Journal of Mathematical Systems, Estimation, and Control
Vol. 5, No. 3, 1995, pp. 1–13

A Hamiltonian Formalism for Optimization Problems*

Leonid Faybusovich

Abstract

We consider dynamical systems that solve general convex optimization problems. We describe in detail their Hamiltonian structure and the phase portrait. In particular, it is proved that these dynamical systems are completely integrable. The Marsden-Weinstein reduction procedure plays a crucial role in our constructions. Dikin-type algorithms are briefly discussed.

Key words: convex optimization, completely integrable Hamiltonian systems

1 Introduction

Consider the following optimization problem:

$$< c, x > \to \max, \tag{1.1}$$

$$< a_i, x >= b_i, i = 1, \dots, r, \tag{1.2}$$

$$f_j(x) \geq 0, j = 1, 2, \dots, m, x \in R^n. \tag{1.3}$$

Here $c, a_1, \dots, a_r \in R^n, b_1, \dots, b_r$ are real numbers and f_j are smooth concave functions. Suppose that for any $\beta > 0$ the problem

$$\beta < c, x > + \sum_{j=1}^{m} \ln f_j(x) \to max, \tag{1.4}$$

$$< a_i, x >= b_i, i = 1, \dots, r, x \in R^n, \tag{1.5}$$

has a unique solution $x(\beta)$ for any $\beta > 0$. It is reasonable to expect that $x(\beta)$ tends to a solution of the problem 1.1-1.3 when $\beta \to +\infty$. This approach to solving 1.1-1.3 (known as a penalty function method) is described

*Received January 18, 1994; received in final form September 2, 1994. Summary appear in Volume 5, Number 3, 1995.

in detail in [6]. The set of points $x(\beta), \beta \geq 0$, is usually called the central trajectory for the problem 1.1-1.3. For the construction of practical algorithms it is necessary to choose a sequence of points $\beta_1 < \beta_2 < \dots$ such that $\beta_k \to +\infty$ and find the way to calculate $x(\beta_k)$ with a sufficient accuracy, knowing $x(\beta_1), \dots, x(\beta_{k-1})$. Tremendous progress in the construction of efficient algorithms of this type has been made recently for the case where $f_1, \dots, f_m$ are linear functions [7]. The corresponding algorithms are known as path-following algorithms of linear programming. Regrettably, in the case of nonlinear functions $f_1, \dots, f_m$ progress is much more modest [8]. The experience of the linear case suggests that in the nonlinear situation it is necessary to consider large step (i.e. $\beta_k - \beta_{k-1}$ are large) procedures. In this situation a trajectory of the corresponding discrete algorithm does not follow closely the central trajectory. In the linear case it turns out to be possible to introduce certain dynamical systems (called affine-scaling vector fields) which are "infinitesimal" versions of large-step algorithms. See e.g. [1]. In the present paper we study the corresponding dynamical systems for the problem 1.1-1.3. Observe that our construction works only for the case of linear cost function (which can be assumed without loss of generality). See [4] for the case of nonlinear cost function and linear constraints. As in the linear case, these dynamical systems admit a Hamiltonian structure and, moreover, turn out to be completely integrable. We study this structure in some detail. In particular, we show that the equality constraints can be handled by means of the Marsden-Weinstein reduction procedure [9]. The Hamiltonian properties of our dynamical systems are quite similar to ones arising in mechanics and optimal control. There are further interesting analogies like the relation between Lagrangian and Hamiltonian approaches in mechanics and between primal and dual problems in optimization (in both cases via the Legendre transform).

