Proceedings of the 31st Conference
on Decision and Control
Tucson, Arizona • December 1992

On the Geometry of Saddle Point Algorithms.

Anthony M. Bloch[1]
Department of Mathematics
The Ohio State University
Columbus, OH 43210

Roger W. Brockett[2]
Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Tudor S. Ratiu[3]
Department of Mathematics
University of California
Santa Cruz, CA 95064

Abstract.

There has been great deal of innovative work in recent years relating discrete algorithms to continuous flows. Of particular interest are flows which are gradient flows or Hamiltonian flows. Hamiltonian flows do not have asymptotically stable equilibria, but a restriction of the system to a certain set of variables may have such an equilibrium. In nonlinear optimization and game theory one is interested in systems with saddle point equilibria. We show here that certain flows with such equilibria can be both Hamiltonian and gradient and we discuss the relationship of such flows with the gradient method for finding saddle points in nonlinear optimization problems. We compare these results with gradient flows associated with the Toda lattice.

0. Introduction.

To those who embrace the Lagrange multiplier point of view for treating constrained finite dimensional optimization problems, it may be said that Hamiltonian systems arise in the study of the calculus of variations because the calculus of variations deals with constraints in the form of differential equations and when one introduces the, necessarily time dependent, Lagrange multipliers associated with these constraints, the simultaneous differential equations consisting of the original system and the multipliers, satisfy a set of equations of the Hamiltonian form. This is all very familiar to students of the maximum principle. The fact that the Hamiltonian point of view replaces the requirement that one search over an (infinite dimensional) space of curves with the requirement that one search over a (finite dimensional) set of boundary values is mitigated by the fact that the mixed boundary conditions are usually difficult to solve. Symes [1982] pointed out a new, and surprisingly different, role for Hamiltonian systems. He pointed out that the QR algorithm for diagonalizing symmetric matrices, can be thought of as being generated by a Hamiltonian flow in the following sense. The QR algorithm proceeds by successive application of a particular operation so as to generate a sequence of symmetric matrices. If L is the symmetric matrix to be diagonalized then there is an operation T such that the difference equation $L(k+1) = T(L(k))$, initialized with L, converges to a diagonal matrix having the same spectrum as L. Symes' observation was that $T(L)$ can be generated from L by solving on the interval $[0,1]$ a certain initial value problem for a Hamiltonian system of differential equations. Because the QR algorithm can be regarded as an algorithm for minimizing the sum of the squares of the off-diagonal elements of a symmetric matrix, subject to the constraint that the eigenvalues are $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, we may say that Symes discovered a role for Hamiltonian systems of the initial value type in solving constrained finite dimensional (i.e. not calculus of variations) optimization problems. In fact, we can be even more specific. The Hamiltonian system of Symes is an integrable system with eigenvalues of L playing the role of conserved (momentum-like) quantities. This is not an isolated story. Deift, Nanda, and Tomei [1983] also developed results of this type and Bayer and Lagarias [1989] showed that the so-called A-Trajectories associated with one version of Karmarkar's algorithm for linear programming are also generated by Hamiltonian systems in a similar way. (For further work in this regard see the references.)

The main goal of this paper is to provide some ways of thinking about how these Hamiltonian methods fit in with the more familiar ways of using differential equations to solve finite dimensional optimization problems. Of course the best known differential equation method for optimization is the continuous descent method for solving unconstrained maximization problems whereby to maximize V one simply integrates the equation $dx/dt = \nabla V$ until an equilibrium is reached. From the point of view of dynamical systems, one of the important aspects of the use of Hamiltonian systems in this context is that (finite-dimensional) Hamiltonian systems do not have asymptotically stable equilibria. To see this, one need only consider the eigenvalue pattern associated with linearization at an equilibrium point. Thus these algorithms succeed for reasons that are different from those that explain the success of gradient algorithms. However, it can (but need not) happen

[1] Supported in part by NSF Grant DMS-90-02136 and NSF PYI Grant DMS-9157556.

[2] Supported in part by AFOSR grant F49620-92-5-0466, by U.S. Army Research Office grant DAAL03-86-K-0171 and by NSF grant CDR-85-00108.

[3] Supported in part by NSF Grant DMS-9142613.

that in the neighborhood of an equilibrium point of a Hamiltonian system of dimension $2n$ there will be an n-dimensional manifold that is both invariant and asymptotically stable in the sense that any trajectory that starts on this manifold remains on it and converges to the equilibrium point. The Symes vector field provides an example of this.

