J. Austral. Math. Soc. Ser. B 34(1993), 495–510

ISOSPECTRAL FLOWS AND LINEAR PROGRAMMING

U. HELMKE [1]

(Received 10 June 1991; revised 15 October 1991)

Abstract

Brockett has studied the isospectral flow $\dot{H} = [H, [H, N]]$, with $[A, B] = AB - BA$, on spaces of real symmetric matrices. The flow diagonalises real symmetric matrices and can be used to solve linear programming problems with compact convex constraints. We show that the flow converges exponentially fast to the optimal solution of the programming problem and we give explicit estimates for the time needed by the flow to approach an ε-neighbourhood of the optimum. An interior point algorithm for the standard simplex is analysed in detail and a comparison is made with a continuous time version of Karmarkar algorithm.

1. Introduction

Isospectral, that is eigenvalue preserving, flows evolving on spaces of symmetric matrices have been intensively studied over the past few years, one of the main initial sources of interest being that they provide an explicit method to solve various completely integrable Hamiltonian systems. The current interest in analog computing and networks has also shifted the focus from the application of combinatorial or algebraic approaches to problems in discrete mathematics, towards continuous time methods (differential equations).

In recent work, Brockett [2] has studied the ordinary differential equation on the set of real symmetric $n \times n$ matrices H:

$$\dot{H} = [H, [H, N]] \tag{1.1}$$

where $[A, B] = AB - BA$ denotes the Lie bracket and $N = \mathrm{diag}(\mu_1, \dots, \mu_n)$, $\mu_1 > \cdots > \mu_n$, is a fixed real diagonal matrix. Brockett proves that (1.1) defines an isospectral flow which diagonalises $H(t)$ asymptotically as $t \to \infty$.

[1] University of Regensburg, Department of Mathematics, 8400 Regensburg, Germany.

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

Furthermore, he shows that (1.1) solves linear programming problems and achieves sorting. His approach to the linear programming problem is similar in spirit to Karmarkar's algorithm, in the sense that in both cases a trajectory is constructed which approaches the optimal solution from the interior of the constraint set. Nevertheless, there seems to be an essential difference between Karmarkar's and Brockett's algorithms. This is due to the fact that the optimising trajectory in Brockett's approach is constructed using a gradient flow evolving on a higher dimensional smooth manifold.

The usefulness of (1.1) could appear moreover limited in view of Brockett's remark that convergence of the gradient linear programming algorithm is not exponential in time. In this note we show that the proposed algorithm does in fact converge exponentially fast to the optimal solution. Explicit bounds for the rate of convergence are given, as well as for the time needed for the trajectory produced from (1.1) to enter an ε-neighbourhood of the optimal solution. In the special case where the convex set is the standard simplex, Brockett's equation, or rather its simplification studied in Section 2, is shown to induce an interior point algorithm. The algorithm in this case is formally very similar to Karmarkar's interior point flow and our result suggests the possibility of common generalisations of these algorithms. Very recently, Faibusovich [3] has constructed a new class of interior point algorithms for linear programming which, in the case of a standard simplex, coincide with the flow studied here. These flows can thus be expected to have similar performance properties to the Karmarkar algorithm.

In this paper no attempt is made to achieve maximal generality. Instead, we rather try to give a thorough analysis of Brockett's algorithm, thus trying to emphasise its basic simplicity and beauty.

We proceed as follows. After reviewing Brockett's approach to linear programming, we develop a simplified set of equations for linear programming (Theorem 2.2). An extension of these ideas leads to an interior point flow on the set of stochastic matrices which is analysed by Theorem 2.3. In Section 3 an interior point flow on the standard simplex is derived (Corollary 3.1) and its connection with the Karmarkar flow is pointed out. Proposition 3.2 and Theorem 3.3 estimate the time required for these algorithms to reach an ε-neighbourhood of the optimum. The connection with the recent work of Faibusovich is made. We conclude with some further remarks in Section 4.

2. The linear programming problem

Let $C(v_1, \dots, v_m) \subset \mathbb{R}^n$ denote the convex hull of m vectors $v_1, \dots, v_m \in \mathbb{R}^n$. Given the compact convex set $C(v_1, \dots, v_m) \subset \mathbb{R}^n$ and a row vector

$c' = (c_1, \ldots, c_n) \in \mathbb{R}^n$, the linear programming problem then asks to find a vector $x \in C(v_1, \ldots, v_m)$ which maximises $c' \cdot x$. Of course, the optimal solution is a vertex point v_{i_*} of the constraint set $C(v_1, \ldots, v_m)$.

Since $C(v_1, \ldots, v_m)$ is not a smooth manifold in any reasonable sense, it is not possible to apply in the usual way steepest descent gradient methods in order to find the optimum. One possibly way to circumvent such technical difficulties would be to replace the constraint set $C(v_1, \ldots, v_m)$ by some suitable compact manifold M so that the optimisation takes place on M instead of $C(v_1, \ldots, v_m)$. A mathematically convenient way here would be to construct a suitable resolution space for the singularities of $C(v_1, \ldots, v_m)$. This is actually the approach taken by Brockett.

Let $\Delta_{m-1} = \{(\eta_1, \ldots, \eta_m) \in \mathbb{R}^m | \eta_i \geq 0, \sum_{i=1}^m \eta_i = 1\}$ denote the standard $(m-1)$-dimensional simplex in $\mathbb{R}^m$. Let

$$T = [v_1, \ldots, v_m]$$

be the real $n \times m$ matrix whose column vectors are the vertices $v_1, \ldots, v_m$. Thus $T: \mathbb{R}^m \to \mathbb{R}^n$ maps the simplex Δ_{m-1} of $\mathbb{R}^m$ linearly onto $C(v_1, \ldots, v_m)$. We can thus use T in order to replace the constraint set $C(v_1, \ldots, v_m)$ by the standard simplex Δ_{m-1}.