2 Dynamical Systems that Solve Optimization Problems

Denote by P the set in R^n defined by constraints 1.2, 1.3 and by $int(P)$ the set $\{x \in P : f_j(x) > 0, j = 1, \dots, m\}$. Throughout this paper we suppose that $int(P) \neq \emptyset$ (not empty), P is compact and the vectors $a_1, \dots, a_r$ (see 1.2) are linearly independent. Denote by T the smallest vector subspace in R^n containing all vectors $a_1, \dots, a_r$. Let $\pi : R^n \to T^\perp$ be the orthogonal projection of R^n onto the orthogonal complement $T^\perp$ of T in R^n relative to the standard scalar product. Let, further,

$$\alpha(x) = \sum_{i=1}^{m} \ln f_i(x), \tag{2.1}$$

$$\nabla\alpha(x) = \psi(x) = \sum_{i=1}^{m} \frac{\nabla f_i(x)}{f_i(x)}, \tag{2.2}$$

$$\nabla^2\alpha(x) = \gamma(x) = \sum_{i=1}^{m} [\frac{\nabla^2 f_i(x)}{f_i(x)} - \frac{\nabla f_i(x)\nabla f_i(x)^T}{f_i^2(x)}]. \tag{2.3}$$

Here α is defined on the set

$$int(Q) = \{x \in R^n : f_i(x) > 0, i = 1, \dots, m\}. \tag{2.4}$$

We think of $\nabla^2\alpha(x)$ as n by n symmetric matrix. Observe that due to the imposed conditions the matrix $\gamma(x)$ is nonpositive definite. We assume throughout this paper that the matrix $\gamma(x)$ is negative definite for any $x \in int(Q)$. Let $\tilde{\psi} = \pi \circ \psi : int(P) \to T^{\perp}$.

Proposition 2.1 *The map $\tilde{\psi}$ is a diffeomorphism of $int(P)$ on $T^{\perp}$.*

Proof: Consider the problem

$$f_\beta(x) = \beta < c, x > +\alpha(x) \to max, \tag{2.5}$$

$$x \in P. \tag{2.6}$$

Here $c \in R^n$. Observe that for any $\beta \in R$ the problem 2.5,2.6 has a unique solution $x(\beta)$. Moreover, $x(\beta) \in int(P)$. Indeed, let $y \in int(P)$ and $P_y = \{x \in P : f_\beta(x) \geq f_\beta(y)\}$. It is clear that $\max\{f_\beta(x) : x \in P\} = \max\{f_\beta(x) : x \in P_y\}$. Suppose that $\tilde{x}$ belongs to the closure of P_y. Let $x_i, i = 1, 2, \dots$ be a sequence of points of P_y such that $x_i \to \tilde{x}, i \to +\infty$. It is clear that $\tilde{x} \in P$. If $\tilde{x} \in P \backslash int(P)$, then $\alpha(x_i) \to -\infty, i \to +\infty$. But then $f_\beta(x_i) \to -\infty, i \to +\infty$ contradicts the fact that $x_i \in P_y$ for any i. Hence, $\tilde{x} \in int(P)$. But f_β is continuous on $int(P)$. Consequently, $f_\beta(\tilde{x}) \geq f_\beta(y)$. We conclude that $\tilde{x} \in P_y$. Hence, P_y is a closed subset of P, i.e. P_y is compact.We proved that P_y is a compact subset of $int(P)$. Since $\nabla^2\alpha$ is strictly concave on $int(P)$, we conclude that the problem 2.5,2.6 has a unique solution $x(\beta) \in int(P)$. It is clear now that

$$\nabla f_\beta(x(\beta)) \in T$$

or $\beta c + \psi(x(\beta)) \in T$. Hence,

$$\beta(\pi c) + \tilde{\psi}(x(\beta)) = 0. \tag{2.7}$$

By 2.7 we see that the map $\tilde{\psi}$ is surjective. Let $z_1, z_2 \in int(P), \tilde{\psi}(z_1) = \tilde{\psi}(z_2) = v$. Consider the problem

$$f(x) = - < v, x > +\alpha(x) \to \max, x \in P. \tag{2.8}$$

We have $\nabla f(z_i) = -v + \psi(z_i) \in T, i = 1,2$. Hence, both z_1, z_2 are solutions to 2.8. This implies $z_1 = z_2$, since f is a strictly concave function on $int(P)$. ∎

Given $c \in R^n$, consider a vector field V_c on $int(P)$:

$$V_c(x) = -\gamma(x)^{-1}c + \gamma(x)^{-1}A^T(A\gamma(x)^{-1}A^T)^{-1}A\gamma(x)^{-1}c. \tag{2.9}$$

Here $\gamma(x)$ was defined in 2.3. It is an invertible matrix according to imposed conditions. Further, A is r by n matrix such that $A^T = [a_1, \dots, a_r]$.

