

June 4, 1990

LIDS-P-1979

**ON THE LINEAR CONVERGENCE OF DESCENT METHODS
FOR CONVEX ESSENTIALLY SMOOTH MINIMIZATION***

by

Zhi-Quan Luo^{†‡} Paul Tseng[†]

– Dedicated to those courageous people who, one year ago, sacrificed their lives in Tiananmen Square, Beijing –

ABSTRACT

Consider the problem of minimizing, over a polyhedral set, the composition of an affine mapping with a strictly convex essentially smooth function. There exist many well-known descent methods for solving this problem (or special cases of), such as the gradient projection algorithm of Goldstein and Levitin and Polyak, a matrix splitting algorithm using regular splitting, and the coordinate descent method, whose rate of convergence remain poorly understood. Rate of convergence analysis for these methods typically require restrictive assumptions such as the cost function be strongly convex. We show, under very mild assumptions on the cost function, that the aforementioned algorithms converge at least linearly. Our results do not require that the cost function be strongly convex or that the optimal solution set be bounded. The key to our analysis lies in a new bound for estimating the distance from a point to the optimal solution set.

KEY WORDS: Linear convergence, differentiable minimization, gradient projection, matrix splitting, coordinate descent.

* This research is partially supported by the U.S. Army Research Office, contract DAAL03-86-K-0171 (Center for Intelligent Control Systems), by the National Science Foundation, grant NSF-DDM-8903385, and by a grant from the Science and Engineering Research Board of McMaster University.

[†] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[‡] Current address: Room 225/CRL, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada.

1 Introduction

Consider the problem of minimizing a strictly convex essentially smooth function subject to linear constraints. This problem contains a number of important optimization problems as special cases, including (strictly) convex quadratic programs, “ $x \ln(x)$ ” entropy minimization problems [Fri75], [Her80], [Jay82], [JoS84], [LaS81], [Pow88], and “ $-\ln(x)$ ” minimization problems [FiM68], [GMSTW86], [JoS84], [Son88]. A popular approach to solving this problem is to dualize the linear constraints to obtain a dual problem of minimizing, over a box, the composition of a strictly convex essentially smooth function with an affine mapping; then apply a feasible descent method to solve the dual problem (see [Cen88], [CeL87], [CoP82], [Hil57], [Kru37], [LaS81], [LiP87], [MaD87], [MaD88a], [Tse90], [TsB87a], [TsB87b] and references therein). Popular choices for the descent method include a gradient projection algorithm of Goldstein [Gol64] and Levitin and Poljak [LeP65], the coordinate descent method, and a matrix splitting algorithm using regular splitting [Kel65], [OrR70], [Pan82].

An outstanding theoretical question associated with the above solution approach concerns the *rate* of convergence of the iterates generated by it. Existing rate of convergence results all require very restrictive assumptions on the problem, such as the cost function be strongly convex (see [Dun81], [Dun87], [LeP65]). Unfortunately, these assumptions do not hold for most applications. In fact, even the convergence of the iterates has been very difficult to establish, owing to the possible unboundedness of the optimal solution set (see [Che84], [LuT89a], [LuT89b]).

Recently, the rate of convergence question for the coordinate descent method was resolved by Luo and Tseng [LuT89b], using a new proof idea based on estimating the distance from each iterate to the optimal solution set. In this paper, we extend their proof idea and results to general descent methods and a more general problem class. In particular, we consider an extension of the above dual problem in which the constraint set is any polyhedral set, not just a box; we give general conditions for a descent method, applied to solve this problem, to be linearly convergent and show that all the aforementioned algorithms (gradient projection, etc.) satisfy these conditions. Thus our results resolve the rate of convergence question for both the gradient projection algorithm and the matrix splitting algorithm using regular splitting. In fact, the line of analysis which we develop here is applicable not only to convex programs, but also to non-convex programs such as symmetric (non-monotone) linear complementarity problems. The key to our results lies in a new bound for estimating the distance from a feasible point to the optimal solution set which, unlike many existing bounds, holds even when the cost function is not strongly convex.

We now formally describe our problem. Let $f : \mathfrak{R}^n \mapsto (-\infty, +\infty]$ be a function of the form

$$f(x) = g(Ex) + \langle q, x \rangle, \quad \forall x, \tag{1.1}$$

where $g : \mathfrak{R}^m \mapsto (-\infty, +\infty]$ is some function, E is some $m \times n$ matrix (possibly with zero columns)

and q is some vector in \mathfrak{R}^n , the n -dimensional Euclidean space. In our notation, all vectors are column vectors and $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product.

We make the following standing assumptions regarding the function g :

Assumption 1.1.

- (a) The effective domain domain of g , denoted by C_g , is nonempty and open;
- (b) g is strictly convex twice continuously differentiable on C_g ;
- (c) $g(t) \rightarrow \infty$ as t approaches any boundary point of C_g .

Assumption 1.1 implies that g is, in the terminology of Rockafellar [Roc70], a strictly convex *essentially smooth* function. Such a function has a number of interesting theoretical properties. For example, its conjugate function is also strictly convex essentially smooth (see [Roc70, Chap. 26]).

It is easily seen from Assumption 1.1 and (1.1) that $C_f \neq \emptyset$, f is convex twice continuously differentiable on C_f , and f tends to ∞ at any the boundary of C_f . Hence, f is convex essentially smooth but, unlike g , not necessarily strictly convex. Also notice that if q is in the row span of E , then f depends on x through Ex only. In general, however, this needs not be the case.

Let X be a polyhedral set in \mathfrak{R}^n . Consider the following convex program associated with f and X :

$$\begin{aligned} & \text{minimize} && f(x) && (1.2) \\ & \text{subject to} && x \in X. \end{aligned}$$

We make the following standing assumptions regarding f and X :

Assumption 1.2.

- (a) The set of optimal solutions for (1.2), denoted by X^* , is nonempty.
- (b) $\nabla^2 g(Ex^*)$ is positive definite for every $x^* \in X^*$.

Part (b) of Assumption 1.2 states that g has a positive curvature on the image of X^* under the affine transformation $x \mapsto Ex$. This condition is guaranteed to hold if g has a positive curvature everywhere on C_g . [There are many important functions that satisfy this latter condition (in addition to Assumption 1.1), the most notable of which are the quadratic function, the exponential function, and the negative of the logarithm function.] Of course, if g is strongly convex and twice differentiable everywhere, then Assumptions 1.1 and 1.2 (b) hold automatically. We remark that the twice differentiability of g is not necessary for our results to hold, but it makes for a simpler statement of the assumptions. In general it suffices that g be differentiable on C_g and that ∇g be “locally” strongly monotone and Lipschitz continuous.

The problem (1.2) contains a number of important problems as special cases. For example, if E is the null matrix, then (1.2) reduces to a linear program. If g is a strictly convex quadratic

function, then (1.2) reduces to the symmetric monotone linear complementarity problem [Man77], [LiP87] (also see Section 5). If g is the function given by $g(t) = \sum_j \ln(t_j)$ for all $t \in (0, \infty)^m$ and $g(t) = \infty$ otherwise, where $\ln(\cdot)$ denotes the natural logarithm, and X is the non-negative orthant, then (1.2) reduces to the Lagrangian dual of a certain linearly constrained logarithmic penalty problem (see, e.g., [CeL87]).

For any $x \in \mathfrak{R}^n$, let $[x]^+$ denote the orthogonal projection of x onto X , i.e.,

$$[x]^+ = \arg \min_{y \in X} \|x - y\|,$$

where $\|\cdot\|$ denotes the Euclidean norm (i.e., $\|x\| = \sqrt{\langle x, x \rangle}$ for all x). Since C_f is nonempty, and f is differentiable on C_f , it is easily seen from the Kuhn–Tucker conditions for (1.2) that X^* comprises all $x \in X \cap C_f$ for which the orthogonal projection of $x - \nabla f(x)$ onto X is x itself, i.e.,

$$X^* = \{ x \in \mathfrak{R}^n \mid x = [x - \nabla f(x)]^+ \}. \quad (1.3)$$

Notice that, since both f and X are closed and convex, then so is X^* (in fact, X^* is a polyhedral set). However, X^* may be unbounded.

This paper proceeds as follows: In Section 2 we derive a new bound on the distance from a point near X^* to X^* . In Section 3 we use this bound to establish general conditions under which a sequence of points converge at least linearly to an optimal solution of (1.2). In Sections 4 to 6 we show that the iterates generated by, respectively, the gradient projection algorithm, the matrix splitting algorithm using regular splitting, and the coordinate descent method all satisfy the convergence conditions outlined in Section 3. In Section 7 we give our conclusion and discuss possible extensions.

We adopt the following notations throughout. For any $l \times k$ matrix A , we denote by A^T the transpose of A , by $\|A\|$ the matrix norm of A induced by the vector norm $\|\cdot\|$ (i.e., $\|A\| = \max_{\|x\|=1} \|Ax\|$), by A_i the i -th column of A and, for any nonempty $I \subseteq \{1, \dots, k\}$, by A_I the submatrix of A obtained by removing all columns $i \notin I$ of A . Analogously, for any k -vector x , we denote by x_i the i -th coordinate of x , and, for any nonempty subset $I \subseteq \{1, \dots, k\}$, 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). Also, for any $i \in \{1, \dots, n\}$ we denote by $\nabla_i f$ the i -th coordinate of ∇f and, for any $I \subseteq \{1, \dots, n\}$, we denote by $\nabla_I f$ the vector function comprising $\nabla_i f, i \in I$.

2 A New Error Bound

In this section we prove a key result that, for all $x \in X$ sufficiently close to the optimal solution set X^* , the distance from x to X^* is of the order $\|x - [x - \nabla f(x)]^+\|$. This result will be used in the rate of convergence analysis of Section 3.

We first need the following lemma which says that the affine mapping $x \rightarrow Ex$ is invariant over X^* . This lemma is a simple consequence of the strict convexity of g .

Lemma 2.1. There exists a $t^* \in \mathfrak{R}^m$ such that

$$Ex^* = t^*, \quad \forall x^* \in X^*.$$

Proof. For any $x^* \in X^*$ and $y^* \in X^*$, we have by the convexity of X^* that $\frac{1}{2}(x^* + y^*) \in X^*$. Then, $f(x^*) = f(y^*) = f(\frac{1}{2}(x^* + y^*))$, so that (using (1.1)) $g(\frac{1}{2}(Ex^* + Ey^*)) = \frac{1}{2}(g(Ex^*) + g(Ey^*))$. Since both $g(Ex^*)$ and $g(Ey^*)$ are finite, so that $Ex^* \in C_g$ and $Ey^* \in C_g$, this together with the strict convexity of g on C_g yields $Ex^* = Ey^*$. **Q.E.D.**

As an immediate corollary of Lemma 2.1, we have, by using the observation [cf. (1.1)]

$$\nabla f(x) = E^T \nabla g(Ex) + q, \quad \forall x \in C_f, \quad (2.1)$$

that ∇f is invariant over X^* . In fact, it is easily seen that

$$\nabla f(x^*) = d^*, \quad \forall x^* \in X^*, \quad (2.2)$$

where we denote

$$d^* = E^T \nabla g(t^*) + q. \quad (2.3)$$

The above invariant property of ∇f on X^* is quite well-known (see, e.g., [Man88]) and in fact holds for more general convex programs.

