JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 72, No. 1, JANUARY 1992

On the Convergence of the Coordinate Descent Method for Convex Differentiable Minimization[1]

Z. Q. LUO[2] AND P. TSENG[3]

Communicated by O. L. Mangasarian

Abstract. The coordinate descent method enjoys a long history in convex differentiable minimization. Surprisingly, very little is known about the convergence of the iterates generated by this method. Convergence typically requires restrictive assumptions such as that the cost function has bounded level sets and is in some sense strictly convex. In a recent work, Luo and Tseng showed that the iterates are convergent for the symmetric monotone linear complementarity problem, for which the cost function is convex quadratic, but not necessarily strictly convex, and does not necessarily have bounded level sets. In this paper, we extend these results to problems for which the cost function is the composition of an affine mapping with a strictly convex function which is twice differentiable in its effective domain. In addition, we show that the convergence is at least linear. As a consequence of this result, we obtain, for the first time, that the dual iterates generated by a number of existing methods for matrix balancing and entropy optimization are linearly convergent.

Key Words. Coordinate descent, convex differentiable optimization, symmetric linear complementarity problems.

1. Introduction

A very important problem in optimization is that of minimizing a convex function of the Legendre type (i.e., a function that is strictly convex

---

[1]This work was partially supported by the U.S. Army Research Office, Contract No. DAAL03-86-K-0171, by the National Science Foundation, Grant No. ECS-8519058, and by the Science and Engineering Research Board of McMaster University.

[2]Assistant Professor, Communications Research Laboratory, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.

[3]Research Associate, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts. Presently, Assistant Professor, Department of Mathematics, University of Washington, Seattle, Washington.

differentiable on an open convex set and whose gradient tends to infinity in norm at the boundary points), subject to linear constraints. As an example, when the cost function is quadratic, this problem has applications in linear programming (Refs. 1, 2), image reconstruction (Ref. 3), and the solution of boundary-value problems (Refs. 4–6). When the cost function is the $x \log(x)$ entropy function, this problem has applications in information theory (Refs. 7, 8), matrix balancing (Refs. 9, 10), image reconstruction (Refs. 11–13), speech processing (Refs. 14–16), and statistical inference (Ref. 17). As a final example, when the cost function is the $-\log(x)$ entropy function, this problem reduces to an analytic centering problem, which plays a key role in many new algorithms for linear programming (Refs. 18–21).

A popular approach to solving the above problem is to dualize the linear constraints to obtain a dual problem of the form

$$\text{minimize} \quad g(Ex) + \langle b, x \rangle, \tag{1a}$$

$$\text{subject to} \quad x \geq 0, \tag{1b}$$

where g is a strictly convex essentially smooth function, E is a matrix, and b is a vector (see Section 5), and then use a coordinate descent method to solve this problem, whereby at each iteration one of the coordinates of x is adjusted in order to minimize the cost function, while the other coordinates are held fixed. Such a method is simple, uses little storage, and in certain cases is highly parallelizable. Methods that follow this approach include a method of Hildreth for quadratic programming (Ref. 22, also see Refs. 1, 3, 4, 23, 24), a method of Kruithof for matrix balancing (Ref. 9, also see Refs. 10, 25–28 and references cited in Ref. 10), as well as a number of related methods for entropy optimization (Refs. 12, 29, and p. 236 in Ref. 14).

An outstanding question concerns the convergence of the iterates generated by the above coordinate descent scheme. Typically, convergence requires the cost function to have bounded level sets and to be strictly convex in some sense (see for example Refs. 30–38), neither of which, unfortunately, holds for the cost function (1a) (e.g., when E has redundant rows). For (1a), it was known, under mild restrictions on the order of coordinate relaxation, that the gradient of the cost function, evaluated at the iterates, converge (Refs. 39–41, also see Refs. 22, 29, 42, 43), but it was not known if the iterates themselves converge or if they are even bounded. The only nontrivial special cases for which the iterates are known to converge, without assuming uniqueness of the optimal solution, are (i) when g is separable and E is the node-arc incidence matrix for a digraph (Ref. 44), and (ii) when g is a strictly convex quadratic function (Ref. 45).

In this paper, we give the first result on the convergence of the iterates generated by the above coordinate descent scheme, applied to solving (1).

In particular, we show that the iterates converge to an optimal solution of (1), provided that g has a positive-definite Hessian and tends to infinity at the boundary of its effective domain, and that the coordinates are iterated in an almost cyclic or a Gauss–Southwell manner. In addition, we show that the rate of convergence is at least linear. These results are rather remarkable, since the optimal solution set may be unbounded and the function g may have a very complicated form. As a corollary, we establish, for the first time, the linear convergence of the dual iterates generated by many known methods (see Section 5). Our results are based on an interesting new fact that the distance from an iterate to the solution set is upper bounded by a constant times the changes in the iterate during the current cycle of coordinate iterations (see Lemma 4.5).

In our notation, all vectors are column vectors, $\mathfrak{R}^k$ denotes the k-dimensional Euclidean space; $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product; and superscript T denotes transpose. For any $x \in \mathfrak{R}^k$, we denote by x_i the ith coordinate of x and, for any $I \subseteq \{1, \ldots, k\}$ (possibly empty), by x_I the vector with components x_i, $i \in I$ (with the x_i's arranged in the same order as in x). We also denote by $\|x\|$ the Euclidean norm of x, i.e.,

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Analogously, for any $k \times m$ matrix A, we denote by $\|A\|$ the matrix norm of A induced by the vector norm $\|\cdot\|$, i.e.,

$$\|A\| = \max_{\|x\|=1} \|Ax\|;$$

by A_i the ith column of A; and, for any $I \subseteq \{1, \ldots, k\}$, by A_I the submatrix of A obtained by removing all columns $i \notin I$ of A. For any proper closed convex function h in R^k, we denote by $\operatorname{dom} h$ the effective domain of h, i.e.,

$$\operatorname{dom} h = \{x \in \mathfrak{R}^k \mid h(x) < \infty\}.$$

If h is differentiable at a point x, we denote by $\nabla h(x)$ the gradient of h at x; by $\nabla_i h(x)$ the ith coordinate of $\nabla h(x)$; and, for any $I \subseteq \{1, \ldots, k\}$, by $\nabla_I h(x)$ the vector with components $\nabla_i h(x)$, $i \in I$.

2. Algorithm Description and Main Result

Consider the following problem [compare with (1)]:

$$\text{minimize} \quad f(x), \tag{2a}$$

$$\text{subject to} \quad x \in \mathscr{X}, \tag{2b}$$

where $\mathscr{X}$ is a box (possibly unbounded) in $\Re^n$ and f is a proper closed convex function in $\Re^n$ of the form

$$f(x) = g(Ex) + \langle b, x \rangle, \tag{3}$$

with g a proper closed convex function in $\Re^m$, E an $m \times n$ matrix having no zero column, and b a vector in $\Re^n$.

We make the following standing assumptions about g and (2).

Assumption A1.

(a) The set of optimal solutions for (2), denoted by $\mathscr{X}^*$, is nonempty.
(b) dom g is open and g is strictly convex twice continuously differentiable on dom g.
(c) $\nabla^2 g(Ex^*)$ is positive definite for all $x^* \in \mathscr{X}^*$.

Part (a) of Assumption A1 is standard. Part (b) of Assumption A1 implies that (dom g, g) is, in the terminology of Rockafellar (Ref. 46), a convex function of the Legendre type. Assumption A1(b) in addition assumes that g is twice differentiable on dom g and that, since g is closed and dom g is assumed open, g tends to infinity at the boundary of dom g. Such a function has a number of nice properties: for example, its conjugate function is also a convex function of the Legendre type. Part (c) of Assumption A1 states that g has a positive curvature on the image of $\mathscr{X}^*$, under the linear transformation $x \mapsto Ex$. This condition is guaranteed to hold if g has a positive curvature everywhere on its effective domain. There are many important functions that satisfy this latter condition [in addition to Assumption A1(b)], the most notable of which are the quadratic function, the exponential function, and the negative of the logarithm function. We will discuss these example functions in detail in Section 5. Notice that, if g is strongly convex and twice differentiable everywhere, then g automatically satisfies both parts (b) and (c) of Assumption A1.

Notice that, since g is differentiable on dom g, then so is f on dom f. Then, by (3) and the chain rule for differentiation, we have

$$\nabla f(x) = E^T \nabla g(Ex) + b. \tag{4}$$

From the Kuhn–Tucker conditions for the convex program (2), it is easily seen that an x belongs to $\mathscr{X}^*$ if and only if $x \in \text{dom} f$ and the orthogonal projection of $x - \nabla f(x)$ onto the feasible set $\mathscr{X}$ is x itself, that is,

$$\mathscr{X}^* = \{x \in \text{dom} f \mid x = [x - \nabla f(x)]^+\}, \tag{5}$$

where $[x]^+$ denotes the orthogonal projection of x onto the feasible set $\mathscr{X}$. Since $\mathscr{X}$ is a box, it can be written explicitly as

$$\mathscr{X} = \{x \in \Re^n \mid l \leq x \leq c\}, \tag{6}$$

for some vectors $l = (l_1, \ldots, l_n) \in [-\infty, \infty)^n$ and $c = (c_1, \ldots, c_n) \in (-\infty, \infty]^n$. Then, $[x]^+$ is simply the n-vector whose ith coordinate is $[x_i]_i^+$, where we let

$$[x_i]_i^+ = \max\{l_i, \min\{c_i, x_i\}\}. \tag{7}$$

Consider the following well-known iterative method for solving (2), whereby at each iteration one of the coordinates of the current iterate is adjusted in order to minimize f over $\mathscr{X}$ along this coordinate direction (see Refs. 34, 47). This method, which we call the coordinate descent method (also known as the coordinate relaxation method or the Gauss–Seidel method, among others), is formally described below.