Proposition 2.2 *Let $x(\beta)$ be (a unique) solution to the problem 2.5,2.6, $\beta \in R$. Then*

$$\frac{dx(\beta)}{d\beta} = V_c(x(\beta)).$$

Remark 2.3 The curves $x(\beta)$ are usually called central trajectories for the problem 1.1-1.3. The point $x(0)$ is called the analytic center of the set P.

Proposition 2.2 follows from the more general.

Proposition 2.4 *For any $x \in int(P)$ the integral curve $y(\beta)$ of 2.9 such that $y(0) = x$ has the form:*

$$y(\beta) = \tilde{\psi}^{-1}(\tilde{\psi}(x) - \beta\pi c), \beta \in R. \tag{2.10}$$

In particular, $int(P)$ is an invariant manifold for V_c.

Proof: It is clear that $y(0) = x$. Further, if $\delta(\beta) = \tilde{\psi}(y(\beta))$, we have by 2.10:

$$\frac{d\delta}{d\beta}(\beta) = -\pi c.$$

On the other hand,

$$\frac{d\delta}{d\beta} = D\tilde{\psi}(y(\beta))\dot{y}(\beta).$$

Further,

$$\epsilon(\beta) = D\tilde{\psi}(y(\beta))V_c(y(\beta)) = \pi\gamma(y(\beta))V_c(y(\beta)) = \pi(-c + A^T\Lambda(\beta)).$$

Here $\Lambda(\beta) = (A\gamma(y(\beta))^{-1}A^T)^{-1}A\gamma(y(\beta))^{-1}c$. But $ImA^T \subset T$. Hence, $\pi(A^T\Lambda(\beta)) = 0$. We conclude that $\epsilon(\beta) = -\pi c$. In other words,

$$-\pi c = D\tilde{\psi}(y(\beta))V_c(y(\beta)) = D\tilde{\psi}(y(\beta))\dot{y}(\beta).$$

Since $D\tilde{\psi}(y)$ is a linear bijection of $T^\perp = KerA$, it follows that $V_c(y(\beta)) = \dot{y}(\beta)$. Indeed, it is clear that $Ay(\beta) = b$. Hence, $\dot{y}(\beta) \in KerA$ and $AV_c(y(\beta)) = 0$. ∎

We now introduce a Riemannian metric g on $int(Q)$:

$$g(x;\xi,\eta) = -\eta^T\gamma(x)\xi. \tag{2.11}$$

Recall that $\gamma(x)$is negative definite on $int(Q)$ due to imposed conditions. It is clear that $int(P) \subset int(Q)$ can be thought of as a Riemannian submanifold.

Proposition 2.5 *The vector field V_c is a gradient flow of the function $f(x) = < c, x >$ on the manifold $int(P)$.*

Proof: Since the vector field V_c is tangent to the manifold $int(P)$, it is sufficient to verify that

$$< c, \xi >= g(x; V_c(x), \xi)$$

for any $\xi \in R^n$ such that $A\xi = 0$ and $x \in int(P)$. But

$$g(x; V_c(x), \xi) =< \xi, \gamma(x)\gamma(x)^{-1}c - \gamma(x)\gamma(x)^{-1}A^T\Lambda(x) >=< \xi, x >,$$

since $< \xi, A^T\Lambda(x) >=< A\xi, \Lambda(x) >= 0$. Here

$$\Lambda(x) = (A\gamma(x)^{-1}A^T)^{-1}A\gamma(x))^{-1}c.$$

∎

Theorem 2.6 *Let $y(\beta), \beta \in R$, be an integral curve of the vector field V_c such that $y(0) \in int(P)$. Then*

$$\lim < c, y(\beta) >=< c, x^* >, \beta \to +\infty,$$

$$\lim < c, y(\beta) >=< c, x_* >, \beta \to -\infty.$$

Here x^* (resp. x_*) is a solution to the problem $< c, x >\to \max$ (resp. $\min$), $x \in P$.

