Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 3, No. 1, pp. 53 – 72, January 2007

LINEAR PROGRAMS AND IMPLICIT FUNCTIONS

Katrin Hauk and Florian Jarre

Dedicated to Professor Masakazu Kojima on the occasion of his 60th birthday.

Abstract: This paper explores the solution of linear programs based on the minimization of convex, differentiable, piecewise quadratic functions. These functions define certain implicit functions which provide bounds for a complexity analysis. One of the approaches is based on an augmented Lagrangian method. The results that are known for the augmented Lagrangian in the case of more general nonlinear programs are strengthened and a link between linear programs and convex conjugates of convex piecewise quadratic functions is established.

Key words: *linear program, piecewise quadratic function, augmented Lagrangian*

Mathematics Subject Classification: *90C05*

---

1 Introduction

We consider the linear program

$$(P) \qquad \text{minimize } \ c^T z \ \text{ s.t. } \ Az = b, \qquad z \geq 0$$

and its dual

$$(D) \qquad \text{maximize } \ b^T y \ \text{ s.t. } \ A^T y \leq c.$$

For later convenience we have denoted the primal variable by z, all other notation follows the standard conventions, i.e. the data is given by a matrix $A \in \mathbb{R}^{m \times n}$ with $n > m$ and the vectors b and c of appropriate dimensions.

Polynomial time methods for linear programs have first been proposed by Khachiyan [13] and Karmarkar [12]. The complexities of the ellipsoid method and the interior-point method in these papers depend on the encoding length of the data A, b, c. Megiddo [15] was able to show that linear programs can be solved in linear time when the dimension is fixed, and Tardos [21] was the first to propose a polynomial time algorithm for combinatorial linear programs whose complexity does not depend on b and c. The layered least squares algorithm by Vavasis and Ye [22] and an improved version by Monteiro and Tsuchiya [16] eliminated the dependence on the vectors b and c for the case of general linear programs. Homogeneous self-dual approaches by Ye, Todd, and Mizuno [23] introduced an elegant approach to circumvent the problem of finding a good starting point. A practical algorithm is described in [14].

Our approach uses a reformulation of the linear program as a convex, differentiable, piecewise quadratic minimization problem as well as an augmented Lagrangian (see e.g. [9, 18]) technique. The complexity analysis is based on a generalized Newton method applied to the piecewise quadratic function f. The number of steps of the generalized Newton method is bounded by exploiting the properties of the conjugate function for f.

Augmented Lagrangian approaches have been successfully applied to nonlinear and non-convex programs, see e.g. [4, 5], and are the subject of ongoing research, see e.g. [7, 19]. The application to nonlinear programs is well understood. It simplifies considerably when applied to linear programs.

Notation

In the remainder of this paper we use the following notation. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ we denote the derivative of f at y by the row vector $Df(y)$ and the gradient by the column vector $\nabla f(y) := Df(y)^T$. Second derivatives are denoted by $D^2 f(y)$ or $\nabla^2 f(y)$, i.e. we do not distinguish between square matrices and bilinear forms. If ∇f is differentiable almost everywhere, the generalized Hessian of f at a point x is given by the convex hull of the limits of $\nabla^2 f(y)$ where $y \to x$ such that $\nabla^2 f(y)$ is well defined; see e.g. [3]. (This definition is not to be confused with other versions of generalized derivatives by Sobolev or Lanczos which are based on partial integration.)

When there is a given set of measure zero (e.g. the set where the second derivative of a given function is not defined) we say a point is in general position, if it does not lie within this set. A point that is generated by some random process with a continuous density function always lies in general position – with probability one.

By e we denote the vector of all ones, and the notation $M \succ 0$ is used to indicate that the symmetric matrix M is positive definite. The pseudo inverse of a matrix M is denoted by $M^\dagger$, see e.g. [8]. Throughout this paper we assume that the matrix A defining the linear program (P) has full row rank m. The columns of A are denoted by a_i for $1 \le i \le n$, the components of c by c_i.

For the remainder of this paper we make the following assumption:

Assumption 1 *We assume from now on that there is no direction y with $A^T y \le 0$ and $b^T y > 0$.*

If there was a y violating Assumption 1 then (D) would not have a finite optimal solution and Algorithm 1 below would identify this case.

$\boxed{2}$ Newton's Method for Certain Piecewise Quadratic Functions

For a real number α we set $\alpha^+ = \max\{0, \alpha\}$ and for a vector $z \in \mathbb{R}^n$ we denote by z^+ the vector with components $(z^+)_i = (z_i)^+$ for $1 \le i \le n$. Using the optimality conditions of (P) and (D), it is straigtforward to see that a point $(\bar z,\ \bar y)$ solves (P) and (D), if, and only if, it minimizes the convex, differentiable, piecewise quadratic function

$$f^{(P),(D)}(z, y) := (c^T z - b^T y)^2 + \|Az - b\|_2^2 + \sum_{i=1}^{n} ((a_i^T y - c_i)^+)^2 + ((-z_i)^+)^2 \qquad (1)$$

and satisfies $f^{(P),(D)}(\bar z, \bar y) = 0$. Next we consider the minimization of $f^{(P),(D)}$ by a generalized Newton approach with line search.

To analyze the generalized Newton path we consider certain convex, differentiable, piecewise quadratic functions f. For simplicity, the function f below is defined on $\mathbb{R}^m$, the transfer of the results for f to $f^{(P),(D)} : \mathbb{R}^{n+m} \to \mathbb{R}$ is straightforward.

We say f is piecewise quadratic on $\mathbb{R}^m$ if $\mathbb{R}^m$ is partitioned into a finite number of polyhedra and f is quadratic on each of these polyhedra. In this section we always consider functions $f : \mathbb{R}^m \to \mathbb{R}$ of the special form

$$f(y) := q(y) + \frac{1}{2}\sum_{i=1}^{n}((a_i^T y - \gamma_i)^+)^2, \tag{2}$$

where q is a convex quadratic function,

$$q(y) = b^T y + \frac{1}{2} y^T H y.$$

Other types of convex, differentiable, piecewise quadratic functions will be considered in Section 3.1.

The indices i of the "plus-squared" terms in (2) are divided into active, weakly active and inactive indices.

Definition 1 *An index i of (2) is called active at y if the i-th component satisfies $a_i^T y - \gamma_i > 0$. It is called weakly active if $a_i^T y - \gamma_i = 0$. Otherwise it is called inactive. Indices are called linearly independent if the associated vectors a_i are linearly independent.*

2.1 The Generalized Newton Path

Next, we consider two types of straightforward generalizations of the Newton step for minimizing f as in (2). In (3) below, we consider the case where the Hessian of f exists but may be singular, and in (5) below, we consider certain points where the Hessian is not defined. The generalized Newton step $\Delta\hat{y}$ for minimizing a convex function f starting at a point $\hat{y}$ is defined as follows: When ∇f is differentiable at $\hat{y}$ we set

$$\Delta\hat{y} := \begin{cases} \lim_{\epsilon\to 0,\ \epsilon>0} -(\nabla^2 f(\hat{y}) + \epsilon I)^{-1}\nabla f(\hat{y}) & \text{if this is finite} \\ \lim_{\epsilon\to 0,\ \epsilon>0} -\epsilon(\nabla^2 f(\hat{y}) + \epsilon I)^{-1}\nabla f(\hat{y}) & \text{else.} \end{cases} \tag{3}$$

(Here I denotes the identity matrix.) Hence, when $\nabla^2 f(\hat{y})$ is invertible $\Delta\hat{y}$ is defined by the first case in (3) and coincides with the Newton step. When $\nabla^2 f(\hat{y})$ is singular and the gradient of f is not contained in the null space of $\nabla^2 f(\hat{y})$, the generalized Newton step is defined by the second case in (3). Using the eigenvalue decomposition of $\nabla^2 f(\hat{y})$ it then follows that $\Delta\hat{y}$ is the orthogonal projection of the negative gradient onto the null space of $\nabla^2 f(\hat{y})$. Finally, if the gradient of f is contained in the null space of $\nabla^2 f(\hat{y})$, then the generalized Newton step is defined again by the first case in (3) and coincides with $-(\nabla^2 f(\hat{y}))^\dagger \nabla f(\hat{y})$, where $\dagger$ denotes the pseudo inverse.

If f is a quadratic function on all of $\mathbb{R}^m$ and the step $\Delta\hat{y}$ is defined by the first case in (3), the minimum of f is given by the step length $t_{max}(\hat{y}) := 1$; the point $\hat{y} + t_{max}(\hat{y})\Delta\hat{y}$ is a minimizer of f. If f is quadratic on all of $\mathbb{R}^m$ and $\Delta\hat{y}$ is defined by the second case in (3), the function f does not have a minimum and $t_{max}(\hat{y}) := \infty$.