Coordinate Descent Method.

Iteration 0. Choose arbitrarily an $x^0 \in \operatorname{dom} f \cap \mathscr{X}$.

Iteration r+1. Given an $x^r \in \operatorname{dom} f \cap \mathscr{X}$, choose an $i \in \{1, 2, \ldots, n\}$, and compute a new iterate $x^{r+1} \in \operatorname{dom} f \cap \mathscr{X}$ satisfying the following system of nonlinear equations:

$$x_i^{r+1} = [x_i^{r+1} - \nabla_i f(x^{r+1})]_i^+, \tag{8}$$

$$x_j^{r+1} = x_j^r, \qquad \forall j \neq i. \tag{9}$$

We claim that the above method is well defined; that is, for every r and any $i \in \{1, \ldots, n\}$, an $x^{r+1} \in \operatorname{dom} f \cap \mathscr{X}$ satisfying (8)–(9) exists. To see this, notice from (7) and (9) that (8) is equivalent to

$$x_i^{r+1} = \arg \min_{l_i \leq x_i \leq c_i} f(x_1^r, \ldots, x_{i-1}^r, x_i, x_{i+1}^r, \ldots, x_n^r). \tag{10}$$

Hence, if the method is not well defined, then there would exist some r and some $i \in \{1, \ldots, n\}$ for which the minimization in (10) is not attained. Let e^i denote the ith coordinate vector in $\Re^n$. Then, either (i) $l_i = -\infty$ and $f(x^r - \lambda e^i)$ is monotonically decreasing with increasing λ or (ii) $c_i = \infty$ and $f(x^r + \lambda e^i)$ is monotonically decreasing with increasing λ. Suppose that Case (i) holds. Case (ii) may be treated analogously. Then, since the set $\{Ex \mid l \leq x \leq c, f(x) \leq f(x^r)\}$ is bounded by Lemma 3.3 given in Section 3, there would hold $Ee^i = 0$, a contradiction of the assumption that E contains no zero column.

It is readily seen from (9) and (10) that the f-value of the iterates generated by the coordinate descent method are monotonically decreasing.

However, these values need not converge to the optimal value of (2). To guarantee that they do converge to this latter value, we will assume that the coordinates are iterated upon according to either one of the following two rules.

Almost Cyclic Rule. There exists an integer $B \geq n$ such that every coordinate is iterated upon at least once every B successive iterations.

Gauss–Southwell Rule. The index i of the coordinate chosen for iteration at the $(r+1)$th iteration satisfies

$$|x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+| \geq \beta \max_j |x_j^r - [x_j^r - \nabla_j f(x^r)]_j^+|, \tag{11}$$

where β is a fixed constant in the interval $(0, 1]$.

The above almost cyclic order of iteration is a direct extension of the classical cyclic order of iteration (see Refs. 25, 37, 48 and p. 158 of Ref. 34). The Gauss–Southwell rule is discussed, for example, in Ref. 39 and p. 158 of Ref. 34. General discussions of coordinate descent methods can be found in Refs. 30, 31, 34, and 47. Alternative orders of iteration are discussed in Refs. 41, 44, and 48.

The main result of this paper is the following theorem.

Theorem 2.1. Let $\{x^r\}$ be a sequence of iterates generated by the coordinate descent method (8)–(9), using either the almost cyclic or the Gauss–Southwell rule. Then, $\{x^r\}$ converges at least linearly to an element of $\mathscr{X}^*$.

The proof of Theorem 2.1 is quite intricate and will be given through a sequence of lemmas, which we present in the following two sections.

The remainder of this paper proceeds as follows. In Section 3, we prove some preliminary technical facts about the problem (2). In Section 4, we use these results to establish Theorem 2.1. In Section 5, we consider dual applications of Theorem 2.1 and relate them to existing methods. In Section 6, we discuss extensions of our work.

3. Technical Preliminaries

In this section, we prove a number of useful facts about the problem (2). These facts, some of which are of independent interests, will be used in Section 4 to prove Theorem 2.1.

First, we have from the strict convexity of g the following lemma which says that the function $x \mapsto Ex$ is invariant over $\mathscr{X}^*$:

Lemma 3.1. There exists a $t^* \in \Re^m$ such that

$$Ex^* = t^*, \qquad \forall x^* \in \mathscr{X}^*.$$

Proof. First notice that, for any $x^* \in \mathscr{X}^*$ and $y^* \in \mathscr{X}^*$, we have by the convexity of $\mathscr{X}^*$ that $(x^* + y^*)/2 \in \mathscr{X}^*$. Then,

$$f(x^*) = f(y^*) = f((x^* + y^*)/2),$$

so that, using (3),

$$g((Ex^* + Ey^*)/2) = (g(Ex^*) + g(Ey^*))/2.$$

Since both Ex^* and Ey^* must be in $\operatorname{dom} g$, this together with the strict convexity of g on $\operatorname{dom} g$ yields $Ex^* = Ey^*$. □

Since $Ex^* = t^*$ for all $x^* \in \mathscr{X}^*$ [cf. Lemma 3.1], we have that ∇f is itself invariant over $\mathscr{X}^*$. In particular, we have from (4) that

$$\nabla f(x^*) = d^*, \qquad \forall x^* \in \mathscr{X}^*, \tag{12}$$

where we let

$$d^* = E^T \nabla g(t^*) + b. \tag{13}$$

Since $\nabla^2 g(t^*)$ is positive definite [cf. Lemma 3.1 and Assumption A1(c)], it follows from the continuity property of $\nabla^2 g$ [cf. Assumption A1(b)] that $\nabla^2 g$ is positive definite in some neighborhood of t^*. This in turn implies that g is strongly convex near t^*, i.e., there exist a positive scalar $\sigma > 0$ and a closed ball $\mathscr{U}^* \subseteq \operatorname{dom} g$ around t^* such that

$$g(z) - g(w) - \langle \nabla g(w), z - w \rangle \geq \sigma \|z - w\|^2, \qquad \forall z \in \mathscr{U}^*, \forall w \in \mathscr{U}^*. \tag{14}$$

By interchanging the role of w with that of z in (14) and adding the resulting relation to (14), we also obtain

$$\langle \nabla g(z) - \nabla g(w), z - w \rangle \geq 2\sigma \|z - w\|^2, \qquad \forall z \in \mathscr{U}^*, \forall w \in \mathscr{U}^*. \tag{15}$$

Since $\nabla^2 g$ is bounded on $\mathscr{U}^*$ [cf. Assumption A1(b)], then ∇g is Lipschitz continuous on $\mathscr{U}^*$, i.e., there exists a scalar constant $\rho > 0$ such that

$$\|\nabla g(z) - \nabla g(w)\| \leq \rho \|z - w\|, \qquad \forall z \in \mathscr{U}^*, \forall w \in \mathscr{U}^*. \tag{16}$$

Next, we state a lemma, due originally to Hoffman (Ref. 49, also see Refs. 50, 51), on the Lipschitz continuity of the solution of a linear system as a function of the right-hand side. This lemma will be used in the proof of Lemmas 3.3, 4.2, and 4.4 to follow.

Lemma 3.2. Let B be any $k \times n$ matrix. Then, there exists a constant $\theta > 0$ depending on B only such that, for any $\bar{x} \in \mathscr{X}$ and any $d \in \Re^k$ such that the linear system $By=d$, $y\in\mathscr{X}$ is consistent, there is a point $\bar{y}$ satisfying $B\bar{y}=d$, $\bar{y}\in\mathscr{X}$, with

$$\|\bar{x}-\bar{y}\|\leq\theta\|B\bar{x}-d\|.$$

By using Lemma 3.2 and the strict convexity of g, we have the following boundedness property of the level sets of f on $\mathscr{X}$.

Lemma 3.3. For any $\zeta\in\Re$, the set $\{Ex\,|\,x\in\mathscr{X}, f(x)\leq\zeta\}$ is a compact subset of dom g.

Proof. See Lemma 9.1 in Ref. 40. For convenience, we include the proof in Section 7. □

4. Proof of Main Result

In this section, we use the results developed in Section 3 to prove Theorem 2.1.

Let $\{x^r\}$ be a sequence of iterates generated by the coordinate descent method (8)–(9), according to either the almost cyclic or the Gauss–Southwell rule. We show below that $\{x^r\}$ converges at least linearly to an optimal solution of (2). Our line of argument is as follows. We first show, by using Lemmas 3.1–3.3 and the strict convexity of g, that $Ex^r\to t^*$ (see Lemma 4.2). This implies that, for all r sufficiently large, the local strong monotonicity and Lipschitz continuity properties of ∇g [cf. (14)–(16)] hold at Ex^r. By using these two properties, we then show the following two key facts: (i) the decrease in the cost per iteration $f(x^r)-f(x^{r+1})$ is at least of the order $\|x^r-x^{r+1}\|^2$ (see Lemma 4.3); and (ii) the cost-to-go $f(x^r)-\upsilon^*$, where υ^* denotes the optimal cost of (2), is at most of the order $\sum_{h=r}^{r+k}\|x^h-x^{h+1}\|^2$, with k some fixed nonnegative integer [see Lemma 4.5 and Eq. (50)]. It readily follows from these two facts that $\{x^r\}$ converges at least linearly to an optimal solution of (2). We remark that the proof of fact (ii) is based on upper bounding the distance from Ex^r to t^* by the error term $\|x^r-[x^r-\nabla f(x^r)]^+\|$ (see Lemma 4.4). Such an error bound, as it turns out, plays a key role in the convergence analysis of other descent methods as well (see Ref. 52).