In 1958 Arrow, Hurwicz and Uzawa published their famous paper on mathematical programming (Arrow et. al [1958]) in which they gave a gradient-like algorithm for solving constrained optimization problems. We may distinguish between two aspects of their work. The most basic (one might even say trivial) part consists of the observation that if one introduces the constraints using Lagrange multipliers and then solves the mixed gradient flow

$$dx/dt = \nabla V + \langle p, \nabla f\rangle$$
$$dp/dt = -f,$$

then it is easy to give natural conditions under which this system converges to the constrained maximum. The treatment of inequality constraints requires more insight and deeper analysis. Because Arrow et al. provide a basic algorithm for constrained optimization and, because the Hamiltonian methods do as well, they should be compared.

In this paper we observe in Section 1 that certain flows with saddle point equilibria can be both Hamiltonian and gradient. To explain this it seems natural to invoke some complex function theory. Saddle point equilibria play a role in nonlinear optimization problems with constraints (as well as in game theory and H^∞ control for example) and we point out in Section 2 the connection with the gradient method of Arrow, Hurwicz and Uzawa [1958]. We then discuss in Section 3 aspects of the Hamiltonian and gradient structures associated with the Toda lattice flow. In this system we start off by studying a Hamiltonian system with no asymptotically stable equilibria, but by restricting to a level set of the integrals of motion, we find a gradient system with a desired stable equilibrium. Details may be found in Bloch, Brockett and Ratiu [1992].

1. Complex Analytic Hamiltonian and Gradient Flows.

Two basic types of flows in dynamical systems theory are gradient and Hamiltonian flows. There are, in general, fundamental differences between such flows, but one can ask when flows are simultaneously Hamiltonian and gradient. An immediate restriction on such flows is that the equilibria must be compatible with both types of system. For example, a Hamiltonian system cannot have an asymptotically stable equilibrium. Thus, as we will see in the next section, the Toda flow in the variables of Flaschka is gradient only when restricted to a level set of its integrals.

However there are flows that are simultaneously Hamiltonian and gradient flows.

Example 1. Consider the flow

$$\begin{aligned}\dot{x} &= -2y\\ \dot{y} &= -2x.\end{aligned} \tag{1.1}$$

This is a Hamiltonian flow on $\mathbb{R}^2$ with Hamiltonian $H(x,y) = x^2 - y^2$ with respect to the usual symplectic structure (see e.g Abraham and Marsden [1978]), and is gradient with gradient function $-G(x,y) = -2xy$ with respect to the usual metric. In other words the flow may be written

$$\begin{aligned}\dot{x} &= \frac{\partial H}{\partial y} = -\frac{\partial G}{\partial x}\\ \dot{y} &= -\frac{\partial H}{\partial x} = -\frac{\partial G}{\partial y}.\end{aligned} \tag{1.2}$$

Note also that at the equilibrium the linearized flow has eigenvalues ± 1.

In fact this is a special case of a general result:

Consider $\mathbb{C}^n = \mathbb{R}^{2n}$ with coordinates $z = (z_1, \cdots, z_n) = x + iy = (x_1, \cdots, x_n) + i(y_1, \cdots, y_n)$. We suppose $\mathbb{R}^{2n}$ is endowed with the usual symplectic structure, the skew symmetric bilinear form $\omega = \Sigma_i (dx_i \wedge dy_i)$, and the usual Riemannian structure $ds^2 = \Sigma_i ((dx_i)^2 + (dy_i)^2)$.

Theorem 1.1. *Let $\mathbb{C}^n = \mathbb{R}^{2n}$ have the usual symplectic and Hamiltonian structures. Let $H(x,y)$ and $G(x,y)$ be real functions with continuous first partial derivatives. Then the Hamiltonian flow of H and the gradient flow of $-G$ are identical if and only if $f(z) = H(x,y) + iG(x,y)$ is analytic on $\mathbb{C}^n$.*

Proof: Write the Hamiltonian and gradient flows in standard form as (1.2). The result then follows from the Cauchy Riemann equations.

We note that such a Hamiltonian or gradient flow cannot have a source or sink.