When $\nabla^2 f(\hat{y})$ is not defined, the generalized Hessian of f at $\hat{y}$ contains several elements. A general analysis of this case is complicated; we only consider the case when there is exactly one weakly active index $\hat{\imath}$ with $a_{\hat{\imath}}^T \hat{y} - \gamma_{\hat{\imath}} = 0$ and assume that $\nabla^2 f(y)$ is positive definite for y near $\hat{y}$ and $a_{\hat{\imath}}^T y - \gamma_{\hat{\imath}} \neq 0$, say $\nabla^2 f(y) =: \tilde{H} \succ 0$ for y near $\hat{y}$ and $a_{\hat{\imath}}^T y - \gamma_{\hat{\imath}} < 0$. For such y

the Newton step Δy is a well defined function of y. (One could use the notation $\Delta y = \Delta(y)$ to indicate that Δy depends on y.) The rank-1-update formula for inverse matrices then implies that the sign of the scalar product of $a_{\hat{i}}$ with Δy is the same for all y near $\hat{y}$ with $a_{\hat{i}}^T y \neq \gamma_{\hat{i}}$, i.e. $sign(a_{\hat{i}}^T \Delta y) \equiv const.$ Namely, if the gradient of f is denoted by $g = g(y)$, and $a = a_{\hat{i}}$, then

$$\begin{aligned} sign(a^T(\tilde{H} + aa^T)^{-1}g) &= sign\Big(a^T\Big(\tilde{H}^{-1} - \frac{\tilde{H}^{-1}aa^T\tilde{H}^{-1}}{1 + a^T\tilde{H}^{-1}a}\Big)g\Big) \\ &= sign\Big(a^T\tilde{H}^{-1}g\Big(1 - \frac{a^T\tilde{H}^{-1}a}{1 + a^T\tilde{H}^{-1}a}\Big)\Big) \\ &= sign(a^T\tilde{H}^{-1}g). \end{aligned} \tag{4}$$

Note that g is a continuous function of y. Hence, if $a^T\tilde{H}^{-1}g \neq 0$, then either $a_{\hat{i}}^T \Delta y > 0$ for all y near $\hat{y}$ with $a_{\hat{i}} y \neq \gamma_{\hat{i}}$ or $a_{\hat{i}}^T \Delta y < 0$ for all such y.

This observation shall be used to generalize the Newton step also for such y near $\hat{y}$ that satisfy $a_{\hat{i}} y = \gamma_{\hat{i}}$. In the sequel we will minimize a function f by following the generalized Newton steps. If a Newton step Δy starts at a point y with $a_i^T y \neq \gamma_i$ for all i and crosses the first weakly active index $a_{\hat{i}}^T(y + t\Delta y) = \gamma_{\hat{i}}$ at some point $\hat{y} = y + \hat{t}\Delta y$ with $\hat{t} < t_{max}(y)$, then it is easy to see that $a_{\hat{i}}^T \tilde{H}^{-1} g(\hat{y}) \neq 0$. (If $\hat{t} = t_{max}(y)$, the minimum is found and the algorithm stops.) Hence we assume $a_{\hat{i}}^T \tilde{H}^{-1} g(\hat{y}) \neq 0$ from now on and based on (5) we may define the generalized Newton step $\Delta\hat{y}$ starting at $\hat{y}$ by

$$\Delta\hat{y} := \begin{cases} \lim_{y \to \hat{y},\ a_{\hat{i}}^T y > \gamma_{\hat{i}}} \Delta y & \text{if } sign(a_{\hat{i}}^T \Delta y) = 1, \\ \lim_{y \to \hat{y},\ a_{\hat{i}}^T y < \gamma_{\hat{i}}} \Delta y & \text{if } sign(a_{\hat{i}}^T \Delta y) = -1. \end{cases} \tag{5}$$

This generalization allows us to define a piecewise linear continuous path based on the relation

$$\dot{y}^+(t) = \frac{\Delta y(t)}{\|\Delta y(t)\|_2}, \tag{6}$$

where $\Delta y(t)$ is the generalized Newton step starting at $y(t)$ and

$$\dot{y}^+(t) := \lim_{\Delta t \to 0,\ \Delta t > 0} \frac{y(t + \Delta t) - y(t)}{\Delta t}.$$

Due to (5), the one sided derivative $\dot{y}^+(t)$ is defined also at points $y(t)$ with exactly one weakly active index $\hat{i}$, as long as the Hessian of f is nonsingular for y near $y(t)$ and $a_{\hat{i}}^T \tilde{H}^{-1} g(y(t)) \neq 0$.

The case when there is exactly one weakly active index at $\hat{y}$ but $\nabla^2 f(y)$ is not positive definite for y near $\hat{y}$ is illustrated in Case 2. in Section 2.2 below. The case when there is more than one weakly active index at $\hat{y}$ is illustrated in Case 1.

We now assume that $\nabla^2 f(y) \succ 0$ everywhere except for such points y that have weakly active constraints *, i.e. for which $\nabla^2 f(y)$ is not defined. We consider the analogue of Newton's method where the generalized Newton direction is updated repeatedly as we encounter weakly active constraints. We assume that

$$\text{exactly one weakly active constraint exists at each iterate}^{\dagger}. \tag{7}$$

*This assumption may not be satisfied for all (z, y) when f is of the form (1). Modifications to account for singular Hessians are tedious and are therefore omitted here.

†Assumption (7) is generically satisfied: Let $\tilde{S}$ be the set of points that have two or more weakly active constraints. Then, $\tilde{S}$ has dimension $n - 2$. The set of points leading – via the generalized Newton path – to $\tilde{S}$ therefore has dimension $n - 1$. A point in general position will lie outside this set.

In this case we may define the generalized Newton path $y(t)$ starting from y^0 by (6). The points on this path form a piecewise linear curve leading from its initial point $y(0) = y^0$ to the minimizer y^* of f if it exists.

Tracing the path is simple: Given an initial point in general position the path crosses just one weakly active index at a time, and the new direction can be computed by a rank-one-update formula in order m^2 operations. The possibility of rank-one-updates for a Newton path has been observed earlier in [6], for example.

The complexity of following the generalized Newton path depends on the number of points with weakly active indices that are crossed by the path. Note that the straight line $[y^0, y^*]$ intersects at most n points with weakly active indices. Unfortunately, as we will see next, the generalized Newton path may pass the same weakly active index multiple times. We indicate an example where the path contains $n^2/4$ or more points with weakly active indices.

$\boxed{2.2}$ A Small Example

We return to the function f of (2). For illustration we consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(y) = -y_1 + 2y_2 + ((y_1)^+)^2 + ((y_1 - y_2)^+)^2.$$

This function has weakly active indices at all points with $y_1 = 0$ or with $y_1 = y_2$. The generalized Newton path starting at $y^0 := (-1, 2)^T$ leads along $y^0 + t(1, -2)^T$ for $0 \le t \le 1$ and then continues along the line $(0, 0)^T + t(-1, -1)^T$ for $t \ge 0$.

Case 1. The derivative of the path is not defined at $(0, 0)^T$; by distinguishing the four cases $y_1 \ge 0$ and/or $y_1 - y_2 \ge 0$, one easily finds that the continuation of the path in $(0, 0)^T$ is uniquely defined as stated above. Hence, the path does not pass through the line $y_1 = 0$ but is "reflected" at this line. As indicated in (5), such a reflection cannot occur, when there is just one weakly active index!

Case 2. When the initial point is changed to $y^0 = (-1, 3)^T$, the path will lead from y^0 to $(0, 1)^T$, then to $(0, 0)^T$, and then along the line $(0, 0)^T + t(-1, -1)^T$ for $t \ge 0$.

Case 3. If, in addition, a "prox-term" is added, $f(y) \longrightarrow f(y) + \epsilon y^T y$, the path will pass through the line $y_1 = 0$ near $(0, 1)^T$, then through the line $y_1 = y_2$, and will then pass the line $y_1 = 0$ a second time for some $y_2 < 0$. Hence, we cannot guarantee that the generalized Newton path will cross the same weakly active index (here $y_1 = 0$) only once.