First, we have from (9) and (10) that $\{f(x^r)\}$ is monotonically decreasing, i.e.,

$$f(x^0)\geq f(x^1)\geq\cdots\geq f(x^r)\geq f(x^{r+1})\geq\cdots. \tag{17}$$

Hence, $\{f(x^r)\}$ is bounded from above, so by Lemma 3.3,

$$\{Ex^r\} \text{ lies in a compact subset of } \operatorname{dom} g. \tag{18}$$

We next have the following result.

Lemma 4.1. $x^{r+1} - x^r \to 0$.

Proof. We will argue by contradiction. Suppose that the claim does not hold. Then, there would exist an $\epsilon > 0$, an $i \in \{1, \ldots, n\}$, and a subsequence $\mathscr{R} \subseteq \{0, 1, \ldots\}$ such that $|x_i^{r+1} - x_i^r| \geq \epsilon$, for all $r \in \mathscr{R}$. Then,

$$\|Ex^{r+1} - Ex^r\| = \|E_i\| \cdot |x_i^{r+1} - x_i^r| \geq \|E_i\| \epsilon,$$

for all $r \in \mathscr{R}$ [cf. (9)]. Since, by (18), both $\{Ex^{r+1}\}$ and $\{Ex^r\}$ lie in a compact subset of $\operatorname{dom} g$, we will (by further passing into a subsequence if necessary) assume that $\{Ex^r\}_{\mathscr{R}}$ and $\{Ex^{r+1}\}_{\mathscr{R}}$ converge to, say, t' and t'', respectively. Then, $t' \neq t''$ and both t' and t'' are in $\operatorname{dom} g$.

Since t' and t'' are in $\operatorname{dom} g$ and, by Theorem 2.10.1 in Ref. 46, g is continuous on $\operatorname{dom} g$, we have that $\{g(Ex^r)\}_{\mathscr{R}} \to g(t')$ and $\{g(Ex^{r+1})\}_{\mathscr{R}} \to g(t'')$, or equivalently, since

$$f(x) = g(Ex) + \langle b, x \rangle, \qquad \text{for all } x,$$

we have

$$\{\langle b, x^r \rangle\}_{\mathscr{R}} \to f^\infty - g(t'), \qquad \{\langle b, x^{r+1} \rangle\}_{\mathscr{R}} \to f^\infty - g(t''), \tag{19}$$

where $f^\infty = \lim_{r \to \infty} f(x^r)$ [cf. (17) and Assumption A1(a)]. Also, for each $r \in \mathscr{R}$, since x^{r+1} differs from x^r in the ith coordinate, then by (8) and (9), x^{r+1} is obtained from x^r by performing a line minimization along the ith coordinate direction in $\mathfrak{R}^n$ [cf. (10)], which together with the convexity of f yields

$$f(x^{r+1}) \leq f((x^{r+1} + x^r)/2) = g((Ex^{r+1} + Ex^r)/2) + \langle b, x^{r+1} + x^r \rangle / 2 \leq f(x^r).$$

Upon passing into the limit as $r \to \infty$, $r \in \mathscr{R}$, and using (19) and the continuity of g on $\operatorname{dom} g$, we then obtain

$$f^\infty \leq g((t'' + t')/2) + f^\infty - [g(t'') + g(t')]/2 \leq f^\infty,$$

a contradiction of the strict convexity of g on $\operatorname{dom} g$, i.e.,

$$g((t' + t'')/2) < (g(t') + g(t''))/2. \qquad \square$$

We next have the following lemma on the convergence of $\{Ex^r\}$, showing among other things that $\{Ex^r\}$ converges to t^*. This latter result is useful, since it enables us to invoke the strong monotonicity and the Lipschitz continuity properties of ∇g around t^*, namely, (14)–(16).

Lemma 4.2.

(a) $Ex^r \to t^*$.

(b) There exists an index r_0 such that, for all $r \geq r_0$,

$$x_i^r = l_i, \quad \text{if } d_i^* > 0; \qquad x_i^r = c_i, \quad \text{if } d_i^* < 0.$$

(c) $x^r - [x^r - \nabla f(x^r)]^+ \to 0$.

Proof.

(a) Since, by (18), $\{Ex^r\}$ lies in a compact subset of dom g, then $\{Ex^r\}$ is bounded and every one of its limit points is in dom g. Let t^∞ be any such limit point (so $t^\infty \in \text{dom}\, g$). We show below that t^∞ is equal to t^*, which, since the choice of t^∞ was arbitrary, would then complete our proof. For convenience, let

$$d^\infty = E^T \nabla g(t^\infty) + b,$$

and let $\mathfrak{R}$ be a subsequence of $\{0, 1, \ldots\}$ such that

$$\{Ex^r\}_{\mathfrak{R}} \to t^\infty. \tag{20}$$

Notice that, since g is continuously differentiable in an open set around t^∞, and since

$$\nabla f(x^r) = E^T \nabla g(Ex^r) + b, \qquad \text{for all } r,$$

this implies that

$$\{\nabla f(x^r)\}_{\mathfrak{R}} \to d^\infty. \tag{21}$$

First, we claim that

$$\{x_i^r\}_{\mathfrak{R}} \to l_i, \qquad \text{if } d_i^\infty > 0, \tag{22}$$

$$\{x_i^r\}_{\mathfrak{R}} \to c_i, \qquad \text{if } d_i^\infty < 0. \tag{23}$$

To see this, suppose that the almost cyclic rule is used. Since $x^{r-1} - x^r \to 0$ (cf. Lemma 4.1), we have

$$\{Ex^{r-j}\}_{\mathfrak{R}} \to t^\infty, \qquad j = 0, \ldots, B-1,$$

which in turn implies [cf. (21)]

$$\{\nabla f(x^{r-j})\}_{\mathfrak{R}} \to d^\infty, \qquad j = 0, \ldots, B-1,$$

so there exists an index $\bar{r}$ such that

$$\nabla_i f(x^{r-j}) > 0, \qquad \text{if } d_i^\infty > 0, j = 0, \ldots, B-1, \tag{24}$$

$$\nabla_i f(x^{r-j}) < 0, \qquad \text{if } d_i^\infty < 0, j = 0, \ldots, B-1, \tag{25}$$

for all $r \in \mathscr{R}$ with $r \geq \bar{r}$. Fix any i with $d_i^\infty > 0$. For each $r \in \mathscr{R}$, let $\tau(r)$ be the largest integer h not exceeding r such that the ith coordinate of x is iterated upon at the hth iteration. Then, for all $r \in \mathscr{R}$,

$$x_i^r = x_i^{\tau(r)}, \tag{26}$$

and [cf. (8)]

$$x_i^{\tau(r)} = [x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_i^+. \tag{27}$$

Since by the almost cyclic rule,

$$r - B + 1 \leq \tau(r) \leq r, \qquad \text{for all } r \in \mathscr{R},$$

it follows from (24) that, if in addition $r \geq \bar{r}$, then $\nabla_i f(x^{\tau(r)}) > 0$, so by (27) $x_i^{\tau(r)} = l_i$. Then by (26), $x_i^r = l_i$. Since the choice of i above was arbitrary, this proves (22). A symmetric argument proves (23). Now, suppose that the Gauss–Southwell rule is used. By an argument analogous to that for (24) and (25), we have that there exists an index $\bar{r}$ such that

$$\nabla_i f(x^{r+j}) > 0, \qquad \text{if } d_i^\infty > 0, j = 0, 1, \tag{28}$$

$$\nabla_i f(x^{r+j}) < 0, \qquad \text{if } d_i^\infty < 0, j = 0, 1, \tag{29}$$

for all $r \in \mathscr{R}$ with $r \geq \bar{r}$. Fix any $r \in \mathscr{R}$ with $r \geq \bar{r}$. Let s denote the index of the coordinate chosen for iteration at the $(r+1)$th iteration, so that, by (11),

$$\beta \max_i |x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+| \leq |x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+|. \tag{30}$$

If $d_s^\infty = 0$, then from the nonexpansive property of the projection operator $[\cdot]_s^+$, we have that

$$|x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+| \leq |\nabla_s f(x^r)|. \tag{31}$$

If $d_s^\infty > 0$, then we have from $\nabla_s f(x^{r+1}) > 0$ [cf. (28)] and $x_s^{r+1} = [x_s^{r+1} - \nabla_s f(x^{r+1})]_s^+$ [cf. (8)] that $x_s^{r+1} = l_s$, which together with $\nabla_s f(x^r) > 0$ [cf. (28)] implies that

$$|x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+| \leq x_s^r - l_s = |x_s^r - x_s^{r+1}|. \tag{32}$$

If $d_s^\infty < 0$, then a symmetric argument [using (29) in place of (28)] shows that (32) holds also. Then, combining (30)–(32) yields

$$\beta \max_i |x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+| \leq \max\left\{ \max_{s \text{ with } d_s^\infty = 0} |\nabla_s f(x^r)|, \|x^r - x^{r+1}\| \right\},$$

which together with Lemma 4.1 and (21) shows that

$$\{x^r - [x^r - \nabla f(x^r)]^+\}_{\mathscr{R}} \to 0.$$

This together with (21) immediately implies (22)–(23).

