 2, \cdots, m\}$ and $\bigcup_{i=1}^{s} I_i = I$. The following definition was modified from that in Lent [11].

DEFINITION 1. A sequence $\{J_k\}_{k=0}^{\infty}$ is *almost cyclic on* Π if

(a) $J_k \in \Pi$ for all $k \geq 0$, and

(b) there exists an integer C, called an *almost cyclicality constant*, such that for all $k \geq 0$, $\Pi \subset \{J_{k+1}, J_{k+2}, \cdots, J_{k+C}\}$.

We now give a PSIRA algorithm for solving the standard problem (2.2).

We will assume throughout that $a_i \neq 0$ for $i \in I \equiv \{1, 2, \cdots, m\}$, where $a_i \in R^n$ forms the i-th row of the matrix A. Denoting $S \equiv \{x | Ax \leq b\}$, the feasible set of (2.2). It will be assumed throughout that $S \neq \emptyset$.

PSIRA Algorithm.

Initialization: $z^{(0)} \in R_+^m$ arbitrary (R_+^m stands for the nonnegative orthant of R^m) and $x^{(0)} \equiv -A^t z^{(0)}$.

Typical step: $x^{(k+1)} = x^{(k)} + \sum_{i \in I} c_i^{(k)} \lambda_i^{(k)} a_i,$

$$z^{(k+1)} = z^{(k)} - \sum_{i \in I} c_i^{(k)} \lambda_i^{(k)} e_i$$

with

$$c_i^{(k)} = \begin{cases} \min\left\{z_i^{(k)}/\lambda_i^{(k)},\ r^{(k)} \dfrac{b_i - \langle a_i, x^{(k)} \rangle}{\|a_i\|^2}\right\} & \text{if } i \in J_k, \\ 0 & \text{if } i \notin J_k. \end{cases}$$

where e_i is the m-vector with 1 in the i-th place and zeros elsewhere; $\{r^{(k)}\}$ is the sequence of *relaxation parameters*, $r^{(k)} > 0$ for all k; $\{\lambda_i^{(k)}\}$ are *convex combination parameters*, $\lambda_i^{(k)} > 0$ for all $i \in J_k$, $\sum_{i \in J_k} \lambda_i^{(k)} = 1$ and $\lambda_i^{(k)} = 0$ for all $i \notin J_k$.

REMARK 1. In fact, in the typical step,

$$x^{(k+1)} = x^{(k)} + \sum_{i \in J_k} c_i^{(k)} \lambda_i^{(k)} a_i,$$

and

$$z^{(k+1)} = z^{(k)} - \sum_{i \in J_k} c_i^{(k)} \lambda_i^{(k)} e_i.$$

REMARK 2. For the case $s = m$, $I_j = \{j\}$ for all $j = 1, 2, \cdots m$, it reduces to the algorithm proposed by Lent and Censor [12]. On the other hand, when $s = 1$ it reduces to the algorithm proposed by Iusem and De Pierro [9].

LEMMA 1. $x^{(k)} = -A^t z^{(k)}$ *for all* k.

Proof. True for $k = 0$ by initialization. By induction, assume true for k.

$$
\begin{aligned}
x^{(k+1)} &= x^{(k)} + \sum_{i\in I} c_i^{(k)} \lambda_i^{(k)} a_i \\
&= -A^t z^{(k)} + A^t \Big(\sum_{i\in I} c_i^{(k)} \lambda_i^{(k)} e_i\Big) \\
&= -A^t z^{(k+1)}.
\end{aligned}
$$

3. **Convergence theorem for PSIRA algorithm.** The following theorem for PSIRA algorithm will be proved.

THEOREM 1. *Assumptions*: (1) $S \equiv \{x \mid Ax \le b\} \ne \emptyset$, $a_i \ne 0$ *for all* $i \in I \equiv \{1, 2, \cdots, m\}$,

(2) $\{J_k\}_{k=0}^{\infty}$ *is an almost cyclic sequence on* Π,

(3) *there exist* ε_1, $\varepsilon_2 > 0$ *such that* $\varepsilon_1 \le r^{(k)} \le 2 - \varepsilon_2$ *for all* k, *and*

(4) *there exists* $\varepsilon_3 > 0$ *such that* $\lambda_i^{(k)} > \varepsilon_3$ *for all* $i \in J_k$ *and for all* k *and* $\sum_{i\in J_k} \lambda_i^{(k)} = 1$ *for all* k.

Conclusion: *The sequence* $\{x_k\}$ *produced by PSIRA algorithm converges to the solution of* (2.2).

In the sequel, we modify the steps in Lent and Censor [12] to develop the necessary theory to prove this theorem.

