Mathematical Programming 64 (1994) 123–147

A globally convergent primal–dual interior point algorithm for convex programming

Renato D.C. Monteiro

Systems and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721, USA

Received 24 September 1991; revised manuscript received 13 April 1993

Abstract

In this paper, we study the global convergence of a large class of primal–dual interior point algorithms for solving the linearly constrained convex programming problem. The algorithms in this class decrease the value of a primal–dual potential function and hence belong to the class of so-called potential reduction algorithms. An inexact line search based on Armijo stepsize rule is used to compute the stepsize. The directions used by the algorithms are the same as the ones used in primal–dual path following and potential reduction algorithms and a very mild condition on the choice of the "centering parameter" is assumed. The algorithms always keep primal and dual feasibility and, in contrast to the polynomial potential reduction algorithms, they do not need to drive the value of the potential function towards $-\infty$ in order to converge. One of the techniques used in the convergence analysis of these algorithms has its root in nonlinear unconstrained optimization theory.

AMS 1980 Subject Classification: Primary 90C25; Secondary 49D45, 65K05.

Keywords: Interior-point methods; Convex programming; Karmarkar's algorithm; Global convergence; Potential function; Potential reduction algorithms; Armijo stepsize rule

1. Introduction

Research in interior point methods have been very intense since the publication of the seminal paper [11] by Karmarkar in 1984. Since then several papers describing interior point methods for various classes of optimizations problems have been published. These include papers in the context of linear programming (e.g., see [3, 6, 12, 17, 24, 27, 33, 38, 40, 41]), quadratic programming (e.g., see [26, 29]), convex programming (e.g., see [8–

Research supported in part by NSF Grant DDM-9109404.

SSDI 0025-5610(93)E0047-I

10, 19, 20, 23, 28, 31, 35, 42, 43]), linear complementarity problems (e.g., see [14, 16, 18, 25, 32]), nonlinear complementarity problems (e.g., see [13, 39]), nonlinear equations (e.g., see [36, 37]) and etc. These methods can be classified according to certain classes depending on some of their aspects. Among these classes, we cite the following ones:

(a) Path following methods (e.g., see [6, 8, 10, 13, 16, 17, 20, 23, 24, 26–28, 33, 36–40, 43]).

(b) Potential reduction methods (e.g., see [9, 14, 18, 19, 32, 41, 42]).

(c) Affine scaling methods (e.g., see [3, 29]).

(d) Projective scaling methods (e.g., see [11]).

One can also distinguish whether these algorithms are primal-only (or, dual-only) algorithms or primal–dual algorithms. The algorithms that we discuss in this work are more related to the algorithms of class (b) mentioned above and they are primal–dual methods.

In this paper, we study the global convergence of a large class of primal–dual (PD) interior point algorithms for solving the convex programming problem

$$(\mathrm{P}) \qquad \min\{f(x) \mid Ax = b,\ x \geqslant 0\}$$

where $f(x)$ is a smooth convex function, A is an $m \times n$ matrix and b is an m-vector. Our class of PD algorithms possesses many properties that are common to the class of algorithms (a) and (b) above. The directions generated are of the same type as those employed in the PD path following and PD potential reduction algorithms. Moreover, our class of algorithms use, at kth iteration, a direction of movement Δw_k which is a convex combination of two other directions, namely, a centering direction Δw_k^{c} and an affine direction Δw_k^{a}. A very mild condition is imposed on the coefficient σ_k of the convex combination $\Delta w_k \equiv \sigma_k \Delta_k^{\mathrm{c}} + (1-\sigma_k)\Delta w_k^{\mathrm{a}}$. Namely, for some fixed scalar $\bar{\sigma} \in (0,1)$, we require that $\sigma_k \in [0, \bar{\sigma}]$ for all k. This property is also present in a class of algorithms discussed by Kojima, Megiddo, Noma and Yoshise [14] where they investigate global convergence for algorithms in which even more freedom is allowed on the choice of the parameters σ_k. The analysis in [14] is done only in the context of linear complementarity problems and no attempt is made to consider the nonlinear case which includes problem (P) as a special case. In Section 2, we present a detailed description of the class of algorithms that we consider in this paper.

The class of algorithms we consider employs a primal–dual potential function as a merit function much like the PD potential reduction algorithms. For this reason, we can say that these algorithms belong to the class of potential reduction algorithms. The stepsize is determined by using an inexact line search procedure, namely, a version of Armijo rule specially tailored to our situation. One aspect of the algorithms we consider is that when viewed in a certain primal–dual space, the next iterate is searched from points lying along a curve, instead of a straight line, passing through the current iterate. This is due to the need of keeping primal and dual feasibility and the fact that the condition for dual feasibility in the case of a general convex function $f(x)$ involves nonlinear constraints. Calculation of the derivatives of these curves require the computation of the Hessian of $f(\cdot)$. For this reason, a line search which does not require derivative computation at intermediate points of the search, like Armijo rule, is specially important for our situation. In Section 3, we

present a detailed analysis of Armijo rule applied to the situation where (1) the merit function is defined only on an open set $D \subseteq \mathbb{R}^n$ and (2) the next iterate is searched among points lying in a curve passing through the current iterate. We then show that if these curves satisfy certain reasonable conditions then no subsequence of the sequence of iterates generated can lie in a compact subset of D (Corollary 3.1). In Section 4, we show that all the assumptions of Corollary 3.1 are satisfied by our class of algorithms and then, using the conclusion of this corollary, we derive the global convergence result that holds for this class of algorithms.

Many algorithms for unconstrained (e.g., steepest descent, Newton, quasi-Newton and conjugate gradient methods) and constrained optimization (e.g., gradient projection methods) have been presented in the literature which use inexact line searches and, in particular, Armijo stepsize selection rule. All these inexact line searches guarantee a sufficient decrease property of the objective or merit function under consideration. If the directions of movement are properly chosen then the iterates generated will eventually converge to a critical point in the unconstrained case and to a Karush–Kuhn–Tucker point in the constrained case (e.g., see Bertsekas [1, pp. 24–26] and Du and Zhang [4]). Hence, the aim of these algorithms is to find either a critical point or a KKT point of the problem. For our class of algorithms, we have another aim in sight. In our case, it is desirable (but not achievable in general) to have the iterates "drive" a primal–dual potential function, the merit function, towards $-\infty$. This is because such behavior of the iterates would guarantee that they converge to the set of optimal solutions of the problem (see Section 2). Many potential reduction algorithms for linear and quadratic programming problems have the property that their iterates drive the potential function towards $-\infty$. For these algorithms, one can show that every iteration reduces the value of the potential function by a fixed positive constant. However, we can not show a similar behavior of the iterates generated by our class of algorithms when applied to the more general problem (P). Nevertheless, we are able to show the following weaker result; namely, no subsequence of the sequence of iterates is contained in a compact set of the primal–dual interior feasible region. This result is sufficient to demonstrate the global convergence of our class of algorithms. The above weaker result is implicitly used in [14] which establishes global convergence of several primal–dual algorithms for linear complementarity problems by bounding the number of iterations necessary to drive the duality gap to a value less than $\varepsilon > 0$. The resulting bound (possibly, exponential) is expressed in terms of the data of the problem and the tolerance ε. In this work, we will not be concerned with bounding the number of iterations as in [14]. Rather, our approach will be more analytical and based on nonlinear programming techniques suitably modified to our situation. As mentioned above, the convergence analysis for our class of algorithms follows from general results involving the Armijo inexact line search presented in Section 3.

It is appropriate to review other algorithms that have been presented for solving convex programming problems or more general classes of problems. Several algorithms have been presented for solving special classes of convex programming problems (e.g., see [8–10, 19, 20, 23, 28, 31, 39, 42, 43]). In all these approaches, some conditions have to be imposed on the behavior of the Hessian of the objective function (and of the nonlinear constraint

functions in case of more general convex programming problems). These conditions vary according to the approach being used but their main aim is the same, namely, to give estimates on the variation of the Hessian function. The final result obtained is that the number of iterations can be bounded by certain problem parameters including the parameters involved in estimating the variation of the Hessian function. The relationship between the several conditions as well as the generality of each condition imposed by the forementioned algorithms are not clearly understood and apparently there exists no paper which studies this issue in detail. One set of conditions for the class of separable convex functions and examples of subsets of separable convex functions which satisfy these conditions have been given in Monteiro and Adler [28]. Mehrotra and Sun [23] discuss a set of conditions for general convex functions and argue that these conditions are satisfied by a large set of convex functions (e.g., uniform convex functions).

While the above algorithms solve special classes of convex programming problems, there are others designed to solve the general convex programming problem. A path following algorithm together with global convergence results is discussed in Kojima, Megiddo and Noma [13] in the context of special classes (e.g., the monotone class) of nonlinear complementarity problems which includes the convex problem (P) as a special case. This algorithm follows a path which may be infeasible but which eventually converges to a pair of primal and dual optimal solutions. Hence, we can view this algorithm as a path following infeasible PD algorithm. Finally, we should mention that Tanabe [36, 37] also describes a framework of path following algorithms for similar classes of problems but apparently no convergence results are given.

The following notation is used throughout our work. $\mathbb{R}^n$, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ denote the n-dimensional Euclidean space, the nonnegative orthant of $\mathbb{R}^n$ and the positive orthant of $\mathbb{R}^n$, respectively. $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote $\mathbb{R}^1$, $\mathbb{R}^1_+$ and $\mathbb{R}^1_{++}$, respectively. $\mathbb{N}$ denotes the set of all nonnegative integers. The vector whose components are all equal to 1 is denoted by e and its dimension is dictated by the appropriate context. If T is a set then T^n denotes the Cartesian product of n copies of T. Vectors are always denoted by lower case letters. The notation x_k may either refer to the kth element of a sequence $\{x_k\}_{k=1}^{\infty}$ or to the kth component of a vector $x \in \mathbb{R}^n$. The distinction will be made clear by the context. If x is a n-vector then $\mathrm{diag}(x)$ denotes the $n \times n$ diagonal matrix with the components of the vector on the diagonal. More notation will be introduced later as the need arises.

2. Description of the problem and the class of algorithms

In the first part of this section, we introduce the problem which will be the subject of our study and then discuss some of its properties. Next, we review basic notions related to path following and potential reduction algorithms, namely, the notions of central path and potential function. We also describe how PD algorithms generate the direction of movement at each iteration and how they compute the next iterate using this direction. The above description will establish a PD framework which is shared by many PD algorithms currently

available in the literature. We then describe our class of algorithms together with a description of the Armijo stepsize rule. Lastly, we state the main result of this paper in Theorem 2.1. The proof of Theorem 2.1 is postponed until the end of Section 4.

The problem we consider is the following minimization problem:

$$\text{(P)} \qquad \min\{f(x) \mid Ax=b,\ x\geqslant 0\} \tag{2.1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a function, A is an $m\times n$ matrix and b is an m-vector. We make the following assumption regarding the function $f(x)$.