Thus there are numerous examples of flows with this dual Hamiltonian and gradient character. In the above example $f(z) = z^2$

Example 2. Let $f(z) = z^3$. Then we obtain the flow

$$\begin{aligned}\dot{x} &= -6xy\\ \dot{y} &= -3x^2 + 3y^2.\end{aligned} \tag{1.3}$$

This has Hamiltonian $x^3 - 3xy^2$ and gradient function $-3x^2y - y^3$.

Example 3. Let $f(z) = e^z$. Then we obtain the flow

$$\begin{aligned}\dot{x} &= -e^x \sin y \\ \dot{y} &= -e^x \cos y. \end{aligned} \tag{1.4}$$

This has Hamiltonian $e^x \cos y$ and gradient function $-e^x \sin y$. Note also that the vector field here has no zeros. This is a feature of the Toda lattice flow in its original physical variables (see section 3).

We note also that there is a natural way of obtaining gradient flows from Hamiltonian flows on Kähler manifolds.

Recall that a real vector space V of even dimension is said to have a complex structure if there is a real linear isomorphism $J : V \to V$ such that $J^2 = -I$. If V has an inner product and ω is a symplectic structure on V then there exists a complex structure J on V and a (possibly different) real inner product g on V such that $g(\xi, \eta) = -\omega(J\xi, \eta)$. Setting $h(\xi, \eta) = g(\xi, \eta) - i\omega(\xi, \eta)$, h is a hermitian inner product on V regarded as a complex vector space (see Abraham and Marsden [1978], p173.). If the relation above holds for a given real inner product g, complex structure J and symplectic form ω, the triple (g, J, ω) is said to be *calibrated.* If only J and ω are given, then they are part of a calibrated triple if and only if $\omega(\xi, J\eta)$ defines a real inner product on V.

To globalize these notions on manifolds, more structure is needed. Given a Riemannian manifold (M, g) on which an almost symplectic structure ω is defined (i.e. ω is nondegenerate but not necessarily closed), one can induce an almost complex structure J, i.e. a smooth complex structure on each tangent space. This does not make M into a complex manifold unless the complex structure is integrable, or, equivalently, there exists an affine connection ∇ whose torsion vanishes and $\nabla J = 0$ (Nelson [1967]). As in the vector space case the triplet (g, J, ω) is said to be calibrated if the calibration holds on each tangent space. A calibrated triple (g, J, ω) defines a *Kähler structure* on M if J is a complex structure and ω is a symplectic form. Given a calibrated triple on M with ∇ the Riemannian connection, if $\nabla J = 0$ or $\nabla\omega = 0$ then M is a Kähler manifold. Conversely, if (g, J, ω) is a Kähler structure on M then both $\nabla J = 0$ and $\nabla\omega = 0$.

Note that if (M, g, J, ω) is a Kähler manifold and $F : M \to \mathbb{R}$ is a smooth function the relation between the gradient vector field ∇F (relative to g) and the Hamiltonian vector field X_F is given by $X_F = -J\nabla F$ or $\nabla F = JX_F$.

In the simple situation above with the usual complex structure and symplectic form, the Kähler metric is the usual metric. Letting i act on the Hamiltonian vector field with Hamiltonian $H(x, y)$ gives the gradient flow of $H(x, y)$ with respect to the usual metric. Thus in example 1 above action by i gives the flow

$$\begin{aligned}\dot{x} &= 2x \\ \dot{y} &= -2y. \end{aligned} \tag{1.5}$$

which is indeed the gradient flow of $x^2 - y^2$.

Similarly, letting i act on the Hamiltonian flow of $-G(x, y) = -2xy$ gives the flow (1.1). Thus we see that another way to interpret Theorem 1.1 is to consider the Hamiltonian flows of both $H(x, y)$ and $-G(x, y)$. Then, letting i act on the Hamiltonian flow of $-G$ gives a gradient flow which equals the Hamiltonian flow of H. (The Toda lattice flow in Moser's variables (see section 3) may be obtained using this construction (see Bloch, Flaschka and Ratiu [1990]).)

On compact manifolds, one cannot get the above gradient/Hamiltonian duality.