Suppose that we are to solve the linear programming problem consisting of maximising $c'x$ over the compact convex set of all $x \in C(v_1, \ldots, v_m)$. Brockett's recipe to solve the problem is this (cf. Theorem 6 in [2]).

THEOREM 2.1. *Let N be the real $m \times m$ matrix defined by*

$$N = \operatorname{diag}(c'v_1, \ldots, c'v_m) \tag{2.2}$$

and assume that N satisfies the genericity conditions $c'v_i \neq c'v_j$ for $i \neq j$. Let $Q = \operatorname{diag}(1, 0, \ldots, 0) \in \mathbb{R}^{m \times m}$. Then for almost all orthogonal matrices $\Theta \in O(m)$ the solution $H(t)$ of the differential equation (with $[A, B] = AB - BA$ the Lie bracket)

$$\begin{aligned} \dot{H} = [H, [H, N]] = H^2N + NH^2 - 2HNH, \\ H(0) = \Theta' Q \Theta \end{aligned} \tag{2.3}$$

converges as $t \to \infty$ to a diagonal matrix $H_\infty = \operatorname{diag}(0, \ldots, 0, 1, 0, \ldots, 0)$, with the entry 1 being at position i_* so that $x = v_{i_*}$ is the optimal vertex of the linear programming problem.

Thus the optimal solution of a linear programming problem can be obtained by applying the linear transformation T to a vector obtained from the diagonal entries of the stable limiting solution of (2.3). Brockett's method, while theoretically appealing, has however a number of shortcomings.

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

First, it works with a huge overparametrisation of the problem. The differential equation (2.3) evolves on the $\frac{1}{2}m(m+1)$-dimensional vector space of real symmetric $m \times m$ matrices H, while the linear programming problem is set up in the n-dimensional space $C(v_1, \dots, v_m)$. Of course, usually n will be much smaller than $\frac{1}{2}m(m+1)$.

Second, convergence to H_∞ is guaranteed only for a generic choice of orthogonal matrices. No explicit description of this generic set of initial data is given.

Finally the method requires the knowledge of the values of the cost functional at all vertex points in order to define the matrix N.

Clearly the last point is the most critical one and therefore Theorem 2.1 should be regarded only as a theoretical approach to the linear programming problem. To overcome this difficulty along with the other difficulties we proceed as follows.

Let $S^{m-1} = \{(\xi_1, \dots, \xi_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \xi_i^2 = 1\}$ denote the set of unit vectors of $\mathbb{R}^m$. We consider the polynomial map

$$f: S^{m-1} \to \Delta_{m-1}$$

defined by

$$f(\xi_1, \dots, \xi_m) = (\xi_1^2, \dots, \xi_m^2). \tag{2.4}$$

Since f maps the vectors $\pm(\xi_1, \dots, \xi_m)$ to the same point on the simplex Δ_{m-1}, we can eliminate this sign ambiguity by passing to the corresponding projective space. This forces uniqueness of the attractor for the flow constructed below. Thus let $\mathbb{RP}^{m-1}$ denote the $(m-1)$-dimensional *projective space* of lines in $\mathbb{R}^m$. We follow the standard notation of using homogeneous co-ordinates to describe the points of $\mathbb{RP}^{m-1}$. Thus for a unit vector $\xi = (\xi_1, \dots, \xi_m) \in \mathbb{R}^m$ let $[\xi_1 : \dots : \xi_m]$ denote the line, passing through $0 \in \mathbb{R}^m$, which is generated by ξ. If ξ, μ are unit vectors then $[\xi_1 : \dots : \xi_m] = [\eta_1 : \dots : \eta_m]$ if and only if $\eta = \pm\xi$. With these notations in mind, the map (2.4) descends to a real algebraic map on the projective space

$$\begin{aligned} F: \mathbb{RP}^{m-1} &\to \Delta_{m-1}, \\ [\xi_1 : \dots : \xi_m] &\mapsto (\xi_1^2, \dots, \xi_m^2). \end{aligned} \tag{2.5}$$

By composing F with the map $T: \Delta_{m-1} \to C(v_1, \dots, v_m)$ we obtain a real algebraic map

$$\begin{aligned} \pi_T = T \circ F: \mathbb{RP}^{m-1} &\to C(v_1, \dots, v_m), \\ [\xi_1 : \cdots : \xi_m] &\mapsto T \begin{pmatrix} \xi_1^2 \\ \vdots \\ \xi_m^2 \end{pmatrix}. \end{aligned} \tag{2.6}$$

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

The linear programming task is to maximise the restriction of the linear functional

$$\begin{aligned} \lambda\colon C(v_1, \dots, v_m) &\to \mathbb{R}, \\ x &\mapsto c'x \end{aligned} \tag{2.7}$$

over $C(v_1, \dots, v_m)$. The idea now is to consider instead of the maximisation of (2.7) the maximisation of the induced smooth function $\lambda \circ \pi_T\colon \mathbb{RP}^{m-1} \to \mathbb{R}$, which is straightforward, since $\mathbb{RP}^{m-1}$ is smooth.