The dual problem to the standard problem (2.2) is

$$
\begin{aligned}
&\max \Psi(z) \equiv -1/2 \, \| A^t z\|^2 - \langle b, z\rangle \\
&\text{such that } z \in R_+^m.
\end{aligned} \tag{3.1}
$$

The standard problem (2.2) and its dual (3.1) are related by the duality theorem [13]

$$
\min_{x\in S} \frac{1}{2} \|x\|^2 = \max_{z\in R_+^m} \Psi(z). \tag{3.2}
$$

LEMMA 2. *Under the assumptions of Theorem* 1, *for the sequences* $\{z^{(k)}\}$, $\{x^{(k)}\}$, $\{c^{(k)}\}$ *produced by the algorithm, we have*

(a) $z^{(k)} \in R_+^m$ *for all* k,

(b) $\Psi(z^{(k+1)}) \ge \Psi(z^{(k)})$ *for all* k,

(c) $\lim_{k\to\infty} [\Psi(z^{(k+1)}) - \Psi(z^{(k)})] = 0$,

(d) $\lim_{k\to\infty} c^{(k)} = 0$, *and*

(e) $\lim_{k\to\infty} [x^{(k+1)} - x^{(k)}] = 0$, $\lim_{k\to\infty} [z^{(k+1)} - z^{(k)}] = 0$.

Proof.

(a) $z^{(0)} \in R_+^m$ by initialization. By induction, assume $z^{(k)} \in R_+^m$. If $i \notin J_k$, $z_i^{(k+1)} = z_i^{(k)} \geq 0$; if $i \in J_k$, $z_i^{(k+1)} = z_i^{(k)} - \lambda_i^{(k)} c_i^{(k)} \geq z_i^{(k)} - z_i^{(k)} = 0$, hence $z^{(k+1)} \in R_+^m$.

(b)
$$\Psi(z^{(k+1)}) - \Psi(z^{(k)}) = -\frac{1}{2}\Big[\Big\|\sum_{i\in I} \lambda_i^{(k)} c_i^{(k)} a_i\Big\|^2 + 2\sum_{i\in I} \lambda_i^{(k)} c_i^{(k)} \langle a_i, x^{(k)}\rangle\Big] + \sum_{i\in I} b_i c_i^{(k)} \lambda_i^{(k)}$$
$$= \sum_{i\in I} \lambda_i^{(k)} c_i^{(k)} (b_i - \langle a_i, x^{(k)}\rangle) - \frac{1}{2}\Big\|\sum_{i\in I} \lambda_i^{(k)} c_i^{(k)} a_i\Big\|^2.$$

Due to convexity of $\|\cdot\|^2$, we have

$$\Big\|\sum_{i\in I} \lambda_i^{(k)} c_i^{(k)} a_i\Big\|^2 \leq \sum_{i\in I} \lambda_i^{(k)}\| c_i^{(k)} a_i\|^2 = \sum_{i\in I} \lambda_i^{(k)}[c_i^{(k)}]^2\|a_i\|^2.$$

Therefore,

$$\Psi(z^{(k+1)}) - \Psi(z^{(k)}) \geq \sum_{i\in I} \lambda_i^{(k)} c_i^{(k)}\|a_i\|^2\Big[\frac{b_i - \langle a_i, x^{(k)}\rangle}{\|a_i\|^2} - \frac{c_i^{(k)}}{2}\Big]$$
$$= \sum_{i\in J_k} \lambda_i^{(k)} c_i^{(k)}\|a_i\|^2\Big[\frac{b_i - \langle a_i, x^{(k)}\rangle}{\|a_i\|^2} - \frac{c_i^{(k)}}{2}\Big].$$

It is easy to check that

$$c_i^{(k)} \frac{b_i - \langle a_i, x^{(k)}\rangle}{\|a_i\|^2} \geq \frac{[c_i^{(k)}]^2}{\gamma^{(k)}} \qquad \text{for all } i \in J_k.$$

The inequality is trivial if $c_i^{(k)} \geq 0$. Otherwise it holds with equality as a consequence of Lemma 2(a). Hence,

$$\Psi(z^{(k+1)}) - \Psi(z^{(k)}) \geq \sum_{i\in J_k} \lambda_i^{(k)}[c_i^{(k)}]^2\|a_i\|^2\Big[\frac{1}{\gamma^{(k)}} - \frac{1}{2}\Big] \geq 0. \tag{3.3}$$

(c) From (b), $\{\Psi(z^{(k)})\}$ is monotonically increasing, and (3.2) shows that it is bounded from above, so the conclusion follows.

