



NEWTON FLOW AND INTERIOR POINT METHODS IN LINEAR PROGRAMMING

JEAN-PIERRE DEDIEU

*MIP. Département de Mathématique, Université Paul Sabatier,
31062 Toulouse cedex 04, France*

MIKE SHUB

*Department of Mathematics, University of Toronto,
100 St. George Street, Toronto, Ontario M5S 3G3, Canada*

Received January 12, 2004; Revised June 17, 2004

We study the geometry of the central paths of linear programming theory. These paths are the solution curves of the Newton vector field of the logarithmic barrier function. This vector field extends to the boundary of the polytope and we study the main properties of this extension: continuity, analyticity, singularities.

Keywords: Linear programming; interior point method; central path; Newton vector field; extension.

1. Introduction

In this paper we take up once again the subject of the geometry of the central paths of linear programming theory. We study the boundary behavior of these paths as in [Megiddo & Shub, 1989], but from a different perspective and with a different emphasis. Our main goal will be to give a global picture of the central paths even for degenerate problems as solution curves of the Newton vector field, $N(x)$, of the logarithmic barrier function which we describe below. See also [Bayer & Lagarias, 1989a, 1989b, 1991]. The Newton vector field extends to the boundary of the polytope. It has the properties that it is tangent to the boundary and restricted to any face of dimension i it has a unique source with unstable manifold dimension equal to i , the rest of the orbits tending to the boundary of the face. Every orbit tends either to a vertex or one of these sources in a face. See Corollary 4.1. This highly cellular structure of the flow lends itself to the conjecture that the total curvature of these central paths may be linearly bounded by the dimension n of the polytope. The orbits may be relatively straight,

except for orbits which come close to an orbit in a face of dimension i which itself comes close to a singularity in a boundary face of dimension less than i . This orbit is then forced to turn almost parallel to the lower dimensional face so its tangent vector may be forced to turn as well. See the two figures at the end of this paper. As this process involves a reduction of the dimension of the face it can only happen for the dimension of the polytope times. So our optimistic conjecture is that the total curvature of a central path is $O(n)$. We have verified the conjecture in an average sense in [Dedieu *et al.*]. It is not difficult to give an example showing that $O(n)$ is the best possible for the worst case. Such an example is worked out in [Megiddo & Shub, 1989]. The average behavior may be however much better. Ultimately we hope that an understanding of the curvature of the central paths may contribute to the analysis of algorithms which use them. In [Vavasis & Ye, 1996] the authors explore similar structure to give an algorithm whose running time depends only on the polytope.

We prove in Corollary 4.1 that the extended vector field is Lipschitz on the closed polytope.

Under a genericity hypothesis we prove in Theorem 5.1 that it extends to be real analytic on a neighborhood of the polytope. Under the same genericity hypothesis we prove in Theorem 5.2 that the singularities are all hyperbolic. The eigenvalues of $-N(x)$ at the singularities are all +1 tangent to the face and -1 transversal to the face. In dynamical systems terminology the vector field is Morse–Smale. The vertices are the sinks. Finally, we mention that in order to prove that $N(x)$ always extends continuously to the boundary of the polytope we prove Lemma 4.2 which may be of independent interest about the continuity of the Moore–Penrose inverse of a family of linear maps of variable rank.

2. The Central Path is a Trajectory of the Newton Vector Field

Linear programming problems are frequently presented in different formats. We will work with one of them here which we find convenient. The polytopes defined in one format are usually affinely equivalent to the polytopes defined in another. So we begin with a discussion of Newton vector fields and how they transform under affine equivalence. This material is quite standard. An excellent source for this fact and linear programming in general is [Renegar, 2001].

Let \mathbb{Q} be an affine subspace of \mathbb{R}^n (or a Hilbert space if you prefer, in which case, assume \mathbb{Q} is closed). Denote the tangent space of \mathbb{Q} by \mathbb{V} . Suppose that U is an open subset of \mathbb{Q} . Let $f : U \rightarrow \mathbb{R}$ be twice continuously differentiable. The derivative $Df(x)$ belongs to $L(\mathbb{V}, \mathbb{R})$, the linear maps from \mathbb{V} to \mathbb{R} . So $Df(x)$ defines a map from U to $L(\mathbb{V}, \mathbb{R})$. The second derivative $D^2f(x)$ is an element of $L(\mathbb{V}, L(\mathbb{V}, \mathbb{R}))$. Thus $D^2f(x)$ is a linear map from a vector space to another isomorphic space and $D^2f(x)$ may be invertible.

Definition 2.1. If f is as above and $D^2f(x)$ is invertible we define the Newton vector field, $N_f(x)$ by

$$N_f(x) = -(D^2f(x))^{-1}Df(x).$$

Note that if \mathbb{V} has a nondegenerate inner product $\langle \cdot, \cdot \rangle$ then the gradient of f , $\text{grad } f(x) \in \mathbb{V}$, and Hessian, $\text{hess } f(x) \in L(\mathbb{V}, \mathbb{V})$, are defined by

$$Df(x)u = \langle u, \text{grad } f(x) \rangle$$

and

$$D^2f(x)(u, v) = \langle u, (\text{hess } f(x))v \rangle.$$

It follows then that $N_f(x) = -(\text{hess } f(x))^{-1}\text{grad } f(x)$.

Now let A be an affine map from \mathbb{P} to \mathbb{Q} whose linear part L is an isomorphism. Suppose U_1 is open in \mathbb{P} and $A(U_1) \subseteq U$. Let $g = f \circ A$.

Proposition 2.1. *A maps the solution curves of N_g to the solution curves of N_f .*

Proof. By the chain rule $Dg(y) = Df(A(y))L$ and

$$D^2g(y)(u, v) = D^2f(A(y))(Lu, Lv).$$

So $u = N_g(y)$ if and only if $D^2g(y)(u, v) = -Dg(y)(v)$ for all v if and only if $D^2f(A(y))(Lu, Lv) = -Df(A(y))Lv$ for all v , i.e. $N_f(A(y)) = L(u)$ or $LN_g(y) = N_fA(y)$. This last is the equation expressing that the vector field N_f is the push forward by the map of the vector field N_g and hence the solution curves of the N_g field are mapped by A to the solution curves of N_f . ■

Now we make explicit the linear programming format used in this paper, define the central paths and relate them to the Newton vector field of the logarithmic barrier function.