THEOREM 2.2. *Let N be defined by* (2.2) *and assume the genericity condition*

$$c'(v_i - v_j) \neq 0 \quad \textit{for all } i \neq j. \tag{2.8}$$

(a) *The gradient vector-field of $\lambda \circ \pi_T$ on $\mathbb{RP}^{m-1}$ is*

$$\dot{\xi} = 2(N - \xi' N \xi)\xi, \tag{2.9}$$

$|\xi|^2 = 1$. *Also,* (2.9) *has exactly m equilibrium points* $[1\colon 0\colon \dots\colon 0], \dots, [0\colon \dots\colon 1]$, *given by the standard basis vectors $e_1, \dots, e_m$ of $\mathbb{R}^m$.*

(b) *The eigenvalues of the linearisation of* (2.9) *at* $[e_i]$ *are*

$$c'(v_1 - v_i), \dots, c'(v_{i-1} - v_i), c'(v_{i+1} - v_i), \dots, c'(v_m - v_i) \tag{2.10}$$

and there is a unique index $1 \leq i_* \leq m$ such that $[e_{i_*}]$ is asymptotically stable.

(c) *Let $X \cong \mathbb{RP}^{m-2}$ be the smooth codimension-one submanifold of $\mathbb{RP}^{m-1}$ defined by*

$$X = \{[\xi_1\colon \dots\colon \xi_m] \mid \xi_{i_*} = 0\}. \tag{2.11}$$

With the exception of initial points contained in X, every solution $\xi(t)$ of (2.9) *converges exponentially fast to the attractor $[e_{i_*}]$ and $\pi_T(\xi(t))$ converges exponentially fast to the optimal solution $\pi_T([e_{i_*}]) = v_{i_*}$ of the linear programming problem, with a bound on the rate of convergence*

$$|\pi_T(\xi(t)) - v_{i_*}| \leq \text{const}\, e^{-2\mu(t-t_0)}$$

where

$$\mu = \min_{j \neq i_*} |c'(v_j - v_{i_*})|. \tag{2.12}$$

PROOF. For any diagonal $m \times m$ matrix $N = \operatorname{diag}(n_1, \dots, n_m)$ consider the smooth function $\varphi\colon \mathbb{R}^m - \{0\} \to \mathbb{R}$ defined by

$$\varphi(x_1, \dots, x_m) = \frac{\sum_{i=1}^m n_i x_i^2}{\sum_{i=1}^m x_i^2}. \tag{2.13}$$

A straightforward computation shows that the gradient of φ at a unit vector $x \in S^{m-1}$ is

$$\nabla\varphi(x) = 2(N - x'Nx)x.$$

Furthermore, if $n_i \neq n_j$ for $i \neq j$, then the critical points of the induced map $\varphi: \mathbb{RP}^{m-1} \to \mathbb{R}$ are the homogeneous co-ordinates of the standard basis vectors, i.e. by $[e_1], \dots, [e_m]$. This proves (a). The Hessian of $\varphi: \mathbb{RP}^{m-1} \to \mathbb{R}$ at $[e_i]$ is readily computed as

$$H_\varphi([e_i]) = 2\,\mathrm{diag}(n_1 - n_i, \dots, n_{i-1} - n_i, n_{i+1} - n_i, \dots, n_m - n_i).$$

Let i_* be the unique index such that $n_{i_*} = \max_{1\le j\le m} n_j$. Thus $H_\varphi([e_i]) < 0$ if and only if $i = i_*$ which proves (b). Let $X\mathbb{C} \subset \mathbb{RP}^{m-2}$ be the closed submanifold of $\mathbb{RP}^{m-1}$ defined by (2.11). Then $\mathbb{RP}^{m-1} - X$ is equal to the stable manifold of $[e_{i_*}]$ and X is equal to the union of the stable manifolds of the other equilibrium points $[e_i]$, $i \neq i_*$. The result follows.

Remarks. (a) From (2.11)

$$\pi_T(X) = X(v_1, \dots, v_{i_*-1}, v_{i_*+1}, \dots, v_m). \tag{2.14}$$

Thus if $n = 2$ and $C(v_1, \dots, v_m) \subset \mathbb{R}^2$ is the convex set, illustrated by the following Figure 2.1 with optimal vertex point v_{i_*} then the shaded region describes the image $\pi_T(X)$ of the set of exceptional initial conditions.

(b) The flow (2.3) with $Q = \mathrm{diag}(1, 0, \dots, 0)$ is equivalent to the gradient flow (2.9) on $\mathbb{RP}^{m-1}$; [4]. Also, (2.9) has an interesting interpretation in neural network theory; see Oja [7]. It is possible to replace the use of the projective space $\mathbb{RP}^{m-1}$ in Theorem 2.2 by the sphere S^{m-1}. This, however,

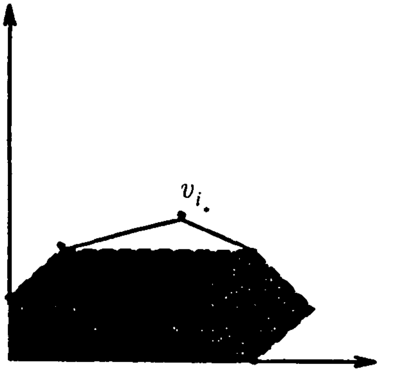

Figure 2.1. The convex set $C(v_1, \dots, v_m)$

would be at the expense of doubling the number of equilibrium points. Also there would then be two (local) attractors $\pm e_{i_*}$. Otherwise, the analysis would go through likewise.