Theorem 1.2. *On a compact manifold which is both Riemannian and symplectic (for example Kähler) there are no vector fields with isolated equilibria which are simultaneously gradient and Hamiltonian.*

Proof. Assume that $X = -\nabla F = X_G$ for some smooth functions $F, G :\to \mathbb{R}$. Since all equilibria of X are isolated, the minimum x_0 of F is an isolated critical point. Since $\frac{dF(x)}{dt} = -\|\nabla F(x)\|^2$, the function $F(x) - F(x_0)$ is a strict Liapunov function. By Liapunov's theorem, x_0 is then asymptotically stable. But this is impossible since G is conserved on the flow of X_G and the level sets of G are necessarily compact. □

The above flows which are both Hamiltonian and gradient are of course flows in an even number of variables. It is possible however, by using the theory of Poisson structures (see e.g Weinstein [1983]) to have flows in an odd number of variables which are both Hamiltonian (in a generalized sense) and gradient.

Recall that for P a smooth manifold, a Poisson structure on P is a skew-symmetric bilinear map on the C^∞ functions on P called the Poisson bracket and written $\{F, G\}$ for $F, G \in C^\infty(P)$, and which satisfies the Jacobi identity and Leibniz's rule. A manifold P equipped with such a structure is called a Poisson manifold.

Associated with any smooth function H on $(P, \{\cdot, \cdot\})$ is a smooth Hamiltonian vector field X_H given by $\langle dF, X_H \rangle = \{F, H\}$ for any smooth function F on P. In local coordinates $x = x_1, \cdots, x_m$ this Hamiltonian flow may be written $\dot{x} = J(x)\text{grad}H(x)$

where grad is the standard gradient. $J(x)$ is a skew-symmetric matrix function with entries $J_{kl}(x) = \{x_k, x_l\}$.

Example 4. To illustrate the above we consider now the following example on $\mathbb{R}^n$(closely related to a probem in network dynamics discussed in Maschke, van der Schaft and Breedveld [1992]): Let $G(x,y) = (-x_1 + x_2)y_1$. Then the gradient flow of G with respect to the usual metric is

$$\begin{aligned}\dot{x}_1 &= -y_1 \\ \dot{x}_2 &= y_1 \\ \dot{y}_1 &= -x_1 + x_2. \end{aligned} \tag{1.6}$$

Now let $F(x,y) = 1/2(-x_1^2 - x_2^2 + y_1^2)$ and let J be the matrix

$$J = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Then this flow is also a Hamiltonian flow given by $\dot{x} = J\text{grad}F$.

In fact, viewed in the right variables this is a simple extension of example 1. Letting, $r = x_1 + x_2, q = 2p_1, p = x_1 - x_2$ the flow becomes

$$\begin{aligned}\dot{q} &= -2p \\ \dot{p} &= -2q \\ \dot{r} &= 0. \end{aligned} \tag{1.7}$$

Thus, the (q,p) flow here is again Hamiltonian with Hamiltonian $q^2 - p^2$ and gradient with gradient function $-2pq$.

In fact (see e.g. Weinstein [1983], Maschke et.al. [1992]) any Poisson system whose Poisson bracket is of constant rank $2n$ may be written locally as a flow of the form $\dot{q}_i = \frac{\partial H}{\partial p_i}, \dot{p}_i = -\frac{\partial H}{\partial q_i}, \dot{r}_j = 0, i = 1 \cdots n, j = 1, \cdots m$. Thus if H is the real part of an analytic function, this flow is again both Hamiltonian and gradient.

We remark here that there is an interesting relationship between the work here and network theory (see e.g. Smale [1972] and Maschke et. al. [1992].) In the theory of nondissipative networks, the flows are gradient with respect to an indefinite metric and can at the same time be Hamiltonian or Poisson. The critical points are typically centers. Here, in contrast we have saddles, and flows which are Hamiltonian and gradient with respect to the standard metric. An indefinite metric may be used however to turn saddles into sinks as in the next section.

2. Saddle Point Algorithms

We now consider the connection of the above analysis with the gradient method in nonlinear programming.

A typical problem is formulated as: Minimize $\theta(x)$, subject to $x \in U \subset \mathbb{R}^n$ and $g_i(x) \leq 0, i = 1, \cdots, n$.

One forms the Lagrangian function ψ associated with the problem defined by

$$\psi(x,u) = \theta(x) + \langle p, g(x) \rangle, \tag{2.1}$$

where $p = (p_1, \cdots, p_m)$ is the vector of Lagrange multipliers.