Assumption 2.1. (a) $f(x)$ is a convex function.

(b) $f(x)\in C^2(\mathbb{R}^n)$, that is $f(x)$ is twice continuously differentiable over $\mathbb{R}^n$.

The Wolfe dual problem (see, e.g. Fletcher [5]) associated with (P) is given by

$$\text{(D)} \qquad \begin{array}{ll} \max & f(x)-\nabla f(x)^{\mathrm{T}}x+b^{\mathrm{T}}y \\ \text{s.t.} & -\nabla f(x)+A^{\mathrm{T}}y\leqslant 0. \end{array}$$

For $x\in\mathbb{R}^n$ and $y\in\mathbb{R}^m$, let $s(x,y)$ denote the slack corresponding to the dual constraint as a function of (x,y), that is,

$$s(x,y)\equiv\nabla f(x)-A^{\mathrm{T}}y. \tag{2.2}$$

It is well-known (e.g., see Fletcher [5, p. 219]) that the existence of an optimal solution $\bar{x}$ for problem (P) implies: (1) the existence of a vector $\bar{y}\in\mathbb{R}^m$ such that $(\bar{x},\bar{y})$ solves problem (D); (2) the primal and dual optimal values are equal, that is, $\nabla f(\bar{x})^{\mathrm{T}}\bar{x}-b^{\mathrm{T}}\bar{y}=0$. Using the fact that $s(\bar{x},\bar{y})=\nabla f(\bar{x})-A^{\mathrm{T}}\bar{y}$ and $A\bar{x}=b$, (2) can be equivalently expressed as $s(\bar{x},\bar{y})^{\mathrm{T}}\bar{x}=0$.

We define the following sets:

$$F_{XY}\equiv\{(x,y)\in\mathbb{R}^n\times R^m \mid Ax=b,\ x\geqslant 0,\ s(x,y)\geqslant 0\},$$

$$F^{\circ}_{XY}\equiv\{(x,y)\in F_{XY} \mid x>0,\ s(x,y)>0\},$$

$$F_{XS}=\{(x,s)\in\mathbb{R}^n\times\mathbb{R}^n \mid (x,s)=(x,s(x,y)) \text{ for some } (x,y)\in F_{XY}\},$$

$$F^{\circ}_{XS}\equiv\{(x,s)\in F_{XS} \mid x>0,\ s>0\}.$$

In general, problem (P) might not have an optimal solution. Assumption 2.2(b) below guarantees the existence of an optimal solution for problem (P) (e.g., see Guler [7], Moré [30], Kojima, Mizuno and Noma [15], McLinden [21]).

Assumption 2.2. (a) rank $(A)=m<n$.

(b) F°_{XY} is non-empty and a point $(x_0,y_0)\in F^{\circ}_{XY}$ is given.

It follows from the discussion above and Assumption 2.2(b) that a pair $(\bar{x},\bar{y})\in F_{XY}$ exists such that $\bar{x}$ solves (P), $(\bar{x},\bar{y})$ solves (D) and $\bar{x}^{\mathrm{T}}\bar{s}=0$ where $\bar{s}\equiv s(\bar{x},\bar{y})$. Equivalently, $(\bar{x},\bar{y},\bar{s})$ satisfies the following Karush–Kuhn–Tucker optimality conditions:

$$x_i s_i = 0, \quad i = 1,\ldots,n, \tag{2.3a}$$

$$Ax = b, \quad x \geq 0, \tag{2.3b}$$

$$-\nabla f(x) + A^{\mathrm{T}} y + s = 0, \quad s \geq 0. \tag{2.3c}$$

Assumption 2.2(a) implies that the correspondence that takes $(x,y) \in F_{XY}$ into $(x, s(x,y)) \in F_{XS}$ is one-to-one and onto. Its inverse takes $(x,s) \in F_{XS}$ into $(x,y) \in F_{XY}$ where $y \equiv (AA^{\mathrm{T}})^{-1} A(\nabla f(x) - s)$. Therefore, an algorithm whose iterates all lie in the set F_{XY}° can equally well be viewed as an algorithm which generates points in the set F_{XS}°. These two ways of viewing a primal–dual algorithm with respect to problem (P) will be used interchangeably throughout our presentation.

An important concept related to problems (P) and (D) is that of the central path. For $\mu > 0$, consider the following system of equations:

$$(\mathrm{Q}(\mu)) \qquad Xs = \mu e, \tag{2.4a}$$

$$Ax = b, \quad x > 0, \tag{2.4b}$$

$$-\nabla f(x) + A^{\mathrm{T}} y + s = 0, \quad s > 0. \tag{2.4c}$$

where $X \equiv \mathrm{diag}(x)$. It can be shown that Assumption 2.1 and Assumption 2.2 imply that $(\mathrm{Q}(\mu))$ has a unique solution which we denote by $w(\mu) \equiv (x(\mu), y(\mu), s(\mu))$. The set of all solutions $w(\mu)$ with $\mu > 0$ is called the central path or, central curve. This path has the property that as μ tends to 0, the curve $(x(\mu), y(\mu), s(\mu))$ converges to the set of solutions of system (2.3). In view of this last property, many algorithms for solving problem (P) have been proposed which approximately traces the central path. These methods are denominated path-following algorithms and the first algorithm for linear programs of this type was proposed by Renegar [33]. His method is a primal algorithm in the sense that it generates points in the set $\{x \in \mathbb{R}^n_{++} \mid Ax = b\}$ that are close to the primal central trajectory $\{x(\mu) \mid \mu > 0\}$. On the other hand, PD algorithms generate points in the set F_{XY}°. Early hints towards obtaining a PD path-following algorithm were given by Megiddo [22] but it was only in Kojima, Mizuno and Yoshise [17] that a first algorithm of this type was described. Later and independently, Kojima, Mizuno and Yoshise [16] and Monteiro and Adler [26, 27] proposed another PD path following algorithm with follows the central path more closely and achieves a better theoretical computational complexity than the algorithm described in Kojima, Mizuno and Yoshise [17]. Since then many other PD path following algorithms have been proposed for different classes of problems (e.g., see [8, 10, 20, 24, 28, 38, 39, 43]).

Another fundamental idea in the development of primal–dual algorithms was the introduction of the primal–dual potential function $\phi: \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}$ defined for all $(x,s) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ as

$$\phi(x,s) \equiv q \log x^{\mathrm{T}} s - \sum_{i=1}^{n} \log x_i s_i \tag{2.5}$$

where q is a fixed scalar such that $q > n$. This function possesses a nice property. Namely, assume that $\{(x_k,s_k)\}_{k=1}^{\infty}$ is a sequence of points in the set F_{XS}° with the property that

$$\lim_{k\to\infty} \phi(x_k,s_k) = -\infty. \tag{2.6}$$

Then, it follows that $\lim_{k\to\infty} x_k^{\mathrm{T}} s_k = 0$. To see that, note that for every $(x,s) \in \mathbb{R}_{++}^{2n}$,

$$\phi(x,s) = (q-n)\log x^{\mathrm{T}}s + n\log n + n\log \frac{x^{\mathrm{T}}s/n}{(\prod_{i=1}^{n} x_i s_i)^{1/n}}$$

$$\geqslant (q-n)\log x^{\mathrm{T}}s + n\log n, \tag{2.7}$$

where the inequality follows from the fact that the arithmetic mean $x^{\mathrm{T}}s/n = (\sum_{i=1}^{n} x_i s_i)/n$ of n positive numbers $x_1 s_1,\ldots,x_n s_n$ is greater than or equal to their geometric mean $(\prod_{i=1}^{n} x_i s_i)^{1/n}$. Clearly, relations (2.6) and (2.7) imply that $\lim_{k\to\infty} x_k^{\mathrm{T}} s_k = 0$. As mentioned in the introduction, we are not able to prove relation (2.6) but relation (2.7) is used in our convergence analysis.

Algorithms that reduce the value of $\phi(x,s)$ at each iteration have been named potential reduction algorithms in Ye [41]. The function $\phi(x,s)$ was introduced independently by Tanabe [36, 37] and Todd and Ye [38]. Algorithms based on this function have been developed for various optimization problems (e.g., see [9, 14, 18, 19, 32, 41, 42]). We should single out the work of Kojima, Megiddo, Noma and Yoshise [14], where a systematic study of PD potential reduction algorithms is undertaken in the context of linear complementarity problems. It is shown there that several versions of primal–dual algorithms converge globally and that some of them also converge in polynomial-time. Polynomial-time versions of the potential reduction method are obtained by showing that the potential function (2.5) is reduced at every iteration by at least an amount $\delta > 0$ independent of k, where δ is not too small (e.g., see [18, 41]).

In spite of being different algorithms, PD path following and PD potential reduction algorithms have an important common property, namely, they use the same type of direction at each iteration. We next describe how these methods generate the direction at each iteration. Let $(x,y,s) = (x_k,y_k,s_k)$ denote the kth iterate where $(x,y) \in F_{XY}^{\circ}$ and $s \equiv s(x,y)$. Choose a scalar $\sigma = \sigma_k \in [0,1]$. Then the direction $\Delta w \equiv (\Delta x,\Delta y,\Delta s) = (\Delta x_k,\Delta y_k,\Delta s_k)$ *associated with the point* $w \equiv (x,y,s)$ is obtained by solving the following system of linear equations

$$S\Delta x + X\Delta s = \sigma(x^{\mathrm{T}}s/n)e - Xs, \tag{2.8a}$$

$$A\Delta x = 0, \tag{2.8b}$$

$$-\nabla^2 f(x)\Delta x + A^{\mathrm{T}}\Delta y + \Delta s = 0, \tag{2.8c}$$

where $X \equiv \mathrm{diag}(x)$ and $S \equiv \mathrm{diag}(s)$. This direction turns out to be the Newton direction at (x,y,s) with respect to system $(\mathrm{Q}(\mu))$ with $\mu = \sigma x^{\mathrm{T}}s/n$. Hence, the full Newton step results in the point $w + \Delta w$ for approximating the point $w(\sigma x^{\mathrm{T}}s/n)$ of the central path. Depending on the value of σ, the direction Δw has different interpretations. For $\sigma = 1$, Δw is viewed as a centering direction since $w + \Delta w$ approximates the point $w(x^{\mathrm{T}}s/n)$ which minimizes the

distance of w to the central path. Here, the "distance" of $w=(x,y,s)$ to the central path is defined by

$$\min_{\mu>0}\|Xs-X(\mu)s(\mu)\|=\min_{\mu>0}\|Xs-\mu e\|=\|Xs-(x^Ts/n)e\|.$$

where $X\equiv\mathrm{diag}(x)$ and $X(\mu)\equiv\mathrm{diag}(x(\mu))$. For $\sigma=0$, Δw is interpreted as an affine scaling direction [29] since it is the primal–dual analogue of the primal affine scaling direction used in Dikin [3]. In this case, the point $w+\Delta w$ aims at approximating the point $w(0)\equiv\lim_{\mu\to 0^+}w(\mu)$. Observe that, when $\sigma=0$, Δw is the Newton direction for the KKT system (2.3). For the case where $\sigma\in(0,1)$, Δw is a convex combination of the affine direction and the centering direction.