(c) For any (skew-)symmetric matrix $\Omega \in \mathbb{R}^{m\times m}$

$$\dot{\xi} = 2(N + \Omega - \xi'(N+\Omega)\xi)\xi \tag{2.15}$$

defines a flow on $\mathbb{RP}^{m-1}$. If $\Omega = -\Omega'$ is skew-symmetric, the functional $\varphi(\xi) = \xi'(N+\Omega)\xi = \xi' N\xi$ is not changed by ω and therefore has the same critical points. Thus, while (2.15) is not the gradient flow of φ (if $\Omega = -\Omega'$), it can still be of interest for the linear programming problem. If $\Omega = \Omega'$ is symmetric, (2.15) is the gradient flow of $\psi(\xi) = \xi'(N+\Omega)\xi$ on $\mathbb{RP}^{m-1}$.

We now study the following specific linear programming problem. Let $\mathscr{P}(n) \subset \mathbb{R}^{n\times n}$ denote the subset of $n\times n$ real stochastic matrices. That is, $P \in \mathscr{P}(n)$ if and only if

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} = [p^1, \dots, p^n] \tag{2.16}$$

with $p_{ij} \geq 0$ and $\sum_{i=1}^n p_{ij} = 1$ for $j = 1, \dots, n$. Thus $\mathscr{P}(n)\mathbb{C} \subset \Delta_{n-1} \times \cdots \times \Delta_{n-1}$ is a compact convex subset of $\mathbb{R}^{n^2}$. Given an $n\times n$ matrix

$$C = \begin{bmatrix} c_1' \\ \vdots \\ c_n' \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{2.17}$$

we consider the task to solve the linear programming problem of maximising the functional

$$\begin{aligned} \lambda\colon \mathscr{P}(n) &\to \mathbb{R}, \\ P &\mapsto \operatorname{tr}(CP) \end{aligned} \tag{2.18}$$

over $\mathscr{P}(n)$. Let

$$M := \mathbb{RP}^{n-1} \times \cdots \times \mathbb{RP}^{n-1} \tag{2.19}$$

denote the n-fold product of real projective $(n-1)$-space and let

$$\pi\colon M \to \mathbb{R} \tag{2.20}$$

denote the map defined by

$$\pi([\xi_{11}: \dots : \xi_{n1}], \dots, [\xi_{1n}: \dots : \xi_{nn}]) := \begin{bmatrix} \xi_{11}^2 & \cdots & \xi_{1n}^2 \\ \vdots & & \vdots \\ \xi_{n1}^2 & \cdots & \xi_{nn}^2 \end{bmatrix} \tag{2.21}$$

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

where ξ_{ij} are homogeneous coordinates with $\sum_{i=1}^n \xi_{ij}^2 = 1$, $j = 1, \dots, n$. Thus $\pi : M \to \mathbb{R}^{n\times n}$ is a smooth real algebraic map of M onto its image $\mathscr{P}(n) = \pi(M)$. The proof of the following theorem is completely similar to that of Theorem 2.2 and is therefore omitted.

THEOREM 2.3. *Suppose the entries of the ith row vector c_i' of C are all pairwise distinct, $i = 1, \dots, n$. Let $C_i = \operatorname{diag}(c_i')$ denote the diagonal $n\times n$-matrix, whose first, ..., nth diagonal entry is given by the first, ..., nth entry of c_i'.*

(a) *The gradient vector field of $\lambda \circ \pi : M \to \mathbb{R}$ is*

$$\dot{\xi}_i = 2(C_i - \lambda \circ \pi([\xi])\xi_i, \tag{2.22}$$

$i = 1, \dots, n$, where $|\xi_i|^2 = 1$, $[\xi] = ([\xi_1], \dots, [\xi_n]) \in M$. Also, (2.22) has exactly n^n critical points $([e_{i_1}], \dots, [e_{i_n}])$, given by arbitrary n-tuples of standard basis vectors $e_{i_1}, \dots, e_{i_n}$ of $\mathbb{R}^n$.

(b) *The eigenvalues of the linearisation of (2.22) at $([e_{i_1}], \dots, [e_{i_n}])$ are equal to $(c_i' = (c_{i1}, \dots, c_{in}))$*

$$2(c_{1\alpha_1} - c_{1i_1}, c_{2\alpha_2} - c_{2i_2}, \dots, c_{n\alpha_n} - c_{ni_n}), \qquad \alpha_1 \neq i_1, \dots, \alpha_n \neq i_n, \tag{2.23}$$

and there exists a unique n-tuple $(\omega_1, \dots, \omega_n)$ of indices such that $([e_{\omega_1}], \dots, [e_{\omega_n}])$ is asymptotically stable.

(c) *Let $X = X_{i_1} \times \cdots \times X_{i_n}$ be the codimension n submanifold of $\mathbb{RP}^{n-1} \times \cdots \times \mathbb{RP}^{n-1}$ defined by $X_i = \{[\xi_1 : \dots : \xi_n] \in \mathbb{RP}^{n-1} | \xi_i = 0\}$, $i = 1, \dots, n$. With the exception of initial points contained in X, every solution of (2.22) converges exponentially fast to $([e_{\omega_1}], \dots, [e_{\omega_n}])$ and $\pi([\xi_1(t)], \dots, [\xi_n(t)])$ converges exponentially to the optimal solution $P_\omega = (e_{\omega_1}, \dots, e_{\omega_1}) \in \mathscr{P}(n)$, with a bound on the rate of convergence*

$$|\pi([\xi(t)]) - P_\omega| \leq \text{const}\, e^{-2\mu(t-t_0)}, \qquad \mu = \min_{1\leq i\leq n,\, j\neq\omega_i} |c_{ij} - c_{i\omega_i}|. \tag{2.24}$$

REMARK. The codimension n submanifold of X, which has to be excluded, is diffeomorphic to $\mathbb{RP}^{n-1} \times \cdots \times \mathbb{RP}^{n-1}$ and is mapped under π into the (distinguished) boundary $\partial\Delta_{m-1} \times \cdots \times \partial\Delta_{m-1}$ of $\mathscr{P}(n)$. Thus (2.22) is really an interior point algorithm.