Let \mathcal{P} be a compact polytope in \mathbb{R}^n defined by m affine inequalities

$$A_i x \geq b_i, \quad 1 \leq i \leq m.$$

Here $A_i x$ denotes the matrix product of the row vector $A_i = (a_{i1}, \dots, a_{in})$ by the column vector $x = (x_1, \dots, x_n)^T$, A is the $m \times n$ matrix with rows A_i and we assume $\text{rank } A = n$. Given $c \in \mathbb{R}^n$, we consider the linear programming problem

$$(LP) \quad \min_{\substack{A_i x \geq b_i \\ 1 \leq i \leq m}} \langle c, x \rangle.$$

Let us denote by

$$f(x) = \sum_{i=1}^m \ln(A_i x - b_i)$$

($\ln(s) = -\infty$ when $s \leq 0$) the logarithmic barrier function associated with the description $Ax \geq b$ of \mathcal{P} . The barrier technique considers the family of

nonlinear convex optimization problems

$$(LP(t)) \quad \min_{x \in \mathbb{R}^n} t\langle c, x \rangle - f(x)$$

with $t > 0$. The objective function

$$f_t(x) = t\langle c, x \rangle - f(x)$$

is strictly convex, smooth, and satisfies

$$\lim_{\substack{x \rightarrow \partial \mathcal{P} \\ x \in \text{Int } \mathcal{P}}} f_t(x) = \infty.$$

Thus, there exists a unique optimal solution $\gamma(t)$ to $(LP(t))$ for any $t > 0$. This curve is called the central path of our problem. Let us denote as D_x the $m \times m$ diagonal matrix $D_x = \text{Diag}(A_i x - b_i)$. This matrix is nonsingular for any $x \in \text{Int } \mathcal{P}$. We also let $e = (1, \dots, 1)^T \in \mathbb{R}^m$,

$$g(x) = \text{grad } f(x) = \sum_{i=1}^m \frac{A_i^T}{A_i x - b_i} = A^T D_x^{-1} e$$

and

$$h(x) = \text{hess } f(x) = -A^T D_x^{-2} A.$$

Since f_t is smooth and strictly convex the central path is given by the equation $\text{grad } f_t(\gamma(t)) = 0$ i.e.

$$g(\gamma(t)) = tc, \quad t > 0.$$

When $t \rightarrow 0$, the limit of $\gamma(t)$ is given by

$$-f(\gamma(0)) = \min_{x \in \mathbb{R}^n} -f(x).$$

It is called the analytic center of \mathcal{P} and denoted by $c_{\mathcal{P}}$.

Lemma 2.1. $g: \text{Int } \mathcal{P} \rightarrow \mathbb{R}^n$ is real analytic and invertible. Its inverse is also real analytic.

Proof. For any $c \in \mathbb{R}^n$ the optimization problem

$$\min_{x \in \mathbb{R}^n} \langle c, x \rangle - f(x)$$

has a unique solution in $\text{Int } \mathcal{P}$ because the objective function is smooth, strictly convex and \mathcal{P} is compact. Thus $g(x) = c$ has a unique solution that is g bijective. We also notice that, for any x , $Dg(x)$ is nonsingular. Thus g^{-1} is real analytic by the inverse function theorem. ■

According to this lemma, the central path is the inverse image by g of the ray $c\mathbb{R}_+$. When c varies

in \mathbb{R}^n we obtain a family of curves. Our aim in this paper is to investigate the structure of this family.

For a subspace $\mathcal{B} \subset \mathbb{R}^m$ we denote by $\Pi_{\mathcal{B}}$ the orthogonal projection $\mathbb{R}^m \rightarrow \mathcal{B}$. Let b_1, \dots, b_r be a basis of \mathcal{B} and let us denote by B the $m \times r$ matrix with columns of the vectors b_i . Then $\Pi_{\mathcal{B}}$, also denoted Π_B , is given by $\Pi_B = B(B^T B)^{-1} B^T = BB^\dagger$ (B^\dagger is the generalized inverse of B equal to $(B^T B)^{-1} B^T$ because B is injective).

Definition 2.2. The Newton vector field associated with g is

$$\begin{aligned} N(x) &= -Dg(x)^{-1} g(x) \\ &= (A^T D_x^{-2} A)^{-1} A^T D_x^{-1} e \\ &= A^\dagger D_x \Pi_{D_x^{-1} A} e. \end{aligned}$$

It is defined and analytic on $\text{Int } \mathcal{P}$.

Note that the expression $A^\dagger D_x \Pi_{D_x^{-1} A} e$ is defined for all $x \in \mathbb{R}^n$ for which $A_i x - b_i$ is not equal to 0 for all i . Thus $N(x)$ is defined by the rational expression in Definition 2.2 for almost all $x \in \mathbb{R}^n$. Later we will prove that this rational expression has a continuous extension to all \mathbb{R}^n .

Lemma 2.2. The central paths $\gamma(t)$, $c \in \mathbb{R}^n$, are the trajectories of the vector field $-N(x)$.

Proof. A central path is given by

$$g(\gamma(t)) = tc, \quad t > 0,$$

for a given $c \in \mathbb{R}^n$. Let us change variable: $t = \exp s$ and $d(s) = \gamma(t)$ with $s \in \mathbb{R}$. Then

$$g(d(s)) = \exp(s)c, \quad s \in \mathbb{R},$$

so that

$$\frac{d}{ds} g(d(s)) = \exp(s)c = g(d(s)).$$

Let us denote $\dot{d}(s) = (d/ds)d(s)$. We have

$$\frac{d}{ds} g(d(s)) = Dg(d(s))\dot{d}(s)$$

thus

$$\dot{d}(s) = Dg(d(s))^{-1} g(d(s)) = -N(d(s))$$

and $d(s)$ is a trajectory of the Newton vector field. Conversely, if $\dot{d}(s) = -N(d(s)) = Dg(d(s))^{-1} g(d(s))$, $s \in \mathbb{R}$, then

$$\frac{d}{ds} g(d(s)) = Dg(d(s))\dot{d}(s) = g(d(s))$$

so that

$$g(d(s)) = \exp(s)g(d(0))$$

which is the central path related to $c = g(d(0))$. \blacksquare

Remark 2.1. The trajectories of $N(x)$ and $-N(x)$ are the same with time reversed. As $t \rightarrow \infty$, $\gamma(t)$ tends to the optimal points of the linear programming problem. So we are interested in the positive time trajectories of $-N(x)$.

Lemma 2.3. *The analytic center $c_{\mathcal{P}}$ is the unique singular point of the Newton vector field $N(x)$, $x \in \text{Int } \mathcal{P}$.*

Proof. $N(x) = 0$ if and only if $g(x) = 0$, that is $x = c_{\mathcal{P}}$. \blacksquare

3. An Analytic Expression for the Newton Vector Field

In this section we compute an analytic expression for $N(x)$ which will be useful later. For any subset $K_n \subset \{1, \dots, m\}$, $K_n = \{k_1 < \dots < k_n\}$, we denote by A_{K_n} the $n \times n$ submatrix of A with rows A_{k_1}, \dots, A_{k_n} , by b_{K_n} the vector in \mathbb{R}^n with coordinates b_{k_1}, \dots, b_{k_n} , and by u_{K_n} the unique solution of the system $A_{K_n}u_{K_n} = b_{K_n}$ when the matrix A_{K_n} is nonsingular. With these notations we have:

Proposition 3.1. *For any $x \in \text{Int } \mathcal{P}$,*

$$N(x) = \frac{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det A_{K_n} \neq 0}} (x - u_{K_n})(\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det A_{K_n} \neq 0}} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}$$