A pair (x^*, p^*), $x^* \in U$, is called a saddle point if for all $x \in U$ and p, $\psi(x,p^*) \leq \psi(x^*,p^*) \leq \psi(x^*,p)$ and $p \geq 0$. It follows that x^* is an optimal solution to the nonlinear programming problem.

To find a saddle point of such a programming problem Arrow, Hurwicz and Uzawa [1958] formulated a gradient method. Let $\psi(x,p)$ be strictly concave and of class C^2 in x and strictly convex and C^2 in p and possess a saddle point (x^*, p^*). (One usually also wishes to keep x and p positive – we consider this below.) To approach the saddle the Arrow-Hurwicz-Uzawa (A-H-U) technique is to use the "gradient" flow

$$\begin{aligned}\dot{x}_i &= \frac{\partial \psi}{\partial x_i} \\ \dot{p}_j &= -\frac{\partial \psi}{\partial p_j} \end{aligned} \tag{2.2}$$

Now how is this related to our gradient flows? Set now $n = m$. First we note that the A-H-U flow is the gradient of ψ not with respect to the usual metric on $\mathbb{R}^{2n}$, but with respect to the metric with matrix

$$\begin{aligned} g_{ij} &= \delta_{ij}, \ i,j = 1, \cdots, n, \\ &= -\delta_{ij}, \ i,j = n+1, \cdots, 2n \end{aligned} \tag{2.3}$$

Now we can show:

Theorem 2.1. *Suppose we have a function $H(x,y)$ with saddle point as in Theorem 1.1. Form a new analytic function $if(x,y)$, by multiplying f by i. Then the A-H-U flow is given by the gradient flow of the imaginary part of if with respect to the metric (2.3) and is equal to the Hamiltonian flow of the real part of if multiplied by the matrix g_{ij} .*

Referring to example 1, note that the A-H-U flow is given by (1.5) with the opposite sign for $\dot{y}$.

The point about the A-H-U flow is that it turns the saddle into a stable equilibrium. This can also be understood rather nicely in the following fundamentally complex sense.

Let Q be a real symmetric $n \times n$ matrix. Then we have

Proposition 2.2. *Consider the function* $f(z) = z^TQz = H(x,y) + iG(x,y) = (x^TQx - y^TQy) + i(y^TQx - x^TQy)$ *on* $\mathbb{C}^n = \mathbb{R}^{2n}$. *The flow* $\dot{z} = \frac{\partial f}{\partial z} = 2Qz$ *is the A-H-U flow and thus is stable if and only if* Q *is negative definite.*

Again we refer to the simple example 1, where $f(z) = z^2$, and the flow $\dot{z} = 2z$ is the A-H-U flow mentioned above.

This result is easy to see, but in fact a more general result is true, which follows again from the Cauchy-Riemann equations:

Theorem 2.3. *Consider the gradient flow of* $F(x,y)$, *the real part of an analytic function* $f(z)$. *Then the smooth A-H-U flow of* F *is given by* $\dot{z} = f'(z)$.

Proof.. We have $f(z) = F(x,y) + iG(x,y)$ where G is the harmonic conjugate of F. Hence

$$f'(z) = F_x + iG_x = F_x - iG_y. \qquad (2.4)$$

Hence $\dot{z} = f'(z)$ does indeed give the Arrow flow.

Now in general, as mentioned earlier, one wishes to keep the variables in the A-H-U flow in the positive quadrant. The full Arrow-Hurwicz-Uzawa algorithm which accomplishes this is given by

$$\begin{aligned} \dot{x}_i &= 0 \quad \text{if} \quad F_{x_i} < 0 \quad \text{and} \quad x_i = 0 \\ \dot{x}_i &= F_{x_i} \qquad \text{otherwise} \\ \dot{y}_i &= 0 \quad \text{if} \quad F_{y_i} > 0 \quad \text{and} \quad y_i = 0 \\ \dot{y}_i &= -F_{y_i} \qquad \text{otherwise.} \end{aligned} \qquad (2.5)$$

The full A-H-U algorithm may also be written compactly in complex form. It is given by

$$\begin{aligned} \dot{z}_i &= i\mathrm{Im}(f_{z_i}) \quad \text{if} \quad \mathrm{Re}(f_{z_i}) < 0 \quad \text{and} \quad \mathrm{Re}(z_i) = 0 \\ \dot{z}_i &= \mathrm{Re}(f_{z_i}) \quad \text{if} \quad \mathrm{Im}(f_{z_i}) < 0 \quad \text{and} \quad \mathrm{Im}(z_i) = 0 \\ \dot{z}_i &= f_{z_i} \qquad \text{otherwise.} \end{aligned} \qquad (2.6)$$

3. Gradient Flows and the Toda Lattice Equations

In this section we give a brief discussion of gradient flows on adjoint orbits of compact Lie groups and the (Hamiltonian) Toda lattice equations and contrast it with the work above. Details may be found in Bloch et.al [1992] and other references below.