The above cases are pictured in Figures 1 – 3. In fact, the negative result of the previous example can be strengthened: By adding $\hat{n} - 1$ further $((\ .\ )^+)^2$-terms, the example can be extended to cross the line $y_1 = 0$ exactly $\hat{n} + 1$ times along a zigzag-line. Then, the term $((y_1)^+)^2$ in the definition of f can be replaced by $\sum_{i=1}^{\hat{n}} \frac{1}{\hat{n}}((y_1 + \epsilon_i)^+)^2$, so that each of these new $((\ .\ )^+)^2$-terms is crossed $\hat{n} + 1$ times. We thus obtain a function f of the form (2) defined with $n = 2\hat{n}$ "$((\ .\ )^+)^2$-terms" and a piecewise linear generalized Newton path that consists of $\hat{n}(\hat{n} + 2) = n + n^2/4$ linear segments.

To estimate the worst-case-complexity for following the generalized Newton path, we like to bound the number of linear segments on the generalized Newton path.

Note that in the situation discussed in Case 1 above, the definition of the generalized Newton path may be difficult. We therefore consider the case of a strictly convex function q in (2), i.e. $H \succ 0$. When applying the generalized Newton method for minimizing f we obtain the following algorithm:

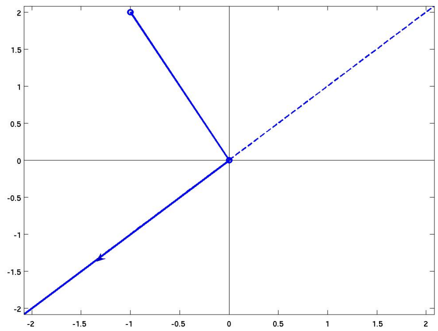

Figure 1: Case 1

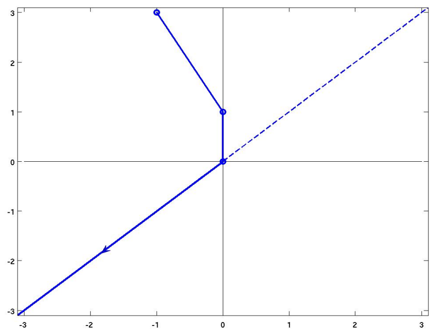

Figure 2: Case 2

Algorithm 1 (Minimizing a strictly convex piecewise quadratic f)

1. *Let a vector* $y^0 \in \mathbb{R}^m$ *in general position be given. Set* $k = 0$.

2. *Compute the generalized Newton step* Δy^k *at* y^k.

3. *Determine the smallest number* $\bar{\lambda}_k \in (0, \infty]$ *such that* $y^k + \bar{\lambda}_k \Delta y^k$ *contains a weakly active index. (Then* f *is quadratic on the line segment* $[y^k, y^k + \bar{\lambda}_k \Delta y^k]$*.)*

   Determine λ_k *minimizing* $f(y^k + \lambda \Delta y^k)$ *for* $\lambda \in (0, \bar{\lambda}_k]$.
   If $\lambda_k = \infty$ *then Stop,* f *does not have a minimum.*

4. *Set* $y^{k+1} := y^k + \lambda_k \Delta y^k$.

5. *If* $\nabla f(y^{k+1}) = 0$ *Stop, else set* $k := k + 1$ *and go to Step 2.*

Note: The case $\lambda_k = \infty$ in *Step 3.* cannot occur when $H \succ 0$.

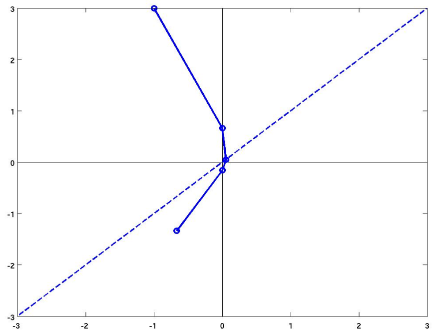

Figure 3: Case 3

$\boxed{2.3}$ The Conjugate of a Differentiable, Piecewise Quadratic, Strictly Convex Function

As in the previous example we will minimize a strictly convex, differentiable, piecewise quadratic function f by tracing the generalized Newton path. In the gradient space the generalized Newton path is a straight line. The link of the primal space and the gradient space is established via the conjugate function f^*. While f is strictly convex and quadratic on each cell of the primal arrangement, f^* is strictly convex and quadratic on each cell of a corresponding dual arrangement. Since the generalized Newton path is a line segment in the gradient space, the number of Newton steps needed to minimize f is the number of cells intersected by the line segment in the dual space. Subsection 2.3 studies in more detail the cell structure while Subsection 2.4 exploits an idea to bound the number of cells intersected by the Newton path.

To simplify the analysis, we assume in this subsection that the function q in (2) is strictly convex [‡], i.e. $H \succ 0$.

Since f is a strictly convex differentiable function, the gradient $v = \nabla f(y)$ is a one to one mapping from $\mathbb{R}^m$ to $\mathbb{R}^m$, and the conjugate function f^* is a strictly convex differentiable function which is given by

$$f^*(v) := \max_{y\in\mathbb{R}^m}\{v^T y - f(y)\}.$$

The function f^* is an implicit function that is closely related to the generalized Newton path. As shown in Theorem 26.6 in [20] it can also be written as

$$f^*(v) = [(\nabla f)^{-1}(v)]^T v - f((\nabla f)^{-1}(v)).$$

Strict monotonicity of ∇f, i.e.

$$[\nabla f(y) - \nabla f(x)]^T (y - x) > 0, \text{ (if } y \neq x)$$

also holds due to strict convexity and differentiability of f (Theorem IV.4.1.4 in [10]). In the sequel, the space $\{v \mid v = \nabla f(y), y \in \mathbb{R}^m\}$ is referred to as the dual space.

[‡]If not, a regularization term $\epsilon y^T y$ may be added to f to obtain a regularized function for which the generalized Newton path is uniquely defined. This path may then be used as a reference path to define the generalized Newton path for f; however, this approach is somewhat tedious and does not seem to be of practical or theoretical importance as long as the open question in Section 2.4 is not answered. It is therefore omitted.

For $J \subset \{1, \ldots, n\}$ let $\mathcal{P}_J$ be the polyhedron

$$\mathcal{P}_J := \{y \mid a_i^T y \geq \gamma_i \text{ for } i \in J, \quad a_i^T y \leq \gamma_i \text{ for } i \notin J\}.$$

By definition f is a quadratic function on each $\mathcal{P}_J$. For $y \in \mathcal{P}_J$, $\nabla f(y)$ is written as follows:

$$\nabla f(y) = \Big(H + \sum_{j \in J} a_j a_j^T\Big) y + b - \sum_{j \in J} \gamma_j a_j.$$

We analyze the gradient of f on each $\mathcal{P}_J$ and define $\tilde{\mathcal{P}}_J$ as the corresponding polyhedron of $\mathcal{P}_J$, i.e.

$$\tilde{\mathcal{P}}_J := \nabla f(\mathcal{P}_J) = \Big(H + \sum_{j \in J} a_j a_j^T\Big) \mathcal{P}_J + b - \sum_{j \in J} \gamma_j a_j.$$

Therefore, $\tilde{\mathcal{P}}_J$ is a polyhedron, and since ∇f is one to one from $\mathbb{R}^m$ to $\mathbb{R}^m$, the union of the polyhedra $\tilde{\mathcal{P}}_J, J \subset \{1, \ldots, n\}$ satisfies

$$\bigcup_{J \subset \{1, \ldots, n\}} \tilde{\mathcal{P}}_J = \mathbb{R}^m.$$

Obviously, for two sets $J, \bar{J} \subset \{1, \ldots, n\}$, $\mathcal{P}_J$ and $\mathcal{P}_{\bar{J}}$ are neighbors, if and only if $\tilde{\mathcal{P}}_J$ and $\tilde{\mathcal{P}}_{\bar{J}}$ are neighbors. It is easily seen that f^* is a continuous strictly convex piecewise quadratic function. On each of the $\tilde{\mathcal{P}}_J$ it is a quadratic function.

In the dual space, the path generated by Algorithm 1 is written as

$$\nabla f(y(t)) = tv^0, \ (t \in [0, 1]),$$

where $v^0 = \nabla f(y^0)$. The number of $\tilde{\mathcal{P}}_J$ intersected by the path is exactly the number of steps needed in Algorithm 1. Since $\tilde{\mathcal{P}}_J := \nabla f(\mathcal{P}_J)$, the number of polyhedra $\tilde{\mathcal{P}}_J$ is the same as the number of $\mathcal{P}_J$ dividing $\mathbb{R}^m$. Since this number is bounded by 2^n, the number of iterations of Algorithm 1 is bounded by 2^n. We summarize the discussion in the following lemma:

Lemma 1 *In Algorithm 1, the Hessian of a strictly convex function f can be updated with order n^2 operations at each step if there is only one weakly active constraint at each iteration. In this case the number of generalized Newton steps is bounded by at most 2^n.*

Remark 1 *By the footnote to Assumption* (7) *the existence of exactly one weakly active constraint at each iteration is guaranteed if the starting point y^0 is given in general position.*

We note that the computation of a generalized Newton step for weakly convex f is somewhat more complicated than the computation of a simplex step. We believe that the upper bound of 2^n generalized Newton steps is overly pessimistic, the worst example we found is given in Section 2.2 which obtains an upper bound of $n + n^2/4$ for even numbers n.