Since $\nabla^2 g(t^*)$ is positive definite [cf. Assumption 2.2], it follows from the continuity property of $\nabla^2 g$ [Assumption 1.1 (b)] that $\nabla^2 g$ is positive definite in some open 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 $\mathcal{U}^* \subseteq C_g$ around t^* such that

$$g(z) - g(y) - \langle \nabla g(y), z - y \rangle \geq \sigma \|z - y\|^2, \quad \forall z \in \mathcal{U}^*, \forall y \in \mathcal{U}^*. \quad (2.4)$$

By interchanging the role of y with that of z in (2.4) and adding the resulting relation to (2.4), we also obtain

$$\langle \nabla g(z) - \nabla g(y), z - y \rangle \geq 2\sigma \|z - y\|^2, \quad \forall z \in \mathcal{U}^*, \forall y \in \mathcal{U}^*. \quad (2.5)$$

Since $\nabla^2 g$ is bounded on \mathcal{U}^* [cf. Assumption 1.1 (b)], then ∇g is Lipschitz continuous on \mathcal{U}^* , i.e., there exists a scalar constant $\rho > 0$ such that

$$\|\nabla g(z) - \nabla g(w)\| \leq \rho \|z - w\|, \quad \forall z \in \mathcal{U}^*, \forall w \in \mathcal{U}^*. \quad (2.6)$$

We next state a lemma on the Lipschitz continuity of the solution of a linear system as a function of the right hand side. This lemma, originally due to [Hof52] (also see [Rob73], [MaS87]), will be used in the proof of Lemmas 2.4, 2.6 and 3.1.

Lemma 2.2. Let B be some $k \times l$ matrix and let $S = \{ y \in \mathfrak{R}^l \mid Cy \geq d \}$, for some $h \times l$ matrix C and some $d \in \mathfrak{R}^h$. Then, there exists a scalar constant $\theta > 0$ depending on B and C only such that, for any $\bar{x} \in S$ and any $e \in \mathfrak{R}^k$ such that the linear system $By = e$, $y \in S$ is consistent, there is a point $\bar{y} \in S$ satisfying $B\bar{y} = e$ and $\|\bar{x} - \bar{y}\| \leq \theta \|B\bar{x} - e\|$.

By using Lemma 2.2 and Assumption 1.1, we can show the following technical fact:

Lemma 2.3. For any $\zeta \in \mathfrak{R}$, the set $\{ Ex \mid x \in X, f(x) \leq \zeta \}$ is a compact subset of C_g .

Proof. See Lemma 9.1 in [Tse89].

Since X is a polyhedral set, we can for convenience express it as

$$X = \{ x \in \mathfrak{R}^n \mid Ax \geq b \},$$

for some $k \times n$ matrix A and some $b \in \mathfrak{R}^k$. Then, for any $x \in X \cap C_f$, the vector $z = [x - \nabla f(x)]^+$ is the unique vector which, together with some multiplier vector $\lambda \in \mathfrak{R}^k$, satisfies the Kuhn–Tucker conditions

$$z - x + \nabla f(x) - A^T \lambda = 0, \quad Az \geq b, \quad \lambda \geq 0, \quad (2.7)$$

$$\lambda_i = 0, \quad \forall i \notin I(x), \quad A_i z = b_i, \quad \forall i \in I(x), \quad (2.8)$$

where we denote

$$I(x) = \{ i \in \{1, \dots, n\} \mid A_i z = b_i \}.$$

We say that an $I \subseteq \{1, \dots, n\}$ is *active* at a vector $x \in X \cap C_f$ if $z = [x - \nabla f(x)]^+$ together with some $\lambda \in \mathfrak{R}^k$ satisfies

$$z - x + \nabla f(x) - A^T \lambda = 0, \quad Az \geq b, \quad \lambda \geq 0, \quad (2.9)$$

$$\lambda_i = 0, \quad \forall i \notin I, \quad A_i z = b_i, \quad \forall i \in I. \quad (2.10)$$

[Clearly, $I(x)$ is active at x for all $x \in X \cap C_f$.]

We next have the following lemma which roughly says that if $x \in X$ is sufficiently close to X^* , then those constraints that are active at x are also active at some optimal solution.

Lemma 2.4. For any $\zeta \in \mathfrak{R}$, there exists an $\epsilon > 0$ such that, for any $x \in X$ with $f(x) \leq \zeta$ and $\|x - [x - \nabla f(x)]^+\| \leq \epsilon$, $I(x)$ is active at some $x^* \in X^*$.

Proof. We argue by contradiction. If the claim does not hold, then for some $\zeta \in \mathfrak{R}$, there would exist an $I \subseteq \{1, \dots, n\}$ and a sequence of vectors $\{x^r\}$ in X satisfying $f(x^r) \leq \zeta$ for all r , $x^r - z^r \rightarrow 0$, where we let $z^r = [x^r - \nabla f(x^r)]^+$ for all r , and $I(x^r) = I$ for all r , and yet there is no $x^* \in X^*$ for which I is active at x^* .

Since $\{f(x^r)\}$ is bounded, it follows from Lemma 2.3 that $\{Ex^r\}$ lies in a compact subset of C_g . Let t^∞ be any such limit point of $\{Ex^r\}$ (so $t^\infty \in C_g$) and let \mathcal{R} be a subsequence of $\{0, 1, \dots\}$ such that

$$\{Ex^r\}_{\mathcal{R}} \rightarrow t^\infty. \quad (2.11)$$

We show below that t^∞ is equal to t^* .

Let $d^\infty = E^T \nabla g(t^\infty) + q$. Then, since $t^\infty \in C_g$ so ∇g is continuous in an open set around t^∞ , we obtain from (2.11) (and using the fact $\nabla f(x^r) = E^T \nabla g(Ex^r) + q$ for all r) that $\{\nabla f(x^r)\}_{\mathcal{R}} \rightarrow d^\infty$. For each $r \in \mathcal{R}$, consider the following linear system in x , z , and λ :

$$z - x - A^T \lambda = -\nabla f(x^r), \quad Az \geq b, \quad \lambda \geq 0,$$

$$\lambda_i = 0, \quad \forall i \notin I, \quad A_i z = b_i, \quad \forall i \in I,$$

$$Ex = Ex^r, \quad z - x = z^r - x^r.$$

The above system is consistent since, by $I(x^r) = I$ and (2.7)–(2.8), (x^r, z^r) together with some $\lambda^r \in \mathfrak{R}^k$ is a solution of it. Then, by Lemma 2.2, it has a solution $(\hat{x}^r, \hat{z}^r, \hat{\lambda}^r)$ whose size is bounded by some constant (depending on A and E only) times the size of the right hand side. Since the right hand side of the above system is clearly bounded as $r \rightarrow \infty$, $r \in \mathcal{R}$, we have that $\{(\hat{x}^r, \hat{z}^r, \hat{\lambda}^r)\}_{\mathcal{R}}$ is bounded. Moreover, every one of its limit points, say $(x^\infty, z^\infty, \lambda^\infty)$, satisfies [cf. $z^r - x^r \rightarrow 0$ and (2.11)]

$$z^\infty - x^\infty - A^T \lambda^\infty = -d^\infty, \quad Az^\infty \geq b, \quad \lambda^\infty \geq 0,$$

$$\lambda_i^\infty = 0, \quad \forall i \notin I, \quad A_i z^\infty = b_i, \quad \forall i \in I,$$

$$Ex^\infty = t^\infty \quad z^\infty - x^\infty = 0.$$

This shows $x^\infty = [x^\infty - \nabla f(x^\infty)]^+$ [cf. (2.7), (2.8) and $d^\infty = E^T \nabla g(Ex^\infty) + q$], so $x^\infty \in X^*$ [cf. (1.3)] and, by Lemma 2.1, $t^\infty = t^*$. Moreover, I is active at x^∞ , so a contradiction is established. **Q.E.D.**

Also, the proof of Lemma 2.4 shows the following:

Lemma 2.5. For any $\zeta \in \mathfrak{R}$ and any $\eta > 0$, there exists an $\epsilon > 0$ such that $\|Ex - t^*\| \leq \eta$ for all $x \in X$ with $f(x) \leq \zeta$ and $\|x - [x - \nabla f(x)]^+\| \leq \epsilon$.

By using Lemmas 2.2, 2.4 and 2.5, we can prove the following intermediate lemma:

Lemma 2.6. For any $\zeta \in \mathfrak{R}$, there exist scalar constants $\delta > 0$ and $\omega > 0$ such that, for any $x \in X$ with $f(x) \leq \zeta$ and $\|x - [x - \nabla f(x)]^+\| \leq \delta$, the following hold:

(a) $Ex \in \mathcal{U}^*$.

(b) There exists an $\lambda \in [0, \infty)^k$ satisfying

$$z - x + \nabla f(x) - (A_I)^T \lambda_I = 0,$$

and an $x^* \in X^*$ and an $\lambda^* \in [0, \infty)^k$ satisfying

$$\nabla f(x^*) - (A_I)^T \lambda_I^* = 0, \quad A_I x^* = b_I,$$

$$\|(x, \lambda) - (x^*, \lambda^*)\| \leq \omega(\|Ex - t^*\| + \|x - z\|),$$

where $I = I(x)$ and $z = [x - \nabla f(x)]^+$.

Proof. Fix any $\zeta \in \mathfrak{R}$. By Lemma 2.5, there exists some $\epsilon' > 0$ such that $Ex \in \mathcal{U}^*$ for all $x \in X$ satisfying $f(x) \leq \zeta$ and $\|x - [x - \nabla f(x)]^+\| \leq \epsilon'$. Choose δ to be the minimum of this ϵ' and the ϵ given in Lemma 2.4.