Corollary 2.7 *If x^* (resp. x_*) is a unique solution to the problem $< c, x >\to \max$ (resp. $\min$), $x \in P$, then $\lim y(\beta) = x^*, \beta \to +\infty$ (resp. $\lim y(\beta) = x_*, \beta \to -\infty$).*

Proof: Observe that $y(\beta)$ is the solution to the optimization problem:

$$< \beta c + d, x > +\alpha(x) \to \max, x \in P, \tag{2.12}$$

where $d = -\tilde{\psi}(y(0))$. Indeed, this easily follows by 2.7,2.10. Fix $\beta > 0$. Set

$$u_i(\beta) = \frac{1}{\beta f_i(y(\beta))}, i = 1, \ldots, m, \mu_\beta(x) =< c + d/\beta, x > .$$

We have:

$$\mu_\beta(x^*) \leq \mu_\beta(x^*) + \sum_{i=1}^m u_i(\beta) f_i(x^*) \leq$$

$$\mu_\beta(y(\beta)) + \sum_{i=1}^m u_i(\beta) f_i(y(\beta)) + < x^* - y(\beta), c + d/\beta + \psi(y(\beta))/\beta > .$$

Here we used a standard inequality

$$\nu(x) - \nu(y) \leq < \nabla\nu(y), x - y >$$

for the concave function

$$\nu(x) = \mu_\beta(x) + \sum_{i=1}^m u_i(\beta) f_i(x)$$

for $x = x^*, y = y(\beta)$. Observe now that $x^* - y(\beta) \in T^\perp$ and

$$\pi(c + d/\beta + \frac{\psi(y(\beta))}{\beta}) = 0.$$

Hence,

$$< x^* - y(\beta), c + d/\beta + \frac{\psi(y(\beta))}{\beta} >= 0.$$

Taking into account

$$\sum_{i=1}^m u_i(\beta) f_i(y(\beta)) = \frac{m}{\beta},$$

we arrive at the inequality

$$< c + d/\beta, x* > \leq < c + d/\beta, y(\beta) > + m/\beta, \beta > 0. \tag{2.13}$$

Similarly,

$$< c + d/\beta, x_* > \geq < c + d/\beta, y(\beta) > + m/\beta, \beta < 0. \tag{2.14}$$

Observe now that $y(\beta)$ belongs to the compact set P. This implies that $\lim \frac{<d, y(\beta)>}{\beta} = 0, |\beta| \to +\infty$. By Proposition 2.5 the function $< c, y(\beta) >$ is monotonically nondecreasing. We conclude from 2.13, 2.14 that

$$\lim_{\beta \to -\infty} < c, y(\beta) > \leq < c, x_* > \leq < c, x^* > \leq \lim_{\beta \to \infty} < c, y(\beta) > .$$

It is obvious , however, that $< c, x^* > \geq \lim < c, y(\beta) >, \beta \to +\infty$; $< c, x_* > \leq \lim < c, y(\beta) >, \beta \to -\infty$. ■

3 A Hamiltonian Structure

Our next goal is to introduce a Hamiltonian structure for the totality of vector fields V_c. Namely, we consider the dynamical systems

$$\dot{x} = V_c(x), \dot{c} = 0 \tag{3.1}$$

on the manifold $M = int(P) \times T^{\perp}$. We introduce a symplectic structure ω on M and prove that 3.1 is a completely integrable Hamiltonian system. We explicitly construct action-angle variables for 3.1 using the Legendre transform $\tilde{\psi}$. Here we follow the pattern of [3], where the linear programming problem in the canonical form was considered.