(d) Let $\|a_i\|^2 \geq a > 0$ for all i. Then, from (3.3) it follows that

$$\Psi(z^{(k+1)}) - \Psi(z^{(k)}) \geq \sum_{i\in J_k} \varepsilon_3[c_i^{(k)}]^2 a\Big[\frac{\varepsilon_2}{2(2-\varepsilon_2)}\Big] \geq \frac{\varepsilon_3\, \varepsilon_2\, a}{4}\, \|c^{(k)}\|^2,$$

hence (c) implies (d).

(e) This is immediate from (d) and the typical step of the algorithm.

Following [12], we introduce a sequence of slack vectors $\{q^{(k)}\}$ which serve to construct a sequence of perturbed contraint sets $S^{(k)}$.

DEFINITION 2. $q^{(0)} = 0$, and

$$q_i^{(k+1)} \equiv \begin{cases} q_i^{(k)} & \text{if } i \notin J_k, \\ -\dfrac{c_i^{(k)}\|a_i\|^2}{\gamma^{(k)}} + b_i - \langle a_i, x^{(k)} \rangle & \text{if } i \in J_k. \end{cases}$$

Next, define the vectors $b^{(k)}$ by

$$b^{(k)} \equiv q^{(k)} + Ax^{(k)},$$

and use them to construct a sequence of pertubed constraint sets $S^{(k)}$ by

$$S^{(k)} \equiv \{x \,|\, Ax \leq b^{(k)}\}.$$

The next lemma is a straight consequence of the definitions of $z^{(k)}$ and $q^{(k)}$.

LEMMA 3. *$q_i^{(k)} z_i^{(k)} = 0$ for all k, i.*

Proof. It is true for $k = 0$. By induction, assume true for k. If $i \notin J_k$, then $q_i^{(k+1)} z_i^{(k+1)} = q_i^{(k)} z_i^{(k)} = 0$. If $i \in J_k$ and $q_i^{(k+1)} \neq 0$, then $c_i^{(k)} = z_i^{(k)}/\lambda_i^{(k)}$ by the definition of $c_i^{(k)}$, therefore $z_i^{(k+1)} = 0$ and $q_i^{(k+1)} z_i^{(k+1)} = 0$. Then we have the following intermediate optimality property.

THEOREM 2. *$x^{(k)}$ produced by the algorithm minimizes $1/2\,\|x\|^2$ over the pertubed constraint set $S^{(k)}$.*

Proof. Since the minimand is convex, the Kuhn-Tucher conditions are sufficient for optimality. They are

(a) $x = -A^t \mu,$

(b) $\mu \geq 0,$

(c) $Ax \leq b^{(k)},$

(d) $\mu_i(\langle a_i, x \rangle - b_i^{(k)}) = 0$ for all i,

Take $x = x^{(k)}$, $\mu = z^{(k)}$.

(a) follows from Lemma 1.

(b) follows from Lemma 2(a).

(c) $q_i^{(k)} \geq 0$ follows from the definition of $q^{(k)}$ and the typical step of the algorithm, hence we have (c)

(d) Since $\langle a_i, x^{(k)}\rangle - b_i^{(k)} = -q_i^{(k)}$, we have (d) by Lemma 3.

LEMMA 4. *Under the assumptions of Theorem* 1, $b^{(k)} \to b$ *as* $k \to \infty$.

Proof. We shall prove convergence componentwise. Let $t \in I$ be fixed, then $t \in I_\alpha$ for some $\alpha \in \{1, 2, \cdots, s\}$. Denote by $u \equiv u(t)$ the index of the last iteration before k with the property $t \in J_u$. Because the almost cyclicality of $\{J_k\}_{k=0}^{\infty}$ we know that $k - C \leq u$, where C is the almost cyclicality constant. (We have assumed implicitly that $k > C$.).

From the definition of $q^{(k)}$ and $b^{(k)}$, we get

$$\begin{aligned} q_t^{(k)} &= q_t^{(k-1)} = \cdots = q_t^{(u+1)}, \\ b_t^{(k)} &= q_t^{(k)} + \langle a_t, x^{(k)}\rangle = q_t^{(u+1)} + \langle a_t, x^{(k)}\rangle \\ &= -\frac{c_t^{(u)}\|a_t\|^2}{\gamma^{(u)}} + b_t + \langle a_t, x^{(k)} - x^{(u)}\rangle . \end{aligned}$$

The first summand on the right hand side tends to zero as $k \to \infty$ because u also tends to infinity and Lemma 2(d) applies. The inner product can be expanded as

$$\langle a_t, x^{(k)} - x^{(u)}\rangle = \sum_{j=u}^{k-1} \langle a_t, x^{(j+1)} - x^{(j)}\rangle$$

with at most C terms in the sum. Therefore because of the convergence of the difference $x^{(j+1)} - x^{(j)}$ to zero (Lemma 2(e)), this inner product tends to 0 with $k \to \infty$, hence we have $b_t^{(k)} \to b_t$ as $k \to \infty$.