Consider any $x \in X$ satisfying the hypothesis of the lemma (with the above choice of δ), and let $z = [x - \nabla f(x)]^+$. Then, by (2.7) and (2.8), there exists some $\lambda \in \mathfrak{R}^k$ satisfying, together with x and z ,

$$z - x + \nabla f(x) - A^T \lambda = 0, \quad Ax \geq b + A(x - z), \quad \lambda \geq 0,$$

$$\lambda_i = 0, \quad \forall i \notin I(x), \quad A_i x = b_i + A_i(x - z), \quad \forall i \in I(x).$$

By Lemma 2.4, there exists an $x^* \in X^*$ such that $I(x)$ is active at x^* , so the following linear system in (x^*, λ^*)

$$d^* - A^T \lambda^* = 0, \quad Ax^* \geq b, \quad \lambda^* \geq 0, \quad Ex^* = t^*,$$

$$\lambda_i^* = 0, \quad \forall i \notin I(x), \quad A_i x^* = b_i, \quad \forall i \in I(x).$$

is consistent. Moreover, every solution (x^*, λ^*) of this linear system satisfies $x^* \in X^*$. Upon comparing the above two systems, we see that, by Lemma 2.2, there exists a solution (x^*, λ^*) to the second system such that

$$\|(x^*, \lambda^*) - (x, \lambda)\| \leq \theta(\|z - x + \nabla f(x) - d^*\| + \|Ex - t^*\| + \|A(x - z)\|),$$

where θ is some scalar constant depending on A and E only. Since our choice of δ also implies that $Ex \in \mathcal{U}^*$, so (2.1) and the Lipschitz condition (2.6) yields $\|\nabla f(x) - d^*\| = \|E^T \nabla g(Ex) - E^T \nabla g(t^*)\| \leq \rho \|E\| \|Ex - t^*\|$, then the above relation implies

$$\|(x^*, \lambda^*) - (x, \lambda)\| \leq \theta((\|A\| + 1)\|z - x\| + (\rho\|E\| + 1)\|Ex - t^*\|).$$

Q.E.D.

For any $x \in \mathfrak{R}^n$, let $\phi(x)$ denote the Euclidean distance from x to X^* , i.e.,

$$\phi(x) = \min_{x^* \in X^*} \|x - x^*\|.$$

By using Lemmas 2.1 and 2.6, we can establish the main result of this section, which roughly says that, for all $x \in X \cap C_f$ sufficiently close to X^* , $\phi(x)$ can be upper bounded by $\|x - [x - \nabla f(x)]^+\|$, the “deficit” of x .

Theorem 2.1. For any $\zeta \in \mathfrak{R}$, there exist scalar constants $\tau > 0$ and $\delta > 0$ such that

$$\phi(x) \leq \tau \|x - [x - \nabla f(x)]^+\|, \quad (2.12)$$

for all $x \in X$ with $f(x) \leq \zeta$ and $\|x - [x - \nabla f(x)]^+\| \leq \delta$.

Proof. Fix any $\zeta \in \mathfrak{R}$ and let δ and ω be the corresponding scalars given in Lemma 2.6.

Consider any $x \in X$ satisfying the hypothesis of the lemma (with the above choice of δ), and let $z = [x - \nabla f(x)]^+$, $I = I(x)$. By Lemma 2.6, $Ex \in \mathcal{U}^*$ and there exists an $\lambda \in [0, \infty)^k$ satisfying

$$z - x + \nabla f(x) - (A_I)^T \lambda_I = 0, \quad A_I z = b_I, \quad (2.13)$$

and an $x^* \in X^*$ and an $\lambda^* \in [0, \infty)^k$ satisfying

$$\nabla f(x^*) - (A_I)^T \lambda_I^* = 0, \quad A_I x^* = b_I, \quad (2.14)$$

$$\|(x, \lambda) - (x^*, \lambda^*)\| \leq \omega(\|Ex - t^*\| + \|x - z\|). \quad (2.15)$$

Also, since $Ex \in \mathcal{U}^*$, we have from (2.5) that

$$2\sigma \|Ex - t^*\|^2 \leq \langle Ex - t^*, \nabla g(Ex) - \nabla g(t^*) \rangle. \quad (2.16)$$

We claim that (2.13)–(2.16) are sufficient to establish our claim. To see this, notice that $Ex^* = t^*$ [cf. Lemma 2.1] and $\nabla f(x) - \nabla f(x^*) = E^T \nabla g(Ex) - E^T \nabla g(t^*)$ [cf. (2.1)], so (2.16) and (2.13)–(2.14) yield

$$\begin{aligned} 2\sigma \|Ex - t^*\|^2 &\leq \langle Ex - t^*, \nabla g(Ex) - \nabla g(t^*) \rangle \\ &= \langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle \\ &= \langle x - x^*, (A_I)^T \lambda_I - z + x - (A_I)^T \lambda_I^* \rangle \\ &= \langle A_I(x - x^*), \lambda_I - \lambda_I^* \rangle + \langle x - x^*, x - z \rangle \\ &= \langle A_I(x - z), \lambda_I - \lambda_I^* \rangle + \langle x - x^*, x - z \rangle \\ &= O(\|x - z\|(\|\lambda - \lambda^*\| + \|x - x^*\|)), \end{aligned}$$

where for convenience we use the notation $\alpha = O(\beta)$ to indicate that $\alpha \leq \kappa\beta$ for some scalar $\kappa > 0$ depending on ζ and the problem data only. Combining the above relation with (2.15) then gives

$$\begin{aligned} \|x - x^*\|^2 &= O\left(\|Ex - t^*\| + \|x - z\|\right)^2 \\ &= O\left(\|Ex - t^*\|^2 + \|x - z\|^2\right) \\ &= O\left(\|x - z\|(\|\lambda - \lambda^*\| + \|x - x^*\|) + \|x - z\|^2\right) \\ &= O\left(2\|x - z\|(\|Ex - t^*\| + \|x - z\|) + \|x - z\|^2\right) \end{aligned}$$

Hence $\|Ex - t^*\|^2$, which is clearly $O(\|x - x^*\|^2)$, must be $O(\|x - z\|(\|Ex - t^*\| + \|x - z\|))$, i.e., there exists a scalar constant $\kappa > 0$ (depending on ζ and the problem data only) such that

$$\|Ex - t^*\|^2 \leq \kappa(\|x - z\|(\|Ex - t^*\| + \|x - z\|)).$$

This is a quadratic inequality of the form $a^2 \leq \kappa(ab + b^2)$, which implies $a \leq \frac{1}{2}(\kappa + \sqrt{\kappa^2 + 4\kappa})b$. Hence we obtain that

$$\|Ex - t^*\| \leq \frac{1}{2}(\kappa + \sqrt{\kappa^2 + 4\kappa})\|x - z\|,$$

which when combined with (2.15) shows $\|x - x^*\| = O(\|x - z\|)$. Since $x^* \in X^*$ so clearly $\phi(x) \leq \|x - x^*\|$, this then completes our proof. **Q.E.D.**

We remark that computable error bounds like the one given in Theorem 2.1 have been quite well studied. In fact, a bound identical to that given in Theorem 2.1 was proposed by Pang for the special case where f is strongly convex [Pan85]. Alternative bounds have also been proposed, for strongly convex programs [MaD88b] and for monotone linear complementarity problems [MaS86]. However, it is unclear whether these alternative bounds are useful for analyzing the rate of convergence of algorithms. On the other hand, notice that the bound in Theorem 2.1 holds only locally, and it would be interesting to see whether this bound can be extended to hold globally on $X \cap C_f$.

3 A General Linear Convergence Result

In this section we give general conditions for a sequence of points in $X \cap C_f$ to converge at least linearly to an optimal solution of (1.2). This result, based in large part on the error bound developed in Section 3, will be used in Sections 4–6 to establish, in some cases for the first time, the linear convergence of a number of well-known algorithms.

The general linear convergence result is the following:

Theorem 3.1. Let v^* denote the optimal value of (1.2). Let $\{x^r\}$ be a sequence of vectors in $X \cap C_f$ satisfying the following two conditions:

$$f(x^r) - v^* \leq \kappa_1 \phi(x^r)^2, \quad \forall r \geq r_0, \quad (3.1)$$

$$\|x^r - [x^r - \nabla f(x^r)]^+\|^2 \leq \kappa_2 (f(x^r) - f(x^{r+1})), \quad \forall r \geq r_0, \quad (3.2)$$

where κ_1, κ_2 and r_0 are some positive scalar constants. Then, $\{f(x^r)\}$ converges at least linearly to v^* . If, in addition, there holds

$$\|x^r - x^{r+1}\|^2 \leq \kappa_3 (f(x^r) - f(x^{r+1})), \quad \forall r \geq r_0, \quad (3.3)$$

for some $\kappa_3 > 0$, then $\{x^r\}$ converges at least linearly to an element of X^* .

Proof. By (3.2), $\{f(x^r)\}$ is monotonically decreasing. Since f is also bounded from below on X [cf. Assumption 1.2 (a)], then $f(x^r) - f(x^{r+1}) \rightarrow 0$, so (3.2) yields $\|x^r - [x^r - \nabla f(x^r)]^+\| \rightarrow 0$. Hence, by Theorem 2.1, there exist scalar constants $\tau > 0$ and $r_1 \geq r_0$ such that

$$\phi(x^r) \leq \tau \|x^r - [x^r - \nabla f(x^r)]^+\|, \quad \forall r \geq r_1. \quad (3.4)$$

By combining (3.1), (3.2) and (3.4), we obtain that, for each $r \geq r_1$, there holds

$$\begin{aligned} f(x^r) - v^* &\leq \kappa_1 \phi(x^r)^2 \\ &\leq \kappa_1 (\tau)^2 \|x^r - [x^r - \nabla f(x^r)]^+\|^2 \\ &\leq \kappa_1 \kappa_2 (\tau)^2 (f(x^r) - f(x^{r+1})). \end{aligned} \quad (3.5)$$

Upon rearranging terms in (3.5), we then obtain

$$f(x^{r+1}) - v^* \leq \left(1 - \frac{1}{\kappa_1 \kappa_2 (\tau)^2}\right) (f(x^r) - v^*),$$

so $\{f(x^r)\}$ converges at least linearly to v^* . If (3.3) holds, then $\|x^{r+1} - x^r\|$ also converges at least linearly, so $\{x^r\}$ converges at least linearly. Let x^∞ denote the limit point of $\{x^r\}$. Then, $x^\infty \in X$ (since X is closed) and, by the lower semicontinuity of f , $f(x^\infty) \leq v^*$. Therefore $x^\infty \in X^*$.

Q.E.D.

[Roughly speaking, condition (3.1) says that the difference in cost between an iterate and the optimal solution nearest to him should grow *at most* quadratically in the distance between them; and condition (3.2) says that the decrease in the cost at each iteration should grow *at least* quadratically in the “deficit” of the current iterate.]