2.4 An Open Problem

We consider the line segment $[0, v^0]$ and assume that the line passes through the interior of k different polyhedra $\tilde{\mathcal{P}}_{J_l}$ $(1 \leq l \leq k)$. Let $0 < t_1 < \ldots < t_l < t_{l+1} < \ldots < t_k \leq 1$ be such that $t_l v^0 \in \tilde{\mathcal{P}}_{J_l}^\circ$ (the interior of $\tilde{\mathcal{P}}_{J_l}$).

To simplify the presentation we therefore assume without loss of generality that $b = 0$ in the definition of q in (2).

The conditions $t_l v^0 \in \tilde{\mathcal{P}}^\circ_{J_l}$ are then equivalent to the system of inequalities

$$-\delta_{l,i} a_i^T H_{J_l}^{-1} t_l v^0 < -\delta_{l,i}\Big(\gamma_i - \sum_{j\in J_l} \gamma_j a_i^T H_{J_l}^{-1} a_j\Big) \tag{8}$$

for $1 \le l \le k$ and $1 \le i \le n$. Here,

$$\delta_{l,i} = \begin{cases} 1 & \text{if } i \in J_l, \\ -1 & \text{else.} \end{cases}$$

If $t_l v^0 \in \tilde{\mathcal{P}}^\circ_{J_l}$ for some i with $i \notin J_l$ (hence, $\delta_{l,i} = -1$), then $t_l v^0 \notin \tilde{\mathcal{P}}^\circ_{J_l \cup \{i\}}$, more precisely, $t_l v^0$ violates the i-th constraint of $\tilde{\mathcal{P}}_{J_l \cup \{i\}}$, i.e.

$$a_i^T H_{J_l\cup\{i\}}^{-1} t_l v^0 < \gamma_i - \sum_{j\in J_l\cup\{i\}} \gamma_j a_i^T H_{J_l\cup\{i\}}^{-1} a_j. \tag{9}$$

Comparing (8) and (9), and using $\delta_{l,i} = -1$ in (9), we may replace J_l in (8) with $J_l \cup \{i\}$, (both, in the summation and in the matrix H_J, but do not change the definition of $\delta_{l,i}$).

Denote the matrix with rows $(H^{-1/2}a_j)^T$ for $j \in J$ by $\hat{A}_J^T$ and $\hat{a}_j := H^{-1/2}a_j$, then (by the update formula for inverse matrices) the vectors $a_i^T H_{J_l}^{-1}$ in the left hand sides of the inequalities in (8) can be written as

$$H_{J_l}^{-1} a_i = (H + A_{J_l} A_{J_l}^T)^{-1} a_i = H^{-1/2}(I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T)\hat{a}_i. \tag{10}$$

With the notation $\hat{v}^0 = H^{-1/2} v^0$ the inequalities (8) can then be written as

$$\begin{aligned} &- \delta_{l,i}\hat{a}_i^T (I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T) t_l \hat{v}^0 \\ &\qquad < -\delta_{l,i}\Big(\gamma_i - \hat{a}_i^T (I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T) \sum_{j\in J_l} \gamma_j \hat{a}_j\Big). \end{aligned} \tag{11}$$

With $\gamma^{J_l} \in \mathbb{R}^{|J_l|}$ being the vector with components γ_i for $i \in J_l$, the right-hand side of (11) can be simplified using the relation

$$\begin{aligned} (I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T) \sum_{j\in J_l} \gamma_j \hat{a}_j &= (I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T) \hat{A}_{J_l} \gamma^{J_l} \\ &= \hat{A}_{J_l}\gamma^{J_l} - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} (\hat{A}_{J_l}^T \hat{A}_{J_l} + I - I)\gamma^{J_l} \\ &= \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \gamma^{J_l}. \end{aligned} \tag{12}$$

(11) can then be written as

$$-\delta_{l,i}\hat{a}_i^T (I - \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \hat{A}_{J_l}^T) t_l \hat{v}^0 < -\delta_{l,i}(\gamma_i - \hat{a}_i^T \hat{A}_{J_l}(I + \hat{A}_{J_l}^T \hat{A}_{J_l})^{-1} \gamma^{J_l})$$

for $1 \le i \le n$ and $1 \le l \le k$.

As pointed out in (9) we may further replace J_l with $J_{l,i} := J_l \cup \{i\}$ in the above inequality. (Here, $J_{l,i} = J_l$ when $i \in J_l$.) We denote by $e_{l,i}$ the unit vector such that $e_{l,i}^T \hat{A}_{J_{l,i}}^T = \hat{a}_i^T$. Using the same argument as in (12), the above inequalities then reduce to

$$-\delta_{l,i} e_{l,i}^T (I + \hat{A}_{J_{l,i}}^T \hat{A}_{J_{l,i}})^{-1} \hat{A}_{J_{l,i}}^T t_l \hat{v}^0 < -\delta_{l,i} e_{l,i}^T (I + \hat{A}_{J_{l,i}}^T \hat{A}_{J_{l,i}})^{-1} \gamma^{J_{l,i}}. \tag{13}$$

Relation (13) can also be written as

$$-\delta_{l,i}e_{l,i}^T(I+\hat{A}_{J_{l,i}}^T\hat{A}_{J_{l,i}})^{-1}(t_l\hat{A}_{J_{l,i}}^T\hat{v}^0-\gamma^{J_{l,i}})<0.$$

Let $D^{(l,i)}$ be the $n\times n$ diagonal matrix with entries $D_{j,j}^{(l,i)}=1$ if $j\in J_{l,i}$ and $D_{j,j}^{(l,i)}=0$ else. Then, the above relation is equivalent to

$$-\delta_{l,i}e_i^TD^{(l,i)}(I+D^{(l,i)}\hat{A}^T\hat{A}D^{(l,i)})^{-1}D^{(l,i)}(t_l\hat{A}^T\hat{v}^0-\gamma)<0.$$

Let $\hat{N}$ be a matrix such that $\hat{N}u=0$ iff $u=\hat{A}^T\hat{v}$ for some $\hat{v}$. The preceding inequality can then be written as

$$-\delta_{l,i}e_i^TD^{(l,i)}(I+D^{(l,i)}\hat{A}^T\hat{A}D^{(l,i)})^{-1}D^{(l,i)}(t_lu-\gamma)<0,\qquad \hat{N}u=0.$$

Let the symmetric $n\times n$-matrix $M^{(l,i)}$ be given by $M^{(l,i)}:=D^{(l,i)}(I+D^{(l,i)}\hat{A}^T\hat{A}D^{(l,i)})^{-1}D^{(l,i)}$. The preceding inequality is then written shortly as

$$-\delta_{l,i}e_i^TM^{(l,i)}(t_lu-\gamma)<0,\qquad \hat{N}u=0.$$

We write this as

$$-\delta_{l,i}e_i^T\left(t_lM^{(l,i)},\ M^{(l,i)}\right)\begin{pmatrix}u\\-\gamma\end{pmatrix}<0,\qquad \hat{N}u=0.\tag{14}$$

We remind that this system is satisfied for all i with $1\le i\le n$ and all l with $1\le l\le k$. We select a subsystem by considering those indices i that are 'dropped from' or 'added to' J_l when moving to J_{l+1}. We make this more precise next:

By definition of t_l, the point t_lv^0 is in the interior of $\tilde{\mathcal{P}}_{J_l}$. Let $\hat{t}_l>t_l$ be maximal such that $\hat{t}_lv^0$ lies in $\tilde{\mathcal{P}}_{J_l}$, and let $i(l)$ be the constraint of $\tilde{\mathcal{P}}_{J_l}$ that is satisfied with equality at $\hat{t}_lv^0$. Hence, if we replace t_l with $\hat{t}_l-\epsilon$ for any sufficiently small $\epsilon>0$, the relation (14) remains valid, and if t_l is replaced with $\hat{t}_l+\epsilon$ for sufficiently small $\epsilon>0$, then (14) remains valid if the sign of the inequality for $i=i(l)$ is reversed. We repeat the corresponding inequalities:

$$\begin{aligned}-\delta_{l,i(l)}e_{i(l)}^T\left((\hat{t}_l-\epsilon)M^{(l,i(l))},\ M^{(l,i(l))}\right)\begin{pmatrix}u\\-\gamma\end{pmatrix}&<0,\\ \delta_{l,i(l)}e_{i(l)}^T\left((\hat{t}_l+\epsilon)M^{(l,i(l))},\ M^{(l,i(l))}\right)\begin{pmatrix}u\\-\gamma\end{pmatrix}&<0.\end{aligned}\tag{15}$$

It is an open question whether this system or (14) will lead to a contradiction when k exceeds a suitable polynomial bound of n. A positive answer to this question would directly imply strong polynomiality of linear programs.