For each $r\in\mathscr{R}$, consider the linear system

$$Ex=Ex^r, \quad x_i=x^r, \quad \forall i \text{ with } d_i^\infty\neq 0,\ x\in\mathscr{X}.$$

This system is clearly consistent, since x^r is a solution. Fix any point $\bar{x}$ in $\mathscr{X}$. By Lemma 3.2, for every $r\in\mathscr{R}$ there exists a solution y^r of the above system satisfying

$$\|\bar{x}-y^r\|\leq\theta\Big(\|E\bar{x}-t^r\|+\sum_{d_i^\infty\neq 0}|\bar{x}_i-x_i^r|\Big),$$

where θ is a constant depending on E only. Since the right-hand side of the above expression is bounded for all $r\in\mathscr{R}$ [cf. (20), (22), (23)], it follows that $\{y^r\}_{\mathscr{R}}$ is also bounded. Then, every limit point of $\{y^r\}_{\mathscr{R}}$, say y^∞, satisfies [cf. (20), (22), (23)]

$$Ey^\infty=t^\infty,\ y_i^\infty=l_i \text{ if } d_i^\infty>0,\ y_i^\infty=c_i \text{ if } d_i^\infty<0,\ y^\infty\in\mathscr{X}.$$

Since

$$Ey^\infty=t^\infty,$$

so that

$$\nabla f(y^\infty)=E^T\nabla g(t^\infty)+b=d^\infty,$$

the above relation then yields

$$y^\infty=[y^\infty-\nabla f(y^\infty)]^+.$$

Hence, by (5), y^∞ is in $\mathscr{X}^*$, and we obtain from Lemma 3.1 that

$$t^\infty=Ey^\infty=t^*.$$

(b) Since $Ex^r\to t^*$, we have that $\nabla f(x^r)\to d^*$ [cf. (4) and (13)], so there exists an index $\bar{r}$ such that

$$\nabla_i f(x^r)>0, \qquad \text{if } d_i^*>0, \tag{33}$$

$$\nabla_i f(x^r)<0, \qquad \text{if } d_i^*<0, \tag{34}$$

for all $r\geq\bar{r}$. We also have, analogously to (22) and (23), that

$$x_i^r\to l_i, \qquad \text{if } d_i^*>0, \tag{35}$$

$$x_i^r\to c_i, \qquad \text{if } d_i^*<0. \tag{36}$$

Fix any i with $d_i^*>0$. It follows from (33) that, whenever x_i is iterated upon after the $\bar{r}$th iteration, say at the rth iteration, so that [cf. (8)]

$$x_i^r=[x_i^r-\nabla_i f(x^r)]_i^+,$$

then

$$x_i^r = l_i.$$

Since x_i^r is unchanged whenever x_i is not iterated upon, it follows that

$$x_i^r = l_i, \qquad \text{for all } r \geq r_i;$$

here, $r_i = \bar{r}$, if $x_i^{\bar{r}} = l_i$; otherwise, r_i is the smallest integer h exceeding $\bar{r}$ such that x_i is iterated upon at the hth iteration. An analogous argument, with (33) replaced by (34), shows that

$$x_i^r = c_i, \qquad \text{for all } r \geq r_i,$$

for all i with $d_i^* < 0$, where r_i is defined analogously as above.

It remains to show that the above r_i's are well defined. If the almost cyclic rule is used, then certainly each r_i is well defined and in fact satisfies

$$r_i \leq \bar{r} + B.$$

Suppose that the Gauss–Southwell rule is used. Fix any i with $d_i^* > 0$. If $x_i^{\bar{r}} = l_i$, then clearly r_i is well defined and equals $\bar{r}$. If $x_i^{\bar{r}} > l_i$, then r_i is also well defined, for otherwise x_i is never iterated upon after the $\bar{r}$th iteration, so $x_i^r = x_i^{\bar{r}} > l_i$ for $r \geq \bar{r}$, a contradiction of (35). A symmetric argument [using (36) in place of (35)] shows that r_i is well defined for any i with $d_i^* < 0$.

(c) From (4), (13), and part (a), we have that $\nabla f(x^r) \to d^*$, which together with part (b) proves the claim. □

The proof of Lemma 4.2(a) is based on one given in Ref. 40 (see proof of Proposition 4.1 therein), specialized to the coordinate descent method (8)–(9). Notice that Lemma 4.2(a) shows that, if E has full column rank so that $\mathscr{X}^*$ has a unique element, then $\{x^r\}$ converges to this element. However, for practical problems, the matrix E need not have full column rank, in which case, as we shall see, proving convergence is much more difficult.

As an immediate consequence of Lemma 4.2(a), we have that there exists an index $r_1 \geq r_0$ such that

$$Ex^r \in \mathscr{U}^*, \qquad \forall r \geq r_1. \tag{37}$$

We now use (37) and the strong convexity property of g on $\mathscr{U}^*$ [namely, (14)] to show that, asymptotically, the cost decreases as the square of the change in the iterate.

Lemma 4.3. We have

$$f(x^r) - f(x^{r+1}) \geq \sigma \min_j \|E_j\|^2 \|x^r - x^{r+1}\|^2, \qquad \forall r \geq r_1. \tag{38}$$

Proof. Fix any $r \geq r_1$, and let i denote the index of the coordinate iterated upon at the $(r+1)$th iteration. Since x^{r+1} is obtained from x^r by minimizing f along the ith coordinate direction [cf. (9), (10)], there holds that

$$\langle \nabla f(x^{r+1}), x^{r+1} - x^r \rangle \leq 0.$$

Hence, by using (3) and (4), we have

$$\begin{aligned} f(x^r) - f(x^{r+1}) &\geq f(x^r) - f(x^{r+1}) + \langle \nabla f(x^{r+1}), x^{r+1} - x^r \rangle \\ &= g(Ex^r) - g(Ex^{r+1}) - \langle \nabla g(Ex^{r+1}), E(x^r - x^{r+1}) \rangle \\ &\geq \sigma \| E(x^r - x^{r+1}) \|^2 \\ &= \sigma \| E_i \|^2 |x_i^r - x_i^{r+1}|^2 \\ &\geq \sigma \min_j \| E_j \|^2 \| x^r - x^{r+1} \|^2, \end{aligned}$$

where the second inequality follows from $Ex^{r+1} \in \mathscr{U}^*$, $Ex^r \in \mathscr{U}^*$ [cf. (37)], and (14); the last equality follows from the observation that x^{r+1} and x^r differ only in their ith coordinate [cf. (9)] and this difference is exactly $x_i^{r+1} - x_i^r$. □

Lemmas 3.1 and 3.2 combined with parts (a) and (c) of Lemma 4.2 yield the following key result. This result, based in part on the local strong convexity property of g [namely, (15)], shows that the distance from Ex^r to t^* is upper bounded by the norm of the natural residual $x^r - [x^r - \nabla f(x^r)]^+$.

Lemma 4.4. There exists a scalar constant κ such that

$$\| Ex^r - t^* \| \leq \kappa \| x^r - [x^r - \nabla f(x^r)]^+ \|, \qquad \text{for all } r \geq r_1.$$

Proof. To simplify the proof, we will assume that $c_i = \infty$ for all i. The case where some of the c_i's are finite can be treated by making a symmetric argument. Then, we have from (5) and (12) that

$$d^* \geq 0. \tag{39}$$

For convenience, let

$$z^r = [x^r - \nabla f(x^r)]^+, \qquad \forall r. \tag{40}$$

For each index set $I \subseteq \{1, \ldots, n\}$, define

$$\mathscr{R}_I = \{ r \in \{1, 2, \ldots\} \mid z_I^r > l_I,\ z_{\bar{I}}^r = l_{\bar{I}} \}, \tag{41}$$

where $\bar{I}$ denotes the complement of I with respect to $\{1, \ldots, n\}$. Since $\{1, 2, \ldots\}$ is clearly the disjoint union of the $\mathscr{R}_I$'s, it suffices to show that

$$\|Ex^r - t^*\| \leq O(\|x^r - z^r\|), \qquad \forall r \in \mathscr{R}_I, r \geq r_1, \tag{42}$$

for any I with $\mathscr{R}_I$ infinite. For convenience, we will use the notation $\alpha \leq O(\gamma)$, for any nonnegative scalars α and γ, to indicate that $\alpha \leq \kappa\gamma$ for some scalar $\kappa > 0$ independent of the iteration index r. Fix any I with $\mathscr{R}_I$ infinite. We show below that (42) holds.

By Lemma 4.2(c) and (40), we have

$$x^r - z^r \to 0. \tag{43}$$

Also, by (40), (41), and using $c_i = \infty$ for all i, we have

$$z_I^r = x_I^r - \nabla_I f(x^r), \qquad \forall r \in \mathscr{R}_I. \tag{44}$$

Since $\nabla f(x^r) \to d^*$ [cf. (4), (13), and Lemma 4.2(a)], then (43) and (44) imply that

$$d_I^* = 0. \tag{45}$$

First, we claim that

$$\text{there exists an } y \in \mathscr{X}^* \text{ satisfying } y_{\bar{I}} = l_{\bar{I}}. \tag{46}$$

To see this, for each $r \in \mathscr{R}_I$, consider the following linear system in x:

$$Ex = Ez^r, \qquad x_{\bar{I}} = l_{\bar{I}}, \qquad x \geq l.$$

The above system is consistent, since z^r is a solution of it. Hence, by Lemma 3.2, it has a solution, say w^r, whose size is bounded by some constant (depending on E only) times the size of the right-hand side. Since the right-hand side of the above system is clearly bounded for all $r \in \mathscr{R}_I$ [cf. Lemma 4.2(a) and (43)], we have that $\{w^r\}_{\mathscr{R}_I}$ is bounded. Moreover, every one of its limit points, say w^∞, satisfies [cf. Lemma 4.2(a) and (43)]

$$Ew^\infty = t^*, \qquad w_{\bar{I}}^\infty = l_{\bar{I}}, \qquad w^\infty \geq l.$$

The first relation, together with (4) and (13), implies $\nabla f(w^\infty) = d^*$ which, when combined with the latter two relations and (39) and (45) (also using $c_i = \infty$ for all i), yields

$$w^\infty = [w^\infty - \nabla f(w^\infty)]^+,$$

so that, by (5), $w^\infty \in \mathscr{X}^*$.