We quote a Hoffman's [8] theorem.

THEOREM 3. *Let* $T = \{y \mid Ay \leq d\}$ *and* $T' = \{y' \mid Ay' \leq d'\}$ *and assume* T' *is nonempty, Then, there exists a constant* α, *which depends only on* A, *such that for any* $y \in T$ *there exists a* $y' \in T'$ *with the property*

$$\|y - y'\| \leq \alpha\|(d - d')^+\|,$$

where the upper plus notation means, for any vector v, $(v^+)_i \equiv \max(0, v_i)$.

Now we adopt some additional notations: x^* will denote the point in S with minimum norm, $\hat{x}^{(k)}$ will denote the point in S closest to $x^{(k)}$, and $\hat{x}_*^{(k)}$ will denote the point in $S^{(k)}$ closest to x^*.

LEMMA 5. *Uuder the assumption of Theorem* 1 *the following hold*:

(a) $\|x^{(k)}\| \to \|x^*\|$ *as* $k \to \infty$,

(b) $\|\hat{x}^{(k)} - x^{(k)}\| \to 0$ *as* $k \to \infty$,

(c) $\|\hat{x}^{(k)}\| \to \|x^*\|$ *as* $k \to \infty$.

Proof. (a) Write

(3.4) $$\|x^*\| \le \|\hat{x}^{(k)}\| \le \|\hat{x}^{(k)} - x^{(k)}\| + \|x^{(k)}\|$$

and

(3.5) $$\|x^{(k)}\| \le \|\hat{x}_*^{(k)}\| \le \|\hat{x}_*^{(k)} - x^*\| + \|x^*\| .$$

Applying Theorem 3 to the first term on the right-hand side of inequalities (3.4) and (3.5), respectively. Then (a) follows from Lemma 4.

(b) Follows from Theorem 3 and Lemma 4.

(c) Follows from (a) and (b).

Proof of Theorem 1. We prove here that $\lim_{k\to\infty} x^{(k)} = x^*$. Write

$$\|x^* - x^{(k)}\| \le \|x^* - \hat{x}^{(k)}\| + \|x^{(k)} - \hat{x}^{(k)}\| .$$

The second term of the right-hand side is taken care of by Lemma 5(b). Concerning the first summand, Lemma 5(c) ensures that, for the sequence $\{\hat{x}^{(k)}\} \subset S$,

$$\|\hat{x}^{(k)}\| \to \|x^*\| \quad \text{as} \quad k \to \infty .$$

Since S is a closed convex set and x^* is the only point in S which has minimum norm, it follows readily that $\hat{x}^{(k)} \to x^*$ as $k \to \infty$, thereby completing the proof.

REFERENCES

1. S. Agmon, *The relaxation method for linear inequalities*, Canad. J. Math., 6 (1954), 382–392.

2. Y. Censor, *Row-action methods for huge and sparse systems and their applications*, SIAM Rev., 23 (1981), 444–466.

3. G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari,* La Ricerca Scientifica, Vol. XVI, Ser III, Anno IX (1938), 326–333.

4. G. T. Herman and A. Lent, *Iterative reconstruction algorithms,* Computers in Biology and Medicine, **6** (1976), 273–294.

5. ———, *A family of iterative quadratic optimization algorithms for pairs of inequalities with application in diagnostic radiology,* Math. Programming Stud., **9** (1978), 15–29.

6. G. T. Herman, A. Lent and P. H. Lutz, *Relaxation methods for image reconstruction,* Comm. ACM, **21** (1978), 152–158.

7. C. Hildreth, *A quadratic programming procedure,* Naval Res. Logist. Quart., **4** (1957), 79–85.

8. A. J. Hoffman, *On approximate solutions of systems of linear inequalities,* J. Res. Nat. Bur. Standards, **49** (1952), 263–265.

9. A. N. Iusem and A. R. De Pierro, *A simultaneous iterative method for computing projections on polyhedra,* SIAM J. Control, **25** (1987), 231–243.

10. S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen,* Bull. Acad. Polon. Sci. Lett., A **35** (1937), 355–357.

11. A. Lent, *Maximum entropy and multiplicative ART,* Image Analysis and evaluation, SPSE Conference Proceedings, R. Show, ed., July 1976, Toronto, Canada, pp. 249–257.

12. A. Lent and Y. Censor, *Extensions of Hildreth's row-action method for quadratic programming,* SIAM J. Control, **18** (1980), 444–454.

13. D. G. Luenberger, *Introduction to Linear and Nonlinear Programming,* Addison-Wesley, Reading, MA, 1973.

Department of Mathematics
National Taiwan University
Taipei, Taiwan 10764
Republic of China