Proof. Let us denote $\Pi = \prod_{l=1}^m (A_l x - b_l)$ and $\Pi_k = \prod_{l \neq k} (A_l x - b_l)$. We already know (Definition 2.2) that $N(x) = (A^T D_x^{-2} A)^{-1} A^T D_x^{-1} e$ with

$$\begin{aligned} (A^T D_x^{-2} A)_{ij} &= \sum_{k=1}^m \frac{a_{ki} a_{kj}}{(A_k x - b_k)^2} \\ &= \frac{1}{\Pi^2} \sum_{k=1}^m a_{ki} a_{kj} \Pi_k^2 \\ &= \frac{1}{\Pi^2} X_{ij} \end{aligned}$$

where X is the $n \times n$ matrix given by $X_{ij} = \sum_{k=1}^m a_{ki} a_{kj} \Pi_k^2$. Moreover

$$\begin{aligned} (A^T D_x^{-1} e)_i &= \sum_{k=1}^m \frac{a_{ki}}{A_k x - b_k} \\ &= \frac{1}{\Pi} \sum_{k=1}^m a_{ki} \Pi_k \\ &= \frac{1}{\Pi} V_i \end{aligned}$$

where V is the n vector given by $V_i = \sum_{k=1}^m a_{ki} \Pi_k$. This gives

$$N(x) = \Pi X^{-1} V.$$

To compute X^{-1} we use Cramer's formula: $X^{-1} = \text{cof}(X)^T / \det(X)$ where $\text{cof}(X)$ denotes the matrix of cofactors: $\text{cof}(X)_{ij} = (-1)^{i+j} \det(X^{ij})$ with X^{ij} the $(n-1) \times (n-1)$ matrix obtained by deleting in X the i th row and j th column. We first compute $\det X$. We have

$$\det X = \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) X_{1\sigma(1)} \cdots X_{n\sigma(n)}$$

where \mathbb{S}_n is the group of permutations of $\{1, \dots, n\}$ and $\epsilon(\sigma)$ the signature of σ . Thus

$$\begin{aligned} \det X &= \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) \prod_{j=1}^n \sum_{k_j=1}^m a_{k_j j} a_{k_j \sigma(j)} \Pi_{k_j}^2 \\ &= \sum_{\substack{1 \leq k_j \leq m \\ 1 \leq j \leq n}} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 a_{k_1 1} \cdots a_{k_n n} \\ &\quad \times \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) a_{k_1 \sigma(1)} \cdots a_{k_n \sigma(n)} \\ &= \sum_{\substack{1 \leq k_j \leq m \\ 1 \leq j \leq n}} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 a_{k_1 1} \cdots a_{k_n n} \det A_{k_1 \dots k_n} \end{aligned}$$

where $A_{k_1 \dots k_n}$ is the matrix with rows $A_{k_1} \dots A_{k_n}$. When two or more indices k_j are equal the corresponding coefficient $\det A_{k_1 \dots k_n}$ is zero. For this reason, instead of this sum taken for n independent indices k_j we consider a set $K_n \subset \{1, \dots, m\}$, $K_n = \{k_1 < \dots < k_n\}$, and all the possible permutations $\sigma \in \mathbb{S}(K_n)$. We obtain

$$\begin{aligned}\det X &= \sum_{\substack{K_n \subset \{1, \dots, m\} \\ \sigma \in \mathbb{S}(K_n)}} \Pi_{\sigma(k_1)}^2 \cdots \Pi_{\sigma(k_n)}^2 \\ &\quad \times a_{\sigma(k_1)1} \cdots a_{\sigma(k_n)n} \det A_{\sigma(k_1) \dots \sigma(k_n)} \\ &= \sum_{K_n \subset \{1, \dots, m\}} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 \\ &\quad \times \sum_{\sigma \in \mathbb{S}(K_n)} \epsilon(\sigma) a_{\sigma(k_1)1} \cdots a_{\sigma(k_n)n} \det A_{k_1 \dots k_n} \\ &= \sum_{K_n \subset \{1, \dots, m\}} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 (\det A_{K_n})^2.\end{aligned}$$

Note that, for any $l = 1, \dots, m$, the product $\Pi_{k_1}^2 \cdots \Pi_{k_n}^2$ contains $(A_l x - b_l)^{2n}$ if $l \notin K_n$ and $(A_l x - b_l)^{2n-2}$ otherwise. For this reason

$$\begin{aligned}\det X &= \Pi^{2n-2} \sum_{K_n \subset \{1, \dots, m\}} (\det A_{K_n})^2 \\ &\quad \times \prod_{l \notin K_n} (A_l x - b_l)^2.\end{aligned}$$

Let us now compute $Y = \text{cof}(X)^T V$. We have

$$\begin{aligned}Y_i &= \sum_{j=1}^n (-1)^{i+j} \det(X^{ji}) V_j \\ &= \sum_{j=1}^n (-1)^{i+j} \det(X^{ji}) \sum_{k=1}^m a_{kj} \Pi_k \\ &= \sum_{k=1}^m \Pi_k \sum_{j=1}^n (-1)^{i+j} \det(X^{ij}) a_{kj}\end{aligned}$$

because X is symmetric. This last sum is the determinant of the matrix with rows $X_1 \dots X_{i-1} A_k X_{i+1} \dots X_n$ so that

$$\begin{aligned}Y_i &= \sum_{k=1}^m \Pi_k \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) X_{1\sigma(1)} \cdots X_{i-1\sigma(i-1)} a_{k\sigma(i)} \\ &\quad \times X_{i+1\sigma(i+1)} \cdots X_{n\sigma(n)}\end{aligned}$$

$$\begin{aligned}&= \sum_{k=1}^m \Pi_k \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) a_{k\sigma(i)} \prod_{\substack{j=1 \\ j \neq i}}^n \sum_{k_j=1}^m a_{k_j j} a_{k_j \sigma(j)} \Pi_{k_j}^2 \\ &= \sum_{k=1}^m \Pi_k \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) a_{k\sigma(i)} \\ &\quad \times \sum_{\substack{k_j=1 \\ 1 \leq j \leq n \\ j \neq i}} a_{k_1 1} a_{k_1 \sigma(1)} \Pi_{k_1}^2 \cdots a_{k_n n} a_{k_n \sigma(n)} \Pi_{k_n}^2 \\ &= \sum_{k=1}^m \Pi_k \sum_{\substack{k_j=1 \\ 1 \leq j \leq n \\ j \neq i}} a_{k_1 1} \cdots a_{k_n n} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 \\ &\quad \times \sum_{\sigma \in \mathbb{S}_n} \epsilon(\sigma) a_{k_1 \sigma(1)} \cdots a_{k \sigma(i)} \cdots a_{k_n \sigma(n)}\end{aligned}$$

which gives

$$\begin{aligned}Y_i &= \sum_{k=1}^m \Pi_k \sum_{\substack{k_j=1 \\ 1 \leq j \leq n \\ j \neq i}} a_{k_1 1} \cdots a_{k_n n} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 \\ &\quad \times \det A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n}.\end{aligned}$$

By a similar argument as before we sum up for any set with $n-1$ elements $K_{n-1} \subset \{1, \dots, m\}$, $K_{n-1} = \{k_1 < \dots < k_{i-1} < k_{i+1} < \dots < k_n\}$ and for any permutation $\sigma \in \mathbb{S}(K_{n-1})$. We obtain as previously

$$\begin{aligned}Y_i &= \sum_{k=1}^m \Pi_k \sum_{K_{n-1} \subset \{1, \dots, m\}} \Pi_{k_1}^2 \cdots \Pi_{k_n}^2 \\ &\quad \times \det A_{k_1 \dots k_{i-1} k_{i+1} \dots k_n}^i \det A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n}\end{aligned}$$

with $A_{k_1 \dots k_{i-1} k_{i+1} \dots k_n}^i$ the matrix with rows A_{k_j} , $j \in K_{n-1}$ and the i th column removed. The quantity $A_l x - b_l$ appears in the product $\Pi_k \Pi_{k_1}^2 \cdots \Pi_{k_n}^2$ with an exponent equal to

- $2n-1$ when $l \neq k$ and $l \notin K_{n-1}$,
- $2n-2$ when $l = k$ and $l \notin K_{n-1}$,
- $2n-3$ when $l \neq k$ and $l \in K_{n-1}$,
- $2n-4$ when $l = k$ and $l \in K_{n-1}$.