We next describe the issues involved in determining the next iterate $(\hat{x},\hat{y},\hat{s})=(x_{k+1},y_{k+1},s_{k+1})$. Define

$$x(\alpha)=x_k(\alpha)\equiv x_k+\alpha\Delta x_k=x+\alpha\Delta x, \tag{2.9a}$$

$$y(\alpha)=y_k(\alpha)\equiv y_k+\alpha\Delta y_k=y+\alpha\Delta y, \tag{2.9b}$$

$$s(\alpha)=s_k(\alpha)\equiv s(x_k(\alpha),y_k(\alpha)), \tag{2.9c}$$

where we recall that $s(\cdot,\cdot)$ is defined in relation (2.2). Then, the next iterate is found as $(\hat{x},\hat{y},\hat{s})=(x(\alpha_k),y(\alpha_k),s_k(\alpha_k))$ where the stepsize α_k is selected so that $(\hat{x},\hat{y})\in F^\circ_{XY}$. clearly, other conditions on α_k must be imposed so that global convergence can be obtained. This important issue will be studied in much more detail later on.

Note that within the set F°_{XY}, a primal–dual algorithm determines the next iterate by moving along the straight line $\alpha\to(x(\alpha),y(\alpha))\equiv(x,y)+\alpha(\Delta x,\Delta y)$. Considering the natural correspondence between F_{XY} and F_{XS}, we can also view the situation in the context of F°_{XS}. It is easy to derive the following expression for $s(\alpha)$ using relations (2.2), (2.8c) and (2.9):

$$s(\alpha)=s+\alpha\Delta s+[\nabla f(x+\alpha\Delta x)-\nabla f(x)-\alpha\nabla^2 f(x)\Delta x] \tag{2.10}$$

where $s\equiv s(x,y)$. Clearly, when $f(x)$ is a quadratic function, it follows that the quantity within the brackets in relation (2.10) vanishes and this shows that $\alpha\to(x(\alpha),s(\alpha))$ is a straight line. In general, however, this line is curved. Therefore, when viewed in the context of the (x,s)-space, the algorithm moves along a curve passing through the current iterate.

For the purpose of future reference, we state the following simple result.

Proposition 2.1. *The following statements hold*:

(a) *For every* $v=(x,s)\in\mathbb{R}^n_{++}\times\mathbb{R}^n_{++}$, *system* (2.8) *has a unique solution which we denote by* $\Delta w(v)\equiv(\Delta x(v),\Delta y(v),\Delta s(v))$.

(b) $\Delta w(v)$ *is a smooth function over* $\mathbb{R}^{2n}_{++}$. (*A function is smooth if it belongs to* C^∞).

(c) *Consider the curve* $w(\alpha)\equiv(x(\alpha),y(\alpha),s(\alpha))$ *defined by relations* (2.9). *Then*, $\dot{w}(0)=\Delta w(v)$.

Proof. It is easy to see that system (2.8) has an associated coefficient matrix which is both nonsingular and smooth for every $v=(x,s)\in\mathbb{R}^{2n}_{++}$. Also, the right hand side of system (2.8) is obviously smooth. These two observations imply both (a) and (b). Obviously, by (2.9a) and (2.9b), we have $\dot{x}(0)=\Delta x(v)$ and $\dot{y}(0)=\Delta y(v)$. Differentiating (2.10) with respect to α and setting $\alpha=0$, we obtain $\dot{s}(0)=\Delta s(v)$. Hence, $\dot{w}(0)=\Delta w(v)$ and (c) follows. □

We have outlined above the framework involved in a large class of primal–dual algorithms. However, a complete description of a PD algorithm from this class involves the specification of the sequence of parameters $\{\sigma_k\}$ as well as the specification of the sequence of stepsizes $\{\alpha_k\}$. By properly choosing these two sequences, many PD algorithms developed in the literature can be obtained as special cases of the above framework.

We will now outline the class of PD algorithms that will be studied in this work for solving the convex problem (P). It is in fact a subclass of the class described above and it is obtained by imposing additional restrictions on $\{\sigma_k\}$ and $\{\alpha_k\}$. By abuse of terminology, we refer to this class of algorithms as a PD algorithm, or more specifically, a potential reduction PD algorithm.

Potential reduction primal–dual algorithm (PRPD algorithm).

Step 0. Let $k=0$ and let $\varepsilon>0$ be a given tolerance for the duality gap. Let $\bar{\sigma}\in[0,1)$ be a fixed scalar. The point (x_0,y_0) as in Assumption 2.2(b) is used to initialize the algorithm.

Step 1. Let $(x,y)=(x_k,y_k)$ and $s=s(x,y)$. If $x^Ts\leqslant\varepsilon$ then stop, else go to Step 2.

Step 2. Choose a scalar $\sigma=\sigma_k\in[0,\bar{\sigma}]$. Compute the direction $\Delta w=(\Delta x,\Delta y,\Delta s)$ by solving the system of linear equations (2.8).

Step 3. Compute the stepsize $\alpha=\alpha_k>0$ in such a way that $(x(\alpha),y(\alpha),s(\alpha))$ as defined by (2.9) satisfies the Armijo stepsize rule described below.

Step 4. Set $(x_{k+1},y_{k+1},s_{k+1})=(x(\alpha),y(\alpha),s(\alpha))$, replace k by $k+1$ and go to Step 1.

The only freedom involved in the above class of algorithms is in the selection of the sequence $\{\sigma_k\}$ in Step 2. If a rule for selecting $\{\sigma_k\}$ is specified then the above 5 steps describe a well-defined algorithm. The sequence of stepsizes $\{\alpha_k\}$ is well-determined by the Armijo rule to be described below. At the end of this section, we state a global convergence result that holds for the above subclass of algorithms. It is interesting at this point to compare our class of algorithms with the one considered by Kojima, Megiddo and Mizuno in the paper [12]. They proved a global convergence result for a class of primal–dual algorithms for linear programs that allows σ to be chosen with the same degree of freedom as in Step 2 above but that uses a different stepsize selection rule. The stepsize selection rule used in [12] seems to be applicable only in the context of linear programs although we can not make any definite claim in this respect. Our stepsize rule is based on the idea of producing a sufficient decrease on the potential function (2.5) and this might be the main reason that makes the rule successful for the more general class of convex programming problems.

We will now discuss our stepsize selection rule in detail. The choice of the stepsize employs a version of Armijo rule in order to obtain a sufficient decrease on the value of the potential function (2.5). Therefore, here as well as in several other papers, the potential function is used in the algorithm as a merit (or, descent) function.

Armijo rule. Let $\mu \in (0,1)$ and $\beta > 1$ be fixed parameters. Let $\{\alpha_k^\circ\}_{k=1}^\infty$ be a sequence of positive scalars to be specified more precisely later. The scalar α_k° will be used as an initial trial value for the kth stepsize α_k. Given the k-iterate $(x,y) = (x_k,y_k) \in F_{XY}^\circ$, *and hence*, $(x,s) \in F_{XS}^\circ$ where $s \equiv s(x,y)$, consider the curve $v(\alpha) \equiv (x(\alpha), s(\alpha))$ defined by relations (2.9). Set $\alpha_k \equiv \alpha_k^\circ \beta^{-m_k}$ where m_k is the smallest nonnegative integer m for which

$$v(\alpha_k^\circ \beta^{-m}) \in \mathbb{R}_{++}^{2n}$$

and

$$\phi(v(\alpha_k^\circ \beta^{-m})) - \phi(v(0)) \leqslant \mu \alpha_k^\circ \beta^{-m} \nabla \phi(v(0))^{\mathrm{T}} \dot{v}(0). \tag{2.11}$$

We will show in Section 4 that the quantity $\nabla\phi(v(0)^{\mathrm{T}}\dot{v}(0)) = (\mathrm{d}/\mathrm{d}\alpha)(\phi \circ v)(0)$ that appears in (2.11) is negative so that Armijo rule guarantees that the potential function $\phi(\cdot)$ is reduced at each iteration (see Corollary 4.1 and the discussion following the statement of the Armijo rule in the more general context of Section 3). The description of the Armijo stepsize rule will be complete once we specify the sequence $\{\alpha_k^\circ\}$ of initial trial stepsizes. We describe below two possible ways of choosing this sequence.

Rule 1. Let $\bar{\alpha} > 0$ be a fixed parameter and set $\alpha_k^\circ \equiv \bar{\alpha}$ for all $k \in \mathbb{N}$.

Rule 2. Let $\eta > 0$ be a fixed parameter. Define $\alpha_k^{\max} \equiv \sup\{\alpha \geqslant 0 \mid x_k + \alpha \Delta x_k \geqslant 0, s_k + \alpha \Delta s_k \geqslant 0\}$ and choose α_k° so that $\alpha_k^\circ \geqslant \eta \alpha_k^{\max}$, for all $k \in \mathbb{N}$.

The following simple result shows that $\alpha_k^{\max}$ is finite for all $k \in \mathbb{N}$ so that Rule 2 is always feasible.

Proposition 2.2. *For all $k \in \mathbb{N}$, we have $0 < \alpha_k^{\max} < \infty$.*

Proof. Clearly $\alpha_k^{\max} > 0$ since $(x_k,s_k) \in \mathbb{R}_{++}^{2n}$. To prove that $\alpha_k^{\max} < \infty$, it is sufficient to show that some component of $(\Delta x_k, \Delta s_k)$ is negative. Assume, by contradiction that $(\Delta x_k, \Delta s_k) \geqslant 0$. This together with the fact that $(\Delta x_k, \Delta s_k)$ satisfies (2.8a) with $x = x_k$, $s = s_k$ and $\sigma = \sigma_k$ imply that

$$0 \leqslant e^{\mathrm{T}}[\sigma_k (x_k^{\mathrm{T}} s_k / n) e - X_k s_k] = (\sigma_k - 1) x_k^{\mathrm{T}} s_k < 0$$

where $X_k \equiv \mathrm{diag}(x_k)$. The last inequality is due to the fact that $\sigma_k < 1$. We have obtained a contradiction and therefore the result follows. □

We are now in a position to state the main result of this work.