$\boxed{2.5}$ Numerical Examples

We have implemented Algorithm 1 with MATLAB in order to test the program for functions of the form (2) and (1). At this point our goal was not to find a competitive numerical algorithm for solving linear programs, but to obtain a better understanding of how many weakly active indices will be intersected by the generalized Newton path minimizing f of the form (2) or (1). To obtain some intuition about the worst-case behavior, we tested a large number of random examples and limited ourselves to small size problems.

The function $f^{(P),(D)}$ in (1) is not strictly convex. When the Hessian of f is singular, the generalized Newton path runs parallel to weakly active constraints, and, as seen in Section

2.2, it will typically run into points with more than one weakly active index. At such points a generalized Newton step is difficult to compute. We therefore added a perturbation ϵI to $\nabla^2 f(y)$ whenever $\nabla^2 f(y)$ was nearly singular. Unfortunately, the numerical results are biassed by rounding errors; the distinction of which constraints are active, weakly active, or incactive becomes unreliable. In several examples the algorithm ended up with very short steps zigzagging between two weakly active indices, a behavior that cannot occur when exact arithmetic is used. In order to obtain a numerical implementation that might be competitive to other algorithms, one would not only need to control rounding errors but also use suitable rank-one update formulae when crossing weakly active indices.

For all numerical experiments we therefore used an exact line search along the (generalized) Newton direction. Since the function f is smooth, it is unlikely that the minimizer of the line search lies at a point with weakly active indices. (The zig-zagging was now indeed reduced to very few cases among 100000 test problems.) The exact line search can be carried out in order nm arithmetic operations. We counted both, the number of iterations (Newton steps) used and the total number of weakly active indices intersected along this path.

For our first set of examples we chose the function (2), where all data vectors b, a_i and γ are uniformly distributed in $[-0.5, 0.5]$, and the Hessian H of q as the product of a matrix Q and its transpose, Q having uniformly distributed entries in $[0, 1]$. The starting point is chosen uniformly distributed in $[-50, 50]$.

In Table 1, the results of the algorithm for such f are listed. We kept the dimension fixed at $n = 30$ and increased m by a factor of $3/2$ for each row. In each row the results are listed for 10000 random examples. The first column displays the values of m. In the second column we list the average number of Newton steps, in the third column the maximum number of Newton steps, in the fourth column the average number of weakly active constraints intersected along the Newton path, and finally, in the last column we list the maximum number of weakly active constraints that were crossed along the path. The algorithm stopped when the norm of the gradient was less than 10^{-12} or when the Newton direction was not a descent direction. We note that in the first two rows the maximum number of Newton steps was higher than the number of intersections with weakly active constraints. This was due to rounding errors in the final iterations.

m	aver. Newt.	max. Newt.	aver. cross.	max. cross
4	3.45	16	2.63	8
6	3.53	15	4.26	12
9	3.58	16	6.63	17
14	3.80	24	10.58	24
21	4.03	11	15.80	35
32	4.22	8	23.89	46
48	4.36	8	35.12	61
72	4.38	7	50.09	81
108	4.33	7	72.07	100
162	4.23	6	102.41	137

Table 1: Random f as in (2)

Note: Table 1 summarizes the results of a total of 100000 test problems. In none of the examples, the number of intersections of weakly active constraints exceeded $2m$. We do know, however, that $m^2/4$ or more intersections are possible for problems that are designed as in Section 2.2.

For our second set of examples we used functions $f^{(P),(D)}$ arising from random linear programs that have primal and dual feasible solutions. Whenever the Hessian of $f^{(P),(D)}$ had a condition number of more than 10^{12}, a regularization term ϵI was added to $f^{(P),(D)}$. The resulting step is a Levenberg-Marquardt step for the convex function $f^{(P),(D)}$. Table 2 lists the results with 1000 random examples for each row. Each problem (P) has $2m$ variables and m linear equality constraints. The resulting primal-dual function $f^{(P),(D)}$ has $4m$ "plus-squared"-terms. Again, the maximum number of crossing weakly active indices during the generalized Newton method is less than twice the number of "plus-squared"-terms.

m	aver. Newt.	max. Newt.	aver. cross.	max. cross
4	3.44	9	4.96	16
6	5.34	11	10.72	27
8	6.66	13	16.06	32
10	8.32	22	23.51	72
12	9.60	23	29.61	63
14	11.35	23	37.95	72
16	12.55	24	44.48	84
18	14.51	28	54.46	104
20	15.80	29	61.20	115

Table 2: $f^{(P),(D)}$ from random linear programs

Finally, in Table 3 we list the results for Klee-Minty problems of the form

$$\max \left\{ \sum_{j=1}^{n} \epsilon^{n-j} x_j \;\middle|\; x_i + 2\sum_{j=1}^{i-1} \epsilon^{i-j} x_j \leq 1 \;\text{ für }\; 1 \leq i \leq n, \;\; x \geq 0 \right\},$$

where $\epsilon = 0.45$. We have implemented both a primal-dual version and a dual-only version minimizing a function $f^{(D)}$ with $2n+1$ "plus-squared"-terms using the information that the optimal value of the above problem is 1. We list the results for the "dual-only" version since this version allowed problems of slightly larger dimension that were not biassed by rounding errors. Here each row lists the results with 1000 different starting points. The Klee-Minty problems were designed specifically to trick a method of completely different nature (the Simplex method). As expected, one would need to find other examples to embarrass the generalized Newton approach as considered in this paper.

n	aver. Newt.	max. Newt.	aver. cross.	max. cross
4	7.48	10	15.72	23
6	9.23	19	21.63	40
8	10.24	23	27.52	58
10	11.88	29	33.07	84
12	13.96	30	40.28	92
14	15.35	28	44.94	79

Table 3: $f^{(D)}$ from Klee-Minty problems

Note: For functions $f = f^{(P),(D)}$ arising from linear programs the observations are very similar as for general f of the form (2). While we do not know whether there might be exponentially many intersections in the worst case, the results inidcate that the average number of intersections might be fairly small.

3 An Implicit Function Derived from the Augmented Lagrangian

The function $f^{(P),(D)}$ of Section 2 is closely related to the augmented Lagrangian function. It does not need any penalty parameter but it depends on $n+m$ unknowns while the augmented Lagrangian can be written as a function of only m variables. In this section we derive further theoretical results based on the augmented Lagrangian.

3.1 The Augmented Lagrangian for Linear Programs: Basic Results

Let a penalty parameter $r > 0$ be given. The augmented Lagrangian for the dual problem (D) is given by

$$\Lambda(y,z,r) := -b^T y + \frac{r}{2}\Big(\Big(A^T y - c + \frac{z}{r}\Big)^+\Big)^T \Big(A^T y - c + \frac{z}{r}\Big)^+ - \frac{z^T z}{2r}.$$

Note: A derivation of the augmented Lagrangian can be found, for example, in [2], p. 395. There are several variants of augmented Lagrangian functions. Other (partially) augmented Lagrangian functions use quadratic penalty terms only for equality constraints and leave simple bounds unmodified. In this case, inequalities are treated via slack variables. Our approach is based on the (fully) augmented Lagrangian as given above, where also inequalities are penalized. We have chosen the dual problem (D) to define Λ so that there is only one type of constraint.

The gradient of Λ with respect to y and z is given by

$$\nabla_y \Lambda(y,z,r) = -b + rA\Big(A^T y - c + \frac{z}{r}\Big)^+$$

and

$$\nabla_z \Lambda(y,z,r) = \Big(A^T y - c + \frac{z}{r}\Big)^+ - \frac{z}{r}.$$

Here, and in the following, by $\nabla_y \Lambda$ we denote the gradient as a column vector and by $D_y\Lambda$ we denote the derivative as a row vector. The next proposition is well known in a more general context; in the case of linear programs it can be stated in a slightly stronger and particularly simple fashion:

Proposition 1 *For fixed $z \in \mathbb{R}^n$ the function $y \mapsto \Lambda(y,z,r)$ is convex, and for fixed $y \in \mathbb{R}^m$ the function $z \mapsto \Lambda(y,z,r)$ is concave. A point $(\bar{y},\bar{z})$ satisfies*

$$\nabla_y \Lambda(\bar{y},\bar{z},r) = 0 \qquad \text{and} \qquad \nabla_z \Lambda(\bar{y},\bar{z},r) = 0, \tag{16}$$

if, and only if, it is an optimal solution of (D) and (P).