Next, we claim, by using (46) and Lemma 3.2, that for every $r \in \mathcal{R}_I$ there exists an $y^r \in \mathcal{X}^*$ satisfying

$$y^r_{\bar{I}} = l_{\bar{I}}, \tag{47}$$

$$\|z^r - y^r\| \le O(\|z^r - x^r\| + \|Ex^r - t^*\|). \tag{48}$$

To see this, fix any $r \in \mathcal{R}_I$. By (46) and Lemma 3.1, the linear system in y,

$$Ey = t^*, \qquad y_{\bar{I}} = l_{\bar{I}}, \qquad y \ge l,$$

is consistent. Moreover, it can be seen [using (4), (13), (39), and (45)] that every solution y of this linear system is in $\mathcal{X}^*$. Since z^r satisfies $z^r_{\bar{I}} = l_{\bar{I}}$ and $z^r \ge l$ [cf. (41)], we have from Lemma 3.2 that there exists a solution y^r to the above linear system with

$$\|z^r - y^r\| \le O(\|Ez^r - t^*\|) \le O(\|z^r - x^r\| + \|Ex^r - t^*\|),$$

from which it readily follows that y^r is in $\mathcal{X}^*$ and satisfies (47), (48).

Fix any $r \in \mathcal{R}_I$ with $r \ge r_1$. We have from (37) that $Ex^r \in \mathcal{U}^*$ and from Lemma 3.1 and $y^r \in \mathcal{X}^*$ that $Ey^r = t^*$, so (4) and the local strong convexity condition (15) imply

$$\begin{aligned}
2\sigma\|Ex^r - t^*\|^2 &\le \langle \nabla f(x^r) - \nabla f(y^r), x^r - y^r \rangle \\
&= \langle x^r_I - z^r_I, x^r_I - y^r_I \rangle + \langle \nabla_{\bar{I}} f(x^r) - \nabla_{\bar{I}} f(y^r), x^r_{\bar{I}} - z^r_{\bar{I}} \rangle \\
&\le \|x^r - z^r\|\,\|x^r - y^r\| + \|\nabla f(x^r) - \nabla f(y^r)\|\,\|x^r - z^r\| \\
&\le (\|x^r - y^r\| + \|E\|\rho\|Ex^r - t^*\|)\|x^r - z^r\| \\
&\le O((\|x^r - z^r\| + \|Ex^r - t^*\|)\|x^r - z^r\|);
\end{aligned}$$

here, the equality follows from $z^r_{\bar{I}} = y^r_{\bar{I}}$ [cf. (41), (47)], (44), and $\nabla_I f(y^r) = d^*_I = 0$ [cf. (12) and (45)]; the third inequality follows from (4) and (16); and the last inequality follows from (48). Using (48) together with the above bound yields

$$\begin{aligned}
\|x^r - y^r\|^2 &\le (\|x^r - z^r\| + \|z^r - y^r\|)^2 \\
&\le O((\|x^r - z^r\| + \|Ex^r - t^*\|)^2) \\
&\le O(\|x^r - z^r\|^2 + \|Ex^r - t^*\|^2) \\
&\le O(\|x^r - z^r\|^2 + (\|x^r - z^r\| + \|Ex^r - t^*\|)\|x^r - z^r\|) \\
&= O(\|x^r - z^r\|^2 + \|Ex^r - t^*\|\,\|x^r - z^r\|).
\end{aligned}$$

Since

$$\|Ex^r - t^*\| \leq \|E\| \, \|x^r - y^r\|$$

(cf. $t^* = Ey^r$), we immediately obtain from the above relation the following quadratic inequality:

$$\|Ex^r - t^*\|^2 \leq O(\|x^r - z^r\|^2 + \|Ex^r - t^*\| \, \|x^r - z^r\|),$$

from which it readily follows that

$$\|Ex^r - t^*\| \leq O(\|x^r - z^r\|).$$

Since the above choice of $r \in \mathscr{R}_I$ with $r \geq r_1$ was arbitrary, this proves (42). □

By using Lemma 4.4, we can prove the following important result, which roughly says that the distance from Ex^r to t^* is upper bounded by the change in the iterate around the rth iteration.

Lemma 4.5.

(a) If $\{x^r\}$ is generated according to the almost cyclic rule, then there exists a scalar constant $\omega > 0$ such that

$$\|Ex^r - t^*\| \leq \omega \sum_{h=r}^{r+B-1} \|x^h - x^{h+1}\|, \qquad \text{for all } r \geq r_1.$$

(b) If $\{x^r\}$ is generated according to the Gauss–Southwell rule, then there exists a scalar constant $\omega > 0$ such that

$$\|Ex^r - t^*\| \leq \omega \|x^{r+1} - x^r\|, \qquad \text{for all } r \geq r_1.$$

Proof.

(a) Suppose that $\{x^r\}$ is generated according to the almost cyclic rule. Fix any coordinate index $i \in \{1, \ldots, n\}$ and, for each iteration index r, let $\tau(r)$ denote the smallest integer h greater than or equal to r such that x_i is iterated on at the hth iteration. Then, by (8),

$$x_i^{\tau(r)} = [x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_i^+, \qquad \forall r,$$

so that, for any $r \geq r_1$, there holds

$$\begin{aligned}
|x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+| &= \Bigg| \sum_{h=r}^{\tau(r)-1} (x_i^h - [x_i^h - \nabla_i f(x^h)]_i^+) \\
&\quad - (x_i^{h+1} - [x_i^{h+1} - \nabla_i f(x^{h+1})]_i^+) \Bigg| \\
&\leq \sum_{h=r}^{\tau(r)-1} |(x_i^h - [x_i^h - \nabla_i f(x^h)]_i^+) \\
&\quad - (x_i^{h+1} - [x_i^{h+1} - \nabla_i f(x^{h+1})]_i^+)| \\
&\leq \sum_{h=r}^{\tau(r)-1} 2|x_i^h - x_i^{h+1}| + |\nabla_i f(x^h) - \nabla_i f(x^{h+1})| \\
&\leq \sum_{h=r}^{r+B-1} 2|x_i^h - x_i^{h+1}| + |\nabla_i f(x^h) - \nabla_i f(x^{h+1})| \\
&\leq \sum_{h=r}^{r+B-1} 2\|x^h - x^{h+1}\| + \|E\|^2 \rho \|x^h - x^{h+1}\|;
\end{aligned}$$

here, the second inequality follows from the triangle inequality and the nonexpansive property of the projection operator $[\cdot]_i^+$; the third inequality follows from the almost cyclic rule [so $\tau(r) \leq r+B$]; and the last inequality follows from (4), (16), and (37). Now use Lemma 4.4.

(b) Suppose that $\{x^r\}$ is generated according to the Gauss–Southwell rule. Fix any $r \geq r_1$. We have from (8) and (11) that

$$x_s^{r+1} = [x_s^{r+1} - \nabla_s f(x^{r+1})]_s^+,$$

for some index $s \in \{1, \ldots, n\}$ with

$$|x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+| \geq \beta |x_j^r - [x_j^r - \nabla_j f(x^r)]_j^+|, \qquad \forall j.$$

Combining the above two relations, we obtain

$$\begin{aligned}
&(\beta/\sqrt{n}) \|x^r - [x^r - \nabla f(x^r)]^+\| \\
&\leq |x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+| \\
&= |x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+ - x_s^{r+1} + [x_s^{r+1} - \nabla_s f(x^{r+1})]_s^+| \\
&\leq 2|x_s^r - x_s^{r+1}| + |\nabla_s f(x^r) - \nabla_s f(x^{r+1})| \\
&\leq 2\|x^r - x^{r+1}\| + \|\nabla f(x^r) - \nabla f(x^{r+1})\| \\
&\leq 2\|x^r - x^{r+1}\| + \|E\|^2 \rho \|x^r - x^{r+1}\|;
\end{aligned}$$

here, the second inequality follows from the triangle inequality and the nonexpansive property of the projection operator $[\cdot]_s^+$; and the last inequality follows from (4), (16), and (37). Now use Lemma 4.4. □

By combining Lemmas 4.2(b), 4.3, and 4.5, we can now prove that $\{x^r\}$ converges at least linearly to an element of $\mathscr{X}^*$. First, by Lemma 4.2(b) and the fact $r_1 \geq r_0$, we have that, for all $r \geq r_1$, there holds

$$x_i^r = l_i, \qquad \text{if } d_i^* > 0, \tag{49a}$$

$$x_i^r = c_i, \qquad \text{if } d_i^* < 0. \tag{49b}$$

Fix any element x^* of $\mathscr{X}^*$, and let

$$I^* = \{i \in \{1, \ldots, n\} \mid d_i^* = 0\}.$$

Then, for any $r \geq r_1$, we have

$$\begin{aligned} f(x^r) - f(x^*) &= g(Ex^r) - g(t^*) + \langle b, x^r - x^* \rangle \\ &= g(Ex^r) - g(t^*) + \langle b_{I^*}, x_{I^*}^r - x_{I^*}^* \rangle \\ &= g(Ex^r) - g(t^*) - \langle E_{I^*} \nabla g(t^*), x_{I^*}^r - x_{I^*}^* \rangle \\ &= g(Ex^r) - g(t^*) - \langle \nabla g(t^*), Ex^r - Ex^* \rangle \\ &\leq \rho \| Ex^r - t^* \|^2; \end{aligned} \tag{50}$$

here, the first equality follows from (3) and Lemma 3.1; the second and the fourth equality follow from (49) and using the fact $x_i^* = l_i$, if $d_i^* > 0$, and $x_i^* = c_i$, if $d_i^* < 0$; the third equality follows from $d_{I^*}^* = E_{I^*} \nabla g(t^*) + b_{I^*} = 0$ [cf. (13) and the definition of I^*]; and the inequality follows from $Ex^r \in \mathscr{U}^*$ [cf. (37)], $Ex^* = t^*$ [cf. Lemma 3.1], and the Lipschitz condition (16).