3. Interior point flows

Brockett's equation (2.3) and its simplified version (2.9) both evolve on a high-dimensional manifold M so that the projection of the trajectories

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

into the polytope leads to a curve which approaches the optimum from the interior. We have also considered in Theorem 2.3 one case where (2.9) actually leads to flow evolving on the convex polytope such that the optimum is approached from all trajectories starting in the interior of the constant set. Such algorithms are called *interior point algorithms*, an example being the celebrated Karmarkar algorithm [5]. Here we like to take such issues a bit further in the case where the constraint set is the standard simplex $\Delta_{m-1} \subset \mathbb{R}^m$.

In this case, (2.9) on $\mathbb{RP}^{m-1}$ becomes

$$\dot{\xi}_i = 2\left(c_i - \sum_{j=1}^{m} c_j \xi_j^2\right)\xi_i, \qquad i = 1, \dots, m, \tag{3.1}$$

$\sum_{i=1}^m \xi_i^2 = 1$. Thus with the substitution $x_i = \xi_i^2$, $i = 1, \dots, m$, we obtain

$$\dot{x}_i = 4\left(c_i - \sum_{i=1}^{m} c_j x_j\right)x_i, \qquad i = 1, \dots, m, \tag{3.2}$$

$x_i \geq 0$, $\sum_{i=1}^m x_i = 1$. Since $\sum_{i=1}^m \dot{x}_i = 0$, (3.2) is a flow on the simplex Δ_{m-1}. The set $X \subset \mathbb{RP}^{m-1}$ of exceptional initial conditions is mapped by the quadratic substitution $x_i = \xi_i^2$, $i = 1, \dots, m$, onto the boundary $\partial\Delta_{m-1}$ of the simplex. Thus Theorem 2.2 implies

COROLLARY 3.1. *Equation* (3.2) *defines a flow on* Δ_{m-1}. *Every solution* $x(t)$ *with initial condition* $x(0)$ *in the interior of* Δ_{m-1} *converges to the optimal solution* e_{i_*} *of the linear programming problem*: *Maximise* $c'x$ *over* $x \in \Delta_{m-1}$, *with an exponential rate of convergence given by*

$$|x(t) - e_{i_*}| \leq \text{const} \cdot e^{-2\mu t}, \qquad \mu = \min_{j \neq i_*}(c_{i_*} - c_j).$$

REMARKS. (a) Equation (3.2) is a Volterra-Lotka type of equation and thus belongs to a well studied class of equations in population dynamics; cf. Schuster *et al.* [8], Zeeman [10].

(b) If the interior $\mathring{\Delta}_{m-1}$ of the simplex is endowed with the Riemannian metric defined on $T_x\mathring{\Delta}_{m-1}$ by

$$\langle\langle \xi, \eta \rangle\rangle = \sum_{i=1}^{m} \frac{\xi_i \eta_i}{x_i}, \qquad x = (x_1, \dots, x_m) \in \mathring{\Delta}_{m-1},$$

then (3.2) is actually 4 times the gradient flow of the linear functional $x \mapsto c'x$ on $\mathring{\Delta}_{m-1}$.

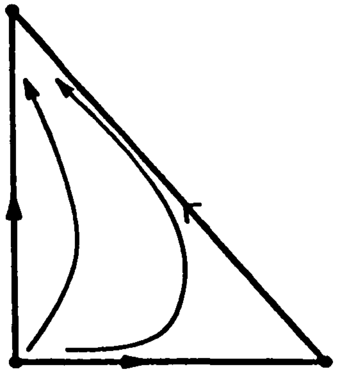

FIGURE 3.1. Phase portrait of (3.2)

(c) In [6] Karmarkar has analysed a class of interior point flows which are the continuous-time versions of the discrete time algorithm described in [5]. In the case of the standard simplex, Karmarkar's equations turn out to be

$$\dot{x}_i = \left(c_i x_i - \sum_{j=1}^{m} c_j x_j^2 \right) x_i , \qquad i = 1, \ldots, m, \tag{3.3}$$

$x_i \geq 0$, $\sum_{i=1}^{m} x_i = 1$. By linearising (3.3) around the optimal solution e_{i_*} it is easily shown that Karmarkar's equation (3.3) converges exponentially fast to e_{i_*} with (up to a factor 4) the *same rate of convergence* as for (3.3). A more general class of equations with the same convergence properties as (3.2) or (3.3) would be

$$\dot{x}_i = \left(c_i f(x_i) - \sum_{j=1}^{m} c_j f(x_j) x_j \right) x_i , \qquad i = 1, \ldots, m_1 , \tag{3.4}$$

with $f: [0, 1] \to \mathbb{R}$ a monotonically increasing C^1 function. Incidentally, the Karmarkar flow (3.3) is just a special case of the equations studied by Zeeman [10].

The following result estimates the time a trajectory of (3.2) needs in order to reach an ε-neighbourhood of the optimal vertex.