Proof. The convexity with respect to y is evident; concavity with respect to z follows from a standard argument. Let (16) be satisfied. Relation $(A^T\bar{y} - c + \frac{\bar{z}}{r})^+ - \frac{\bar{z}}{r} = 0$ implies $A^T\bar{y} \le c$ (dual feasibility), $\bar{z} \ge 0$, and $\bar{z}_i = 0$ if $(A^T\bar{y})_i < c_i$ (complementarity). The relation

$$0 = \nabla_y \Lambda(\bar{y},\bar{z},r) = -b + A(r(A^T\bar{y} - c) + \bar{z})^+$$

implies that $(r(A^T\bar{y} - c) + \bar{z})^+$ is feasible for (P), and by complementarity it follows furthermore that $(r(A^T\bar{y} - c) + \bar{z})^+ = \bar{z}$. Hence, $(\bar{y},\bar{z})$ is an optimal solution of (D) and (P).

Likewise, when $(\bar{y}, \bar{z})$ is an optimal solution of (D) and (P), relation (16) follows. $\square$

For given (y, z, r) let $\sigma \in \mathbb{R}^n$ be defined by

$$\sigma_i := \sigma_i(y, z, r) := \begin{cases} 1 & \text{if } (A^T y - c + \frac{z}{r})_i \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

and let

$$\Sigma := \mathrm{Diag}(\sigma)$$

be the $n \times n$ diagonal matrix with diagonal entries $\Sigma_{ii} = \sigma_i$. Let (y, z, r) be given such that $(A^T y - c + \frac{z}{r})_i \neq 0$ for all i. Then the function $\Lambda(\ .\ ,\ .\ , r)$ is twice differentiable at (y, z) and the second derivatives of Λ with respect to y and z are given by

$$\nabla_y^2 \Lambda(y, z, r) = r A \Sigma A^T \succeq 0$$

and

$$\nabla_z^2 \Lambda(y, z, r) = -\frac{1}{r}(I - \Sigma) \preceq 0.$$

(The latter, along with the differentiability of Λ, also implies concavity of Λ with respect to z.)

Let z be fixed arbitrarily. By Assumption 1 the function $y \mapsto \Lambda(y, z, r)$ is bounded below and due to its piecewise quadratic structure, the solution set $Y(z)$ of the problem

$$\mathrm{minimize}_{y \in \mathbb{R}^m} \quad \Lambda(y, z, r) \tag{18}$$

is nonempty. (On each of the finitely many polyhedra on which Λ is quadratic there exists at least one minimizer y; the ones with the smallest value of Λ solve (18).) Hence, by Assumption 1 there always exists

$$y \in Y(z) := \mathrm{argmin}_y \{\Lambda(\ .\ , z, r)\}. \tag{19}$$

Conversely, if problem (18) has a solution for some $z \in \mathbb{R}^n$, then Assumption 1 must hold. This is readily verified: If Assumption 1 does not hold then there is a vector Δy with $b^T \Delta y > 0$ and $A^T \Delta y \leq 0$. Then for any y and $\lambda > 0$ we have

$$\Lambda(y + \lambda \Delta y, z, r) \leq \Lambda(y, z, r) - \lambda b^T \Delta y \overset{\lambda \to \infty}{\longrightarrow} -\infty$$

so that $\Lambda(\ .\ , z, r)$ does not have a minimum. #

Since for fixed y, the function Λ is concave with respect to z, also

$$\varphi(z) := \Lambda(Y(z), z, r) = \min_{y \in \mathbb{R}^m} \{\Lambda(y, z, r)\}$$

is a concave function of z (the minimum of concave functions is concave).

To avoid set-valued functions we define the point

$$y(z) := \mathrm{argmin}\{\|y\|_2^2 \mid y \in Y(z)\}. \tag{20}$$

The constraint $y \in Y(z)$ is equivalent to the equation $\nabla_y \Lambda(y, z, r) = 0$. Note that for fixed z, the set $Y(z) = \{y \mid \nabla_y \Lambda(y, z, r) = 0\}$ of minimizers of the convex function $\Lambda(\ .\ , z, r)$ is a convex set. On the other hand, by definition of $\nabla_y \Lambda$,

$$Y(z) = \Big\{ y \mid -b + rA\Big(A^T y - c + \frac{z}{r}\Big)^+ = 0 \Big\}. \tag{21}$$

While it is not evident from representation (21) that $Y(z)$ is convex for fixed z, this representation is certainly piecewise linear. Convexity and piecewise linearity imply that $Y(z)$ is a convex polyhedron. Hence, it can be written as

$$Y(z) = \{y \mid B_z y \le \tilde{b}_z\},$$

where the matrix B_z and the vector $\tilde{b}_z$ depend on z. From (21) it also follows that the constraints of $Y(z)$ are piecewise linear also with respect to z implying that B_z and $\tilde{b}_z$ can be written as piecewise linear functions of z. The KKT-conditions of

$$y(z) := \operatorname{argmin}\{\|y\|_2^2 \mid B_z y \le \tilde{b}_z\}$$

imply that $y(z)$ is a piecewise linear function of B_z and $\tilde{b}_z$ and hence a piecewise linear function of z. Thus $\varphi(z) = \Lambda(y(z), z, r)$ is a piecewise quadratic function of z. Note that continuity of φ follows from the concavity of φ.

Moreover, for any $z \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ the function $y(z)$ posseses a directional derivative $y'(z, d)$. It follows that the derivative of φ is given by

$$D_z\varphi(z) = \underbrace{D_y\Lambda(y(z), z, r)}_{=0}\, y'(z,\ .\ ) + D_z\Lambda(y(z), z, r) = D_z\Lambda(y(z), z, r). \tag{22}$$

Hence, the following observation holds:

Proposition 2 *The function φ is differentiable everywhere. To solve the linear programs (P) and (D) it suffices to find a point z such that $D_z\varphi(z) = 0$.*

The proposition is evident as $D_z\varphi(z) = 0$ implies $D_z\Lambda(y(z), z, r) = 0$ and by definition of $y(z)$, also $D_y\Lambda(y(z), z, r) = 0$. #

$\boxed{3.2}$ The Structure of the Implicit Function φ

We consider the case where $\nabla_y^2\Lambda(y(z), z, r) \succ 0$. In this case, $Y(z) = \{y(z)\}$ contains exactly one element, and by the implicit function theorem, its total derivative $D_z y(z) =: \dot{y}(z)$ exists. Taking the derivative with respect to z of the equation $\nabla_y\Lambda(y(z), z, r) \equiv 0$ yields

$$D_y^2\Lambda(y(z), z, r)\dot{y}(z) + D_z(\nabla_y\Lambda(y(z), z, r)) = 0.$$

The second term on the left hand side is given by $D_z(\nabla_y\Lambda(y(z), z, r)) = A\Sigma$. We obtain

$$\dot{y}(z) = -(D_y^2\Lambda(y(z), z, r))^{-1}A\Sigma = -\frac{1}{r}(A\Sigma A^T)^{-1}A\Sigma.$$

From this and (22) we derive

$$\begin{aligned} D^2\varphi(z) &= D_y(\nabla_z\Lambda(y(z), z, r))\dot{y}(z) + D_z^2\Lambda(y(z), z, r) \\ &= -\frac{1}{r}\Sigma A^T(A\Sigma A^T)^{-1}A\Sigma - \frac{1}{r}(I - \Sigma) \preceq 0. \end{aligned} \tag{23}$$

The piecewise linear function $\nabla\varphi$ is differentiable almost everywhere. Whenever it is differentiable its derivative satisfies relation (23). This confirms the earlier observation that φ is concave for all $z \in \mathbb{R}^n$ and all $r > 0$. Due to the piecewise linear-quadratic structure of φ it follows that φ is unbounded above when the primal linear program (P) does not have an optimal solution. (Indeed, if φ is bounded above, due to the piecewise quadratic structure it must have a maximum z^{opt}. Since $\nabla\varphi(z^{opt}) = 0$ it follows that z^{opt} solves (P) which is a contradiction.)