Let G_u be a compact Lie group with Lie algebra $\mathcal{G}_u$ and let $\kappa(\ ,\ )$ denote the Killing form on $\mathcal{G}_u$. Consider the flow (discussed in Brockett [1988] in the unitary case)

$$\dot{L}(t) = [L(t), [L(t), N]]. \qquad (3.1)$$

for L and N in $\mathcal{G}_u$.

This flow is in fact a gradient flow:

Proposition 3.1. *The flow (3.1) is the gradient vector field of* $H(L) = \kappa(L,N)$, κ *the Killing form, on the adjoint orbit* $\mathcal{O}$ *of* $\mathcal{G}_u$ *containing the initial condition* $L(0) = L_0$, *with respect to the normal metric* $\langle\ ,\ \rangle_N$ *on* $\mathcal{O}$.

The normal metric is defined for example in Atiyah [1982]. For the proof of the proposition see Bloch et. al. [1992].

A key property of the flow (3.1) is that it is *isospectral*, i.e. the eigenvalues of L are preserved. Further, we can show that the generalized (Hamiltonian) Toda lattice equations may be written in this form and hence are a gradient flow on their isospectral set (the level set to their integrals). (For $sl(n)$ L is a tridiagonal symmetric matrix and N is diagonal. We recall (see Moser [1974], Flaschka [1976]) that the Toda lattice equations are the integrable equations of motion for a lattice of particles on the line interacting under an exponential potential. Flaschka found an ingenious change of variables which converted the equations to Lax pair form.

We note that the Toda lattice is gradient on its restriction to the isospectral set, since it is the restriction of a gradient vector field on a G_u orbit. There appears to be no appropriate metric off the isospectral set. Further there exits a stable equilibrium (when L is diagonal). Similarly, Moser's form of the gradient flow (see Moser [1974]) also occurs on the isospectral set. For the precise relationship between Moser' flow and the double bracket flow see Bloch, Flaschka and Ratiu [1990]. One can in fact show that Moser's flow is the gradient of a linear functional on complex projective space with respect to the Kähler metric, which in this instance coincides with the normal metric. This is in fact of particular interest in connection with the nonlinear programming problems discussed earlier:

One may view complex projective space $\mathbb{C}P^l$ as a coadjoint orbit of the special unitary group $U(l+1)$. (Simply consider an orbit through an element of $u(l+1)$ of the form $i\mathrm{diag}(1,0,\cdots,0)$). The elements of the orbit are then rank one projection matrices multiplied by i. In this case the Kähler and normal metrics on the orbit coincide and the gradient flow of $\mathrm{Tr}(\Lambda P)$, Λ a real diagonal matrix with distinct entries λ_i, P a projection matrix, is given by

$\dot{P} = [P, [P, \Lambda]]$. Now P being a rank one projection matrix is of the form $z\bar{z}^T$, $z \in \mathbb{C}^{l+1}$, $\sum |z_i|^2 = 1$. Hence $\mathrm{Tr}(\Lambda P) = \sum \lambda_i |z_i|^2$, and the flow minimizes this quadratic form subject to the above constraint. In fact the variables $p_{ii} = |z_i|^2$ may be viewed as lying in the standard polytope. In general the Toda flows may be related to flows in a convex polytope and thus to linear programming. This can be seen through the theory of convexity of the image of the momentum map (see Kostant [1973], Atiyah [1982] and Guillemin and Sternberg [1983]). For details on this work see Brockett [1988], [1992], Bloch, Brockett and Ratiu, [1992], Bloch, Flaschka and Ratiu [1990].