In this latter case, two rows of the matrix $A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n}$ are equal and its determinant is zero. Thus, each term $A_l x - b_l$ appears at least $2n-3$

times so that

$$Y_i = \Pi^{2n-3} \sum_{\substack{k=1 \\ K_{n-1}}}^m (A_k x - b_k) \prod_{\substack{l \neq k \\ l \notin K_{n-1}}} (A_l x - b_l)^2 \det A_{k_1 \dots k_{i-1} k_{i+1} \dots k_n}^i \det A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n}.$$

The i th component of the Newton vector field is equal to $N(x)_i = \Pi Y_i / \det X$ so that

$$N(x)_i = \frac{\sum_{\substack{k=1 \\ K_{n-1}}}^m (A_k x - b_k) \prod_{\substack{l \neq k \\ l \notin K_{n-1}}} (A_l x - b_l)^2 \det A_{k_1 \dots k_{i-1} k_{i+1} \dots k_n}^i \det A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n}}{\sum_{K_n} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}.$$

Instead of a sum taken for k and K_{n-1} in the numerator we use a subset $K_n \subset \{1, \dots, m\}$ equal to the union of k and K_{n-1} . Notice that $\det A_{k_1 \dots k_{i-1} k k_{i+1} \dots k_n} = 0$ when $k \in K_{n-1}$ so that this case is not considered. Conversely, for a given $K_n = \{k_1 \dots k_n\}$, we can write it in n different ways as a union of $k = k_j$ and $K_{n-1} = K_n \setminus \{k_j\}$. For these reasons we get

$$N(x)_i = \frac{\sum_{K_n} \left(\sum_{j=1}^n (A_{k_j} x - b_{k_j}) \det A_{K_n}^{ji} \det A_{K_n, i, j} \right) \prod_{l \notin K_n} (A_l x - b_l)^2}{\sum_{K_n} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}$$

with $A_{K_n}^{ji}$ the matrix obtained from A_{K_n} in deleting the j th row and i th column, and $A_{K_n, i, j}$ obtained from A_{K_n} in removing the line A_j , and in reinserting it as the i th line, the other lines remaining with the same ordering. Note that $\det A_{K_n, i, j} = (-1)^{i+j} \det A_{K_n}$ thus

$$N(x)_i = \frac{\sum_{K_n} \left(\sum_{j=1}^n (A_{k_j} x - b_{k_j}) (-1)^{i+j} \det A_{K_n}^{ji} \right) \det A_{K_n} \prod_{l \notin K_n} (A_l x - b_l)^2}{\sum_{K_n} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}.$$

In fact this sum is taken for the sets K_n such that A_{K_n} is nonsingular, otherwise, the coefficient $\det A_{K_n}$ vanishes and the corresponding term is zero.

According to Cramer's formulas, the expression $(-1)^{i+j} \det A_{K_n}^{ji} / \det A_{K_n}$ is equal to $(A_{K_n}^{-1})_{ij}$. Thus

$$\begin{aligned} & \sum_{j=1}^n (A_{k_j} x - b_{k_j}) (-1)^{i+j} \det A_{K_n}^{ji} \\ &= (A_{K_n}^{-1} (A_{K_n} x - b_{K_n}))_i \\ &= x_i - (A_{K_n}^{-1} b_{K_n})_i = x_i - u_{K_n, i}. \end{aligned}$$

We get

$$N(x)_i = \frac{\sum_{K_n} (x_i - u_{K_n, i}) (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}{\sum_{K_n} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}$$

and we are done. ■

4. Extension to the Faces of \mathcal{P}

Our aim is to extend the Newton vector field, defined in the interior of \mathcal{P} , to its different faces.

Let \mathcal{P}_J be the face of \mathcal{P} defined by

$$\begin{aligned}\mathcal{P}_J = \{x \in \mathbb{R}^n : A_i x = b_i \text{ for any } i \in I \\ \text{and } A_i x \geq b_i \text{ for any } i \in J\}.\end{aligned}$$

Here I is a subset of $\{1, 2, \dots, m\}$ containing m_I integers, $J = \{1, 2, \dots, m\} \setminus I$ and $m_J = m - m_I$.

Definition 4.1. The face \mathcal{P}_J is regularly described when the relative interior of the face is given by

$$\begin{aligned}ri - \mathcal{P}_J = \{x \in \mathbb{R}^n : A_i x = b_i \text{ for any } i \in I \\ \text{and } A_i x > b_i \text{ for any } i \in J\}.\end{aligned}$$

The polytope is regularly described when all its faces have this property.

We assume here that \mathcal{P} is regularly described. This definition avoids, for example, in the description of a \mathcal{P}_J a hyperplane defined by two inequalities: $A_i x \geq b_i$ and $A_i x \leq b_i$ instead of $A_i x = b_i$. Note that every face of a regularly described \mathcal{P} has a unique regular description, the set I consists of all indices i such that $A_i x = b_i$ on the face. The affine hull of \mathcal{P}_J is denoted by

$$\begin{aligned}F_J = \{x = (x_1, \dots, x_n)^T \in \mathbb{R}^n : \\ A_i x = b_i \text{ for any } i \in I\}\end{aligned}$$

which is parallel to the vector subspace

$$\begin{aligned}G_J = \{x = (x_1, \dots, x_n)^T \in \mathbb{R}^n : \\ A_i x = 0 \text{ for any } i \in I\}.\end{aligned}$$

We also let

$$\begin{aligned}E_J = \{y = (y_1, \dots, y_m)^T \in \mathbb{R}^m : \\ y_i = 0 \text{ for any } i \in I\}.\end{aligned}$$

E_I is defined similarly.

Let us denote by A_J (resp. A_I) the $m_J \times n$ (resp. $m_I \times n$) matrix whose i th row is A_i , $i \in J$ (resp. $i \in I$). A_J defines a linear operator $A_J : \mathbb{R}^n \rightarrow \mathbb{R}^{m_J}$. We also let

$$b_J : G_J \rightarrow \mathbb{R}^{m_J}, \quad b_J = A_J|_{G_J}$$

so that

$$b_J^T : \mathbb{R}^{m_J} \rightarrow G_J, \quad b_J^T = \Pi_{G_J} A_J.$$

Here, for a vector subspace E , Π_E denotes the orthogonal projection onto E . Let $D_{x,J}$ (resp. $D_{x,I}$) be the diagonal matrix with diagonal entries

$A_i x - b_i$, $i \in J$ (resp. $i \in I$). It defines a linear operator $D_{x,J} : \mathbb{R}^{m_J} \rightarrow \mathbb{R}^{m_J}$.