Suppose that $\{x^r\}$ is generated according to the almost cyclic rule. Then, (50) together with Lemma 4.5(a) yield that, for all $r \geq r_1$,

$$\begin{aligned} f(x^r) - f(x^*) &\leq \rho \omega^2 B \sum_{k=1}^{B} \| x^{r+k} - x^{r+k-1} \|^2 \\ &\leq \eta \sum_{k=1}^{B} (f(x^{r+k-1}) - f(x^{r+k})) \\ &= \eta (f(x^r) - f(x^{r+B})), \end{aligned}$$

where the second inequality follows from Lemma 4.3 with

$$\eta = \rho \omega^2 B \Big/ \left(\sigma \min_j \| E_j \|^2 \right).$$

Upon rearranging the terms in the above relation, we then obtain

$$f(x^{r+B})-f(x^*)\leq(1-1/\eta)(f(x^r)-f(x^*)), \qquad \forall r\geq r_1,$$

so that $\{f(x^r)\}$ converges at least linearly to $f(x^*)$. Then, (38) implies that $\{\|x^r-x^{r+1}\|^2\}$ converges at least linearly to zero, so that $\{x^r\}$ is a Cauchy sequence and converges at least linearly. See, for example, Ref. 47 on the notion of linear rate of convergence. Since f is lower semicontinuous, then the f-value of the point to which $\{x^r\}$ converges is at most $f(x^*)$. Since $x^r\in\mathscr{X}$ for all r, then this point is also in $\mathscr{X}$, so it must be an optimal solution of (2).

Suppose that $\{x^r\}$ is generated according to the Gauss–Southwell rule. Then, an argument analogous to the above, with Lemma 4.5(a) replaced by Lemma 4.5(b), shows that $\{x^r\}$ also converges at least linearly to an optimal solution of (2).

5. Dual Applications to Quadratic Programming and Entropy Optimization

As we noted in Section 1, an important application of the coordinate descent method is to the solution of problems with strictly convex costs and linear constraints (see, for example, Refs. 10, 11, 14, 17, 23, 25–27, 39, 44, 48, 53). In this section, we consider a number of such problems, including those that arise in matrix balancing and, more generally, in entropy optimization. By using Theorem 2.1, we establish, for the first time, the linear convergence of the dual iterates generated by a number of known methods for solving these problems.

Consider the following convex program:

$$\text{minimize} \quad h(y), \tag{51a}$$

$$\text{subject to} \quad E^T y\geq b, \tag{51b}$$

where h is a proper closed convex function in $\Re^m$, E is an $m\times n$ matrix having no zero column, and b is an n-vector. We remark that our results easily extend to problems with both linear equality and inequality constraints. We make the following standing assumptions about h and (51).

Assumption A2.

(a) The conjugate function of h (Ref. 46) given by

$$h^*(t)=\sup_y\{\langle t,y\rangle-h(y)\},$$

satisfies (i) dom h^* is open, (ii) h^* is strictly convex twice continuously differentiable on dom h^*, and (iii) $\nabla^2 h^*(t)$ is positive definite for all t in dom h^*.

(b) (51) has an optimal solution.

(c) The relative interior of dom h intersects $\{y \in \Re^m \mid E^T y \geq b\}$.

Notice that part (a) of Assumption A2 implies that (dom h^*, h^*) is a convex function of the Legendre type, so that, by Theorem 26.5 in Ref. 46, (dom h, h) must also be a convex function of the Legendre type. Part (c) of Assumption A2 is a constraint qualification condition that ensures the existence of a Kuhn–Tucker vector associated with the constraints $E^T y \geq b$.

By attaching a nonnegative Lagrange multiplier vector p to the constraints $E^T y \geq b$ in (51), we obtain the following dual functional

$$q(p) = \min_y h(y) + \langle p, b - E^T y \rangle = -h^*(Ep) + \langle b, p \rangle.$$

The dual problem then is to maximize $q(p)$ subject to $p \geq 0$ (see Chapter 28 in Ref. 46), or equivalently,

$$\text{minimize} \quad h^*(Ep) - \langle b, p \rangle, \tag{52a}$$

$$\text{subject to} \quad p \geq 0. \tag{52b}$$

The problem (52) is clearly of the form (2). Moreover, by parts (b) and (c) of Assumption A2, (51) has an optimal solution and the relative interior of dom h intersects $\{y \in \Re^m \mid E^T y \geq b\}$, so it follows from Corollary 28.2.2 in Ref. 46 that there exists a Kuhn–Tucker vector associated with the constraints $E^T y \geq b$. By Corollary 28.4.1 in Ref. 46, this vector is also an optimal solution of the dual problem (52), so the optimal solution set for (52) is nonempty. This together with Assumption A2(a) and the fact that h^* is a proper closed convex function implies that h^* and (52) satisfy Assumption A1; so, by Theorem 2.1, the coordinate descent method (8)–(9) applied to solve (52), according to either the almost cyclic rule or the Gauss–Southwell rule, generates iterates that converge at least linearly to an optimal solution of (52). The optimal solution of (51), which is unique since h is strictly convex, can be recovered by using the observation that, for any optimal solution p^* of (52), $\nabla h^*(Ep^*)$ is the optimal solution of (51); see Theorem 23.5 in Ref. 46.

Below, we apply the above convergence result to a number of known methods based on solving the dual program (52) using the coordinate descent method.

Application 5.1. *Convex Quadratic Programming.* Consider the special case of (51) where h is a strictly convex quadratic function, i.e.,

$$h(y) = \langle y, Qy \rangle / 2 + \langle q, y \rangle,$$

with Q an $m \times m$ symmetric positive-definite matrix and q an m-vector. For the above h, its conjugate function can be verified to be

$$h^*(t) = \langle t - q, Q^{-1}(t - q) \rangle / 2,$$

which clearly satisfies conditions (i)–(iii) of Assumption A2(a). If in addition the problem (51) is feasible, so the set $\{y \in \Re^m \mid E^T y \geq b\}$ is nonempty, then parts (b) and (c) of Assumption A2 also hold [(51) has an optimal solution, since it is feasible and since h, being in fact strongly convex, has bounded level sets], and we can conclude that the iterates generated by applying the coordinate descent method (8)–(9), according to either the almost cyclic or the Gauss–Southwell rule, to solve this special case of (52) converge at least linearly to an optimal solution of the problem. This result significantly improves upon those obtained previously for this application of the coordinate descent method, also known as Hildreth's method, most of which only showed that the iterates, premultiplied by E, converge [see Refs. 1, 3, 22, 24, 53; also see Lemma 4.2(a)]. If the order of relaxation is cyclic, then the convergence of the iterates follows from a result in Ref. 45 on the convergence of matrix splitting algorithms using regular splitting. In any case, our rate of convergence result appears to be new.

During the revision of this paper, we learned that a weaker rate of convergence result has been obtained by Iusem and de Pierro (Ref. 54), showing that, under the almost cyclic order of relaxation, the iterates, premultiplied by E, converge at least linearly.

Application 5.2. *Logarithmic Function Optimization.* Consider the special case of (51) where h is the $-\log(y)$ function, i.e.,

$$h(y) = \begin{cases} -\sum_{j=1}^{m} \log(y_j), & \text{if } y > 0, \\ \infty, & \text{otherwise.} \end{cases}$$

We can also allow positive weights on the $\log(y_j)$ terms or have a linear function of y added to h. In this case, the conjugate function of h can be verified to be (see p. 106 of Ref. 46)

$$h^*(t)=\begin{cases}-\sum_{j=1}^{m}\log(-t_j)-m, & \text{if } t<0,\\ \infty, & \text{otherwise,}\end{cases}$$

which clearly satisfies conditions (i)–(iii) of Assumption A2(a). If in addition, the set $\{y\in\Re^m \mid E^T y\geq b, y>0\}$ is nonempty and bounded, then it can be seen that parts (b) and (c) of Assumption A2 also hold[4] and, once again, we can conclude that the iterates generated by applying the coordinate descent method (8)–(9), according to either the almost cyclic or the Gauss–Southwell rule, to solve this special case of (52) converge at least linearly to an optimal solution of the problem. This result significantly improves upon that obtained in Ref. 29 which only consider the equality constrained case and only showed that the iterates, premultiplied by E, converge [also see Lemma 4.2(a)].