It turns out that, for our applications (see Sections 4–6), conditions (3.2) and (3.3) are relatively easy to verify. The difficulty lies in verifying that (3.1) holds. To help us with this endeavor, we develop below, by using Lemmas 2.2, 2.3 and 2.5, a sufficiency condition for (3.1) to hold. This condition, though more restrictive than (3.1), is much easier to verify for the algorithms considered in this paper.

Lemma 3.1. Suppose that $\{x^r\}$ satisfies (3.2) and (3.3) for some scalars κ_1, κ_2 and r_0 , and the following holds

$$x^{r+1} = [x^r - \alpha^r \nabla f(x^r) + e^r]^+, \quad \forall r \geq r_1, \quad (3.6)$$

for some scalar r_1 , some bounded sequence of scalars $\{\alpha^r\}$ bounded away from zero and some sequence of n -vectors $e^r \rightarrow 0$. Then $\{x^r\}$ also satisfies (3.1) (possibly with a different value for r_0).

Proof. Since f is bounded from below on X [cf. Assumption 1.2 (a)] and (3.3) and (3.2) hold, then we must have

$$x^r - x^{r-1} \rightarrow 0, \quad (3.7)$$

$$x^r - [x^r - \nabla f(x^r)]^+ \rightarrow 0. \quad (3.8)$$

We claim that there exists an $r_2 \geq r_1$ such that

$$\langle \nabla f(x^*), x^r - x^* \rangle = 0, \quad \forall x^* \in X^*, \quad (3.9)$$

for all $r \geq r_2$. To see this, let X be expressed as $X = \{x \in \mathfrak{R}^n \mid Ax \geq b\}$, for some $k \times n$ matrix A and some k -vector b , and, for every $r \geq r_1$, let I^r denote the set of indices $i \in \{1, \dots, n\}$ such that $A_i x^r = b_i$. Consider an $I \subseteq \{1, \dots, n\}$ for which the index set $\mathcal{R}_I = \{r \in \{1, 2, \dots\} \mid I^r = I\}$ is infinite. Then, there exists an $r_2 \geq r_1$ such that every integer $r \geq r_2$ belongs to \mathcal{R}_I for some such I . Hence it suffices to show that (3.9) holds for all $r \in \mathcal{R}_I$, for any I with \mathcal{R}_I infinite. We show this below.

Fix any I such that \mathcal{R}_I is infinite. Our argument will follow closely the proof of Lemma 2.4. Since $\{f(x^r)\}$ is monotonically decreasing [cf. (3.3)] so that it is bounded, we have by Lemma 2.3 that $\{Ex^r\}$ lies in a compact subset of C_g . Let t^∞ be any limit point of $\{Ex^{r-1}\}_{\mathcal{R}_I}$ (so $t^\infty \in C_g$) and let \mathcal{R}' be any subsequence of \mathcal{R}_I such that

$$\{Ex^{r-1}\}_{\mathcal{R}'} \rightarrow t^\infty. \quad (3.10)$$

Let $d^\infty = E^T \nabla g(t^\infty) + q$. Then, since $t^\infty \in C_g$ so ∇g is continuous in an open set around t^∞ , we obtain from (3.10) (and using the fact $\nabla f(x^r) = E^T \nabla g(Ex^r) + q$ for all r) that $\{\nabla f(x^{r-1})\}_{\mathcal{R}'} \rightarrow d^\infty$.

For each $r \in \mathcal{R}'$, consider the following system of linear equations in x , z , and λ :

$$z - x - A^T \lambda = e^{r-1} - \alpha^{r-1} \nabla f(x^{r-1}), \quad Az \geq b, \quad \lambda \geq 0,$$

$$\lambda_i = 0, \quad \forall i \notin I, \quad A_i z = b_i, \quad \forall i \in I,$$

$$Ex = Ex^{r-1}, \quad z - x = x^r - x^{r-1}.$$

The above system is consistent since, by $I^r = I$ and (3.6), it is satisfied by $x = x^{r-1}$, $z = x^r$, and some λ in \mathfrak{R}^k [cf. (2.7)–(2.8)]. Then, by Lemma 2.2, it has a solution $(\hat{x}^r, \hat{z}^r, \hat{\lambda}^r)$ whose size is bounded by some constant (depending on A and E only) times the size of the right hand side. Since the right hand side of the above system is clearly bounded as $r \rightarrow \infty$, $r \in \mathcal{R}'$ [cf. (3.7), (3.10), $e^r \rightarrow 0$, and the boundedness of $\{\alpha^r\}$], we have that $\{(\hat{x}^r, \hat{z}^r, \hat{\lambda}^r)\}_{\mathcal{R}'}$ is also bounded. Moreover, every one of its limit points, say $(x^\infty, z^\infty, \lambda^\infty)$, satisfies together with some $\alpha^\infty > 0$ the following conditions [cf. (3.7), $e^r \rightarrow 0$, and the boundedness hypothesis on $\{\alpha^r\}$]

$$z^\infty - x^\infty - A^T \lambda^\infty = -\alpha^\infty d^\infty, \quad Az^\infty \geq b, \quad \lambda^\infty \geq 0, \quad (3.11)$$

$$\lambda_i^\infty = 0, \quad \forall i \notin I, \quad A_i z^\infty = b_i, \quad \forall i \in I, \quad (3.12)$$

$$Ex^\infty = t^\infty \quad z^\infty - x^\infty = 0. \quad (3.13)$$

This shows $x^\infty = [x^\infty - \alpha^\infty \nabla f(x^\infty)]^+$ [cf. (2.7), (2.8) and $d^\infty = E^T \nabla g(t^\infty) + q$], so $x^\infty \in X^*$ [cf. (1.3)]. Fix any $r \in \mathcal{R}_I$. From (3.11)–(3.13) we also have that $A_I x^\infty = b_I$ and $\alpha^\infty \nabla f(x^\infty) = (A_I)^T \lambda_I^\infty$. Since $A_I x^r = b_I$ (cf. $I = I^r$), we thus obtain

$$\langle \nabla f(x^\infty), x^r - x^\infty \rangle = \frac{1}{\alpha^\infty} \langle \lambda_I^\infty, A_I (x^r - x^\infty) \rangle = 0. \quad (3.14)$$

Since x^∞ belongs to the convex set X^* and f is constant on X^* , then we must also have

$$\langle \nabla f(x^\infty), x^* - x^\infty \rangle = 0, \quad \forall x^* \in X^*,$$

which when combined with (3.14) and (2.2) yields (3.9).

Now, since (3.8) holds, then Lemma 2.5 implies $Ex^r \rightarrow t^*$, so there exists an $r_3 \geq r_2$ such that $Ex^r \in \mathcal{U}^*$ for all $r \geq r_3$. Fix any $r \geq r_3$. By the Mean Value Theorem, there exists a ξ lying on the line segment joining x^r with any $x^* \in X^*$ such that $f(x^r) - f(x^*) = \langle \nabla f(\xi), x^r - x^* \rangle$. This combined with (3.9) then yields

$$\begin{aligned} f(x^r) - f(x^*) &= \langle \nabla f(\xi), x^r - x^* \rangle \\ &= \langle \nabla f(\xi) - \nabla f(x^*), x^r - x^* \rangle + \langle \nabla f(x^*), x^r - x^* \rangle \\ &\leq \rho \|E\|^2 \|x^r - x^*\|^2 + \langle \nabla f(x^*), x^r - x^* \rangle \\ &= \rho \|E\|^2 \|x^r - x^*\|^2, \end{aligned}$$

where the inequality follows from (2.1) and the Lipschitz condition (2.6) [recall that $Ex^r \in \mathcal{U}^*$]. Since the above relation holds for all $x^* \in X^*$, then by choosing x^* to be the one nearest to x^r (in the Euclidean norm) then yields (3.1) (with $\kappa_1 = \rho \|E\|^2$). **Q.E.D.**

A few remarks about the condition (3.6) are in order. Condition (3.6) roughly says that the iterates $\{x^r\}$ should eventually identify those constraints that are active at some optimal solution. To see why this helps us to show (3.1), consider the case when X is simply a box (i.e., bound constraints). In this case, (3.6) translates to say that, for all r sufficiently large, those coordinates x_i^r of x^r for which $d_i^* > 0$ (respectively, $d_i^* < 0$) become fixed at the upper (respectively, lower) bound of x_i , which is also the bound which is active for any optimal solution. Then, it follows that, for each such r , there holds $\langle d^*, x^r - x^* \rangle = 0$, for all $x^* \in X^*$ [compare with (3.9)] from which (3.1) readily follows. The scalars $\{\alpha^r\}$ can be thought of as stepsizes and are introduced to take care of algorithms which incorporate stepsizes into their iterations (such as the gradient projection algorithm and algorithms that employ a line search step). The vectors e^r carry no meaning in themselves and are introduced mainly as a convenient tool to simplify the analysis.

By using Lemma 3.1, we immediately obtain the following useful corollary of Theorem 3.1:

Corollary 3.1. Suppose that $\{x^r\}$ satisfies the conditions (3.2), (3.3) and (3.6) for some scalars $\kappa_1, \kappa_2, r_0, r_1$, some bounded sequence of scalars $\{\alpha^r\}$ bounded away from zero and some sequence of n -vectors $e^r \rightarrow 0$. Then $\{f(x^r)\}$ converges at least linearly to the optimal value of (1.2) and $\{x^r\}$ converges at least linearly to an element of X^* .

4 Linear Convergence of Gradient Projection Algorithm

In this section, we make (in addition to Assumptions 1.1 and 1.2) the following assumptions on f :

Assumption 4.1.

(a) $C_f = \mathfrak{R}^n$.

(b) ∇f is Lipschitz continuous on \mathfrak{R}^n , i.e.,

$$\|\nabla f(y) - \nabla f(x)\| \leq \lambda \|y - x\|, \quad \forall x, \forall y, \quad (4.1)$$

where λ is some Lipschitz constant.

Consider the following algorithm of Goldstein [Gol64] and of Levitin and Polyak [LeP65] applied to solve this special case of (1.2):

Gradient Projection Algorithm

At the r -th iteration we are given an $x^r \in X$ (x^0 is chosen arbitrarily), and we compute a new iterate x^{r+1} in X given by:

$$x^{r+1} = [x^r - \alpha^r \nabla f(x^r)]^+, \quad (4.2)$$

where α^r is some positive stepsize.

For convenience we will assume that $\{\alpha^r\}$ satisfies the condition given in [Gol64] and in [LeP65]:

$$\epsilon \leq \alpha^r \leq \frac{2}{\lambda + 2\epsilon}, \quad \forall r, \quad (4.3)$$

where ϵ is some positive scalar constant (see discussion after Theorem 4.1 on extension of our result to more general stepsize rules).