Since the faces of the polytope are regularly described, for any $x \in ri - \mathcal{P}_J$, $D_{x,J}$ is nonsingular. \mathcal{P}_J is associated with the linear program

$$(LP_J) \quad \min_{x \in \mathcal{P}_J} \langle c, x \rangle.$$

The barrier function

$$f_J(x) = \sum_{i \in J} \ln(A_i x - b_i)$$

is defined for any $x \in F_J$ and finite in $ri - \mathcal{P}_J$ the relative interior of \mathcal{P}_J . The barrier technique considers the family of nonlinear convex optimization problems $(LP_J(t))$

$$\min_{x \in F_J} t \langle c, x \rangle - f_J(x)$$

with $t > 0$. The objective function

$$f_{t,J}(x) = t \langle c, x \rangle - f_J(x)$$

is smooth, strictly convex and

$$\lim_{x \rightarrow \partial \mathcal{P}_J} f_{t,J}(x) = \infty,$$

thus $(LP_J(t))$ has a unique solution $\gamma_J(t) \in ri - \mathcal{P}_J$ given by

$$Df_{t,J}(\gamma_J(t)) = 0.$$

For any $x \in ri - \mathcal{P}_J$, the first derivative of f_J is given by

$$Df_J(x)u = \sum_{i \in J} \frac{A_i u}{A_i x - b_i} = \langle A_J^T D_{x,J}^{-1} e_J, u \rangle$$

with $u \in G_J$ and $e_J = (1, \dots, 1)^T \in \mathbb{R}^{m_J}$. We have

$$\begin{aligned}g_J(x) = \text{grad } f_J(x) &= \Pi_{G_J} A_J^T D_{x,J}^{-1} e_J \\ &= b_J^T D_{x,J}^{-1} e_J.\end{aligned}$$

The second derivative of f_J at $x \in ri - \mathcal{P}_J$ is given by

$$\begin{aligned}D^2 f_J(x)(u, v) &= - \sum_{i \in J} \frac{(A_i u)(A_i v)}{(A_i x - b_i)^2} \\ &= - \langle A_J^T D_{x,J}^{-2} A_J v, u \rangle \\ &= - \langle b_J^T D_{x,J}^{-2} b_J v, u \rangle\end{aligned}$$

for any $u, v \in G_J$ so that

$$Dg_J(x) = \text{hess } f_J(x) = -b_J^T D_{x,J}^{-2} b_J.$$

To \mathcal{P}_J we associate the Newton vector field given by

$$N_J(x) = -Dg_J(x)^{-1} g_J(x), \quad x \in ri - \mathcal{P}_J.$$

We have:

Lemma 4.1. *For any $x \in ri - \mathcal{P}_J$ this vector field is defined and*

$$\begin{aligned} N_J(x) &= (b_J^T D_{x,J}^{-2} b_J)^{-1} b_J^T D_{x,J}^{-1} e_J \\ &= b_J^\dagger D_{x,J} \Pi_{\text{im}(D_{x,J}^{-1} b_J)} e_J \in G_J. \end{aligned}$$

Proof. We first have to prove that $Dg_J(x)$ is nonsingular and that $N_J(x) \in G_J$. This second point is clear. For the first, we take $u \in G_J$ such that $Dg_J(x)u = 0$. This gives $A_J u = b_J u = 0$ which implies $Au = 0$ because $u \in G_J$ that is $A_I u = 0$. Since A is injective we get $u = 0$. By the same argument we see that b_J is injective so that $b_J^T b_J$ is nonsingular. The first expression for $N_J(x)$ comes from the description of g_J and Dg_J . We have

$$\begin{aligned} N_J(x) &= (b_J^T D_{x,J}^{-2} b_J)^{-1} b_J^T D_{x,J}^{-1} e_J \\ &= (b_J^T b_J)^{-1} b_J^T D_{x,J} D_{x,J}^{-1} b_J \\ &\quad \times (b_J^T D_{x,J}^{-2} b_J)^{-1} b_J^T D_{x,J}^{-1} e_J \\ &= b_J^\dagger D_{x,J} \Pi_{\text{im}(D_{x,J}^{-1} b_J)} e_J \in G_J. \quad \blacksquare \end{aligned}$$

The curve $\gamma_J(t)$, $0 < t < \infty$, is the central path of the face \mathcal{P}_J . It is given by

$$\gamma_J(t) \in F_J \quad \text{and} \quad Df_J(\gamma_J(t)) - tc = 0$$

that is

$$x \in F_J, \quad A_J^T D_{x,J}^{-1} e_J - tc \in G_J^\perp \quad \text{and} \quad \gamma_J(t) = x$$

or, projecting on G_J ,

$$\begin{aligned} A_i x = b_i, \quad i \in I, \quad b_J^T D_{x,J}^{-1} e_J - t \Pi_{G_J} c = 0 \\ \text{and} \quad \gamma_J(t) = x. \end{aligned}$$

When $t \rightarrow 0$, $\gamma_J(t)$ tends to the analytic center $\gamma_J(0)$ of \mathcal{P}_J defined as the unique solution of the convex program

$$-f_J(\gamma_J(0)) = \min_{x \in F_J} -f_J(x).$$

The analytic center is also given by

$$A_i x = b_i, \quad i \in I, \quad b_J^T D_{x,J}^{-1} e_J = 0 \quad \text{and} \quad \gamma_J(0) = x$$

so that $\gamma_J(0)$ is the unique singular point of N_J in the face \mathcal{P}_J .

We now investigate the properties of this extended vector field: continuity, derivability and so on. We shall investigate the following abstract problem: for any $y \in \mathbb{R}^m$ we consider the linear operator

$$\mathcal{D}_y : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

given by the $m \times m$ diagonal matrix $\mathcal{D}_y = \text{Diag}(y_i)$. Let P be a vector subspace in \mathbb{R}^m . Then, for any $y \in \mathbb{R}^m$ with nonzero coordinates, the operator

$$\mathcal{D}_y \circ \Pi_{\mathcal{D}_y^{-1}(P)} : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

is well defined. Can we extend its definition to any $y \in \mathbb{R}^m$? The answer is yes and proved in the following

Lemma 4.2. *Let $\bar{y} \in E_J$ be such that $\bar{y}_i \neq 0$ for any $i \in J$.*

Then $\mathcal{D}_{\bar{y}}|_{E_J} : E_J \rightarrow E_J$ is nonsingular and

$$\lim_{y \rightarrow \bar{y}} \mathcal{D}_y \circ \Pi_{\mathcal{D}_y^{-1}(P)} = \mathcal{D}_{\bar{y}}|_{E_J} \circ \Pi_{(\mathcal{D}_{\bar{y}}|_{E_J})^{-1}(P \cap E_J)}.$$

Proof. To prove this lemma we suppose that $I = \{1, 2, \dots, m_1\}$ and $J = \{m_1 + 1, \dots, m_1 + m_2 = m\}$.