Theorem 2.1. *Assume that the sequence* $\{(x_k,y_k,s_k)\}_{k=1}^{\infty}$ *is generated by the PRPD Algorithm where the sequence of stepsizes* $\{\alpha_k\}_{k=1}^{\infty}$ *is selected according to Armijo rule and the sequence of initial stepsizes* $\{\alpha_k^\circ\}_{k=1}^{\infty}$ *is selected by means of either Rule* 1 *or Rule* 2 *described above. Then, the following two statements hold:*

(a) $\lim_{k\to\infty} x_k^{\mathrm{T}} s_k = 0$, *and;*

(b) *The sequence* $\{(x_k,y_k,s_k)\}_{k=1}^{\infty}$ *has a limit point and if* $(\bar{x},\bar{y},\bar{s})$ *is one such limit point then* $\bar{x}$ *solves* (P), $(\bar{x},\bar{y})$ *solves* (D) *and* $\bar{x}^{\mathrm{T}}\bar{s}=0$.

Our main objective from now on will be to provide a proof of Theorem 2.1. In Section 3, we analyze Armijo rule in a broader context where a general function defined in an open set is considered and where the iterates can be obtained by moving along curves instead of straight lines. The results obtained in Section 3 is then used in Section 4 to provide a proof of Theorem 2.1.

3. Analysis of the Armijo rule

In the previous section, we saw that when the PRPD Algorithm is viewed on the (x,s)-space, the next iterate is selected from among points of a curve passing through the current iterate. The next iterate is then selected by means of the Armijo stepsize rule applied to the potential function (2.5) whose domain of definition is an open subset of $\mathbb{R}^{2n}$, namely, the positive orthant $\mathbb{R}^{2n}_{++}$. The objective of the present section is to present the necessary machinery in order to understand Armijo rule when applied to the above situation. This will be done by considering a general function $h: D\subseteq\mathbb{R}^n\to\mathbb{R}$ and a general iterative algorithm in which the next iterate $x_{k+1}\in D$ is obtained from the current iterate $x_k\in D$ through a recurrence relation of the form $x_{k+1}=x_k(\alpha_k)$ where $x_k(\cdot)$ is a curve such that $x_k(0)=x_k$. We note that the notation used in this section is independent of the other sections so that the reader may see here a notation referring to an entity different from the one it used to refer in the other sections.

Before stating the main result, we first need to introduce the necessary conditions on the function $h(x)$ and on the iterative algorithm. Regarding the function $h(x)$, we make the following assumption throughout this section.

Assumption 3.1. The function $h: D\to\mathbb{R}$ is continuously differentiable on the open set $D\subseteq\mathbb{R}^n$.

In analogy with the definition of a gradient method (or, gradient algorithm) defined in several textbooks (e.g., see Bertsekas [1, p. 20]), we define the following more general notion of a gradient method.

Definition 3.1. A sequence $\{x_k\}_{k=1}^{\infty}$ is said to be generated by a gradient method with respect to the function $h(x)$ if given $x_k\in D$, we find the next iterate x_{k+1} as

$$x_{k+1} = x_k(\alpha_k)$$

where $x_k : I_k = [0, \eta_k) \to \mathbb{R}$ is a curve differentiable at 0 such that $x_k(0) = x_k$ and

$$x_k(\alpha) \equiv x_k \text{ for all } \alpha \in I_k, \quad \text{if } \nabla h(x_k) = 0,$$
$$\nabla h(x_k)^T \dot{x}_k(0) < 0, \quad \text{if } \nabla h(x_k) \neq 0$$

and the selection of the stepsize $\alpha_k \in [0, \eta_k)$ must ensure that $x_{k+1} \in D$ and $h(x_{k+1}) \leqslant h(x_k)$. A sequence with the above properties will be called a gradient sequence with respect to $h(\cdot)$.

We note that we do not require that $x_k(I_k) \subseteq D$ in the above definition. Clearly, it follows from the definition that if $\nabla h(x_k) = 0$ for some $k \in \mathbb{N}$ then $x_j = x_k$ for all $j > k$. In order to obtain relevant results regarding a gradient method, an additional notion regarding the curves $x_k(\alpha)$ is needed.

Definition 3.2. Consider a sequence of curves $\{x_k : [0, \eta_k) \to \mathbb{R}^n\}_{k=1}^{\infty}$ as in Definition 3.1. The subsequence $\{x_k(\cdot)\}_{k \in K}$ is said to be uniformly differentiable at 0 if the following two conditions hold:

(a) $\delta_K \equiv \inf_{k \in K} \|\dot{x}_k(0)\| \eta_k > 0$.

(b) For all $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) \in [0, \delta_K)$ such that for all $k \in K$,

$$\|x_k(\alpha) - x_k(0) - \alpha \dot{x}_k(0)\| \leqslant \varepsilon \alpha \|\dot{x}_k(0)\|$$

for all $\alpha \in \mathbb{R}$ such that $\alpha \|\dot{x}_k(0)\| \in [0, \delta)$.

Several observations regarding the above definition are in order. First, the terminology of uniform differentiability introduced above might have already been used in the literature in connection with a notion distinct from the one above. However, the author is not aware of any such case. Second, condition (a) of Definition 3.2 implies that $\dot{x}_k(0) \neq 0$ for all $k \in K$. Third, condition (a) of Definition 3.2 is automatically satisfied if $I_k \equiv \mathbb{R}_+$ and $\dot{x}_k(0) \neq 0$ for all $k \in K$. Fourth, if $I_k = \mathbb{R}_+$ and $x_k(\alpha)$ is an affine function of α such that $\dot{x}_k(0) \neq 0$, for all $k \in \mathbb{N}$, then any subsequence of $\{x_k(\cdot)\}_{k=1}^{\infty}$ satisfies the notion of uniform differentiability at 0. Fifth, observe that $x_k(\alpha)$ is well-defined, that is, $\alpha \in I_k$, whenever α satisfies $\alpha \|\dot{x}_k(0)\| \in [0, \delta_K)$. Sixth, let $\{\tau_k\}_{k=1}^{\infty}$ be a sequence of positive scalars and assume that the subsequence $\{x_k : I_k \equiv [0, \eta_k) \to \mathbb{R}^n\}_{k \in K}$ is uniformly differentiable at 0. Define the curves $y_k : J_k \equiv [0, \tilde{\eta}_k) \to \mathbb{R}^n$ where, for all $k \in \mathbb{N}$, $\tilde{\eta}_k \equiv \eta_k / \tau_k$ and $y_k(\alpha) \equiv x_k(\tau_k \alpha)$ for all $\alpha \in J_k$. Then, one can easily verify that $\{y_k\}_{k \in K}$ is uniformly differentiable at 0. In other words, the notion of uniform differentiability is invariant under a sequence of changes of variable of the form $\alpha \to \tau_k \alpha$, $k \in \mathbb{N}$. However, the notion of uniform differentiability may not be invariant under

scalar multiplications of the form $z_k(\alpha) \equiv \tau_k x_k(\alpha)$, $k \in \mathbb{N}$. Seventh, in addition to the above conditions, assume also that $\tau_k \equiv \|\dot{x}_k(0)\|^{-1}$ for all $k \in \mathbb{N}$. In this case, it is easy to see that $\{y_k(\cdot)\}_{k=1}^{\infty}$ as defined above satisfies $\|\dot{y}_k(0)\| = 1$ for all $k \in \mathbb{N}$ and that the notion of uniform differentiability reduces to the following two simpler conditions:

(a′) $\tilde{\delta}_K \equiv \inf_{k \in K} \tilde{\eta}_k > 0$.

(b′) For all $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) \in [0, \tilde{\delta}_K)$ such that for all $k \in K$,

$$\|y_k(\alpha) - y_k(0) - \alpha \dot{y}_k(0)\| \leq \varepsilon\alpha \quad \text{for all } \alpha \in [0, \delta).$$

Lastly, it can also be shown that if $0 < \inf_{k \in \mathbb{N}} \|\dot{x}_k(0)\| \leq \sup_{k \in \mathbb{N}} \|\dot{x}_k(0)\| < \infty$ then $\{x_k(\cdot)\}_{k \in K}$ is uniformly differentiable at 0 if and only if $\{x_k(\cdot)\}_{k \in K}$ satisfies (a′) and (b′) above with $y_k(\cdot)$ replaced by $x_k(\cdot)$.

The following result is a consequence of Definition 3.2 and will be useful later in the proof of Theorem 3.1 and Theorem 3.2. For the purpose of stating this result, we introduce the following notation. For an infinite set $K \subseteq \mathbb{N}$ and an integer $\bar{k} \in \mathbb{N}$, let $K(\bar{k})$ denote the set

$$K(\bar{k}) \equiv \{k \in K \mid k \geq \bar{k}\}. \tag{3.1}$$

Lemma 3.1. *Consider a sequence of curves* $\{x_k : I_k \to \mathbb{R}^n\}_{k=1}^{\infty}$ *as in Definition* 3.1. *Assume that the subsequence* $\{x_k(\cdot)\}_{k \in K}$ *is uniformly differentiable at* 0 *and let* $\{\tau_k\}_{k \in K}$ *be a subsequence of positive scalars such that* $\lim_{k \in K} \tau_k \|\dot{x}_k(0)\| = 0$. *Then the following statements hold*:

(a) *there exists* $\bar{k} \in \mathbb{N}$ *such that* $\tau_k \in I_k$ *for all* $k \in K(\bar{k})$;

(b) $\lim_{k \in K(\bar{k})} \|x_k(\tau_k) - x_k(0)\| = 0$;

(c) $\lim_{k \in K(\bar{k})} \|\Delta x_k(\tau_k) - \dot{x}_k(0)\| / \|\dot{x}_k(0)\| = 0$ *where* $\Delta x_k(\tau) \equiv [x_k(\tau) - x_k(0)]/\tau$, *for* $\tau \in I_k$;

(d) *in addition, if we assume that there exists a compact subset* G *of the open set* D *such that* $\{x_k(0)\}_{k \in K} \subseteq G$ *then there exists* $\hat{k} \geq \bar{k}$ *such that* $x_k(\tau_k) \in D$ *for all* $k \in K(\hat{k})$.

Proof. Since $\lim_{k \in K} \tau_k \|\dot{x}_k(0)\| = 0$, it follows that there exists $\bar{k} \in \mathbb{N}$ such that $\tau_k \|\dot{x}_k(0)\| < \delta_K$, for all $k \in K(\bar{k})$, where $\delta_K \equiv \inf_{k \in K} \|\dot{x}_k(0)\| \eta_k > 0$ (see Definition 3.2(a)). Hence, it follows that $\tau_k < \eta_k$ for all $k \in K(\bar{k})$ and this shows statement (a). We leave the proof of (b) and (c) to the reader. Statement (d) is an immediate consequence of (b) and the fact that the distance $\operatorname{dist}(G, \mathbb{R}^n \setminus D)$ between the sets G and the complement $\mathbb{R}^n \setminus D$ of D with respect to $\mathbb{R}^n$, that is, $\operatorname{dist}(G, \mathbb{R}^n \setminus D) \equiv \inf\{\|x - y\| \mid x \in G, y \in \mathbb{R}^n \setminus D\}$, is positive due to the fact that G is compact and D is open. □

We next state Armijo rule with respect to the function $h(x)$ and a general gradient algorithm as defined above.