PROPOSITION 3.1. *Let* $0 < \varepsilon < 1$ *and* $\mu = \min_{j \neq i_*} (c_{i_*} - c_j)$. *Then for any initial condition* $x(0)$ *in the interior of* Δ_{m-1} *the solution* $x(t)$ *of* (3.2) *is contained in an* ε*-neighbourhood of the optimum vertex* e_{i_*} *if* $t \geq t_\varepsilon$ *where*

$$t_\varepsilon = \frac{|\log(\min_{1 \leq i \leq m} x_i(0) \varepsilon^2)|}{4\mu}. \tag{3.5}$$

Proof. We first show that every solution $x(t)$ of (3.2) is of the form

$$x(t) = \frac{e^{4tN}x(0)}{\langle e^{4tN}x(0)\rangle} \tag{3.6}$$

where $N = \operatorname{diag}(c_1, \dots, c_m)$ and $\langle e^{4tN}x(0)\rangle = \sum_{j=1}^{m} e^{4tc_j}x_j(0)$.

In fact, by differentiating the right-hand side of (3.6) one sees that both sides satisfy the same conditions (3.2) with identical initial conditions. Thus (3.6) holds. Using (3.6) one has

$$\begin{aligned}\|x(t) - e_{i_*}\|^2 &= \|x(t)\|^2 - 2\frac{\langle e^{4tN}x(0), e_{i_*}\rangle}{\langle e^{4tN}x(0)\rangle} + 1 \\ &\leq 2 - 2\frac{e^{4tc_{i_*}}x_{i_*}(0)}{\langle e^{4tN}x(0)\rangle}.\end{aligned} \tag{3.7}$$

Now

$$\sum_{j=1}^{m} e^{4t(c_j - c_{i_*})}x_j(0) \leq e^{-4\mu t} + x_{i_*}(0) \tag{3.8}$$

and thus

$$\|x(t) - e_{i_*}\|^2 \leq 2 - 2\left(1 + e^{-4\mu t}x_{i_*}(0)^{-1}\right)^{-1}.$$

Therefore $\|x(t) - e_{i_*}\| \leq \varepsilon$ if

$$\left(1 + e^{-4\mu t}x_{i_*}(0)^{-1}\right)^{-1} \geq 1 - \frac{\varepsilon^2}{2},$$

i.e., if

$$t \geq \left|\frac{\log\left(\frac{\varepsilon^2 x_{i_*}(0)}{2-\varepsilon^2}\right)}{4\mu}\right| \geq \frac{|\log(\varepsilon^2 \min_{1\leq i\leq m} x_i(0))|}{4\mu}.$$

This proves the result.

Note that for the initial condition $x(0) = \frac{1}{m}(1, \dots, 1)$ the estimate (3.5) becomes

$$t_\varepsilon \geq \frac{|\log \frac{\varepsilon^2}{m}|}{4\mu}, \tag{3.9}$$

and (3.9) gives an effective lower bound for (3.5) valid for all $x(0) \in \mathring{\Delta}_{m-1}$.

We can use Proposition 3.2 to obtain an explicit estimate for the time needed in either Brockett's flow (2.3) or for (2.9) that the projected interior point trajectory $\pi_T(x(t))$ enters an ε-neighbourhood of the optimal solution.

Thus let N, T be defined by (2.1), (2.2) with (2.8) understood and let v_{i_*} denote the optimal solution for the linear programming problem of maximising $c'x$ over the convex set $C(v_1, \dots, v_m)$.

Theorem 3.1. *Let $0 < \varepsilon < 1$, $\mu = \min_{j \neq i_*}(c'v_{i_*} - c'v_j)$, and let X be defined by (2.11). Then for all initial conditions $\xi(0) \in \mathbb{RP}^{m-1} - X$ the projected trajectory $\pi_T(\xi(t)) \in C(v_1, \ldots, v_m)$ of the solution $\xi(t)$ of (2.9) is in an ε-neighbourhood of the optimal vertex v_{i_*} for all $t \geq t_\varepsilon$ with*

$$t_\varepsilon = \frac{\left|\log \frac{\varepsilon^2}{m\|T\|^2}\right|}{4\mu}. \tag{3.10}$$

Proof. By (2.6), $\pi_T(\xi(t)) = Tx(t)$ where $x(t) = (\xi_1(t)^2, \ldots, \xi_m(t)^2)$ satisfies

$$\dot{x}_i = 4\left(c'v_i - \sum_{j=1}^{m} c'v_j x_j\right) x_i, \qquad i = 1, \ldots, m.$$

Proposition (3.2) implies $\|x(t) - e_{i_*}\| < \varepsilon/\|T\|$ for $t \geq |\log \frac{\varepsilon^2}{m\|T\|^2}|4\mu$ and hence

$$\|\pi_T(\xi(t)) - v_{i_*}\| \leq \|T\| \cdot \|x(t) - e_{i_*}\| < \varepsilon.$$

In the above, we were mainly concerned with interior point flows evolving on the standard simplex Δ_{m-1}. Thus these flows, such as (3.2)–(3.4), enable us to do sorting of lists of real numbers rather than solving more general linear programming problems. In recent work [3], L. Faibusovich has developed a rather complete theory of such interior point flows for linear programming, leading in fact to different flows than those studied by Karmarkar [6]. We briefly sketch Faibusovich's approach and stress its connection with the present work.

Let us consider the task of optimising a C^1 function $\varphi: \mathbb{R}^n \to \mathbb{R}$ over the convex constraint set P defined by the conditions

(i) $x \geq 0$, $x \in \mathbb{R}^n$,

(ii) $Ax = b$.