Let us denote $p = \dim P$. P is identified to an $n \times p$ matrix with rank $P = p$. We also introduce the following matrices:

$$\mathcal{D}_y = \begin{pmatrix} D_{y,1} & 0 \\ 0 & D_{y,2} \end{pmatrix}, \quad P = \begin{pmatrix} U & 0 \\ V & W \end{pmatrix}.$$

The different blocks appearing in these two matrices have the following dimensions: $D_{y,1} : m_1 \times m_1$, $D_{y,2} : m_2 \times m_2$, $U : m_1 \times p_1$, $V : m_2 \times p_1$, $W : m_2 \times p_2$. We also suppose that the columns of $\begin{pmatrix} 0 \\ W \end{pmatrix}$ are a basis for $P \cap E_J$ and those of $\begin{pmatrix} U \\ V \end{pmatrix}$ a basis of the orthogonal complement of $P \cap E_J$ in P that is $(P \cap E_J)^\perp \cap P$. Let us notice that $p_2 \leq m_2$ and rank $W = p_2$ and also that $p_1 \leq m_1$ and rank $U = p_1$. Let us prove this last assertion. Let U_i , $1 \leq i \leq p_1$ be the columns of U . If $\alpha_1 U_1 + \dots + \alpha_{p_1} U_{p_1} = 0$, we have

$$\begin{aligned} \alpha_1 \begin{pmatrix} U_1 \\ V_1 \end{pmatrix} + \dots + \alpha_{p_1} \begin{pmatrix} U_{p_1} \\ V_{p_1} \end{pmatrix} \\ = \begin{pmatrix} 0 \\ \alpha_1 V_1 + \dots + \alpha_{p_1} V_{p_1} \end{pmatrix}. \end{aligned}$$

The left-hand side of this equation is in $(P \cap E_J)^\perp \cap P$ and the right-hand side in $P \cap E_J$. Thus this vector is equal to 0 and since $\text{rank } P = p$ we get $\alpha_1 = \dots = \alpha_{p_1} = 0$.

For every subspace X in \mathbb{R}^m with $\dim X = p$ identified with an $m \times p$ rank p matrix we have

$$\Pi_X = X(X^T X)^{-1} X^T.$$

This gives here

$$\Pi_{\mathcal{D}_y^{-1} P}$$

$$\begin{aligned} &= \begin{pmatrix} D_{y,1}^{-1} U & 0 \\ D_{y,2}^{-1} V & D_{y,2}^{-1} W \end{pmatrix} \\ &\times \begin{pmatrix} U^T D_{y,1}^{-2} U + V^T D_{y,2}^{-2} V & V^T D_{y,2}^{-2} W \\ W^T D_{y,2}^{-2} V & W^T D_{y,2}^{-2} W \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} U^T D_{y,1}^{-1} & V^T D_{y,2}^{-1} \\ 0 & W^T D_{y,2}^{-1} \end{pmatrix} \end{aligned}$$

and

$$\Pi_{E_J} = \begin{pmatrix} 0 & 0 \\ 0 & I_{m_2} \end{pmatrix}.$$

We also notice that

$$\mathcal{D}_y \Pi_{\mathcal{D}_y^{-1} P} = \mathcal{D}_y \Pi_{E_I} \Pi_{\mathcal{D}_y^{-1} P} + \mathcal{D}_y \Pi_{E_J} \Pi_{\mathcal{D}_y^{-1} P}.$$

We have

$$\lim_{y \rightarrow \bar{y}} \mathcal{D}_y \Pi_{E_I} \Pi_{\mathcal{D}_y^{-1} P} = 0.$$

This is a consequence of the two following

$$\|\Pi_{E_I} \Pi_{\mathcal{D}_y^{-1} P}\| \leq 1$$

because it is the product of two orthogonal projections and

$$\lim_{y \rightarrow \bar{y}} \mathcal{D}_y \Pi_{E_I} = \mathcal{D}_{\bar{y}} \Pi_{E_I} = 0.$$

We now have to study the limit

$$\lim_{y \rightarrow \bar{y}} \mathcal{D}_y \Pi_{E_J} \Pi_{\mathcal{D}_y^{-1} P}.$$

Let us denote $A = U^T D_{y,1}^{-2} U$. The following identities hold:

$$\begin{aligned} &\mathcal{D}_y \Pi_{E_J} \Pi_{\mathcal{D}_y^{-1} P} \\ &= \begin{pmatrix} D_{y,1} & 0 \\ 0 & D_{y,2} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_{m_2} \end{pmatrix} \begin{pmatrix} D_{y,1}^{-1} U & 0 \\ D_{y,2}^{-1} V & D_{y,2}^{-1} W \end{pmatrix} \\ &\times \begin{pmatrix} U^T D_{y,1}^{-2} U + V^T D_{y,2}^{-2} V & V^T D_{y,2}^{-2} W \\ W^T D_{y,2}^{-2} V & W^T D_{y,2}^{-2} W \end{pmatrix}^{-1} \begin{pmatrix} U^T D_{y,1}^{-1} & V^T D_{y,2}^{-1} \\ 0 & W^T D_{y,2}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ V & W \end{pmatrix} \left[\begin{pmatrix} A & 0 \\ 0 & I_{m_2} \end{pmatrix} \begin{pmatrix} I_{m_1} + A^{-1}(V^T D_{y,2}^{-2} V) & A^{-1}(V^T D_{y,2}^{-2} W) \\ W^T D_{y,2}^{-2} V & W^T D_{y,2}^{-2} W \end{pmatrix} \right]^{-1} \begin{pmatrix} U^T D_{y,1}^{-1} & V^T D_{y,2}^{-1} \\ 0 & W^T D_{y,2}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ V & W \end{pmatrix} \begin{pmatrix} I_{m_1} + A^{-1}(V^T D_{y,2}^{-2} V) & A^{-1}(V^T D_{y,2}^{-2} W) \\ W^T D_{y,2}^{-2} V & W^T D_{y,2}^{-2} W \end{pmatrix}^{-1} \begin{pmatrix} A^{-1}(U^T D_{y,1}^{-1}) & A^{-1}(V^T D_{y,2}^{-1}) \\ 0 & W^T D_{y,2}^{-1} \end{pmatrix}. \end{aligned}$$

We will prove later that

$$\lim_{y \rightarrow \bar{y}} A^{-1} = \lim_{y \rightarrow \bar{y}} \left(U^T D_{y,1}^{-1} \right) = 0$$

when $y \rightarrow \bar{y}$. Since

$$\lim_{y \rightarrow \bar{y}} D_{y,2} = D_{\bar{y},2}$$

is a nonsingular matrix we get

$$\begin{aligned} \lim_{y \rightarrow \bar{y}} \mathcal{D}_y \Pi_{E_J} \Pi_{\mathcal{D}_y^{-1} P} \\ = \begin{pmatrix} 0 & 0 \\ V & W \end{pmatrix} \begin{pmatrix} I_{m_1} & 0 \\ W^T D_{\bar{y},2}^{-2} V & W^T D_{\bar{y},2}^{-2} W \end{pmatrix}^{-1} \\ \times \begin{pmatrix} 0 & 0 \\ 0 & W^T D_{\bar{y},2}^{-1} \end{pmatrix} \\ = \begin{pmatrix} 0 & 0 \\ 0 & W(W^T D_{\bar{y},2}^{-2} W)^{-1} W^T D_{\bar{y},2}^{-1} \end{pmatrix} \end{aligned}$$

and this last matrix represents the operator

$$\mathcal{D}_{\bar{y}}|_{E_J} \circ \Pi_{(\mathcal{D}_{\bar{y}}|_{E_J})^{-1}(P \cap E_J)}$$

as announced in this lemma.