Armijo stepsize rule. Select fixed parameters $\mu \in (0,1)$ and $\beta > 1$. Assume that a sequence $\{\alpha_k^\circ\}_{k=1}^\infty$ of positive scalars is given. The scalar α_k° will be used as an initial trial value for the kth stepsize α_k. We say that a gradient sequence $\{x_k\}_{k=1}^\infty$ with respect to $h(\cdot)$ has been generated according to Armijo stepsize rule if for all $k \in \mathbb{N}$, the stepsize α_k is determined as follows: $\alpha_k \equiv \alpha_k^\circ \beta^{-m_k}$ where m_k is the smallest nonnegative integer m for which $\alpha \equiv \alpha_k^\circ \beta^{-m}$ satisfies

$$\alpha \in I_k,$$

$$x_k(\alpha) \in D$$

and

$$h(x_k(\alpha)) - h(x_k) \leqslant \mu\alpha \nabla h(x_k)^T d_k$$

where $d_k \equiv \dot{x}_k(0)$.

The existence of the integer m_k as in the above definition follows from the following remark: for each $k \in \mathbb{N}$, there exists a positive scalar ε_k such that any $\alpha \in [0, \varepsilon_k]$ satisfies the above three relations. This observation follows immediately once we note that $[h(x_k(\alpha)) - h(x_k)]/\alpha$ converges to $h(x_k)^T d_k$ as α tends to 0 and use the fact that $\mu \in (0,1)$ and $\nabla h(x_k)^T d_k < 0$.

We next state one of the main results of this section. We note that, for this result, we do not need to impose any condition on the way the sequence $\{\alpha_k^\circ\}_{k=1}^\infty$ is selected. However, as we will see later, significant and applicable results are obtained only when we choose this sequence appropriately.

Theorem 3.1. *Let $h: D \subseteq \mathbb{R}^n \to \mathbb{R}$ satisfy Assumption 3.1 and assume that $\{x_k\}_{k=1}^\infty$ is a gradient sequence with respect to $h(x)$ that has been generated according to Armijo stepsize rule. Let $K \subseteq \mathbb{N}$ be an infinite index set and let G be a compact subset of D. Assume that the subsequence $\{x_k(\cdot)\}_{k \in K}$ is uniformly differentiable at 0. If we let $d_k \equiv \dot{x}_k(0)$ for all $k \in \mathbb{N}$ then the following two conditions:*

$$\{x_k\}_{k \in K} \subseteq G, \tag{3.2}$$

$$\inf_{k \in K} |\nabla h(x_k)^T d_k| / \|d_k\| > 0, \tag{3.3}$$

imply

$$\lim_{k \in K} \alpha_k^\circ \|d_k\| = 0. \tag{3.4}$$

Proof. By Armijo rule, the sequence $\{h(x_k)\}_{k=1}^\infty$ is monotonically decreasing. This fact together with relation (3.2) and the fact that G is compact imply that $\{h(x_k)\}_{k=1}^\infty$ converges, and consequently,

$$\lim_{k\to\infty} h(x_{k+1}) - h(x_k) = 0. \tag{3.5}$$

By Armijo rule, we obtain

$$h(x_{k+1}) - h(x_k) \leqslant \mu\alpha_k \nabla h(x_k)^T d_k = -\mu\rho_k \alpha_k \|d_k\| \tag{3.6}$$

where we define $\rho_k \equiv |\nabla h(x_k)^T d_k| / \|d_k\|$ for all $k \in \mathbb{N}$. Then, it follows from (3.5) and (3.6) that $\lim_{k\to\infty} \rho_k \alpha_k \|d_k\| = 0$. This fact together with (3.3) imply

$$\lim_{k\in K} \alpha_k \|d_k\| = 0. \tag{3.7}$$

We claim that $\alpha_k = \alpha_k^\circ$ for all $k \in K$ sufficiently large. Note that the validity of this claim together with (3.7) proves (3.4) as well as the theorem. We show the claim by contradiction. Indeed, assume that $\alpha_k < \alpha_k^\circ$ for all $k \in K'$ where K' is an infinite subset of K. Define $\tilde{\alpha}_k \equiv \beta\alpha_k$ for all $k \in \mathbb{N}$. From the way the stepsize α_k is determined by Armijo rule, for all $k \in K'$ it follows that either

$$\text{both } \tilde{\alpha}_k \in I_k \text{ and } x_k(\tilde{\alpha}_k) \in D \text{ can not hold,} \tag{3.8}$$

or

$$h(x_k(\tilde{\alpha}_k)) - h(x_k) > \mu\tilde{\alpha}_k \nabla h(x_k)^T d_k. \tag{3.9}$$

Relation (3.7) and the definition of $\tilde{\alpha}_k$ clearly imply that $\lim_{k\in K} \tilde{\alpha}_k \|d_k\| = 0$ which together with the uniform differentiability of the subsequence $\{x_k(\cdot)\}_{k\in K}$ at 0 show that the assumptions of Lemma 3.1 are satisfied. Hence, from statements (a) and (d) of Lemma 3.1, it follows that, for some $\hat{k} \in \mathbb{N}$, $\tilde{\alpha}_k \in I_k$ and $x_k(\tilde{\alpha}_k) \in D$ for every $k \in K'(\hat{k})$ (see (3.1)), or equivalently, (3.8) is violated for every $k \in K'(\hat{k})$. Thus, (3.9) must hold for every $k \in K''$ where $K'' \equiv K'(\hat{k})$. Note also that statements (b) and (c) of Lemma 3.1 imply

$$\lim_{k\in K''} \|x_k(\tilde{\alpha}_k) - x_k\| = 0 \tag{3.10}$$

and

$$\lim_{k\in K''} \| r_k - d_k\| / \|d_k\| = 0 \quad \text{where } r_k \equiv [x_k(\tilde{\alpha}_k) - x_k] / \tilde{\alpha}_k. \tag{3.11}$$

Applying the mean value theorem on the left hand side of (3.9), we obtain for all $k \in K''$ that

$$\nabla h(u_k)^T r_k > \mu \nabla h(x_k)^T d_k, \tag{3.12}$$

where μ_k is some point in the line segment joining x_k to $x_k(\tilde{\alpha}_k)$ and hence, $\lim_{k\in K''} \|u_k - x_k\| = 0$ due to (3.10). Using Cauchy–Schwartz inequality, relation (3.12) and the fact that $\nabla h(x_k)^T d_k < 0$ for all $k \in \mathbb{N}$, we obtain that, for all $k \in K''$,

$$\nabla h(u_k)^T (r_k - d_k) / \|d_k\| + \|\nabla h(u_k) - \nabla h(x_k)\|$$

$$\geq [\nabla h(u_k)^T(r_k - d_k) + [\nabla h(u_k) - \nabla h(x_k)]^T d_k]/\|d_k\|$$
$$= [\nabla h(u_k)^T r_k - \nabla h(x_k)^T d_k]/\|d_k\|$$
$$> (1-\mu)\,|\nabla h(x_k)^T d_k|/\|d_k\|$$
$$= (1-\mu)\rho_k. \tag{3.13}$$

Using the fact that G is compact, that $\nabla h(x)$ is continuous in $D \supseteq G$, that $\lim_{k\in K''}\|u_k - x_k\| = 0$ and relation (3.2), one can easily show that $\{\|\nabla h(u_k)\|\}_{k\in K''}$ is bounded and that $\lim_{k\in K''}\|\nabla h(u_k) - \nabla h(x_k)\| = 0$. These facts together with (3.11) imply that the first expression in (3.13) converges to 0 as $k \in K''$ tends to ∞. This, in turn, implies that $\lim_{k\in K''}\rho_k = 0$ contradicting (3.3). Hence, the claim must hold and the theorem is proved. □

For the purpose of stating the implications of Theorem 3.1, we first need to introduce some conditions on the way the sequence $\{\alpha_k^\circ\}_{k=1}^\infty$ is chosen. Let $\eta_i > 0$ be fixed scalars, $i = 1,2,3,4$. Consider an index $k \in \mathbb{N}$. We consider the following four conditions for choosing α_k°:

(1) choose α_k° so that $\alpha_k^\circ \geq \eta_1$, or;

(2) define $J_k \equiv \mathbb{R}_+ \setminus \{\alpha \geq 0 \mid x_k + \alpha d_k \in D\}$ where $d_k = \dot{x}_k(0)$ and let $\alpha_k^{\max} \equiv \inf J_k$; choose α_k° so that $\alpha_k^\circ \geq \eta_2 \alpha_k^{\max}$, or;

(3) define $\tilde{J}_k \equiv \mathbb{R}_+ \setminus \{\alpha \mid \alpha \in I_k, x_k(\alpha) \in D\}$ and let $\tilde{\alpha}_k^{\max} \equiv \inf \tilde{J}_k$; choose α_k° so that $\alpha_k^\circ \geq \eta_3 \tilde{\alpha}_k^{\max}$, or;

(4) choose α_k° so that $\alpha_k^\circ \geq \eta_4 \|d_k\|^{-1}$.

We note that conditions (2) or (3) can be used only if $\alpha_k^{\max} < \infty$ or $\tilde{\alpha}_k^{\max} < \infty$, respectively. (We are assuming that the infimum of the empty set is ∞.) Otherwise, either conditions (1) or (4) should be used. In the result below, we do not specify any rule for deciding between the above conditions. We allow different conditions to be used at distinct iterations of the algorithm. However, for those indices $k \in \mathbb{N}$ for which condition (1) is used, we need to assume that the norm of $d_k \equiv \dot{x}_k(0)$ is bounded away from 0. We use the following notation in the theorem below. For $K \in \mathbb{N}$ and $i \in \{1,2,3,4\}$, let $K_i \subseteq K$ denote the set of those indices $k \in K$ for which the ith condition is used for the selection of α_k°.