Here $A \in \mathbb{R}^{m \times n}$, $\operatorname{rk} A = m < n$ and $x \geq 0$ means that all components of x are non-negative. By linearity of the constraint (ii), the tangent space of P at an element x coincides with the kernel of A, that is,

$$T_x P = \{\xi \in \mathbb{R}^n | A\xi = 0\}, \qquad x \in P.$$

Given an interior point $x \in \mathring{P}$ and tangent vectors $\xi, \eta \in T_x P$, an inner product on $T_x P$ is defined by

$$\langle\langle \xi, \eta \rangle\rangle := \xi' D(x)^{-1} \eta, \qquad x \in \mathring{P}, \tag{3.11}$$

where $D(x) = \operatorname{diag}(x_1, \ldots, x_n)$. This defines a Riemannian metric on the set $\mathring{P} = \{x \in \mathbb{R}^n | Ax = b, x > 0\}$ of interior points of P.

It is now easy to compute the gradient of $\varphi\colon \mathring{P} \to \mathbb{R}$ with respect to the Riemannian metric $\langle\langle\ \rangle\rangle$. In fact, the gradient of φ at $x \in \mathring{P}$ is characterised by the property

$$\langle\langle \operatorname{grad}\varphi(x), \xi\rangle\rangle = \nabla\varphi(x)'\xi \qquad \forall \xi \in T_x P \tag{3.12}$$

where

$$\nabla\varphi(x)' = \left(\frac{\partial\varphi}{\partial x_1}(x), \dots, \frac{\partial\varphi}{\partial x_n}(x)\right).$$

Using (3.11) it follows that (3.12) is equivalent to $D(x)^{-1}\operatorname{grad}\varphi(x) - \nabla\varphi(x)$ being orthogonal to the kernel of A. Thus

$$D(x)^{-1}\operatorname{grad}\varphi(x) - \nabla\varphi(x) = A'\lambda$$

for a uniquely determined $\lambda \in \mathbb{R}^m$. Since $A\operatorname{grad}\varphi(x) = 0$ thus

$$\lambda = -(AD(x)A')^{-1}AD(x)\nabla\varphi(x)$$

and therefore

$$\begin{aligned}\operatorname{grad}\varphi(x) &= D(x)\nabla\varphi(x) - D(x)A'(AD(x)A')^{-1}AD(x)\nabla\varphi(x)\\ &= (I - D(x)A'(AD(x)A')^{-1}A)D(x)\nabla\varphi(x).\end{aligned} \tag{3.13}$$

Note that $AD(x)A'$ is positive definite, and hence is invertible, if A has full row rank and $x \in \mathring{P}$.

The associated gradient interior point flow for maximising $\varphi(x)$ subject to $x \in P$ thus is

$$\dot{x} = (I - D(x)A'(AD(x)A')^{-1}A)D(x)\nabla\varphi(x). \tag{3.14}$$

This coincides with the interior point flow first derived and studied by Faibusovich [3]. If $A = (1, \dots, 1)$ and $b = 1$, then the polytope P is equal to the simplex Δ_{m-1}. Let us also consider the special case where $\varphi(x) = \sum_{j=1}^{n} f_j(x_j)$. Then (3.14) is equivalent to

$$\dot{x}_i = \left(f_i'(x_i) - \sum_{j=1}^{n} f_j'(x_j)x_j\right)x_i, \qquad i = 1, \dots, n, \tag{3.15}$$

which is of the form (3.4). If $\varphi(x) = c'x$, the flow (3.15) coincides with (3.2) (up to constant factor), and for $\varphi(x) = \sum c_j x_j^2$, (3.15) is just Karmarkar's flow (3.3) on the simplex.

4. Further remarks and conclusions

As can be seen from the preceding sections, the "gradient method" for linear programming consists of the following program:

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

(i) To find a smooth compact manifold M and a smooth map $\pi: M \to \mathbb{R}^n$, which maps M onto the convex constraint set C.

(ii) To solve the gradient flow of the smooth function $\lambda \circ \pi: M \to \mathbb{R}$ and determine its stable equilibria points.

In the cases discussed here we had $M = \mathbb{RP}^{m-1}$ (or $M = \mathbb{RP}^{m-1} \times \cdots \times \mathbb{RP}^{m-1}$) and $\pi: \mathbb{RP}^{m-1} \to \mathbb{R}^n$ was the composition of the linear map $T: \Delta_{m-1} \to C(v_1, \dots, v_m)$ with the smooth map from $\mathbb{RP}^{m-1} \to \Delta_{m-1}$ defined by (2.4).

Note that for any smooth map $\pi: M \to \mathbb{R}^m$ with $\pi(M) = C$, the boundary points $y \in \partial C$ are necessarily critical values of π. Furthermore, if $\lambda: \mathbb{R}^n \to \mathbb{R}$ is a generic smooth function with no critical points in C, then the critical points of $\lambda \circ \pi: M \to \mathbb{R}$ are all contained in $\pi^{-1}(\partial C)$, thus reflecting the fact that λ assumes its minima and maxima on the boundary of C.

In a similar way other discrete optimisation problems may be investigated. Certainly the previous results say nothing about the actual convergence speed of these algorithms when implemented on a computer. Furthermore, there is in general no canonical choice for the compact manifolds M. In particular, it would be of interest to know the minimal dimension of M (this asks for a nonlinear realisation theory for linear programming!). The problem is even of interest if $n = 2$, i.e. for C a given compact subset of $\mathbb{R}^2$. Here the minimal dimension is two.