We begin by defining a symplectic structure on $int(Q) \times R^n$. Recall that

$$int(Q) = \{x \in R^n : f_i(x) > 0, i = 1, \ldots, m\}$$

and we defined a Riemannian metric $g(x; \xi, \eta) = - < \xi, \gamma(x)\eta >$ (see 2.11, 2.3) on $int(Q)$. Given $(x, c) \in int(Q) \times R^n$, set

$$\Omega(x, c; (\xi_1, \eta_1), (\xi_2, \eta_2)) = g(x; \xi_1, \eta_2) - g(x; \xi_2, \eta_1), \tag{3.2}$$

$(\xi_i, \eta_i) \in R^n \times R^n, i = 1, 2$. Recall (see e.g. [9]) that there is a canonical symplectic structure on $R^n \times R^n$:

$$\omega_{can}(x, y; (\xi_1, \eta_1), (\xi_2, \eta_2)) =< \xi_2, \eta_1 > - < \xi_1, \eta_2 >,$$

$x, y, \xi_i, \eta_i \in R^n, i = 1, 2$. By Proposition 2.1 (under the assumption that Q is compact) the map $\psi \times Id_{R^n} : int(Q) \times R^n \to R^n \times R^n$ is a diffeomorphism (see 2.2). We claim that Ω is a pullback of ω_{can} under this map.

Proposition 3.1 *We have:*

$$(\psi \times Id_{R^n})^*(\omega_{can}) = \Omega. \tag{3.3}$$

In particular, Ω is a symplectic two-form.

Proof: Let $\Gamma = (\psi \times Id_{R^n})^*(\omega_{can})$. We have:

$$\Gamma(x, c; (\xi_1, \eta_1), (\xi_2, \eta_2)) = \omega_{can}(\psi(x), c; D\psi(x)\xi_1, \eta_1), (D\psi(x)\xi_2, \eta_2)) =$$

$$< \gamma(x)\xi_2, \eta_1 > - < \gamma(x)\xi_1, \eta_2 >= \Omega(x, c; (\xi_1, \eta_1), (\xi_2, \eta_2)).$$

Hence, $\Gamma = \Omega$. Here we used $D\psi(x) = \gamma(x)$. ∎

Let H be a smooth function on the manifold $N = int(Q) \times R^n$. The corresponding Hamiltonian vector field X_H is determined by the condition [9]:

$$\Omega(x, c; X_H(x, c), (\xi, \eta)) = D_x H\xi + D_c H\eta \tag{3.4}$$

for any $(\xi, \eta) \in R^n \times R^n$. From 3.4 we easily obtain for X_H :

$$\dot{x} = -\gamma(x)^{-1} D_c H, \dot{c} = \gamma(x)^{-1} D_x H. \tag{3.5}$$

In particular, for the case $H(x,c) =< c, c > /2$, we arrive at the Hamiltonian system:

$$\dot{x} = -\gamma(x)^{-1} c, \dot{c} = 0, \tag{3.6}$$

which coincides with 3.1 for the situation considered (no equality constraints). Let H_1, H_2 be two smooth functions on N. To compute their Poisson bracket observe that [9]:

$$\{H_1, H_2\}(x,c) = \Omega(X_{H_1}, X_{H_2}).$$

Using 3.5, we easily obtain:

$$\{H_1, H_2\}(x,c) =$$

$$< D_x H_2(x,c), \gamma(x)^{-1} D_c H_1(x,c) > - < D_x H_1(x,c), \gamma(x)^{-1} D_c H_2(x,c) > . \tag{3.7}$$

Consider now the Hamiltonians $h_i : N \to R, h_i(x,c) =< a_i, c >, i = 1, \ldots, r$. Here $a_i, i = 1, \ldots, r$ are linearly independent vectors in R^n. It is clear by 3.7 that the corresponding Hamiltonians are in involution (i.e. $\{h_i, h_j\} = 0, i, j = 1, \ldots, r$). This enable us to define a Hamiltonian action [9] of the abelian group R^r on N. More precisely, according to 3.5 the Hamiltonian vector fields X_i corresponding to Hamiltonians h_i have the form:

$$\dot{x} = \gamma(x)^{-1} a_i, \dot{c} = 0, i = 1, \ldots, r. \tag{3.8}$$

Denote $-\gamma(x)^{-1} a_i$ by $W_i(x), i = 1, \ldots, r$. Introduce a simultaneous flow $G : int(Q) \times R^r \to int(Q)$ in the following way:

$$G(x; t_1, \ldots, t_r) = \psi^{-1}(\psi(x) - a_1 t_1 - \ldots - a_m t_m). \tag{3.9}$$