The gradient projection algorithm has been studied very extensively (see, e.g., [Ber76], [CaM87], [Che84], [Dun84], [Dun87], [GaB84], [Gol64], [Gol74], [LeP65], [McT72]) but very little is known about its rate of convergence. Typically, some type of strong convexity assumption must be imposed on f to establish a rate of convergence result for this algorithm (see [Dun81], [Dun87], [LeP65]). We show below, by using the results of Section 3, that such assumptions are not necessary for our problem. More precisely, we show, under no additional assumption, that the iterates generated by (4.2)–(4.3) converge at least linearly to an optimal solution of (1.2). To the best of our knowledge, this is the first rate of convergence result for the gradient projection algorithm that does not require f to be strongly convex in any sense. (The convergence of the iterates has been shown by [Che84].) Our argument is based on showing that the iterates satisfy the convergence conditions outlined in Corollary 3.1.

Theorem 4.1. The iterates $\{x^r\}$ generated by the gradient projection algorithm (4.2)–(4.3) satisfy the hypothesis of Corollary 3.1 and hence converge at least linearly to an element of X^* .

Proof. It is well-known (and not difficult to show) that $\{x^r\}$ satisfies

$$f(x^r) - f(x^{r+1}) \geq \epsilon \|x^r - x^{r+1}\|^2, \quad \forall r,$$

(see, for example, [Gol64], [LeP65]) so (3.3) holds with $\kappa_3 = 1/\epsilon$.

Now we show that (3.2) holds. It is well-known that, for any $x \in X$ and any $d \in \mathfrak{R}^n$, the quantity $\|x - [x - \alpha d]^+\|$ is monotonically increasing with $\alpha > 0$ and the quantity $\|x - [x - \alpha d]^+\|/\alpha$ is monotonically decreasing with $\alpha > 0$ (see Lemma 1 in [GaB84] or Lemma 2.2 in [CaM87]), so that

$$\|x - [x - \alpha d]^+\| \geq \min\{1, \alpha\} \|x - [x - d]^+\|, \quad \forall \alpha > 0.$$

Applying the above bound with $x = x^r$ and $d = \nabla f(x^r)$ to (4.2) then yields

$$\|x^r - x^{r+1}\| = \|x^r - [x^r - \alpha^r \nabla f(x^r)]^+\| \geq \min\{1, \alpha^r\} \|x^r - [x^r - \nabla f(x^r)]^+\|, \quad \forall r.$$

Since $\alpha^r \geq \epsilon$ for all r [cf. (4.3)], this together with (3.3) implies that (3.2) holds.

Finally, it is easily seen from (4.2) that (3.6) holds with α^r as given and with $e^r = 0$ for all r .

Q.E.D.

We remark that Theorem 4.1 still holds when (4.3) is replaced by the more general condition that

$$\langle \nabla f(x^r), x^r - [x^r - \alpha^r \nabla f(x^r)]^+ \rangle \leq \kappa (f(x^r) - f([x^r - \alpha^r \nabla f(x^r)]^+)), \quad \forall r,$$

and

$$\epsilon \leq \alpha^r \leq \beta, \quad \forall r,$$

for some positive scalar constants κ , ϵ , and β . This is an important generalization since the stepsize rules used in practice, such as the Armijo-like rule of Bertsekas [Ber76], do not satisfy (4.3) but they do satisfy the condition above (see [Ber76], [Dun81]).

5 Linear Convergence of Matrix Splitting Algorithm

In this section we make (in addition to Assumptions 1.1 and 1.2) the following assumption on f :

Assumption 5.1. f is a convex quadratic function, i.e.,

$$f(x) = \frac{1}{2}\langle x, Mx \rangle + \langle q, x \rangle, \quad \forall x, \quad (5.1)$$

where M is some $n \times n$ symmetric positive semi-definite matrix, and q is some n -vector.

[Such f is of the form (1.1) because it is well-known that any symmetric positive semi-definite matrix can be expressed as $E^T E$ for some matrix E .] If in addition X is the non-negative orthant in \mathbb{R}^n , then the problem (1.2) reduces to the well-known symmetric monotone linear complementarity problem.

Let (B, C) be a *regular splitting* of M (see, e.g., [OrR70], [Kel65], [LiP87]), i.e.,

$$M = B + C, \quad B - C \text{ is positive definite}, \quad (5.2)$$

and consider the following algorithm for solving this special case of (1.2):

Matrix Splitting Algorithm

At the r -th iteration we are given an $x^r \in X$ (x^0 is chosen arbitrarily), and we compute a new iterate x^{r+1} in X satisfying

$$x^{r+1} = [x^{r+1} - Bx^{r+1} - Cx^r - q + h^r]^+, \quad (5.3)$$

where h^r is some n -vector.

[One simple choice for (B, C) is $B = \mu I$ and $C = M - \mu I$, where μ is a fixed scalar greater than $\|M\|/2$.]

The problem of finding an x^{r+1} satisfying (5.3) may be viewed as an affine variational inequality problem, whereby x^{r+1} is the vector $y \in X$ which satisfies the variational inequality:

$$\langle By + Cx^r + q - h^r, z - y \rangle \geq 0, \quad \forall z \in X. \quad (5.4)$$

Because B is positive definite ($2B$ is the sum of a positive definite matrix $B - C$ and a positive semi-definite matrix M), the above variational inequality problem always admits a unique solution (see, e.g., [BeT89], [KiS80]). Thus, the iterates $\{x^r\}$ are well-defined. [Methods for solving the above variational inequality problem can be found in, for example, [BeT89]. In the special case where X is a box, one can choose B so that the problem can be solved very easily (see, e.g., [LiP87]).]

The vector h^r can be thought of as an “error” vector arising as a result of an inexact computation of x^{r+1} . [This idea of introducing an error vector is taken from Mangasarian [Man90].] Let γ denote the smallest eigenvalue of the symmetric part of $B - C$ (which by hypothesis is positive) and let ϵ be a fixed scalar in $(0, \gamma/2]$. We will consider the following restriction on h^r governing how fast h^r tends to zero:

$$\|h^r\| \leq (\gamma/2 - \epsilon)\|x^r - x^{r+1}\|, \quad \forall r. \quad (5.5)$$

[Notice that the above restriction on h^r is practically enforceable and, in fact, can be used as a termination criterion for whatever method being used to compute x^{r+1} .]

In the special case where X is a box, the above matrix splitting algorithm has been very well studied (see [Man77], [Pan82], [Pan84], [Pan86], [LuT89a]). But, even in this case, very little is known about its convergence or its rate of convergence. Only very recently was it shown, under no additional assumption on the problem, that the iterates generated by this algorithm indeed converge (see [LuT89a]). Below we improve on the result of [LuT89a] by showing that these iterates converge at least linearly. Moreover, our result holds for the general case where X is any polyhedral set, not just a box. We remark that, from a theoretical standpoint, the general polyhedral case is no harder to treat (using our analysis) than the box case. However, from a practical standpoint, the box case is typically easier to work with.

Theorem 5.1. Let $\{x^r\}$ be iterates generated by the matrix splitting algorithm (5.2)–(5.3), (5.5). Then $\{x^r\}$ satisfies the hypothesis of Corollary 3.1 and hence converges at least linearly to an element of X^* .

Proof. We first verify that (3.3) holds. Fix any r . Since x^{r+1} satisfies the variational inequality (5.4), then, by plugging in x^r for z and x^{r+1} for y in (5.4), we obtain

$$\langle x^{r+1} - x^r, Bx^{r+1} + Cx^r + q - h^r \rangle \leq 0.$$

Also, from $M = B + C$ [cf. (5.2)] and our choice of f [cf. (5.1)] we have that

$$f(x^{r+1}) - f(x^r) = \langle x^{r+1} - x^r, Bx^{r+1} + Cx^r + q \rangle + \langle x^{r+1} - x^r, (C - B)(x^{r+1} - x^r) \rangle / 2.$$

Combining the above two relations then gives

$$\begin{aligned} f(x^{r+1}) - f(x^r) &\leq \langle x^{r+1} - x^r, h^r \rangle + \langle x^{r+1} - x^r, (C - B)(x^{r+1} - x^r) \rangle / 2 \\ &\leq \|x^{r+1} - x^r\| \|h^r\| - \gamma \|x^{r+1} - x^r\|^2 / 2 \\ &\leq -\epsilon \|x^{r+1} - x^r\|^2, \end{aligned}$$

where the last inequality follows from (5.5). Hence (3.3) holds with $\kappa_3 = 1/\epsilon$.

We now show that (3.2) holds. From (5.3) we have that

$$\|x^r - [x^r - \nabla f(x^r)]^+\| = \|x^r - [x^r - Mx^r - q]^+\|$$

$$\begin{aligned}
&= \|x^r - [x^r - Mx^r - q]^+ - x^{r+1} + [x^{r+1} - Bx^{r+1} - Cx^r - q + h^r]^+\| \\
&\leq \|x^r - x^{r+1}\| + \|[x^r - Mx^r - q]^+ - [x^{r+1} - Bx^{r+1} - Cx^r - q + h^r]^+\| \\
&\leq 2\|x^r - x^{r+1}\| + \|Mx^r - Bx^{r+1} - Cx^r + h^r\| \\
&\leq 2\|x^r - x^{r+1}\| + \|B(x^r - x^{r+1})\| + \|h^r\| \\
&\leq (2 + \|B\| + \gamma/2)\|x^r - x^{r+1}\|,
\end{aligned}$$

where the second inequality follows from the nonexpansive property of the projection operator $[\cdot]^+$, the third inequality follows from $M = B + C$, and the last inequality follows from (5.5). This together (3.3) shows that (3.2) holds.

Finally, we show that (3.6) holds with $\alpha^r = 1$ for all r and some sequence of n -vectors $e^r \rightarrow 0$. From (5.3), $\nabla f(x^r) = Mx^r + q$ [cf. (5.1)] and $M = B + C$ [cf. (5.2)] we have

$$x^{r+1} = [x^r - \nabla f(x^r) + B(x^r - x^{r+1}) + h^r]^+, \quad \forall r,$$

so (3.6) holds with $\alpha^r = 1$ and $e^r = B(x^r - x^{r+1}) + h^r$ for all r . Since f is bounded from below on X (cf. Assumption 1.2 (a)), then (3.3) implies $x^r - x^{r+1} \rightarrow 0$. Hence $h^r \rightarrow 0$ [cf. (5.5)] and therefore $e^r \rightarrow 0$. **Q.E.D.**

Notice that we can allow the matrix splitting (B, C) to vary from iteration to iteration, provided that the eigenvalues of the symmetric part of $B - C$ are bounded away from zero.