Application 5.3. *Entropy Optimization.* Consider the special case of (51) where h is the $y\log(y)$ entropy function, i.e.,

$$h(y)=\begin{cases}\sum_{j=1}^{m} y_j\log(y_j), & \text{if } y\geq 0,\\ \infty, & \text{otherwise.}\end{cases}$$

In this case, the conjugate function of h can be verified to be the exponential function

$$h^*(t)=\sum_{j=1}^{m} e^{t_j-1},$$

which clearly satisfies conditions (i)–(iii) of Assumption A2(a). If in addition the set $\{y\in\Re^m \mid E^T y\geq b, y>0\}$ is nonempty, then, since h has compact level sets, it can be seen that parts (b) and (c) of Assumption A2 also hold and, once again, we can conclude that the iterates generated by applying the coordinate descent method (8)–(9) to solve this special case of (52), according to either the almost cyclic or the Gauss–Southwell rule, converge at least linearly to an optimal solution of the problem. As a corollary, we obtain that the dual iterates generated by a well known matrix balancing method of Kruithof (Ref. 9, also see Refs. 25–27 and p. 408 in Ref. 31), which is effectively the cyclic coordinate descent method applied to the special case

[4]In fact, it is easily seen that the boundedness of $\{y\in\Re^m \mid E^T y\geq b, y>0\}$ is also necessary for the primal problem (51) to have an optimal solution.

of (52) where the nonnegativity constraints $p \geq 0$ are removed (i.e., the primal problem is an equality constrained problem) and E is the node-arc incidence matrix for some bipartite graph, converge at least linearly. This result significantly improves upon those obtained previously for this method (see Refs. 10, 26), which only showed that the dual iterates, premultiplied by E, converge [also see Lemma 4.2(a)].

6. Extensions

It can be verified that Theorem 2.1 holds if each x_i comprises, instead of a single coordinate, a block of coordinates (and the blocks may overlap), provided that the columns of E corresponding to the coordinates in each block are linearly independent.

The iteration (8)–(9) can also be extended to allow under/over-relaxation of the coordinates. In particular, consider an extension of the coordinate descent method in which (8) and (9) are replaced by the iteration

$$x^{r+1} = \omega^r x^r + (1-\omega^r)\tilde{x}^r,$$

where $\tilde{x}^r$ is the vector in $\mathscr{X} \cap \operatorname{dom} f$ obtained by applying (8)–(9) to x^r, that is, $\tilde{x}^r$ satisfies

$$\tilde{x}_i^r = [\tilde{x}_i^r - \nabla_i f(\tilde{x}^r)]_i^+,$$

$$\tilde{x}_j^r = x_j^r, \qquad \forall j \neq i,$$

and $\omega^r > 0$ is some relaxation factor. If $\omega^r = 1$, then the above iteration reduces to (8)–(9). Under suitable restrictions on the ω^r's, so that a sufficient decrease in the cost is achieved at every iteration, it can be shown that the iterates generated by the above method, according to either the almost cyclic rule or the Gauss–Southwell rule, also converge at least linearly to an optimal solution of (2); cf. Theorem 2.1.

Finally, we mention that the so-called Bregman's algorithm for linearly constrained strictly convex programs (Refs. 10, 42, 55, 56) may be viewed as a dual (block) coordinate descent method in the sense discussed in Section 5 (see Ref. 39). Thus, it readily follows from our convergence result that Bregman's algorithm is linearly convergent, provided that Assumption A2 holds. This latter result significantly improves upon previous convergence results for this algorithm, which only assert that the primal iterates converge (see Refs. 39, 42, 55, 56).

7. Appendix: Proof of Lemma 3.3

Fix any $\zeta \in \Re$ and consider the set

$$\{Ex \mid x \in \mathscr{X}, f(x) \leq \zeta\}. \tag{53}$$

If the set (53) is empty, then there is nothing to prove. Hence, we will assume that it is nonempty.

We first show that the set (53) is bounded. We will argue by contradiction. Suppose that this set is not bounded. Then, the convex set $\{(t, x, \zeta) \mid t = Ex, x \in \mathscr{X}, f(x) \leq \zeta\}$ in $\Re^{m+n+1}$ would have a direction of recession $(v, u, 0)$ satisfying $v \neq 0$; see Theorem 8.3 in Ref. 46. Thus, $v = Eu$ and, for any $x \in \operatorname{dom} f \cap \mathscr{X}$, there holds

$$x + \lambda u \in \mathscr{X}, \quad f(x + \lambda u) \leq f(x), \qquad \text{for all } \lambda \geq 0.$$

Choose x to be an element of $\mathscr{X}^*$. Then, there furthermore holds

$$f(x + \lambda u) = f(x), \qquad \text{for all } \lambda \geq 0,$$

so that

$$g(Ex + \lambda v) + \langle b, x + \lambda u \rangle = g(Ex) + \langle b, x \rangle,$$

or equivalently,

$$g(Ex + \lambda v) = g(Ex) - \lambda \langle b, u \rangle, \qquad \forall \lambda \geq 0.$$

Since $v \neq 0$ and [cf. (3) and $x \in \operatorname{dom} f$] $Ex \in \operatorname{dom} g$, then the above relation contradicts the strict convexity of g on $\operatorname{dom} g$ [cf. Assumption A1(b)].

Now, we show that the set (53) is a closed subset of $\operatorname{dom} g$. To show this, it suffices to show that, for any sequence of vectors $\{w^r\}$ in $\mathscr{X}$ such that $\{f(w^r)\}$ is bounded, there holds that $\{g(Ew^r)\}$ is bounded; so, by the lower semicontinuity property of g, every limit point of $\{Ew^r\}$ is in $\operatorname{dom} g$. To see this, consider any sequence $\{w^r\}$ in $\mathscr{X}$ such that $\{f(w^r)\}$ is bounded. Then, by the result shown above, $\{Ew^r\}$ is bounded, so that $\{g(Ew^r)\}$ is bounded from below. Therefore, if $\{g(Ew^r)\}$ is not bounded, then there must exist a subsequence $\mathscr{R}$ of $\{0, 1, \ldots\}$ such that $\{g(Ew^r)\}_{\mathscr{R}} \to \infty$. This in turn implies [since $f(w^r) = g(Ew^r) + \langle b, w^r \rangle$ is bounded] that

$$\{\langle b, w^r \rangle\}_{\mathscr{R}} \to -\infty. \tag{54}$$

By further passing into a subsequence if necessary, let us assume that, for each $i \in \{1, \ldots, n\}$, either $\{w_i^r\}_{\mathscr{R}}$ is bounded or $\{w_i^r\}_{\mathscr{R}} \to \infty$ or $\{w_i^r\}_{\mathscr{R}} \to -\infty$. Let I denote the set of indices i such that $\{w_i^r\}_{\mathscr{R}}$ is bounded and let $\bar{x}$ be any point in $\mathscr{X}$. For each $r\in\mathscr{R}$, consider the linear system

$$Ey=Ew^r, \qquad y_i=w_i^r, \qquad \forall i\in I,\ y\in\mathscr{X}.$$

This system is clearly consistent, since w^r is a solution. By Lemma 3.2, there exists a solution y^r of this system whose size is upper bounded by some scalar constant (depending on E only) times the size of the right-hand side. Since the right-hand side of the above system is clearly bounded for all $r\in\mathscr{R}$, it follows that $\{y^r\}_{\mathscr{R}}$ is also bounded. Let $z^r=w^r-y^r$. Then, $Ez^r=0$, $z_I^r=0$, for all $r\in\mathscr{R}$, and [cf. (54)] $\langle b, z^r\rangle<0$, for all $r\in\mathscr{R}$ sufficiently large. Moreover, for each $i\notin I$, we have either $\{w_i^r\}_{\mathscr{R}}\to\infty$ or $\{w_i^r\}_{\mathscr{R}}\to-\infty$, so that either (i) $c_i=\infty$ and $z_i^r>0$ for all $r\in\mathscr{R}$ sufficiently large, or (ii) $l_i=-\infty$ and $z_i^r<0$ for all $r\in\mathscr{R}$ sufficiently large. Hence, for any $r\in\mathscr{R}$ sufficiently large, z^r is a feasible direction of unbounded cost; i.e., for any x satisfying $l\le x\le c$ and $f(x)<\infty$, we have

$$l\le x+\lambda z^r\le c, \qquad \text{for all } \lambda>0,$$

and

$$f(x+\lambda z^r)=g(Ex)+\langle b, x+\lambda z^r\rangle=f(x)+\lambda\langle b, z^r\rangle\to-\infty, \qquad \text{as } \lambda\to\infty.$$

This contradicts the hypothesis [cf. Assumption A1(a)] that (2) has an optimal solution. □

References

1. MANGASARIAN, O. L., *Sparsity-Preserving SOR Algorithms for Separable Quadratic and Linear Programming*, Computers and Operations Research, Vol. 11, pp. 105–112, 1984.
2. MANGASARIAN, O. L., and DE LEONE, R., *Parallel Gradient Projection Successive Overrelaxation for Symmetric Linear Complementarity Problems and Linear Programs*, Annals of Operations Research, Vol. 14, pp. 41–59, 1988.
3. HERMAN, G. T., and LENT, A., *A Family of Iterative Quadratic Optimization Algorithms for Pairs of Inequalities, with Application in Diagnostic Radiology*, Mathematical Programming Study, Vol. 9, pp. 15–29, 1978.
4. COTTLE, R. W., and GOHEEN, M. S., *A Special Class of Large Quadratic Programs*, Nonlinear Programming 3, Edited by O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, New York, pp. 361–390, 1978.
5. COTTLE, R. W., GOLUB, G. H., and SACHER, R. S., *On the Solution of Large, Structured Linear Complementarity Problems: The Block-Partitioned Case*, Applied Mathematics and Optimization, Vol. 4, pp. 347–363, 1978.