It is clear that $G(x; \vec{0}) = x$ for any $x \in int(Q)$. By Proposition 2.4:

$$\frac{\partial G}{\partial t_i}(x; \vec{t}) = W_i(G(x; \vec{t})), i = 1, \ldots r. \tag{3.10}$$

The above mentioned Hamiltonian action of R^r on N has the form:

$$\vec{t} \bullet (x,c) = (G(x; \vec{t}), c), \vec{t} \in R^r, x \in int(Q), c \in R^n. \tag{3.11}$$

The Hamiltonians $h_1, \ldots, h_r$ define the so-called moment map $\Phi : N \to R^r$:

$$\Phi(x,c) = (h_1(x,c), \ldots, h_r(x,c)) = (< a_1, c >, \ldots, < a_r, c >). \tag{3.12}$$

It is obvious that the set $int(Q) \times T^{\perp} = \Phi^{-1}(0)$ is an invariant manifold under the action of R^r on N. The Marsden-Weinstein reduction procedure [9] enables one to define a canonical symplectic structure on the factor-manifold $\Phi^{-1}(0)/R^r$. Our immediate goal is to identify this factor-manifold with $int(P) \times T^{\perp}$.

Lemma 3.2 *Given $x \in int(Q)$, suppose that $G(x;\vec{t}) \in int(P), G(x;\vec{u}) \in int(P)$ for some $\vec{t},\vec{u} \in R^r$. Then $\vec{t} = \vec{u}$.*

Proof: Recall that $\pi : R^n \to T^{\perp}$ is the orthogonal projection and $\tilde{\psi} = \pi\psi$ is a diffeomorphism of $int(P)$ onto $T^{\perp}$ (see Proposition 2.1). Let $\mu = G(x;\vec{u}) \in int(P), \nu = G(x;\vec{t}) \in int(P)$. By 3.9 $\psi(\mu) = \psi(x) - u_1a_1 - \ldots - u_ra_r$; $\psi(\nu) = \psi(x) - t_1a_1 - \ldots - t_ra_r$. Hence, $\pi\psi(\mu) = \pi\psi(\nu) = \pi\psi(x)$. This implies by Proposition 2.1 that $\mu = \nu$. Hence, $\psi(\nu) = \psi(\mu)$ and consequently, $u_1a_1 + \ldots + u_ra_r = t_1a_1 + \ldots t_ra_r$ or $t_i = u_i, i = 1,\ldots,r$, since the vectors $a_1,\ldots a_r$ are linearly independent. ∎

Lemma 3.3 *For any $x \in int(Q)$ there exists a unique $\vec{t}(x) \in R^r$ such that $G(x;\vec{t}(x)) \in int(P)$. Moreover, the map $x \to \vec{t}(x)$ is smooth.*

Proof: Set $y(x) = \tilde{\psi}^{-1}(\pi\psi(x))$. It is clear that $y(x)$ is a smooth function from $int(Q)$ into $int(P)$. We have: $\pi\psi(x) = \tilde{\psi}(y(x)) = \pi\psi(y(x))$. In other words, $\psi(x) - \psi(y(x)) \in T$. Consequently, $\psi(y(x)) = \psi(x) - t_1(x)a_1 - \ldots - t_r(x)a_r$ for some smooth functions $t_i(x)$. We claim that $G(x;\vec{t}(x)) = y(x) \in int(P)$. Indeed, $G(x;\vec{t}(x)) = \psi^{-1}(\psi(x) - t_1(x)a_1 - \ldots t_r(x)a_r)$. Hence, $\psi G(x;\vec{t}(x)) = \psi(y(x))$ or $G(x;\vec{t}(x)) = y(x)$. The uniqueness of $\vec{t}(x)$ follows by Lemma 3.2. ∎