6 Linear Convergence of Coordinate Descent Methods

In this section we make (in addition to Assumptions 1.1 and 1.2) the following assumption:

Assumption 6.1.

(a) E has no zero column.

(b) X is a box in \mathfrak{R}^n , i.e., $X = \prod_{i=1}^n [l_i, c_i]$, for some $l_i \in [-\infty, \infty)$ and some $c_i \in (-\infty, \infty]$, $i = 1, \dots, n$.

Let $[\cdot]_i^+$ denote the orthogonal projection onto the i -th interval $[l_i, c_i]$, for all i . Consider the classical coordinate descent method for solving this special case of (1.2):

Coordinate Descent Method

At the r -th iteration we are given an $x^r \in X \cap C_f$ (x^0 is chosen arbitrarily), we choose some coordinate index $i \in \{1, \dots, n\}$ and compute a new iterate x^{r+1} satisfying

$$x_i^{r+1} = [x_i^{r+1} - \nabla_i f(x^{r+1})]_i^+, \quad (6.1)$$

$$x_j^{r+1} = x_j^r, \quad \forall j \neq i. \quad (6.2)$$

The above method may be viewed as a Gauss–Seidel method whereby the cost function f is minimized with respect to a coordinate x_i over $[l_i, c_i]$ (with the other coordinates held fixed) at each iteration, that is,

$$x_i^{r+1} = \arg \min_{l_i \leq x_i \leq c_i} f(x_1^r, \dots, x_{i-1}^r, x_i, x_{i+1}^r, \dots, x_n^r). \quad (6.3)$$

General discussions of coordinate descent methods can be found in, for example, the books [Aus76], [BeT89], [Glo84], [Lue73], [OrR70], [Pol71], [Zan69].

We claim that (6.3) [or, equivalently, (6.1)] is well-defined. To see this, suppose that, for some r and i , the minimum in (6.3) is not attained. Let e^i denote the i -th coordinate vector in \mathfrak{R}^n . Then since f is convex, either (i) $l_i = -\infty$ and $f(x^r - \alpha e^i)$ is monotonically decreasing with increasing α or (ii) $c_i = \infty$ and $f(x^r + \alpha e^i)$ is monotonically decreasing with increasing α . Suppose that case (i) holds. [Case (ii) may be treated analogously.] Then, since the set $\{Ex \mid x \in X, f(x) \leq f(x^r)\}$ is bounded by Lemma 2.3, there must hold $Ee^i = 0$, a contradiction of Assumption 6.1 (a).

To ensure that the iterates generated by the coordinate descent method (6.1)–(6.2) are convergent, we need to impose some rule on the order in which coordinates are iterated upon. Consider the following two popular rules (see, e.g., [Lue73], [SaS73], [Tse88]):

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 -th iteration satisfies

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

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

It has been shown (see [LuT89]) that $\{x^r\}$ generated by the coordinate descent method (6.1)–(6.2), using either the almost cyclic or the Gauss–Southwell rule, is linearly convergent. Below we will prove this same result, but using a different and, in some sense, simpler argument. In particular, we will show that $\{x^r\}$ satisfies the convergence conditions of Corollary 3.1.

We first need the following technical lemma (also see [LuT89, Lemma A.3] or [Tse89, proof of Proposition 4.1]):

Lemma 6.1. Let $\{x^r\}$ be iterates generated by the coordinate iterations (6.1)–(6.2). Then the following hold:

- (a) $\{Ex^r\}$ lies in a compact subset of C_g .
- (b) $x^{r+1} - x^r \rightarrow 0$.
- (c) $\nabla f(x^{r+1}) - \nabla f(x^r) \rightarrow 0$.

[For completeness, a proof of Lemma 6.1 is given in Appendix A.]

Theorem 6.1. Let $\{x^r\}$ be iterates generated by the coordinate descent method (6.1)–(6.2) according to the Gauss–Southwell rule. Then, $\{x^r\}$ satisfies the hypothesis of Corollary 3.1 and hence converges at least linearly to an element of X^* .

Proof. For any r we have from (6.1) and the description of the Gauss–Southwell rule that

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

where s is some coordinate index for which

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

Upon squaring both sides in the above relation and summing over all $j = 1, \dots, n$, we obtain

$$\begin{aligned} \frac{\beta}{\sqrt{n}} \|x^r - [x^r - \nabla f(x^r)]^+\| &\leq |x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+| \\ &= |x_s^r - [x_s^r - \nabla_s f(x^r)]_s^+ - x_s^{r+1} + [x_s^{r+1} - \nabla_s f(x^{r+1})]_s^+| \\ &\leq 2|x_s^r - x_s^{r+1}| + |\nabla_s f(x^r) - \nabla_s f(x^{r+1})| \\ &\leq 2\|x^r - x^{r+1}\| + \|\nabla f(x^r) - \nabla f(x^{r+1})\|, \end{aligned} \tag{6.4}$$

for all r , where the second inequality follows from the triangle inequality and the nonexpansive property of the projection operator $[\cdot]_s^+$, so parts (b) and (c) of Lemma 6.1 imply that $x^r - [x^r - \nabla f(x^r)]^+ \rightarrow 0$. Then, since $\{f(x^r)\}$ is also bounded, we have from Lemma 2.5 that

$$Ex^r \rightarrow t^*, \quad (6.5)$$

so there exists a scalar r_1 such that

$$Ex^r \in \mathcal{U}^*, \quad \forall r \geq r_1, \quad (6.6)$$

where \mathcal{U}^* is the closed ball around t^* given in (2.4).

We first show that (3.3) holds. Fix any $r \geq r_1$. Since x^{r+1} is obtained from x^r by minimizing f along the i -th coordinate for some $i \in \{1, \dots, n\}$ [cf. (6.3) and (6.2)], there holds

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

so (1.1) and (2.1) yield

$$\begin{aligned} f(x^r) - f(x^{r+1}) &\geq f(x^r) - f(x^{r+1}) + \langle \nabla f(x^{r+1}), x^{r+1} - x^r \rangle \\ &= g(Ex^r) - g(Ex^{r+1}) - \langle \nabla g(Ex^{r+1}), E(x^r - x^{r+1}) \rangle \\ &\geq \sigma \|E(x^r - x^{r+1})\|^2 \\ &= \sigma \|E_i\|^2 |x_i^r - x_i^{r+1}|^2 \\ &\geq \sigma \min_j \|E_j\|^2 \|x^r - x^{r+1}\|^2, \end{aligned} \quad (6.7)$$

where the second inequality follows from $Ex^{r+1} \in \mathcal{U}^*$, $Ex^r \in \mathcal{U}^*$ [cf. (6.6) and (2.4)]; the last equality follows from the observation that x^{r+1} and x^r differ only in their i -th coordinate and this difference is exactly $x_i^{r+1} - x_i^r$ [cf. (6.2)]. Hence (3.3) holds with $\kappa_3 = 1/(\sigma \min_j \|E_j\|^2)$.

We next show that (3.2) holds. We have from (6.4) that, for all $r \geq r_1$, there holds

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

where the second inequality follows from the Lipschitz condition for ∇g on \mathcal{U}^* [cf. (6.6), (2.6)]. This together with (3.3) shows that (3.2) holds.

Finally, we show that (3.6) holds. let

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

and, for each r , we define e^r to be the n -vector whose i -th coordinate is $\nabla_i f(x^r) + x_i^{r+1} - x_i^r$ if $i \in I^*$ and 0 otherwise. We claim that (3.6) holds with this choice of $\{e^r\}$ and $\alpha^r = 1$ for all r . To

see this, notice from (6.5) and the continuity of ∇g at t^* [also using (2.1)–(2.2)] that $\nabla f(x^r) \rightarrow d^*$, which together with the definition of e^r and Lemma 6.1 (b) implies that

$$e_i^r - \nabla_i f(x^r) = x_i^{r+1} - x_i^r \quad \forall r, \quad \forall i \in I^*, \quad (6.8)$$

$$e_i^r - \nabla_i f(x^r) > 0, \quad \forall r \geq r_2, \quad \forall i \text{ with } d_i^* < 0, \quad (6.9)$$

$$e_i^r - \nabla_i f(x^r) < 0, \quad \forall r \geq r_2, \quad \forall i \text{ with } d_i^* > 0, \quad (6.10)$$

for some $r_2 \geq r_1$. Eq. (6.8) implies that, for every $i \in I^*$, there holds

$$x_i^{r+1} = [x_i^r - \nabla_i f(x^r) + e_i^r]_i^+, \quad (6.11)$$

for all r . What about those $i \notin I^*$? It can be seen from (6.9)–(6.10) that each coordinate of x^r not index by I^* is eventually fixed at one of its two bounds. More precisely, there exists an $r_3 \geq r_2$ such that, for all i with $d_i^* < 0$ ($d_i^* > 0$), there holds $x_i^r = c_i < \infty$ ($x_i^r = l_i > -\infty$) for all $r \geq r_3$. This, combined with (6.9) and (6.10), then implies that (6.11) holds for all $i \notin I^*$ and all $r \geq r_3$. Hence (3.6) holds, with $\alpha^r = 1$ and the above choice of e^r , for all $r \geq r_3$. Also, since $\nabla_i f(x^r) \rightarrow d_i^* = 0$ for all $i \in I^*$ and [cf. Lemma 6.1 (b)] $x^r - x^{r+1} \rightarrow 0$, we have that $e^r \rightarrow 0$. **Q.E.D.**

Theorem 6.2. Let $\{x^r\}$ be generated by the coordinate descent method (6.1)–(6.2) according to the almost cyclic rule. Then, the sequence of vectors $\{y^k\}$ given by

$$y^k = x^{kB}, \quad \forall k,$$

satisfies the hypothesis of Corollary 3.1 and hence converges at least linearly to an element of X^* .

Proof. Fix any coordinate index $i \in \{1, \dots, n\}$ and, for each iteration index r , let $\tau(r)$ denote the smallest integer h greater than or equal to $r + 1$ such that x_i is iterated on at the $(h - 1)$ -th iteration. Then, by (6.1),

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

for all r , so that

$$\begin{aligned} |x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+| &= \left| \sum_{h=r}^{\tau(r)-1} (x_i^h - [x_i^h - \nabla_i f(x^h)]_i^+) - (x_i^{h+1} - [x_i^{h+1} - \nabla_i f(x^{h+1})]_i^+) \right| \\ &\leq \sum_{h=r}^{\tau(r)-1} |(x_i^h - [x_i^h - \nabla_i f(x^h)]_i^+) - (x_i^{h+1} - [x_i^{h+1} - \nabla_i f(x^{h+1})]_i^+)| \\ &\leq \sum_{h=r}^{\tau(r)-1} 2|x_i^h - x_i^{h+1}| + |\nabla_i f(x^h) - \nabla_i f(x^{h+1})| \\ &\leq \sum_{h=r}^{r+B-1} 2|x_i^h - x_i^{h+1}| + |\nabla_i f(x^h) - \nabla_i f(x^{h+1})|, \end{aligned} \quad (6.12)$$

where the last inequality follows from the almost cyclic rule (so $\tau(r) \leq r + B$). The above relation together with parts (a) and (b) of Lemma 6.1 yields $x_i^r - [x_i^r - \nabla_i f(x^r)]_i^+ \rightarrow 0$. Since our choice of i was arbitrary, this holds for all i and hence

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

This together with Lemma 2.5 shows that there exists a scalar r_1 such that (6.6) holds.