To achieve the proof of this lemma we have to show that

$$\lim A^{-1} = \lim A^{-1} \left(U^T D_{y,1}^{-1} \right) = 0$$

with $A = U^T D_{y,1}^{-2} U$. In fact it suffices to prove $\lim A^{-1} = 0$ because

$$\begin{aligned} \|A^{-1}(U^T D_{y,1}^{-1})\|^2 \\ = \|A^{-1}(U^T D_{y,1}^{-1})(A^{-1}(U^T D_{y,1}^{-1}))^T\| \\ = \|A^{-1}\|. \end{aligned}$$

Since U is full rank, the matrix $U^T U$ is positive definite so that

$$\mu = \min \text{Spec}(U^T U) > 0.$$

Let us denote \succ the ordering on square matrices given by the cone of non-negative matrices. We have

$$D_{y,1}^{-2} \succ \frac{1}{\max y_i^2} I_{m_1} \succ \frac{1}{\|y\|^2} I_{m_1}$$

so that

$$U^T D_{y,1}^{-2} U \succ \frac{1}{\|y\|^2} U^T U \succ \frac{\mu}{\|y\|^2} I_{p_1}.$$

Taking the inverses changes this inequality in the following

$$0 \prec (U^T D_{y,1}^{-2} U)^{-1} \prec \frac{\|y\|^2}{\mu} I_{p_1} \rightarrow 0$$

when $y \rightarrow \bar{y}$ and we are done. ■

Corollary 4.1. *The vector field $N(x)$ extends continuously to all of R^n . Moreover it is Lipschitz on compact sets. When all the faces of the polytope \mathcal{P} are regularly described, the continuous extension of $N(x)$ to the face \mathcal{P}_J of \mathcal{P} equals $N_J(x)$. Consequently any orbit of $N(x)$ in the polytope \mathcal{P} tends to one of the singularities of the extended vector field, i.e. either to a vertex or an analytic center of one of the faces.*

Proof. It is a consequence of Definition 2.2, Lemmas 4.1 and 4.2 and the equality $A_J^\dagger y = A^\dagger y$ for any $y \in E_J$ that $N(x)$ extends continuously to all of R^n and equals $N_J(x)$ on \mathcal{P}_J . Moreover a rational function which is continuous on R^n has bounded partial derivatives on compact sets and hence is Lipschitz. Now we use the characterization of the vectorfield restricted to the face to see that any orbit which is not the analytic center of a face tends to the boundary of the face and any orbit which enters a small enough neighborhood of a vertex tends to that vertex. ■

Remark 4.1. We have shown that $N(x)$ is Lipschitz. We do not know an example where it is not analytic and wonder as to what its order of smoothness is, in general. In the next section we will show it is analytic generically.

5. Analyticity and Derivatives

In Sec. 3 we gave the following expression for the Newton vector field:

$$N(x) = \frac{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (x - u_{K_n})(\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}$$

for any $x \in \text{Int } \mathcal{P}$. Under a mild geometric assumption, the denominator of this fraction never vanishes so that $N(x)$ may be extended in a real analytic vector field.

Theorem 5.1. *Suppose that for any $x \in \partial \mathcal{P}$ contained in the relative interior of a codimension d face of \mathcal{P} , we have $A_{k_i} x = b_{k_i}$ for exactly d indices in $\{1, \dots, m\}$ and $A_l x > b_l$ for the other indices. In that case the line vectors A_{k_i} , $1 \leq i \leq d$, are linearly*

independent. Moreover, for such an x

$$\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2 \neq 0$$

so that $N(x)$ extends analytically to a neighborhood of \mathcal{P} .

Proof. Under this assumption, for any $x \in \mathcal{P}$, there exists a subset $K_n \subset \{1, \dots, m\}$ such that the submatrix A_{K_n} is nonsingular and $A_l x - b_l > 0$ for any $x \notin K_n$. ■

Our next objective is to describe the singular points of this extended vector field.

Theorem 5.2. *Under the previous geometric assumption, the singularities of the extended vector field are: the analytic center of the polytope and the analytic centers of the different faces of the polytope, including the vertices. Each of them is hyperbolic: if $x \in \partial\mathcal{P}$ is the analytic center of a codimension d face \mathcal{F} of \mathcal{P} , then the derivative $DN(x)$ has $n-d$ eigenvalues equal to -1 with corresponding eigenvectors contained in the linear space \mathcal{F}_0 parallel to \mathcal{F} and d eigenvalues equal to 1 with corresponding eigenvectors contained in a complement of \mathcal{F}_0 .*

Proof. The first part of this theorem, about the -1 eigenvalues, is the consequence of two facts.

The first one is a well-known fact about the Newton operator: its derivative is equal to $-\text{id}$ at a zero (if $N(x) = 0$, then $DN(x) = D(-Dg(x)^{-1}g(x)) = D(-Dg(x)^{-1})g(x) - Dg(x)^{-1}Dg(x) = -\text{id}$). The second fact is proved in Sec. 4: the restriction of $N(x)$ to a face is the Newton vector field associated with the restriction of $g(x)$ to this face.

We have now take care of the 1 eigenvalues. To simplify the notations we suppose that $A_i x = b_i$ for $1 \leq i \leq d$, $A_i x > b_i$ when $i+1 \leq i \leq m$, and $N(x) = 0$. N is analytic and its derivative in the direction v is given by

$$DN(x)v = \frac{\text{Num}}{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2}$$

with

$$\begin{aligned} \text{Num} = & \sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} v(\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2 \\ & + \sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (x - u_{K_n})(\det A_{K_n})^2 \\ & \times \sum_{\substack{l_0 \notin K_n \\ l_0 \neq l_0}} 2A_{l_0}v(A_{l_0}x - b_{l_0}) \prod_{\substack{l \notin K_n \\ l \neq l_0}} (A_l x - b_l)^2 \end{aligned}$$

which gives

$$DN(x)v = v + \frac{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (x - u_{K_n})(\det A_{K_n})^2 \sum_{l_0 \notin K_n} 2A_{l_0}v(A_{l_0}x - b_{l_0}) \prod_{\substack{l \notin K_n \\ l \neq l_0}} (A_l x - b_l)^2}{\sum_{\substack{K_n \subset \{1, \dots, m\} \\ \det K_n \neq 0}} (\det A_{K_n})^2 \prod_{l \notin K_n} (A_l x - b_l)^2} = v + Mv$$

where M is, up to a constant factor (i.e. constant in v), the $n \times n$ matrix equal to

$$\sum_{\substack{K_n \\ \det K_n \neq 0 \\ l_0 \notin K_n}} (\det A_{K_n})^2 (A_{l_0}x - b_{l_0}) \left(\prod_{\substack{l \notin K_n \\ l \neq l_0}} (A_l x - b_l)^2 \right) (x - u_{K_n}) A_{l_0}$$

which is also equal to

$$\sum_{\substack{\{1, \dots, d\} \subset K_n \\ \det K_n \neq 0 \\ d+1 \leq l_0 \leq m \\ l_0 \notin K_n}} \frac{(\det A_{K_n})^2}{A_{l_0}x - b_{l_0}} \left(\prod_{l \notin K_n} (A_l x - b_l)^2 \right) (x - u_{K_n}) A_{l_0}$$

because $A_i x = b_i$ when $1 \leq i \leq d$ and $A_i x > b_i$ otherwise.