The overall picture we have developed here is as follows. The Toda flow is Hamiltonian and the level sets of the integrals of motion are Lagrangian submanifolds of the phase space. Also the level sets lie on orbits of $\mathcal{G}_u$. These level sets are invariant submanifolds for the gradient flow of the function $\kappa(L, N)$ with respect to the normal metric. On the level sets the Hamiltonian and gradient flows are identical for suitable N. Thus we have a picture of the connection between the Hamiltonian and gradient flows which is different from, but related to, that sketched in Section 1. In both cases there is an interesting relationship to algorithms.

Acknowledgement: We would like to thank P. Crouch for a useful conversation about circuits.

References

R. A. Abraham and J. E. Marsden [1978], *Foundations of Mechanics,* Benjamin Cummings.

K. J. Arrow, L. Hurwicz, and H. Uzawa (eds.) [1958], *Studies in Nonlinear Programming,* Stanford University Press.

G. Ammar and C. Martin [1986], *The geometry of matrix eigenvalue methods,* Acta. Appl. Math. **5**, 239-278.

M. F. Atiyah [1982], *Convexity and commuting Hamiltonians,* Bull. London Math. Soc. **16**, 1-15.

D. A. Bayer and J. Lagarias [1989], *The nonlinear geometry of linear programming I and II,* Trans. Amer. Math. Soc. **314** 499-526, 527-581.

A. M. Bloch [1990], *Steepest descent, linear programming and Hamiltonian flows,* Contemporary Math. AMS **114**, 77-88.

A. M. Bloch, H. Flaschka and T. Ratiu [1990], *A convexity theorem for isospectral sets of Jacobi matrices in a compact Lie algebra,* Duke Math. J. **61** (1), 41-66.

A. M. Bloch, R. W. Brockett and T. Ratiu [1992], *Completely Integrable Gradient Flows,* Communications in Mathematical Physics **147**, 57-74.

R. W. Brockett [1989], *Least squares matching problems,* Linear Algebra and its Applications, **122/123/124**, 761-777.

R. W. Brockett [1988], *Dynamical systems that sort lists and solve linear programming problems,* in "Proc. 27th IEEE Conf. on Decision and Control," IEEE, 1988, pp. 799-803, subsequently published in Linear Algebra and its Applications **146** (1991), 79-91.

R. W. Brockett [1992], *Differential geometry and the design of gradient algorithms,* Proc. of Symposia in Pure Mathematics (ed. R. Green, S. Yau), to appear.

C.I. Byrnes and J. C. Willems, *Least squares estimation, linear programming and momentum maps,*, IMA. J.

P. Deift, T. Nanda and C. Tomei [1983], *Differential equations for the symmetric eigenvalue problem,* SIAM J. on Numerical Analysis **20**, 1-22.

L. Faybusovich [1991] *Hamiltonian structure of dynamical systems which solve linear programming problems,* Physica D

H. Flaschka [1976], *The Toda lattice,* Phys. Rev. B **9**, 1924-1925.

V. Guillemin and S. Sternberg [1982], *Convexity properties of the moment mapping,* Invent. Math. **67**, 491-513.

U. Helmke [1991], *Isospectral Flows on Symmetric Matrices and the Riccati Equation*, Syst. Contr. Letters **16**, 159-165.

B. Kostant [1973], *On Convexity, the Weyl group and the Iwasawa decomposition,* Ann. Sci. Ec. Norm. Sup. **6**, 413-455.

J. Lagarias [1992], *Monotonicity properties of the generalized Toda flow and QR flow,* SIAM J. Matrix Analysis and Applications (to appear).

B. M. Maschke, A. J. van der Schaft, and P. C. Breedveld, *An intrinsic formulation of network dynamics: non-standard Poisson structures and gyrators,* J. Franklin Inst., to appear.

J. Moser [1974], *Finitely many mass points on the line under the influence of an exponential potential,* Batelles Recontres, Springer Notes in Physics, 417-497.

E. Nelson [1967], *Tensor Analysis,* Math. Notes, Princeton University Press.

S. Smale [1972], *On the mathematical foundations of circuit theory,* J. Diff. Geom. **7**, 193-210.

W. W. Symes [1982], *The QR algorithm and scattering for the nonperiodic Toda lattice,* Physica D **4**, 275-280.

D. S. Watkins and L. Elsner [1990], *On Rutishauser's approach to self-similar flows,* SIAM J. Matrix Anal. Appl. **11**, 301-311.

A. Weinstein [1983], *The local structure of Poisson manifolds,* J. Diff. Geom. **18**, 523-557.