First we show that (3.3) holds. It is easily seen from (6.6) and the proof of Theorem 6.1 that (6.7) holds for all $r \geq r_1$, so that

$$f(x^r) - f(x^{r+1}) \geq \sigma \min_j \|E_j\|^2 \|x^r - x^{r+1}\|^2,$$

for all $r \geq r_1$, which when summed from $r = kB$ to $r = kB + B - 1$ (k is any integer exceeding $k_1 = r_1/B$) gives

$$f(y^k) - f(y^{k+1}) \geq \sigma \min_j \|E_j\|^2 \sum_{r=kB}^{kB+B-1} \|x^r - x^{r+1}\|^2.$$

Then, by applying the Cauchy–Schwartz inequality

$$B \sum_{j=1}^B (a_j)^2 \geq \left(\sum_{j=1}^B a_j \right)^2, \quad (6.13)$$

with $a_j = \|x^{kB+j-1} - x^{kB+j}\|$, $j = 1, \dots, B$, we obtain

$$\begin{aligned} f(y^k) - f(y^{k+1}) &\geq \frac{\sigma \min_j \|E_j\|^2}{B} \left(\sum_{r=kB}^{kB+B-1} \|x^r - x^{r+1}\| \right)^2 \\ &\geq \frac{\sigma \min_j \|E_j\|^2}{B} \left\| \sum_{r=kB}^{kB+B-1} (x^r - x^{r+1}) \right\|^2 \\ &= \frac{\sigma \min_j \|E_j\|^2}{B} \|y^k - y^{k+1}\|^2. \end{aligned}$$

Hence (3.3) holds for y^k .

Now we show that (3.2) holds for $\{y^k\}$. By (6.6), there holds $Ey^k \in \mathcal{U}^*$ for all $k \geq k_1$. Then, for each integer $k \geq k_1$ (so $Ex^r \in \mathcal{U}^*$ for all $r \geq kB$), we have from (6.12) that

$$\begin{aligned} |y_i^k - [y_i^k - \nabla_i f(y^k)]_i^+| &= |x_i^{kB} - [x_i^{kB} - \nabla_i f(x^{kB})]_i^+| \\ &\leq \sum_{h=kB}^{kB+B-1} 2|x_i^h - x_i^{h+1}| + |\nabla_i f(x^h) - \nabla_i f(x^{h+1})| \\ &\leq \sum_{h=kB}^{kB+B-1} 2\|x^h - x^{h+1}\| + \rho \|E\|^2 \|x^h - x^{h+1}\| \\ &= (2 + \rho \|E\|^2) \sum_{h=kB}^{kB+B-1} \|x^h - x^{h+1}\|, \end{aligned}$$

where the second inequality follows from (2.1) and the Lipschitz continuity property of ∇g on \mathcal{U}^* [cf. (6.6), (2.6)]. Upon squaring both sides in the above relation and applying the bound (6.13) to the right hand side, we then obtain

$$|y_i^k - [y_i^k - \nabla_i f(y^k)]_i^+|^2 \leq B(2 + \rho\|E\|^2)^2 \sum_{h=kB}^{kB+B-1} \|x^h - x^{h+1}\|^2,$$

which together with (6.7) yields

$$\begin{aligned} |y_i^k - [y_i^k - \nabla_i f(y^k)]_i^+|^2 &\leq \frac{B(2 + \rho\|E\|^2)^2}{\sigma \min_j \|E_j\|^2} \sum_{h=k}^{kB+B-1} (f(x^h) - f(x^{h+1})) \\ &= \frac{B(2 + \rho\|E\|^2)^2}{\sigma \min_j \|E_j\|^2} (f(y^k) - f(y^{k+1})). \end{aligned}$$

Since the choice of i was arbitrary, the above relation holds for all $i \in \{1, \dots, n\}$, which when summed over all i yields

$$\|y^k - [y^k - \nabla f(y^k)]^+\|^2 \leq n \frac{B(2 + \rho\|E\|^2)^2}{\sigma \min_j \|E_j\|^2} (f(y^k) - f(y^{k+1})).$$

Hence (3.2) holds for $\{y^k\}$.

Finally, we claim that (3.6) holds for $\alpha^r = 1$ for all r and some sequence of n -vectors $e^r \rightarrow 0$. The proof of this is analogous to that given in the proof of Theorem 6.1 and for brevity is omitted.
Q.E.D.

7 Conclusion and Extensions

In this paper we have presented a general framework for establishing the linear convergence of descent methods for solving (1.2) and have applied it to three well-known algorithms: the gradient projection algorithm of Goldstein and Levitin and Polyak, the matrix splitting algorithm using regular splitting, and the coordinate descent method. The key to this framework lies in a new bound for estimating the distance from a point to the optimal solution set.

There are a number of directions in which our results can be extended. For example, the bound of Theorem 2.1 can be shown to hold (locally) for any quadratic (possibly non-convex) function f , which enables us to extend the results of Section 5 to symmetric non-monotone linear complementarity problems. In fact, the same bound can be shown to hold (locally) for non-symmetric linear complementarity problems as well. [An example given by Mangasarian and Shiau [MaS86, Example 2.10] shows that this bound does not hold globally for non-symmetric problems.] We hope to report on these extensions in the future. Finally, it would be worthwhile to find other descent methods, other than those treated here, to which our linear convergence framework can be fruitfully applied.

8 Appendix A

In this appendix we prove Lemma 6.1. Let $\{x^r\}$ be a sequence of iterates generated by the coordinate descent method (6.1)–(6.2). Since $\{f(x^r)\}$ is monotonically decreasing and, by Assumption 1.2 (a), bounded from below, then it converges. Let

$$f^\infty = \lim_{r \rightarrow \infty} f(x^r).$$

(a) Since $\{f(x^r)\}$ is bounded, Lemma 2.3 implies that $\{Ex^r\}$ lies in a compact subset of C_g .

(b) We will argue by contradiction. If the claim does not hold, then there would exist an $\epsilon > 0$, an $i \in \{1, \dots, n\}$, and a subsequence $\mathcal{R} \subseteq \{0, 1, \dots\}$ such that $|x_i^{r+1} - x_i^r| \geq \epsilon$ for all $r \in \mathcal{R}$. Then, the i -th coordinate of x must be iterated upon at the r -th iteration for all $r \in \mathcal{R}$ and so $\|Ex^{r+1} - Ex^r\| = \|E_i\| \cdot |x_i^{r+1} - x_i^r| \geq \|E_i\|\epsilon$ for all $r \in \mathcal{R}$. Since both $\{Ex^r\}$ and $\{Ex^{r+1}\}$ are bounded [cf. part (a)], we will (by further passing into a subsequence if necessary) assume that $\{Ex^r\}_{\mathcal{R}}$ and $\{Ex^{r+1}\}_{\mathcal{R}}$ converge to, say, t' and t'' respectively. Then, $t' \neq t''$ and, by part (a), both t' and t'' are in C_g .

Since t' and t'' are in C_g and g is continuous on C_g (see [Roc70, Theorem 10.1]), we have that $\{g(Ex^r)\}_{\mathcal{R}} \rightarrow g(t')$ and $\{g(Ex^{r+1})\}_{\mathcal{R}} \rightarrow g(t'')$ or, equivalently (since $f(x) = g(Ex) + \langle q, x \rangle$ for all x and $f(x^r) \rightarrow f^\infty$),

$$\{\langle q, x^r \rangle\}_{\mathcal{R}} \rightarrow f^\infty - g(t'), \quad \{\langle q, x^{r+1} \rangle\}_{\mathcal{R}} \rightarrow f^\infty - g(t''). \quad (8.1)$$

Also, for each $r \in \mathcal{R}$, since x^{r+1} is obtained by performing a line minimization of f along the i -th coordinate direction from x^r [cf. (6.3)], the convexity of f then yields

$$f(x^{r+1}) \leq f((x^{r+1} + x^r)/2) = g((Ex^{r+1} + Ex^r)/2) + \langle q, x^{r+1} + x^r \rangle/2 \leq f(x^r), \quad \forall r \in \mathcal{R}.$$

Upon passing into the limit as $r \rightarrow \infty$, $r \in \mathcal{R}$, and using (8.1) and the continuity of g on C_g , we obtain

$$f^\infty \leq g\left(\frac{1}{2}(t'' + t')\right) + f^\infty - \frac{1}{2}(g(t'') + g(t')) \leq f^\infty,$$

a contradiction of the strict convexity of g on C_g , i.e., $g(\frac{1}{2}(t' + t'')) < \frac{1}{2}(g(t') + g(t''))$. Hence $x^r - x^{r+1} \rightarrow 0$.