To prove our theorem we have to show that $\dim \ker M \geq d$. This gives at least d independent vectors v_i such that $Mv_i = 0$, that is, $DN(x)v_i = v_i$; thus 1 is an eigenvalue of $DN(x)$ and its multiplicity is $\geq d$. In fact it is exactly d because we already have the eigenvalue -1 with multiplicity $n - d$. The inequality $\dim \ker M \geq d$ is given by $\text{rank } M \leq n - d$. Why is it true? M is a linear combination of rank 1 matrices $(x - u_{K_n})A_{l_0}$ so that the rank of M is less than or equal to the dimension of the system of vectors $x - u_{K_n}$ with K_n as before. Since $\{1, \dots, d\} \subset K_n$, $A_i x = b_i$ when $1 \leq i \leq d$, and $Au_{K_n} = b_{K_n}$ we have $A(x - u_{K_n}) = (0, \dots, 0, y_{d+1}, \dots, y_m)^T$. From the hypothesis, the line vectors A_1, \dots, A_d defining the face \mathcal{F} are independent, thus the set of vectors $u \in \mathbb{R}^n$ such that the vector $Au \in \mathbb{R}^m$ begins by d zeros has dimension $n - d$ and we are done. ■

Remark 5.1. The last theorem implies that $N(x)$ is Morse–Smale in the terminology of dynamical systems. Recall also that we are really interested in

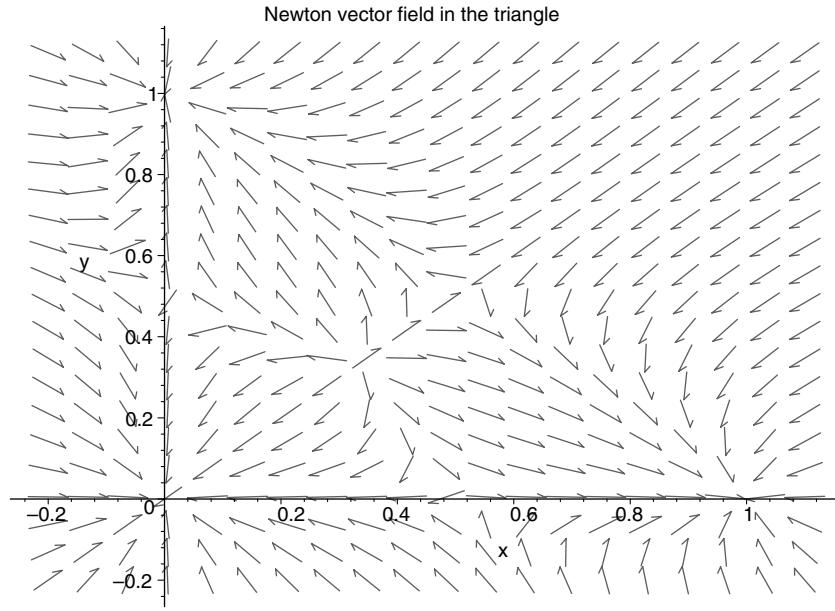
the positive time trajectories of $-N(x)$. For $-N(x)$ the eigenvalues at the critical points are multiplied by -1 so in the faces the critical points of $-N(x)$ are sources and their stable manifolds are transverse to the faces.

6. Example

Let us consider the case of a triangle in the plane. Since the Newton vector field is affinely invariant (Proposition 2.1) we may only consider the triangle with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. A dual description is given by the three inequalities $x \geq 0$, $y \geq 0$, $-x - y \geq -1$ which correspond to the following data:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix},$$

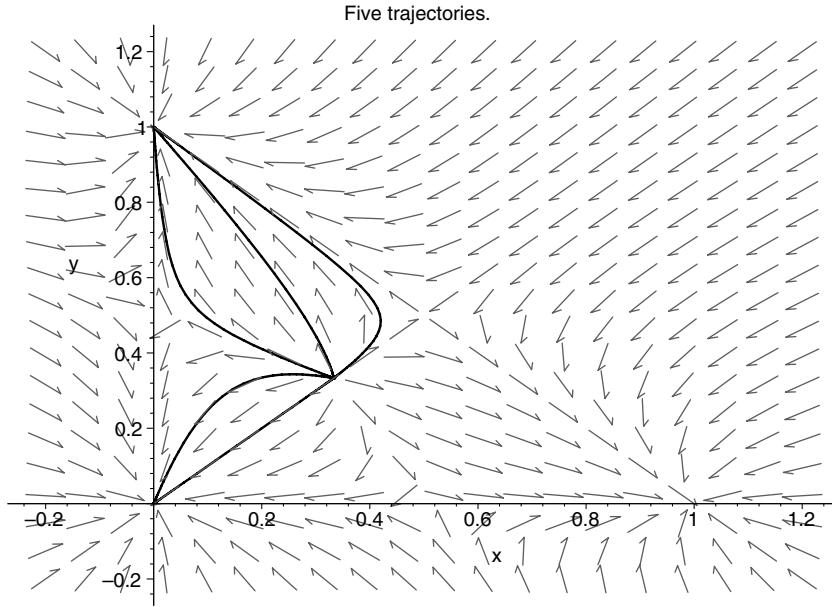
$$D_{(x,y)} = \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & 1 - x - y \end{pmatrix}.$$



The corresponding Newton vector field is given by the rational expressions

$$N(x, y) = \begin{bmatrix} \frac{xz^2 - x^2z + xy^2 - x^2y}{z^2 + y^2 + x^2} \\ \frac{x^2y - xy^2 + yz^2 - y^2z}{z^2 + y^2 + x^2} \end{bmatrix}$$

with $z = 1 - x - y$. This vector field is analytic on the whole plane. The singular points are the three vertices, the midpoints of the three sides and the center of gravity. The arrows in the figure are for $-N(x)$ and the critical points are clearly sources in their faces.



References

- Bayer, D. & Lagarias, J. [1989a] "The non-linear geometry of linear programming I: Affine and projective scaling trajectories," *Trans. Amer. Math. Soc.* **314**, 499–526.
- Bayer, D. & Lagarias, J. [1989b] "The non-linear geometry of linear programming II: Legendre transform coordinates and central trajectories," *Trans. Amer. Math. Soc.* **314**, 527–581.
- Bayer, D. & Lagarias, J. [1991] "Karmarkar's linear programming algorithm and Newton's method," *Math. Progr.* **A50**, 291–330.
- Dedieu, J.-P., Malajovich, G. & Shub, M. "On the curvature of the central path of linear programming theory," to appear.
- Megiddo, N. & Shub, M. [1989] "Boundary behaviour of interior point algorithms in linear programming," *Math. Oper. Res.* **14**, 97–146.
- Renegar, J. [2001] *A Mathematical View of Interior-Point Methods in Convex Optimization* (SIAM, Philadelphia).
- Vavasis, S. & Ye, Y. [1996] "A primal-dual accelerated interior point method whose running time depends only on A ," *Math. Progr.* **A74**, 79–120.