Theorem 3.2. *Let the same assumptions of Theorem* 3.1 *hold. Assume that α_k° is selected according to either conditions* (1), (2), (3) *or* (4) *above. Then, the following three relations are incompatible.*

$$\{x_k\}_{k\in K} \subseteq G, \tag{3.14}$$

$$\inf_{k\in K} |\nabla h(x_k)^T d_k|/\|d_k\| > 0, \tag{3.15}$$

$$\inf_{k\in K_1} \|d_k\| > 0. \tag{3.16}$$

Proof. Assume that both (3.14) and (3.15) hold and let us show that (3.16) can not hold. By Theorem 3.1, we know that (3.14) and (3.15) imply that $\lim_{k\in K}\alpha_k^\circ \|d_k\| = 0$. Since K

is infinite and $K=K_1\cup K_2\cup K_3\cup K_4$, it follows that one of the sets K_i with $i\in\{1,2,3,4\}$ is infinite. These two observations imply the existence of an index $i\in\{1,2,3,4\}$ satisfying

$$K_i \text{ is infinite,} \tag{3.17a}$$

$$\lim_{k\in K_i} \alpha_k^\circ \|d_k\| = 0. \tag{3.17b}$$

We next show that any $i\in\{2,3,4\}$ does not satisfy (3.17). We first show that $i=2$ does not satisfy (3.17). Indeed, let $k\in K_2$ be given. Using the definition of $\alpha_k^{\max}$ (see condition (2)) and the fact that D is an open set, we conclude that $y_k\equiv x_k+\alpha_k^{\max}d_k\in\mathbb{R}^n\setminus D$. Since $x_k\in G$ and $\operatorname{dist}(G,\mathbb{R}^n\setminus D)=\zeta>0$, we obtain

$$\alpha_k^\circ\|d_k\|\geqslant\eta_2\alpha_k^{\max}\|d_k\|=\eta_2\|y_k-x_k\|\geqslant\eta_2\zeta,\quad \text{for all } k\in K_2,$$

where the first inequality is due to the fact that $\alpha_k^\circ\geqslant\eta_2\alpha_k^{\max}$ (see condition (2)). Hence, $i=2$ does not satisfy (3.17). Next, we show that $i=3$ does not satisfy (3.17). Indeed, assume by contradiction that (3.17) holds for $i=3$. Since $\alpha_k^\circ\geqslant\eta_3\tilde{\alpha}_k^{\max}$ for all $k\in K_3$ (see condition (3)), we obtain $\lim_{k\in K_3}\tilde{\alpha}_k^{\max}\|d_k\|=0$. Hence, the assumptions of Lemma 3.1 are satisfied with $\tau_k\equiv\tilde{\alpha}_k^{\max}$ for all $k\in K_3$. By (a) and (d) of Lemma 3.1, it follows that there exists $\hat{k}\in\mathbb{N}$ such that $\tilde{\alpha}_k^{\max}\in\{\alpha\,|\,\alpha\in I_k, x_k(\alpha)\in D\}$, that is $\tilde{\alpha}_k^{\max}\notin\tilde{J}_k$, for all $k\in K_3(\hat{k})$. However, it is easily seen from the definition of $\tilde{\alpha}_k^{\max}$ that $\tilde{\alpha}_k^{\max}\in\tilde{J}_k$ for all $k\in K_3$. Hence, we obtain a contradiction and therefore (3.17) does not hold for $i=3$. Clearly, (3.17) can not hold for $i=4$ since $\alpha_k^\circ\geqslant\eta_4\|d_k\|^{-1}$ for all $k\in K_4$ (see condition (4)). Hence, (3.17) must hold for $i=1$. This implies that $\lim_{k\in K_1}\|d_k\|=0$ since $\alpha_k^\circ\geqslant\eta_1$ for all $k\in K_1$ (see condition (1)). Therefore, (3.16) does not hold. □

The two rules described after the Armijo rule in Section 2 for selecting α_k° fit into the framework of the above conditions. Rule 1 imposes condition (1) at every iteration while Rule 2 imposes condition (2) at every iteration. Observe that if Rule 2 is used then K_1 is empty for all $K\subseteq\mathbb{N}$, and consequently, (3.16) is always true. Thus, in this case, Theorem 3.2 reduces to the incompatibility of (3.14) and (3.15) only. The following special case of Theorem 3.2 is in fact the result we will use later to prove Theorem 2.1.

Corollary 3.1. *Let the same assumptions of Theorem* 3.1 *hold and assume that* α_k° *is selected according to either conditions* (1), (2), (3) *or* (4). *Then, the following three conditions are incompatible.*

$$\{x_k\}_{k\in K}\subseteq G, \tag{3.18}$$

$$\sup_{k\in K}\|d_k\|<\infty \tag{3.19}$$

and

$$\inf_{k\in K}|\nabla h(x_k)^{\mathrm{T}}d_k|>0 \tag{3.20}$$

Proof. We will show that the validity of (3.18), (3.19) and (3.20) implies the validity of (3.14), (3.15) and (3.16). Hence, by Theorem 3.2, relations (3.18), (3.19) and (3.20) must be incompatible which proves the result. Indeed, it is obvious that (3.18), (3.19) and (3.20) imply (3.14) and (3.15). Relation (3.18) together with the continuity of $\nabla h(\cdot)$ over $D \supseteq G$ imply that $M \equiv \sup_{k \in K} \| \nabla h(x_k) \| < \infty$. By Cauchy-Schwartz inequality, it follows that $\|d_k\| \geq |\nabla h(x_k)^T d_k| / \|\nabla h(x_k)\| \geq |\nabla h(x_k)^T d_k| / M$, for all $k \in K$. This and relation (3.20) then imply $\inf_{k \in K} \|d_k\| > 0$ and, in particular, relation (3.16). □

4. Proof of the main result

In this section, we give the proof of Theorem 2.1. The proof of this result will follow as a consequence of several preliminary results, some of which were presented in Section 3. The other results are presented in this section.

The next result is a consequence of more powerful theorems obtained in other papers (e.g., see Guler [7], Kojima, Mizuno and Noma [15], McLinden [21]). However, for the sake of completeness, we include a more direct and simpler proof of this result here. Consider the mapping $J: \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}^n_+$ defined as $J(x,s) = (x_1 s_1, \ldots, x_n s_n)$ for all $(x,s) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ and let $\tilde{J}: F_{XS} \to \mathbb{R}^n_+$ denote the restriction of J to the set F_{XS}.

Lemma 4.1. *If $C \subseteq \mathbb{R}^n_+$ is a compact set then $\tilde{J}^{-1}(C) \subseteq F_{XS}$ is also a compact set.*

Proof. By Assumption 2.2(b), there exists a point $(x_0, s_0) \in F^{\circ}_{XS}$. Now using the definition of F_{XS} it is easy to verify that for all $(x,s) \in F_{XS}$,

$$(s - s_0)^T (x - x_0) = (\nabla f(x) - \nabla f(x_0))^T (x - x_0) \geq 0,$$

where the last inequality is due to the convexity of $f(x)$ (see Assumption 2.1(a)). Hence, it follows that

$$(s_0)^T x + (x_0)^T s \leq (x_0)^T s_0 + x^T s \tag{4.1}$$

for all $(x,s) \in F_{XS}$. Since C is bounded, it follows from the definition of $\tilde{J}$ that the set $\{x^T s \mid (x,s) \in \tilde{J}^{-1}(C)\}$ is bounded. Since $(x_0, s_0) > 0$, relation (4.1) implies that $\tilde{J}^{-1}(C)$ is also bounded. It is trivial to verify that F_{XS} is a closed subset of $\mathbb{R}^{2n}$. Since $\tilde{J}^{-1}(C)$ is obviously closed with respect to F_{XS}, it follows that $\tilde{J}^{-1}(C)$ is also a closed subset of $\mathbb{R}^{2n}$. Hence, $\tilde{J}^{-1}(C)$ is both bounded and closed as a subset of $\mathbb{R}^{2n}$ and hence, it is compact. □

Lemma 4.2. *Consider the function $P: \mathbb{R}^n_{++} \to \mathbb{R}$ defined by*

$$P(u) \equiv q \log e^T u - \sum_{i=1}^{n} \log u_i \tag{4.2}$$

for all $u \in \mathbb{R}^n_{++}$, where $q > n$ is a fixed scalar. Then, for all $\gamma > 0$ and $\eta \in \mathbb{R}$, the set

$$L(\gamma,\eta)\equiv\{u\in\mathbb{R}^n_{++}\mid e^Tu\geq\gamma \text{ and } P(u)\leq\eta\} \tag{4.3}$$

is a compact subset of $\mathbb{R}^n_{++}$.

Proof. We first claim that there exist scalars δ_1, $\delta_2\in\mathbb{R}_{++}$ such that $L(\gamma,\eta)\subseteq[\delta_1,\delta_2]^n$. (Here, by convention, if $\delta_1>\delta_2$ then $[\delta_1,\delta_2]\equiv\emptyset$. The possibility that $\delta_1>\delta_2$ exists only if $L(\gamma,\eta)=\emptyset$.) Indeed, by using the same argument used to derive (2.7), we obtain

$$P(u)\geq(q-n)\log e^Tu+n\log n. \tag{4.4}$$

For every $u\in L(\gamma,\eta)$, we have $P(u)\leq\eta$ and hence relation (4.4) implies

$$\delta_2\equiv\exp[(\eta-n\log n)/(q-n)]\geq e^Tu.$$

Hence, for every $u\in L(\gamma,\eta)$ and every $i=1,\ldots,n$, we have $u_i\leq\delta_2$. Using this fact and expressions (4.2) and (4.3), we obtain that for every $i=1,\ldots,n$ and $u\in L(\gamma,\eta)$,

$$u_i=\exp\left[q\log e^Tu-P(u)-\sum_{j\neq i}\log u_j\right]$$

$$\geq\exp[q\log\gamma-\eta-(n-1)\log\delta_2]\equiv\delta_1.$$

With the scalars δ_1 and δ_2 as defined above, the claim obviously follows. Since the inverse image of a closed set under a continuous function is again a closed set, it follows that $L(\gamma,\eta)$ is closed with respect to $\mathbb{R}^n_{++}$. Since the compact set $[\delta_1,\delta_2]^n\subseteq\mathbb{R}^n_{++}$ contains $L(\gamma,\eta)$, it follows that $L(\gamma,\eta)$ is compact. □

As a consequence of the previous two lemmas, we obtain the following result.

Lemma 4.3. *For every scalars* $\gamma>0$ *and* $\eta\in\mathbb{R}$, *the set*

$$\tilde{L}(\gamma,\eta)\equiv\{(x,s)\in F^\circ_{XS}\mid x^Ts\geq\gamma \text{ and } \phi(x,s)\leq\eta\} \tag{4.5}$$

is compact.

Proof. The result easily follows by Lemma 4.1, Lemma 4.2 and the fact that $\tilde{L}(\gamma,\eta)=\tilde{J}^{-1}(L(\gamma,\eta))$. □

In order to apply Corollary 3.1 in the proof of Theorem 2.1, it is necessary first to derive bounds on the quantity $(d/d\alpha)(\phi\circ v)(0)=\nabla\phi(v(0)^T\dot{v}(0)$ that appears on the right hand side of (2.11) (see (3.20) of Corollary 3.1). First we prove the following lemma.