Observe that $\Sigma A^T(A\Sigma A^T)^{-1}A\Sigma = \Sigma$ when there are exactly $e^T\sigma = m$ linearly independent columns a_i of A with $\sigma_i = 1$. In this case we obtain

$$D^2\varphi(z) = -\frac{1}{r}I. \tag{24}$$

For such points, the Powell-update rule (see [18]) for z

$$z^{k+1} = z^k + r\nabla\varphi(z^k) = z^k + r\nabla_z\Lambda(y(z^k), z^k, r) = (r(A^Ty(z^k) - c) + z^k)^+$$

coincides with the Newton step for maximizing φ. When $e^T\sigma > m$ the matrix $D^2\varphi(z)$ is not invertible. In this case $\Sigma A^T(A\Sigma A^T)^{-1}A\Sigma$ is a projection matrix and $D^2\varphi(z)$ has the eigenvalue zero of multiplicity $e^T\sigma - m$, and the eigenvalue $-\frac{1}{r}$ of multiplicity $n + m - e^T\sigma$. This in turn implies that the Powell-update Δz is too short, a line search minimizing the unknown distance $\|z + \alpha\Delta z - z^{opt}\|_2$ would return a step $\alpha\Delta z$ with $\alpha \geq 1$.

Remark 2 *If (24) was true for all $z \in \mathbb{R}^n$, the Powell-update would return an optimal solution z^{opt} of (P) in one step. Of course, this is generally not the case. However, when (P) and (D) have unique optimal solutions z^{opt} and y^{opt}, z is fixed, and r is sufficiently large, say $r \geq \bar{r}$, then $y(z)$ is close to y^{opt}. Then, each inactive constraint $\bar{i}$ of (D) with $a_{\bar{i}}^Ty^{opt} < c_{\bar{i}}$ induces an inactive index $\bar{i}$ with $a_{\bar{i}}^Ty(z) - c_{\bar{i}} + \frac{z}{r} < 0$. The remaining m indices must be active, so that (24) holds at z. In fact, (24) holds on the entire line segment $[z, z^{opt}]$ and the Powell-update does return the optimal solution z^{opt} of (P) in one step.*

The closeness of $y(z)$ to y^{opt} follows in a straightforward fashion from Pietrzykowskis theorem (see [17] or Thm.11.1.5 in [11]) which states that for a constrained problem with a strict (local) minimizer, the minimizers of the penalty problem converge to the minimizer of the constrained problem. Here, the perturbation $\frac{z}{r}$ of the constraints tends to zero for large r, and uniqueness of y, z allows the use of the implicit function theorem.

We summarize the results of this section in Proposition 3.

Proposition 3 *The function φ is concave, piecewise linear-quadratic, and differentiable for all $z \in \mathbb{R}^n$ and all $r > 0$; its second derivative multiplied by "$-r$" is an orthogonal projection whenever it is defined. (P) has an optimal solution if, and only if, φ has a maximum. The latter is the case if, and only if, φ is bounded above. In this case each maximizer of φ is an optimal solution of the linear program (P).*

3.3 Conjugate Functions of φ

The Powell update for z is closely related to the Newton step for maximizing φ. As φ is piecewise quadratic, the complexity of Newton's method for maximizing φ can again be related to the conjugate function of $-\varphi$. We recall the definition of the implicit function φ,

$$\varphi(z) = \min_{y\in\mathbb{R}^m}\{\Lambda(y, z, r)\}.$$

We assume for the moment that the set of optimal solutions of (D) is bounded. To simplify the notation we also assume $r = 1$ from now on; (this can be done without loss of generality). We obtain

$$\varphi(z) = \min_{y\in\mathbb{R}^m} \Big\{ -b^Ty + \frac{1}{2}\sum_{i=1}^n((a_i^Ty - c_i + z_i)^+)^2 - z_i^2\Big\}.$$

Since φ is concave the convex conjugate function of $-\varphi$ is given by

$$(-\varphi)^*(\tilde{z}) = \max_{z\in\mathbb{R}^n} \Big\{\tilde{z}^T z + \min_{y\in\mathbb{R}^m} \Big\{ -b^T y + \frac{1}{2}\sum_{i=1}^n ((a_i^T y - c_i + z_i)^+)^2 - z_i^2\Big\}\Big\}. \tag{25}$$

For a given $\tilde{z} \in \mathbb{R}^n$ we define the function $l = l_{\tilde{z}}$ of the variables y and z by

$$l(y,z) = \tilde{z}^T z - b^T y + \frac{1}{2}\sum_{i=1}^n (((a_i^T y - c_i + z_i)^+)^2 - z_i^2).$$

As noted before, l is convex with respect to y and concave with respect to z. Since the set of optimal solutions of (D) is bounded, there does not exist a $y \neq 0$ with $b^T y \geq 0$ and $A^T y \leq 0$. This implies that $\lim_{\|y\|\to\infty} l(y,z) = \infty$. Now assume that $\tilde{z}$ is given such that there exists a y^0 with $A^T y^0 < c - \tilde{z}$. Assume $A^T y^0 \leq c - \tilde{z} - \epsilon e$ for some $\epsilon > 0$. Then, when $z_i \to +\infty$, the i-th component in l can be bounded above by

$$\begin{aligned} \tilde{z}_i z_i + \frac{1}{2}(((a_i^T y - c_i + z_i)^+)^2 - z_i^2) &= \tilde{z}_i z_i + \frac{1}{2}((a_i^T y - c_i + z_i)^2 - z_i^2) \\ &\leq \frac{1}{2}(a_i^T y - c_i)^2 - \epsilon z_i \to -\infty. \end{aligned}$$

For $z_i \to -\infty$, the i-th component in l tends to $-\infty$ as well. Hence, $\lim_{\|z\|\to\infty} l(y,z) = -\infty$. Hence, assumptions $(H1)$ to $(H4)$ of Theorem VII,4.3.1 in [10] are satisfied, and there exists a saddle point of $l = l_{\tilde{z}}$ so that the order of the minimization and the maximization may be interchanged. We then obtain from (25)

$$(-\varphi)^*(\tilde{z}) = \min_{y\in\mathbb{R}^m} \Big\{ -b^T y + \max_{z\in\mathbb{R}^n} \Big\{\tilde{z}^T z + \frac{1}{2}\sum_{i=1}^n ((a_i^T y - c_i + z_i)^+)^2 - z_i^2\Big\}\Big\}. \tag{26}$$

Let $\hat{c} = \hat{c}(y) := A^T y - c$. The inner maximization in (26) with respect to z then implies

$$\tilde{z} = z - (\hat{c} + z)^+,$$

or, equivalently,

$$z_i = \begin{cases} \tilde{z}_i & \text{if } \tilde{z}_i < -\hat{c}(y)_i, \\ \geq \tilde{z}_i & \text{if } \tilde{z}_i = -\hat{c}(y)_i, \\ \text{undefined} & \text{if } \tilde{z}_i > -\hat{c}(y)_i. \end{cases}$$

Hence, the maximum is finite if, and only if, $\hat{c}(y) \leq -\tilde{z}$. Note that in case of $\hat{c}(y)_i = -\tilde{z}_i$ we have

$$\tilde{z}_i z_i + \frac{1}{2}\left(((a_i^T y - c_i + z_i)^+)^2 - z_i^2\right) = \frac{1}{2}\tilde{z}_i^2$$

for all $z_i \geq \tilde{z}_i$. Hence, we may replace $z_i = \tilde{z}_i$ for all i, and the function $(-\varphi)^*$ reduces to

$$\begin{aligned} (-\varphi)^*(\tilde{z}) &= \min_{y:\ A^T y - c \leq -\tilde{z}} \Big\{ -b^T y + \tilde{z}^T\tilde{z} + \frac{1}{2}\sum_{i=1}^n (((a_i^T y - c_i + \tilde{z}_i)^+)^2 - \tilde{z}_i^2)\Big\} \\ &= \min_{y:\ A^T y - c \leq -\tilde{z}} \Big\{ -b^T y + \frac{1}{2}\tilde{z}^T\tilde{z}\Big\} = \frac{1}{2}\tilde{z}^T\tilde{z} - \max_{y:\ A^T y \leq c - \tilde{z}} \{b^T y\}. \end{aligned} \tag{27}$$

This function is piecewise quadratic, but not differentiable everywhere since $-\varphi$ is not strictly convex (see again Theorem 26.3 in [20]). Note that the optimal value (not the optimal solution) of the maximization problem in (27) is a continuous function of the data $(A, b, c, \tilde{z})$ whenever it is finite.