(i) If $C_2 = \langle 0, e_1, e_2 \rangle \subset \mathbb{R}^2$ is the standard 2-simplex of $\mathbb{R}^2$,

$$\pi: \mathbb{RP}^2 \to C_2 \subset \mathbb{R}^2,$$
$$[\xi_0 : \xi_1 : \xi_2] \mapsto \left(\frac{\xi_1^2}{|\xi|^2}, \frac{\xi_2^2}{|\xi|^2} \right)$$

defines a real algebraic (finite) map from the projective plane onto the standard 2-simplex $C_2 \subset \mathbb{R}^2$.

(ii) If $Q = \{(x, y) \in \mathbb{R}^2 \mid |x| \le 1, \ |y| \le 1\}$ has 4 vertices, there exists a smooth real algebraic (finite) map from the 2-torus onto Q.

$$\pi: S^1 \times S^1 \to Q,$$
$$(z, w) \mapsto (\operatorname{Re} z, \operatorname{Re} w).$$

(iii) Let C_6 be the hexagon imbedded into $\mathbb{R}^3$ by $C_6 = \{(\lambda_{\pi(1)}, \lambda_{\pi(2)}, \lambda_{\pi(3)}) | \pi$ a permutation$\}$, where $\lambda_1 > \lambda_2 > \lambda_3$ are fixed real numbers. Let $\mathrm{Jac}(\lambda_1, \lambda_2, \lambda_3)$ denote the set of real symmetric 3×3 Jacobi matrices

$$H = \begin{bmatrix} a_1 & b_1 & 0 \\ b_1 & a_2 & b_2 \\ 0 & b_2 & a_3 \end{bmatrix}$$

having fixed eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Then $\mathrm{Jac}(\lambda_1, \lambda_2, \lambda_3)$ can be shown to be a compact orientable two-manifold (of genus 2), Tomei [9], and

$$\pi \colon \mathrm{Jac}(\lambda_1, \lambda_2, \lambda_3) \to C_6, \quad H \mapsto (a_1, a_2, a_3)$$

is a surjection onto C_6. The gradient flow of the induced linear functional $H \mapsto \sum_{i=1}^{3} c_i a_i$ on $\mathrm{Jac}(\lambda_1, \lambda_2, \lambda_3)$ with respect to a certain Riemannian metric is then identical with Brockett's flow (1.1) for $N = \mathrm{diag}(c_1, c_2, c_3)$, if $(c_1, c_2, c_3) = (1, 2, 3)$ [1].

These are all cases in which I know how to *construct* explicitly a map from a smooth compact surface onto a given polytope.

One can always construct a smooth map $\pi \colon X \to C(v_1, \dots, v_m)$ from an amalgated sum $M = \mathbb{RP}^2 \times \cdots \times \mathbb{RP}^2$ of real projective planes onto $C(v_1, \dots, v_m)$. But this map π would in general not be finite.

Another possibility could be to use complex variable theory to obtain a uniformisation of a given convex subset of $\mathbb{R}^2$ by the upper half-plane as in Figure 4.1.

While the uniformising map S can be explicitly written down by means of the Schwarz-Christoffel formula, it is much harder to find the inverse images of the vertex points. Thus this approach does not really lead to an effective way of computing the induced gradients on the upper half plane. Also the higher dimensional analogy of this approach would require higher dimensional uniformisation theory which is known to be deep and difficult.

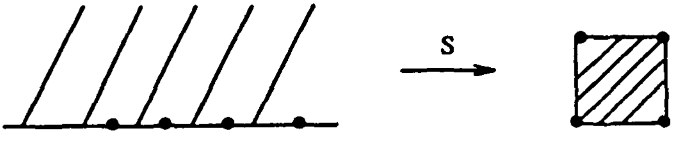

FIGURE 4.1. Example of maps S

https://doi.org/10.1017/S0334270000009048 Published online by Cambridge University Press

Acknowledgement

This work was completed during two months in 1991 which the author spent at the Department of Systems Engineering, A. N. U. Canberra. It is a pleasure to thank the members of that department, and in particular my host, Professor John Moore, for generating such a lively atmosphere for research, and for their hospitality. I would also like to thank Leonid Faibusovich for helpful and enthusiastic discussions and Roger Brockett for providing me with a copy of [6].

References

[1] A. M. Bloch, "Steepest descent, linear programming and Hamiltonian flows", *Contemporary Math.* **114** (1990) 77–88.

[2] R. W. Brockett, "Dynamical systems that sort lists and solve linear programming problems", *Proc. 27th IEEE Conference on Decision and Control*, (1988) 779–803.

[3] L. Faibusovich, "Hamiltonian structure of dynamical systems which solve linear programming problems", preprint (1990).

[4] U. Helmke, "Isospectral flows on symmetric matrices and the Riccati equation", *Systems and Control Letters* **16** (1991) 159–166.

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

[6] N. Karmarkar, "Riemannian geometry underlying interior point methods for linear programming", *Physica* D53 (1991) 217–232.

[7] E. Oja, "A simplified neuron model as a principal component analyzer", *J. Math. Biology* **15** (1982) 267–273.

[8] P. Schuster, K. Sigmund and R. Worff, "Dynamical systems under constant organisation I: Topological analysis of a family of nonlinear differential equations", *Bull. Math. Biophys.* **40** (1978) 743–769.

[9] C. Tomei, "The topology of isospectral manifolds of tridiagonal matrices", *Duke Math. J.* **51** (1984) 981–996.

[10] E. C. Zeeman, "Population dynamics from game theory", in *Global Theory of Dynamical Systems*, eds. Z. Nitecki and C. Robinson, Lecture Notes in Mathematics **819** (1980), Springer-Verlag, Berlin, 471–497.