Given $x \in int(Q)$, denote by $\varphi(x)$ the point $G(x;\vec{t}(x))$. In this way , we define a smooth map $\varphi : int(Q) \to int(P)$, which enable us to identify $\Phi^{-1}(0)/R^r$ with $int(P) \times T^{\perp}$. The map

$$\varphi \times Id_{T^{\perp}} : int(Q) \times T^{\perp} \to int(P) \times T^{\perp} \tag{3.13}$$

plays the role of the canonical projection under this identification . The manifold $\Phi^{-1}(0)/R^r$ is endowed with a natural symplectic structure ω [9] which is under our identifications has the form:

$$\omega(\varphi(x), c; (D\varphi(x)\xi_1, \eta_1), (D\varphi(x)\xi_2, \eta_2)) = \Omega(x, c; (\xi_1, \eta_1), (\xi_2, \eta_2)), \tag{3.14}$$

$x \in int(Q), \xi_i \in R^n, \eta_i \in T^{\perp}, c \in T^{\perp}$. Consider the Hamiltonian $H(x,c) = < c,c > /2$ on $int(Q) \times R^n$. It is clear by 3.5 that the corresponding Hamiltonian vector field has the form:

$$\dot{x} = W_c(x), \dot{c} = 0.$$

Since $\{H, h_i\} = 0, i = 1,\ldots r, H$ can be lifted to the Hamiltonian $\tilde{H}$ on $\Phi^{-1}(0)/R^r$. Under our identifications, $\tilde{H}(x,c) = \frac{<c,c>^2}{2}, x \in int(P), c \in T^{\perp}$.

Theorem 3.4 *The Hamiltonian vector field corresponding to the Hamiltonian $\tilde{H}$ relative to the symplectic form ω has the form 3.1 (see 2.9).*

Proof: Suppose that $\mu(t)$ is an integral curve of W_c. Consider the curve $\nu(t) = \varphi(\mu(t))$. By Proposition 2.4 $\mu(t) = \psi^{-1}(\psi(\mu(0)) - ct)$. On the other hand, by the very definition of φ we have:

$$\psi\varphi(x) = \psi(x) + t_1(x)a_1 + \ldots t_r(x)a_r, \tag{3.15}$$

$x \in int(Q)$ and some $t_1(x), \ldots t_r(x)$. In particular,

$$\psi\varphi(\mu(t)) = \psi(\mu(0)) - ct + t_1(\mu(t))a_1 + \ldots t_r(\mu(t))a_r.$$

Hence, $\tilde{\psi}\varphi(\mu(t)) = \tilde{\psi}(\mu(0)) - \pi ct$. On the other hand, $\tilde{\psi}(\mu(0)) = \tilde{\psi}(\varphi(\mu(0))$. We see by Proposition 2.4 that $\varphi(\mu(t))$ is the integral curve of V_c with the initial condition $\varphi(\mu(0))$. The result follows by general considerations [9]. ∎

Theorem 3.5 *The map $\tilde{\psi} \times Id_{T^\perp}$ is a symplectic diffeomorphism of $(int(P) \times T^\perp, \omega)$ onto $(T^\perp \times T^\perp, \omega_{can})$. Under this diffeomorphism the Hamiltonian system $\dot{x} = V_c(x), \dot{c} = 0$ goes to the dynamical system $\dot{x} = -\pi c, \dot{c} = 0$.*

Proof: Set $\Delta = (\tilde{\psi} \times Id_{T^\perp})^* \omega_{can}$. We have

$$\Delta(\varphi(x), c; (D\varphi(x)\xi_1, \eta_1), (D\varphi(x)\xi_2, \eta_2) =$$

$$\omega_{can}(\tilde{\psi}\varphi(x), c; (D(\tilde{\psi} \circ \varphi)(x)\xi_1, \eta_1), (D(\tilde{\psi} \circ \varphi)(x)\xi_2, \eta_2).$$

Here $\xi_i \in R^n, \eta_i \in T^\perp$. But $\tilde{\psi}(\varphi(x)) = \pi\psi(x)$ by 3.15. Hence,

$$\Delta(\varphi(x), c; (D\varphi(x)\xi_1, \eta_1), (D\varphi(x)\xi_2, \eta_2) =$$

$$\omega_{can}((D\psi(x)\xi_1, \eta_1), (D\psi(x)\xi_2, \eta_2)).$$

Here we used the fact that $\eta_i \in T^\perp$. In other words,

$$(\psi \times Id_{R^n})^* \omega_{can}((x, c); (\xi_1, \eta_1), (\xi_2, \eta_2)) = \Omega(x, c; (\xi_1, \eta_1), (\xi_2, \eta_2))$$

by Proposition 3.1. The result follows by 3.14. ∎

Remark 3.6 Theorem 3.5 shows that our Hamiltonian system $\dot{x} = V_c(x)$, $\dot{c} = 0$ is completely integrable and provides a construction of action-angle variables.