Lemma 4.4. *Consider the function* $P(u)$ *defined in* (4.2) *and define*

$$V(u)=V(u,\sigma)\equiv\sigma(e^Tu/n)e-u \quad \text{for all } u\in\mathbb{R}^n. \tag{4.6}$$

Then, for all $u\in\mathbb{R}^n_{++}$,

$$\nabla P(u)^{\mathrm{T}} V(u) \leqslant -(1-\sigma)(q-n).$$

Proof. We have $\nabla P(u) = (q/e^{\mathrm{T}}u)e - U^{-1}e$, where $U \equiv \mathrm{diag}(u)$. Hence,

$$\begin{aligned}
\nabla P(u)^{\mathrm{T}} V(u) &= [(q/e^{\mathrm{T}}u)e - U^{-1}e]^{\mathrm{T}}\,[\sigma(e^{\mathrm{T}}u/n)e - u] \\
&= q\sigma - q - \frac{\sigma}{n}\left(\sum_{i=1}^{n} u_i\right)\left(\sum_{i=1}^{n} u_i^{-1}\right) + n \\
&\leqslant q\sigma - q - \sigma n + n \\
&= -(1-\sigma)(q-n)
\end{aligned}$$

where in the inequality above we have used the fact that $(\sum_{i=1}^{n} u_i)\ (\sum_{i=1}^{n} u_i^{-1})/n \geqslant n$. This relation can be shown as follows. Let $u_{\mathrm{ave}} \equiv e^{\mathrm{T}}u/n$. Then,

$$\begin{aligned}
0 &\leqslant \sum_{i=1}^{n} \left(\frac{u_i^{1/2}}{u_{\mathrm{ave}}^{1/2}} - \frac{u_{\mathrm{ave}}^{1/2}}{u_i^{1/2}}\right)^2 \\
&= \sum_{i=1}^{n} u_i/u_{\mathrm{ave}} - 2n + u_{\mathrm{ave}} \sum_{i=1}^{n} u_i^{-1} \\
&= -n + \left(\sum_{i=1}^{n} u_i\right)\left(\sum_{i=1}^{n} u_i^{-1}\right)\Big/ n
\end{aligned}$$

so that the validity of the relation follows. □

Lemma 4.5. *For $(x,s) \in \mathbb{R}^{2n}_{++}$, let $(\Delta x, \Delta y, \Delta s)$ denote the solution of system (2.8). Then,*

$$\nabla \phi(x,s)^{\mathrm{T}}(\Delta x, \Delta s) \equiv \nabla_x \phi(x,s)^{\mathrm{T}} \Delta x + \nabla_s \phi(x,s)^{\mathrm{T}} \Delta s \leqslant -(1-\sigma)(q-n).$$

Proof. Since $\phi(x,s) = P(J(x,s))$, it follows that $\nabla \phi(x,s) = \nabla J(x,s) \nabla P(J(x,s))$ for all $(x,s) \in \mathbb{R}^{2n}_{++}$. Here we are adopting the convention that the gradient of a single-valued function is a column vector and the gradient of a multi-valued function is the matrix whose ith column is the gradient of the ith component function. Observe that (2.8a) is equivalent to $\nabla J(x,s)^{\mathrm{T}}(\Delta x, \Delta s) = V(J(x,s))$ where V is defined by relation (4.6). Hence, for all $(x,s) \in \mathbb{R}^{2n}_{++}$,

$$\begin{aligned}
\nabla \phi(x,s)^{\mathrm{T}}(\Delta x, \Delta s) &= \nabla P(J(x,s))^{\mathrm{T}} \nabla J(x,s)^{\mathrm{T}}(\Delta x, \Delta s) \\
&= \nabla P(J(x,s))^{\mathrm{T}} V(J(x,s)) \\
&\leqslant -(1-\sigma)(q-n),
\end{aligned}$$

where the inequality follows from Lemma 4.4. □

The following result is an immediate consequence of Proposition 2.1(c) and Lemma 4.5.

Corollary 4.1. *Let* (x,y) *and* $(\Delta x,\Delta s)$ *be as in Lemma* 4.5. *Let* $v(\alpha)\equiv(x(\alpha),s(\alpha))$ *where* $x(\alpha)$ *and* $s(\alpha)$ *are defined by relations* (2.9). *Then,* $\nabla\phi(v(0))^{\mathrm{T}}\dot{v}(0) \leqslant -(1-\sigma)(q-n)$. □

We need one more lemma before giving the proof of Theorem 2.1.

Lemma 4.6. *Let a sequence* $\{(x_k,s_k)\}_{k=1}^{\infty}\subseteq\mathbb{R}^n_{++}$ *be given. Assume that for some infinite index set* $K\subseteq\mathbb{N}$ *and some compact set* $G\subseteq\mathbb{R}^{2n}_{++}$, *there exists* $\{(x_k,s_k)\}_{k\in K}\subseteq G$. *For all* $k\in\mathbb{N}$, *let* $(\Delta x_k,\Delta y_k,\Delta s_k)$ *denote the solution of system* (2.8) *with* x *and* s *replaced by* x_k *and* s_k, *respectively. Consider the sequence of curves* $\{(x_k(\cdot),s_k(\cdot))\}_{k=1}^{\infty}$ *defined by relations* (2.9). *Then*:

(a) *The subsequence* $\{(\Delta x_k,\Delta y_k,\Delta s_k)\}_{k\in K}$ *is bounded. As a consequence, it follows that* $\sup_{k\in K}\|(\dot{x}_k(0),\dot{s}_k(0))\|<\infty$.

(b) *The subsequence* $\{(x_k(\cdot),s_k(\cdot))\}_{k\in K}$ *is uniformly differentiable at* 0. (*Here, we view each curve* $(x_k(\cdot),s_k(\cdot))$ *as being defined over* $I_k\equiv\mathbb{R}_+$.)

Proof. The first part of statement (a) is an immediate consequence of Proposition 2.1(b) and the fact that a compact set is mapped by a continuous function into a bounded set. The second part of (a) follows immediately from Proposition 2.1(c). We next show statement (b). Define $v_k(\alpha)\equiv(x_k(\alpha),s_k(\alpha))$, for $k\in\mathbb{N}$. Observe that condition (a) of Definition 3.2 is clearly satisfied in this case since $\eta_k\equiv\infty$ and $\dot{v}_k(0)\neq 0$, for all $k\in\mathbb{N}$. To show condition (b) of Definition 3.2, let $\varepsilon>0$ be given. From the assumptions of the lemma, it follows that $\{x_k\}_{k\in K}\subseteq G_x$, where G_x is some compact subset of $\mathbb{R}^n$. Since, by Assumption 2.1(b), $f(x)\in C^2$ over the set $\mathbb{R}^n$, it follows that there exists a scalar $\delta>0$ such that $\|\nabla^2 f(x+h)-\nabla^2 f(x)\|\leqslant\varepsilon$ for all $x\in G_x$ and $h\in\mathbb{R}^n$ satisfying $\|h\|\leqslant\delta$. We claim that the scalar δ fulfills condition (b) of Definition 3.2. Indeed, let $\alpha\in\mathbb{R}_+$ be such that $\alpha\|\dot{v}_k(0)\|<\delta$. Then, $\alpha\Delta x_k=\alpha\|\dot{x}_k(0)\|<\delta$, and consequently, we obtain

$$\|\nabla^2 f(x_k+\alpha\Delta x_k)-\nabla^2 f(x_k)\|\leqslant\varepsilon \tag{4.7}$$

due to the way δ is defined. Using (2.9) and (2.10), it is easy to see that

$$v_k(\alpha)-v_k(0)-\alpha\dot{v}_k(0)=(0,\nabla f(x_k+\alpha\Delta x_k)-\nabla f(x_k)-\alpha\nabla^2 f(x_k)\Delta x_k). \tag{4.8}$$

Hence, we obtain

$$\begin{aligned}\|v_k(\alpha)-v_k(0)-\alpha\dot{v}_k(0)\| &= \|\nabla f(x_k+\alpha\Delta x_k)-\nabla f(x_k)-\alpha\nabla^2 f(x_k)\Delta x_k\|\\ &= \left\|\alpha\int_0^1[\nabla^2 f(x_k+\theta\alpha\Delta x_k)-\nabla^2 f(x_k)]\Delta x_k\,\mathrm{d}\theta\right\|\\ &\leqslant \alpha\|\dot{x}_k(0)\|\int_0^1\|\nabla^2 f(x_k+\theta\alpha\Delta x_k)-\nabla^2 f(x_k)\|\,\mathrm{d}\theta\\ &\leqslant \alpha\|\dot{v}_k(0)\|\,\varepsilon\end{aligned}$$

where the first equality is due to (4.8), the second equality is due to the first order integral

Taylor formula, the third inequality is due to a well-known inequality regarding integrals and the equality $\Delta x_k = \dot{x}_k(0)$ and the last inequality follows from (4.7) and the inequality $\|\dot{x}_k(0)\| \leq \|\dot{v}_k(0)\|$. We have thus proved that $\{v_k(\cdot)\}_{k\in K}$ is uniformly differentiable at 0. □

We are now in a position to give the proof of Theorem 2.1.

Proof of Theorem 2.1. It is easy to see that statement (b) follows from Lemma 4.1 and statement (a). To show statement (a), assume by contradiction that $\limsup_{k\to\infty} x_k^{\mathrm{T}} s_k > 0$. Then, there exists an infinite index set $K \subseteq \mathbb{N}$ and a scalar $\gamma > 0$ such that $x_k^{\mathrm{T}} s_k \geq \gamma$, for all $k \in K$. Let $v_k \equiv (x_k, s_k)$ and $v_k(\alpha) \equiv (x_k(\alpha), s_k(\alpha))$ for all $k \in \mathbb{N}$, where we recall that $x_k(\alpha)$ and $s_k(\alpha)$ are defined by relations (2.9). From the definition (4.5) of the set $\tilde{L}(\gamma,\eta)$ and the fact that $\{\phi(v_k)\}_{k=1}^{\infty}$ is monotonically decreasing, it follows that $\{v_k\}_{k\in K} \subseteq \tilde{L}(\gamma,\eta)$ where $\eta \equiv \phi(v_0)$. This last fact, Lemma 4.6 and the fact that $\tilde{L}(\gamma,\eta)$ is a compact subset of $F_{XS}^{\circ} \subseteq \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ (see Lemma 4.3) imply that $\{v_k(\cdot)\}_{k\in K}$ is uniformly differentiable at 0 and $\sup_{k\in K} \|\dot{v}_k(0)\| < \infty$. By Corollary 4.1, we also know that

$$\liminf_{k\in K} |\nabla\phi(v_k)^{\mathrm{T}} \dot{v}_k(0)| \geq (1-\sigma_k)(q-n) > 0 \geq (1-\bar{\sigma})(q-n) > 0,$$

where $\bar{\sigma}$ is the fixed scalar specified in step 0 of the PRPD algorithm. Letting $D = \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$, $G = \tilde{L}(\gamma,\eta)$, $h = \phi$ in Corollary 3.1, the assertions above show that the assumptions as well as relations (3.18), (3.19) and (3.20) of Corollary 3.1 are satisfied. But according to the conclusion of Corollary 3.1, this is impossible. Hence, we obtain a contradiction and statement (a) follows. □

5. Remarks

In this section, we make the following observation regarding our approach in this paper.