6. Dembo, R. S., and Tulowitzki, U., *On the Minimization of Quadratic Functions Subject to Box Constraints*, Yale University, School of Organization and Management, Working Paper Series B, No. 78, 1983; revised 1984.
7. Arimoto, S., *An Algorithm for Computing the Capacity of Arbitrary DMCs*, IEEE Transactions on Information Theory, Vol. IT-18, pp. 14–20, 1972.
8. Blahut, R., *Computation of Channel Capacity and Rate Distortion Functions*, IEEE Transactions on Information Theory, Vol. IT-18, pp. 460–473, 1972.
9. Kruithof, J., *Calculation of Telephone Traffic*, De Ingenieur (E. Electrotechnik 3), Vol. 52, pp. E15–25, 1937.
10. Lamond, B., and Stewart, N. F., *Bregman's Balancing Method*, Transportation Research, Vol. 15B, pp. 239–248, 1981.
11. Censor, Y., *Parallel Application of Block-Iterative Methods in Medical Imaging and Radiation Therapy*, Mathematical Programming, Series B, Vol. 42, pp. 307–325, 1988.
12. Lent, A., *A Convergent Algorithm for Maximum Entropy Image Restoration with a Medical X-Ray Application*, Image Analysis and Evaluation, Edited by R. Shaw, Society of Photographic Scientists and Engineers, Washington, DC, pp. 249–257, 1977.
13. Powell, M. J. D., *An Algorithm for Maximizing Entropy Subject to Simple Bounds*, Mathematical Programming, Vol. 42, pp. 171–180, 1988.
14. Frieden, B. R., *Image Enhancement and Restoration*, Picture Processing and Digital Filtering, Edited by T. S. Huang, Springer-Verlag, Berlin, Germany, pp. 177–248, 1975.
15. Jaynes, E. T., *On the Rationale of Maximum Entropy Methods*, Proceedings of the IEEE, Vol. 70, pp. 939–952, 1982.
16. Johnson, R., and Shore, J. E., *Which Is the Better Entropy Expression for Speech Processing—S log S or log S?*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, pp. 129–136, 1984.
17. Darroch, J. N., and Ratcliff, D., *Generalized Iterative Scaling for Log-Linear Models*, Annals of Mathematical Statistics, Vol. 43, pp. 1470–1480, 1972.
18. Freund, R. M., *Projective Transformations for Interior Point Methods, Part 1: Basic Theory and Linear Programming*, Massachusetts Institute of Technology, Operations Research Center, Working Paper No. 180–88, 1988.
19. Huard, P., *Resolution of Mathematical Programming with Nonlinear Constraints by the Method of Centers*, Nonlinear Programming, Edited by J. Abadie, North-Holland, Amsterdam, Holland, pp. 207–219, 1967.
20. Karmarkar, N., *A New Polynomial-Time Algorithm for Linear Programming*, Combinatorica, Vol. 4, pp. 373–395, 1984.
21. Sonnevend, G., *New Algorithms in Convex Programming Based on a Notion of Center for Systems of Analytic Inequalities and on Rational Extrapolation*, Trends in Mathematical Optimization, Edited by K. H. Hoffman, J. B. Hiriart-Urruty, J. Zowe, and C. Lemarechal, Birkhäuser-Verlag, Basel, Switzerland, pp. 311–326, 1988.
22. Hildreth, C., *A Quadratic Programming Procedure*, Naval Research Logistics Quarterly, Vol. 4, pp. 79–85, 1957; see also *Erratum*, Vol. 4, p. 361, 1957.

23. Cottle, R. W., and Pang, J.-S., *On the Convergence of a Block Successive Overrelaxation Method for a Class of Linear Complementarity Problems*, Mathematical Programming Study, Vol. 17, pp. 126–138, 1982.
24. Cryer, C. W., *The Solution of a Quadratic Programming Problem Using Systematic Overrelaxation*, SIAM Journal on Control and Optimization, Vol. 9, pp. 385–392, 1971.
25. Bachem, A., and Korte, B., *On the RAS Algorithm*, Computing, Vol. 23, pp. 189–198, 1979.
26. Bregman, L. M., *Proof of the Convergence of Sheleikhovskii's Method for a Problem with Transportation Constraints*, USSR Computational Mathematics and Mathematical Physics, Vol. 1, pp. 191–204, 1967.
27. Schneider, M. H., and Zenios, S. A., *A Comparative Study of Algorithms for Matrix Balancing*, Operations Research, Vol. 38, pp. 439–455, 1990.
28. Zenios, S. A., and Iu, S. L., *Vector and Parallel Computing for Matrix Balancing*, Annals of Operations Research, Vol. 22, pp. 161–180, 1990.
29. Censor, Y., and Lent, A., *Optimization of log x Entropy over Linear Equality Constraints*, SIAM Journal on Control and Optimization, Vol. 25, pp. 921–933, 1987.
30. Auslender, A., *Optimisation: Méthodes Numériques*, Masson, Paris, France, 1976.
31. Bertsekas, D. P., and Tsitsiklis, J. N., *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.
32. D'Esopo, D. A., *A Convex Programming Procedure*, Naval Research Logistics Quarterly, Vol. 6, pp. 33–42, 1959.
33. Glowinski, R., *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, New York, 1984.
34. Luenberger, D. G., *Linear and Nonlinear Programming*, Addison-Wesley, Reading, Massachusetts, 1973.
35. Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York, New York, 1971.
36. Powell, M. J. D., *On Search Directions for Minimization Algorithms*, Mathematical Programming, Vol. 4, pp. 193–201, 1973.
37. Sargent, R. W. H., and Sebastian, D. J., *On the Convergence of Sequential Minimization Algorithms*, Journal of Optimization Theory and Applications, Vol. 12, pp. 567–575, 1973.
38. Zangwill, W. I., *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.
39. Tseng, P., *Dual Ascent Methods for Problems with Strictly Convex Costs and Linear Constraints: A Unified Approach*, SIAM Journal on Control and Optimization, Vol. 28, pp. 214–242, 1990.
40. Tseng, P., *Descent Methods for Convex Essentially Smooth Optimization*, Journal of Optimization Theory and Applications, Vol. 70, pp. 109–135, 1991.
41. Tseng, P., and Bertsekas, D. P., *Relaxation Methods for Problems with Strictly Convex Costs and Linear Constraints*, Mathematics of Operations Research, Vol. 16, pp. 462–481, 1991.

42. BREGMAN, L. M., *The Relaxation Method of Finding the Common Point Convex Sets and Its Application to the Solution of Problems in Convex Programming*, USSR Computational Mathematics and Mathematical Physics, Vol. 7, pp. 200–217, 1967.
43. PANG, J. S., *On the Convergence of Dual Ascent Methods for Large-Scale Linearly Constrained Optimization*, University of Texas at Dallas, School of Management, Technical Report, 1984.
44. BERTSEKAS, D. P., HOSEIN, P. A., and TSENG, P., *Relaxation Methods for Newtork flow Problems with Convex Arc Costs*, SIAM Journal on Control and Optimization, Vol. 25, pp. 1219–1243, 1987.
45. LUO, Z. Q., and TSENG, P., *On the Convergence of a Matrix-Splitting Algorithm for the Symmetric Monotone Linear Complementarity Problem*, SIAM Journal on Control and Optimization, Vol. 29, pp. 1037–1060, 1991.
46. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
47. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.
48. TSENG, P., and BERTSEKAS, D. P., *Relaxation Methods for Problems with Strictly Convex Separable Costs and Linear Constraints*, Mathematical Programming, Vol. 38, pp. 303–321, 1987.
49. HOFFMAN, A. J., *On Approximate Solutions of Systems of Linear Inequalities*, Journal of Research of the National Bureau of Standards, Vol. 49, pp. 263–265, 1952.
50. ROBINSON, S. M., *Bounds for Errors in the Solution Set of a Perturbed Linear Program*, Linear Algebra and Its Applications, Vol. 6, pp. 69–81, 1973.
51. MANGASARIAN, O. L., and SHIAU, T. H., *Lipschitz Continuity of Solutions of Linear Inequalities: Programs and Complementarity Problems*, SIAM Journal on Control and Optimization, Vol. 25, pp. 583–595, 1987.
52. LUO, Z. Q., and TSENG, P., *On the Linear Convergence of Descent Methods for Convex Essentially Smooth Minimization*, SIAM Journal on Control and Optimization (to appear).
53. LIN, Y. Y., and PANG, J. S., *Iterative Methods for Large Convex Quadratic Programs: A Survey*, SIAM Journal on Control and Optimization, Vol. 18, pp. 383–411, 1987.
54. IUSEM, A. N., and DE PIERRO, A. R., *On the Convergence Properties of Hildreth's Quadratic Programming Algorithm*, Mathematical Programming, Series A, Vol. 47, pp. 37–51, 1990.
55. DE PIERRO, A. R., and IUSEM, A. N., *A Relaxed Version of Bregman's Method for Convex Programming*, Journal of Optimization Theory and Applications, Vol. 51, pp. 421–440, 1986.
56. CENSOR, Y., and LENT, A., *An Iterative Row-Action Method for Interval Convex Programming*, Journal of Optimization Theory and Applications, Vol. 34, pp. 321–352, 1981.