(c) Since $\{Ex^r\}$ lies in a compact subset of C_g and, by Assumption 1.1 (b), ∇g is uniformly continuous on this subset, we then obtain from part (b) that $\nabla g(Ex^{r+1}) - \nabla g(Ex^r) \rightarrow 0$, so (2.1) yields

$$\nabla f(x^{r+1}) - \nabla f(x^r) \rightarrow 0.$$

References

- [Aus76] Auslender, A., *Optimisation: Méthodes Numériques*, Masson, Paris, New York, Barcelona, Milan (1976).
- [Ber76] Bertsekas, D. P., "On the Goldstein–Levitin–Polyak Gradient Projection Method," *IEEE Trans. Automat. Control*, 21 (1976), 174–184.
- [Ber82] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY (1982).
- [BeT89] Bertsekas, D. P. and Tsitsiklis, J. N., *Parallel and Distributed Computation: Numerical Methods*, Prentice–Hall, Englewood Cliffs, NJ (1989).
- [CaM87] Calamai, P. H. and Moré, J. J., "Projected Gradient Methods for Linearly Constrained Problems", *Math. Prog.*, 39 (1987), 93–116.
- [Che84] Cheng, Y. C., "On The Gradient–Projection Method for Solving the Nonsymmetric Linear Complementarity Problem", *J. Appl. Math. and Optim.*, 43 (1984), 527–540.
- [Cen88] Censor, Y., "Parallel Application of Block-Iterative Methods in Medical Imaging and Radiation Therapy," *Math. Prog.: Series B*, 42 (1988), 307–325.
- [CeL87] Censor, Y. and Lent, A., "Optimization of "log x" Entropy over Linear Equality Constraints," *SIAM J. Contr. & Optim.*, 25 (1987), 921–933.
- [CoP82] Cottle, R. W. and Pang, J.-S., "On the Convergence of a Block Successive Over-Relaxation Method for a Class of Linear Complementarity Problems", *Math. Prog. Study*, 17, North–Holland, Amsterdam (1982), 126–138.
- [Cry71] Cryer, C. W., "The Solution of a Quadratic Programming Problem Using Systematic Overrelaxation", *SIAM J. Contr. & Optim.*, 9 (1971), 385–392.
- [Dun81] Dunn, J. C., "Global and Asymptotic Convergence Rate Estimates for a Class of Projected Gradient Processes", *SIAM J. Contr. & Optim.*, 19 (1981), 368–400.
- [Dun86] Dunn, J. C., "On the Convergence of Projected Gradient Processes to Singular Critical Points", *J. Optim. Theory & Applic.*, 55 (1987), 203–216.
- [FiM68] Fiacco, A. V., and McCormick, G. P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, NY (1968).
- [Fri75] Frieden, B. R., "Image Enhancement and Restoration," in (T. S. Huang, ed.) *Picture Processing and Digital Filtering*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo (1975), 177–248.

- [GaB82] Gafni, E. M., and Bertsekas, D. P., “Convergence of a Gradient Projection Method”, Laboratory for Information and Decision Systems Report No. P-1201, M. I. T., Cambridge, MA (1982).
- [GaB84] Gafni, E. M., and Bertsekas, D. P., “Two-Metric Projection Methods for Constrained Optimization”, *SIAM J. Contr. & Optim.*, 22 (1984), 936–964.
- [GMSTW86] Gill, P. E., Murray, W., Saunders, M. A., Tomlin, J. A., and Wright, M. H., “On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar’s Projective Methods”, *Math. Prog.*, 36 (1986), 183–209.
- [Glo84] Glowinski, R., *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo (1984).
- [Gol64] Goldstein, A. A., “Convex Programming in Hilbert Space”, *Bull. Am. Math. Soc.*, 70 (1964), 709–710.
- [Gol74] Goldstein, A. A., “On Gradient Projection”, *Proc. 12-th Ann. Allerton Conference on Circuits and Systems*, Allerton Park, Ill. (1974), 38–40.
- [Her80] Herman, G. T. *Image Reconstruction from Projection: The Fundamentals of Computed Tomography*, Academic Press, New York, NY (1980).
- [Hil57] Hildreth, C., “A Quadratic Programming Procedure”, *Naval Res. Logist. Quart.*, 4 (1957), 79–85; see also “Erratum”, *Naval Res. Logist. Quart.*, 4 (1957), 361.
- [Hof52] Hoffman, A. J., “On Approximate Solutions of Systems of Linear Inequalities,” *J. Res. Natl. Bur. Standards*, 49 (1952), 263–265.
- [Jay82] Jaynes, E. T., “On the Rationale of Maximum Entropy Methods,” *Proc. IEEE*, 70 (1982), 939-952.
- [JoS84] Johnson, R. and Shore, J. E., “Which is the Better Entropy Expression for Speech Processing: $-S \log S$ or $\log S$?” *IEEE Trans. Acoust., Speech, Signal Process.* ASSP-32 (1984), 129-136.
- [Kel65] Keller, H. B., “On the Solution of Singular and Semidefinite Linear Systems by Iteration”, *SIAM Journal of Numerical Analysis*, 2 (1965), 281–290.
- [KiS80] Kinderlehrer, D. and Stampacchia, G., *An Introduction to Variational Inequalities and Applications*, Academic Press, New York, NY (1980).
- [Kru37] Kruithof, J., “Calculation of Telephone Traffic,” *De Ingenieur (E. Electrotechnik 3)*, 52 (1937), E15-25.

- [LaS81] Lamond, B. and Stewart, N. F., "Bregman's Balancing Method," *Transpn. Res.-B*, 15B (1981), 239-248.
- [LeP65] Levitin, E. S. and Poljak, B. T., "Constrained Minimization Methods", *Z. Vycisl. Mat. i Mat. Fiz.*, 6 (1965), 787-823. English translation in *USSR Comput. Math. Phys.*, 6 (1965), 1-50.
- [LiP87] Lin, Y. Y. and Pang, J.-S., "Iterative Methods for Large Convex Quadratic Programs: A Survey", *SIAM J. Contr. & Optim.*, 25 (1987), 383-411.
- [Lue73] Luenberger, D. G., *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA (1973).
- [LuT89a] Luo, Z.-Q. and Tseng, P., "On the Convergence of a Matrix-Splitting Algorithm for the Symmetric Monotone Linear Complementarity Problem," Laboratory for Information and Decision Systems Report No. P-1884, M. I. T., Cambridge, MA (July 1989; revised May 1990); to appear in *SIAM J. Contr. & Optim.*
- [LuT89b] Luo, Z.-Q. and Tseng, P., "On the Convergence of the Coordinate Descent Method for Convex Differentiable Minimization", Laboratory for Information and Decision Systems Report No. P-1924, M. I. T., Cambridge, MA (December 1989; revised January 1990).
- [McT72] McCormick, G. P. and Tapia, R. A., "The Gradient Projection Method under Mild Differentiability Conditions", *SIAM J. Contr.*, 10 (1972), 93-98.
- [Man77] Mangasarian, O. L., "Solution of Symmetric Linear Complementarity Problems by Iterative Methods", *J. Optim. Theory & Appl.*, 22 (1977), 465-485.
- [Man88] Mangasarian, O. L., "A Simple Characterization of Solution Sets of Convex Programs", *Operations Research Letters* 7 (1988), 21-26.
- [Man90] Mangasarian, O. L., "Convergence of Iterates of an Inexact Matrix Splitting Algorithm for the Symmetric Monotone Linear Complementarity Problem", Computer Sciences Technical Report #917, University of Wisconsin, Madison, WI (March, 1990).
- [MaD87] Mangasarian, O. L. and De Leone, R., "Parallel Successive Overrelaxation Methods for Symmetric Linear Complementarity Problems and Linear Programs", *J. Optim. Theory & Appl.*, 54 (1987), 437-446.
- [MaD88a] Mangasarian, O. L. and De Leone, R., "Parallel Gradient Projection Successive Overrelaxation for Symmetric Linear Complementarity Problems and Linear programs," in *Ann. Oper. Res.*, 14 (1988), 41-59.

- [MaD88b] Mangasarian, O. L. and De Leone, R., “Error Bounds for Strongly Convex Programs and (Super)Linearly Convergent Iterative Schemes for the Least 2–Norm Solution of Linear Programs”, *Applied Mathematics and Optimization*, 17 (1988), 1–14.
- [MaS86] Mangasarian, O. L. and Shiau, T.-H., “Error Bounds for Monotone Linear Complementarity Problems”, *Math. Prog.*, 36 (1986), 81–89.
- [MaS87] Mangasarian, O. L. and Shiau, T.-H., “Lipschitz Continuity of Solutions of Linear Inequalities, Programs and Complementarity Problems”, *SIAM J. Contr. & Optim.* 25 (1987), 583–595.
- [Mor89] Moré, J. J., “Gradient Projection Techniques for Large–Scale Optimization Problems,” *Proc. of the 28–th Conference on Decision and Control*, Tampa, Florida (December 1989).
- [OrR70] Ortega, J. M. and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY (1970).
- [Pan82] Pang, J.-S., “On the Convergence of a Basic Iterative Method for the Implicit Complementarity Problem”, *J. Optim. Theory & Appl.*, 37 (1982), 149–162.
- [Pan84] Pang, J.-S., “Necessary and Sufficient Conditions for the Convergence of Iterative Methods for the Linear Complementarity Problem”, *J. Optim. Theory & Appl.*, 42 (1984), 1–17.
- [Pan85] Pang, J.-S., “A Posteriori Error Bounds for the Linearly–Constrained Variational Inequality Problem”, School of Management Report, University of Texas, Dallas, TX (June 1985).
- [Pan86] Pang, J.-S., “More Results on the Convergence of Iterative Methods for the Symmetric Linear Complementarity Problem”, *J. Optim. Theory & Appl.*, 49 (1986), 107–134.
- [Pol71] Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York, NY (1971).
- [Pow88] Powell, M. J. D., “An Algorithm for Maximizing Entropy Subject to Simple Bounds,” *Math. Prog.*, 42 (1988), 171–180.
- [Rob73] Robinson, S. M., “Bounds for Error in the Solution Set of a Perturbed Linear Program”, *Lin. Algeb. & Appl.*, 6 (1973), 69–81.
- [Roc70] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, NJ (1970).
- [SaS73] Sargent, R. W. H. and Sebastian, D. J., “On the Convergence of Sequential Minimization Algorithms”, *J. Optim. Theory & Appl.*, 12 (1973), 567–575.

- [Son88] Sonnevend, G., “New Algorithms in Convex Programming Based on a Notion of “Center” (for Systems of Analytic Inequalities) and on Rational Extrapolation,” in (K.-H. Hoffman, J.-B. Hiriart-Urruty, J. Zowe, C. Lemarechal, eds.) *Trends in Mathematical Optimization*, Birkhäuser-Verlag, Basel, Switzerland (1988), 311-326.
- [Tse89] Tseng, P., “Descent Methods for Convex Essentially Smooth Minimization,” Laboratory for Information and Decision Systems Report No. P-1896, M. I. T., Cambridge, MA (1989; revised 1990); to appear in *J. Optim. Theory & Appl.*
- [Tse90] Tseng, P., “Dual Ascent Methods for Problems with Strictly Convex Costs and Linear Constraints: A Unified Approach”, *SIAM J. Contr. & Optim.*, 28 (1990), 214–242.
- [TsB87a] Tseng, P. and Bertsekas D. P., “Relaxation Methods for Problems with Strictly Convex Separable Costs and Linear Constraints,” *Math. Prog.*, 38 (1987), 303-321.
- [TsB87b] Tseng, P. and Bertsekas D. P., “Relaxation Methods for Problems with Strictly Convex Costs and Linear Constraints,” Laboratory for Information and Decision Systems Report No. P-1717, M. I. T., Cambridge, MA (1987); to appear in *Math. Oper.*
- [Zan69] Zangwill, W. I., *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, NJ (1969).