Assumption 2.1 can be somewhat relaxed to include more general convex functions. We could replace Assumption 2.1 by the following set of assumptions:

(1) Assume that $f: \mathbb{R}^n \to (-\infty, \infty]$ is a closed convex function, that is the epigraph of f is a closed convex set. Assume also that the effective domain of f, that is the set $\mathrm{dom}(f) \equiv \{x \in \mathbb{R}^n \mid f(x) < \infty\}$, contains $\mathbb{R}^n_{++}$.

(2) Assume that $f(x)$ is twice continuously differentiable over $\mathbb{R}^n_{++}$.

With the set of assumptions above replacing Assumption 2.1, one can show that Theorem 2.1 is still valid. There is one minor technical difficulty that arises when the above weaker assumptions are considered. Namely, the constraints of the Wolfe dual problem may not be defined at points where the optimal solutions of (P) occur since we are not assuming that the gradient of f exists over the set $\mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$. In this case, one can instead work with the following more general dual problem

$$\begin{aligned} \max \quad & f(x) - x^{\mathrm{T}} s \\ \text{s.t.} \quad & A^{\mathrm{T}} y + s \in \partial f(x), \\ & s \geqslant 0, \quad x \in \operatorname{dom}(f). \end{aligned}$$

For points $x \in \mathbb{R}^n_{++}$, the above constraint involving the subgradient $\partial f(x)$ reduces to the constraint involving the gradient $\nabla f(x)$ of the Wolfe dual (D). This follows from the fact that $\partial f(x) = \{\nabla f(x)\}$ whenever f is finite and differentiable at x (see Rockafellar [34, Theorem 25.1]). One can show using Theorem 27.4 of Rockafellar [34] that if $\bar{x}$ is an optimal solution for (P) then there exists $(\bar{y},\bar{s}) \in \mathbb{R}^m \times \mathbb{R}^n$ such that $(\bar{x},\bar{y},\bar{s})$ solves the above dual and satisfies $\bar{x}^{\mathrm{T}}\bar{s} = 0$.

Next we mention a classical example (Dembo [2, p. 247]) of a convex programming problem (P) which satisfies the conditions outlined above but not the conditions of Assumption 2.1. In this example, $f(x)$ is defined over the whole $\mathbb{R}^n_+$ and $\nabla^2 f(x)$ is defined for every $x \in \mathbb{R}^n_{++}$ but not for all $x \in \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$.

Example. For $x \in \mathbb{R}^4_+$, let

$$f(x) \equiv \sum_{i=1}^{4} x_i \log x_i + x_3 \log 4 - (x_3 + x_4) \log (x_3 + x_4)$$

and define

$$A \equiv \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & \frac{1}{2} & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}, \qquad b \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

It can be shown f is a convex function and that problem (P) has the following properties:

(a) it has the unique optimal solution $\bar{x} \equiv (\frac{1}{2},\frac{1}{2},0,0)$;

(b) f is not differentiable at $\bar{x}$;

(c) F°_{XS} is non-empty.

Acknowledgement

The author wishes to thank Suvrajeet Sen for many useful discussions which have been invaluable toward the development of this work and to express his gratitude to the two anonymous referees for their patient reading of a prior version of this manuscript and for their very useful comments. This research was partially funded by the National Science Foundation (NSF) under Grant No. DDM-9109404. Their financial support is gratefully acknowledged.

References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, London, 1982).

[2] R.S. Dembo, "Dual to primal conversion in geometric programming," *Mathematical Programming* 26 (1978) 243–252.

[3] I.I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Mathematics Doklady* 8 (1967) 674–675.

[4] D.-Z. Du and X.-S Zhang, "Global convergence of Rosen's gradient projection method," *Mathematical Programming* 44 (1989) 357–366.

[5] R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1987).

[6] C.C. Gonzaga, "An algorithm for solving linear programming problems in O (n^3L) operations," in: N. Megiddo, ed., *Progress in Mathematical Programming: Interior Point Algorithms and Related Methods* (Springer, New York, 1989) pp. 1–28.

[7] O. Guler, "Existence of interior points and interior paths in nonlinear monotone complementarity problems," *Mathematics of Operations Research* 18 (1993) 128–147.

[8] D. den Hertog, C. Roos and T. Terlaky, "On the classical logarithmic barrier function method for a class of smooth convex programming problems," *Journal of Optimization Theory and Applications* 73 (1992) 1–25.

[9] D. den Hertog, C. Roos and T. Terlaky, "A large-step analytic center method for a class of smooth convex programming problems", *SIAM Journal on Optimization* 2 (1992) 55–70.

[10] F. Jarre, G. Sonnevend and J. Stoer, "On the complexity of a numerical algorithm for solving generalized convex quadratic programs by following a central path, in: J.C. Lagarias and M.J. Todd, editors, *Mathematical Developments Arising from Linear Programming: Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988,* volume 114 of *Contemporary Mathematics* (American Mathematical Society, Providence, Rhode Island, USA, 1990) pp. 233–242.

[11] N. Karmarkar, "A new polynomial time algorithm for linear programming," *Combinatorica* 4 (1984) 373–395.

[12] M. Kojima, N. Megiddo and S. Mizuno, "Theoretical convergence of large-step primal–dual interior point algorithms for linear programming," *Mathematical Programming* 59 (1993) 1–21.

[13] M. Kojima, N. Megiddo and T. Noma, "Homotopy continuation methods for nonlinear complementarity problems," *Mathematics of Operations Research* 16 (1991) 754–774.

[14] M. Kojima, N. Megiddo, T. Noma and A. Yoshise, "A unified approach to interior point algorithms for linear complementarity problems," in: *Lecture Notes in Computer Science No. 538* (Springer, Berlin, 1991).

[15] M. Kojima, S. Mizuno and T. Noma, "Limiting behavior of trajectories generated by a continuation method for monotone complementarity problems," *Mathematics of Operations Research* 15 (1990) 662–675.

[16] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," *Mathematical Programming* 44 (1989) 1–26.

[17] M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming: Interior Point Algorithms and Related Methods* (Springer, New York, 1989) pp. 131–158.

[18] M. Kojima, S. Mizuno and A. Yoshise, "An O($\sqrt{n}L$) *iteration potential reduction algorithm for linear complementarity problems,*" *Mathematical Programming* 50 (1991) 331–342.

[19] K.O. Kortanek, F. Potra and Y. Ye, "On some efficient interior point methods for nonlinear convex programming," *Linear Algebra and its Applications* 152 (1991) 169–189.

[20] K.O. Kortanek and J. Zhu, "A polynomial barrier algorithm for linearly constrained convex programming problems," *Mathematics of Operations Research* 18 (1993) 116–127.

[21] L. McLinden, "The complementarity problem for maximal monotone multifunctions," in: R.W. Cottle, F. Giannessi and J.-L. Lions, eds., *Variational Inequalities and Complementarity Problems* (Wiley, New York, 1980) pp. 251–270.

[22] N. Megiddo, "Pathways to the optimal set in linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming: Interior Point Algorithms and Related Methods* (Springer, New York, 1989) pp. 131–158.

[23] S. Mehrotra and J. Sun, "An interior point algorithm for solving smooth convex programs based on Newton's method," in: J.C. Lagarias and M.J. Todd, ed., *Contemporary Mathematics: Proceedings of the AMS – IMS–SIAM Joint Summer Research Conference on "Mathematical Developments Arising from Linear Programming", held at Bowdoin College, Brunswick, Maine on July 25–31, 1988, vol. 114* (AMS, Province, RI, 1990) pp. 265–284.

[24] S. Mizuno, M.J. Todd and Y. Ye, "On adaptive step primal-dual interior-point algorithms for linear programming," *Mathematics of Operations Research* 18 (1993) 945–981.

[25] S. Mizuno, A. Yoshise and T. Kikuchi, "Practical polynomial time algorithms for linear complementarity problems," *Journal of the Operations Research Society of Japan* 32 (1989) 75–92.

[26] R.D.C. Monteiro and I. Adler, "Interior path following primal–dual algorithms. Part II: Convex quadratic programming," *Mathematical Programming* 44 (1989) 43–66.

[27] R.D.C. Monteiro and I. Adler, "Interior path following primal–dual algorithms. Part I: Linear programming," *Mathematical Programming* 44 (1989) 27–41.

[28] R.D.C. Monteiro and I. Adler, "An extension of kamarkar type algorithm to a class of convex separable programming problems with global linear rate of convergence," *Mathematics of Operations Research* 15 (1990) 408–422.

[29] R.D.C. Monteiro, I. Adler and Mauricio G.C. Resende, "A polynomial-time primal–dual affine scaling algorithm for linear and convex quadratic programming and its power series extension," *Mathematics of Operations Research* 15 (1990) 191–214.

[30] J.J. Moré, "Class of functions and feasibility conditions in nonlinear complementarity problem," *Mathematical Programming* 6 (1974) 327–338.

[31] Y.E. Nesterov and A.S. Nemirovsky, "A general approach to the design of optimal methods for smooth convex functions minimization," *Ekonomika i Matem. Metody* 24 (1988) 509–517. (In Russian; English transl. Matekon: Translations of Russian and East European Math. Economics.)

[32] J.S. Pang, "Iterative descent methods for a row sufficient linear complementarity problem," *SIAM Journal on Matrix Analysis and Applications* 12 (1991) 611–624.

[33] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* 40 (1988) 59–93.

[34] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[35] G. Sonnevend, "An analytical centre for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," in: A. Prekopa, ed., *Lecture Notes in Control and Information Sciences, No. 84* (Springer, Berlin, 1985) pp. 866–876.

[36] K. Tanabe, "Complementarity-enforcing centered Newton method for mathematical programing: global theory," in: K. Tone, ed., *New Methods for Linear Programming* (The Institute of Statistical Mathematics, Tokyo, 1987) pp. 118–144.

[37] K. Tanabe, "Centered Newton method for mathematical programming," in: M. Iri and K. Yajima, eds., *Systems Modelling and Optimization* (Springer, New York, 1988) pp. 197–206.

[38] M.J. Todd and Y. Ye, "A centered projective algorithm for linear programming," *Mathematics of Operations Research* 15 (1990) 508–529.

[39] P. Tseng, "Global linear convergence of a path following algorithm for some monotone variational inequality problems." *Journal of Optimization Theory and Applications.* 75 (1992) 265–279.

[40] P.M. Vaidya, "An algorithm for linear programming which requires $O((m+n)n_2+(m+n)L)$ arithmetic operations," *Mathematical Programming* 47 (1990) 175–202.

[41] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programmig," *Mathematical Programming* 50 (1991) 239–258.

[42] F.A. Potra and Y. Ye, A quadratically convergent polynomial algorithm for solving entropy optimization problems, *SIAM Journal on Optimization* 3 (1993) 843–861.

[43] J. Zhu, "A path following algorithm for a class of convex programming problems," *ZOR – Methods and Models of Operations Research* 36 (1992) 359–377.