The conjugate function of $(-\varphi)^*$ in (27) is given by

$$\begin{aligned}(-\varphi)^{**}(z) &= \max_{\tilde{z}} \left\{ z^T\tilde{z} - \frac{1}{2}\tilde{z}^T\tilde{z} + \max_y\{b^Ty \mid A^Ty \le c - \tilde{z}\}\right\} \\ &= -\min_{\tilde{z},y}\left\{ -b^Ty - z^T\tilde{z} + \frac{1}{2}\tilde{z}^T\tilde{z} \mid A^Ty \le c - \tilde{z}\right\}.\end{aligned}$$

We thus obtain another representation of $(-\varphi)^{**}(z) = -\varphi(z)$ as a solution of a convex quadratic program with linear constraints.

Since φ is not convex but concave, the segment $[0, \tilde{z}^0]$ on which the gradient of φ^* corresponds to the generalized Newton path of φ is given by

$$-\tilde{z}^0 = \nabla\varphi(z^0) = (A^Ty(z^0) - c + z^0)^+ - z^0 \ge A^Ty(z^0) - c.$$

We write this as $A^Ty(z^0) \le c - \tilde{z}^0$. By our assumption, $A^Ty \le c$ has a feasible solution and by convexity, $A^Ty \le c - t\tilde{z}^0$ has a feasible solution for $t \in [0, 1]$, so that formula (27) is applicable along the line $t\tilde{z}^0$ for $t \in [0, 1]$.

We consider the polyhedra (in the $\tilde{z}$-space) in which φ^* is quadratic. These polyhedra are bounded by the manifolds at which the active indices of strictly complementary solutions y of $\max_{y:\ A^Ty\le c-\tilde{z}}\ \{b^Ty\}$ in (27) are changing. Unfortunately, there may be exponentially many points along the line $c - \tilde{z}$ where φ^* changes the quadratic representation.

Let $z \in \mathbb{R}^n$ be given in *general position* such that

$$(P_1) \qquad \text{minimize} \ \ (c + z)^Tx \ \ \text{s.t.} \ \ x \in \mathcal{P}$$

has a finite optimal solution, i.e. such that

$$(D_1) \qquad \text{maximize} \ \ b^Ty \ \ \text{s.t.} \ \ y \in \mathcal{D}_1 := \{y \ \mid \ A^Ty \le c + z\}$$

is feasible. In this case, (P_1) and (D_1) also have a unique optimal primal dual solution.

Consider the function $\phi : [0, 1] \to \mathbb{R}$ defined by

$$\phi(t) := \underbrace{\min\{(c + tz)^Tx \mid \ Ax = b, \ x \ge 0\}}_{(P_t)} = \underbrace{\max\{b^Ty \mid \ A^Ty \le c + tz\}}_{(D_t)}. \tag{28}$$

As indicated, we refer to the parameterized problems by (P_t) and (D_t). The function ϕ is concave and piecewise linear.

Concavity follows directly from the definition of (D_t); if $y(t)$ is an optimal solution for (D_t), then $\lambda y(t_1) + (1 - \lambda)y(t_2)$ is feasible for $(D_{\lambda t_1+(1-\lambda)t_2})$, and hence the optimal value $\phi(\lambda t_1 + (1 - \lambda)t_2)$ is at least $\lambda\phi(t_1) + (1 - \lambda)\phi(t_2)$. #

Following the generalized Newton path for φ is identical to following the path of ϕ, and as shown in [1], this path may have an exponential number of linear segments.

4 Concluding Remark

We recalled the equivalence of minimizing a certain convex, differentiable, piecewise linear function f with the problem of solving a linear program. We defined a generalized Newton path for minimizing f. This path is piecewise linear. The gradient $\nabla f(z)$ of this path forms a straight line from $\nabla f(z^0)$ to zero. We therefore considered the convex conjugate function f^* of f. The number of piecewise quadratic segments of the implicit function f^* along a given line therefore corresponds to the number of (generalized) Newton steps with line search for minimizing f. Closely related is another implicit function defined by the augmented Lagrangian. This function has a slightly different structure, and there are known examples where Newton's method for minimizing this function may take an exponential number of steps. While the discussion in this paper concentrated on linear programs, similar considerations seem possible for convex quadratic objective functions.

Acknowledgment

The second author would like to thank Michael Saunders and Franz Rendl for very stimulating discussions leading to this paper and to Kurt Anstreicher for helpful comments on an earlier version of this paper. Both authors are indebted to two anonymous referees for valuable comments and criticism.

References

[1] N. Amenta and G. Ziegler, Shadows and slices of polytopes, in *Proceedings of the Twelfth Annual Symposium on Computational Geometry*, ACM Press, Philadelphia, P.A, 1996, pp. 10–19.

[2] D.P. Bertsekas, Nonlinear Programming (2nd edition), Athena Scientific, Nashua, N.H, 1999.

[3] F.H. Clarke, Optimization and Nonsmooth Analysis (2nd edition), Classics in Applied Mathematics 5, SIAM, Philadelphia, P.A, 1990.

[4] A.R. Conn, N.I.M. Gould, A. Sartenaer and P.L. Toint, Convergence properties of an augmented lagrangian algorithm for optimization with a combination of general equality and nonlinear constraints, *SIAM J. Optim.* 6 (1996) 674–703.

[5] A.R. Conn, N.I.M. Gould and P.L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM J. Numerical Anal.* 28 (1991) 545–572.

[6] M.C. Ferris, T.S. Munson and D. Ralph, A homotopy method for mixed complementarity problems based on the PATH solver, in *Numerical Analysis 1999, Research Notes in Mathematics*, D.F. Griffiths and G.A. Watson (eds.), Chapman and Hall, London, 2000, pp. 143–167.

[7] M.P. Friedlander and M.A. Saunders, A globally convergent linearly constrained Lagrangian method for nonlinear optimization, *SIAM J. Optim.* (2006) to appear.

[8] G.H. Golub and C.F. Van Loan, Matrix Computations, 2nd Edn., The John Hopkins University Press, Baltimore, M.D, 1989.

[9] M.R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.* 4 (1969) 303–320.

[10] J.B. Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithms, Volume 1, Springer, Berlin, 1996.

[11] F. Jarre and J. Stoer, Mathematische Optimierung, Springer, Berlin, 2004.

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

[13] L.G. Khachiyan, A polynomial algorithm in linear programming, *Soviet Mathematics Doklady* 20 (1979) 191–194.

[14] M. Kojima, N. Megiddo and S. Mizuno, Theoretical convergence of large-step primaldual interior point algorithms for linear programming, *Math. Program.* 59 (1992) 1–21.

[15] N. Megiddo, Linear programming in linear time when the dimension is fixed, *J. ACM* 31 (1984) 114–127.

[16] R. Monteiro and T. Tsuchiya, A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming, *SIAM J. Optim.* 13(4) (2003) 1054–1079.

[17] T. Pietrzykowski, The potential method for conditional maxima in the locally compact metric spaces, *Numer. Math.* 14(4) (1970) 325–329.

[18] M.J.D. Powell, A method for nonlinear constraints, in *minimization problems*, R. Fletcher (ed), Optimization, Academic Press, New York, 1969, pp. 283–298.

[19] F. Rendl, Solving Semidefinite Programs using Bundle Methods and the Augmented Lagrangian Approach, Plenary talk at Veszprem Optimization Conference: Advanced Algorithms (VOCAL), Veszprem, Hungary, December 13-15, 2004.

[20] R.T. Rockafellar, *Convex Analysis*, Princeton mathematical series 28, Princeton University Press, Princeton, N.J, 1970.

[21] E. Tardos, A strongly polynomial algorithm to solve combinatorial linear programs, *Oper. Res.* 34(2) (1986) 250–256.

[22] S. Vavasis and Y. Ye, A primal dual accelerated interior point method whose running time depends only on A, *Math. Program.* 74 (1996) 79–120.

[23] Y. Ye, M.J. Todd and S. Mizuno, An $O(\#nL)$-iteration homogeneous and self-dual linear programming algorithm, *Math. Oper. Res.* 19 (1994) 53–67.

Manuscript received 21 December 2005
revised 13 September 2006
accepted for publication 26 September 2006

Katrin Hauk
Institut für Mathematik, Universität Düsseldorf, Universitätsstr. 1,
D–40225 Düsseldorf, Federal Republic of Germany
E-mail address: hauk@opt.uni-duesseldorf.de

Florian Jarre
Institut für Mathematik, Universität Düsseldorf, Universitätsstr. 1,
D–40225 Düsseldorf, Federal Republic of Germany
E-mail address: jarre@opt.uni-duesseldorf.de