Theorem 3.7 *The pair $(int(P) \times T^\perp, \omega)$ is a symplectic submanifold of $(int(Q) \times R^n, \Omega)$. In particular,*

$$\omega = \Omega|_{int(P) \times T^\perp}.$$

Proof: By 3.14 it is sufficient to prove that

$$\begin{aligned}\Delta_1 &= \Omega(x,c;(\xi_1.\eta_1),(\xi_2,\eta_2)) = \Delta_2 = \\ &\quad \Omega(\varphi(x),c;(D\varphi(x)\xi_1,\eta_1),D\varphi(x)\xi_2,\eta_2)) \qquad (3.16)\end{aligned}$$

for any $x \in int(Q), c \in T^{\perp}, \xi_i \in R^n, \eta_i \in T^{\perp}$. By Proposition 3.1 $\Omega = (\psi \times Id_{R^n})^*$. Hence,

$$\Delta_2 = \omega_{can}((D(\psi\circ\varphi)(x)\xi_1,\eta_1),D(\psi\circ\varphi)(x)\xi_2,\eta_2)) =$$
$$\omega_{can}((D\psi(x)\xi_1,\eta_1),(D\psi(x)\xi_2,\eta_2)) =$$
$$(\psi \times Id_{R^n})^*\omega_{can}(x,c;(\xi_1,\eta_1),(\xi_2,\eta_2)) = \Delta_1.$$

Here we used 3.14 and the condition $\eta_i \in T^{\perp}, i = 1,2$. ■

Remark 3.8 Thus in the situation considered the reduced manifold can be realized as a symplectic submanifold of the initial manifold.

4 Concluding Remarks

In the present paper we have considered dynamical systems that solve general (smooth) convex programming problems. We have analyzed in detail their Hamiltonian structure. The crucial part of our analysis is based on the notion of the Legendre transform. In particular, the underlying symplectic form is obtained as a pullback of the canonical symplectic structure via the Legendre transform. Dynamical systems corresponding to the problems with equality constraints are obtained by applying the Marsden-Weinstein reduction procedure. In particular, their solutions are projections of solutions of dynamical systems corresponding to the problems without inequality constraints. The reduced manifold turned out to be the symplectic submanifold of the initial manifold. The Legendre transform also provides a linearization of our dynamical systems and the explicit construction of action-angle variables. At present time it is not quite clear how to generalize the interior-point methodology for the case of nonlinear equality constraints. Our Hamiltonian framework may be helpful in this situation. It seems that it is natural to use the universal Hamiltonian $< c, c > /2$ and the introduced symplectic structure to construct constrained Hamiltonian systems in this more general situation.

The essential part of the classical book [6] is devoted to the analysis of one particular type of trajectories of our dynamical systems-central trajectories. While this is a very important class of trajectories, the consideration of large-step algorithms requires the analysis of all trajectories.

Observe that we considered here only logarithmic barrier functions. More general classes of barrier functions can be analyzed using the same technique (see [5] for the case of entropy barrier functions).

We now briefly outline the idea behind one particular class of large-step algorithms. Given $\xi \in R^n$, denote $[g(x;\xi,\xi)]^{1/2}$ by $\| \xi \|_x$. Suppose that there exists $\epsilon > 0$ such that for any $x \in int(Q)$

$$x + \xi \in int(Q), \| \xi \|_x < \epsilon. \tag{4.1}$$

If 4.1 holds and $x \in int(P)$, then

$$\nu(x) = x + \mu \frac{V_c(x)}{\| V_c(x) \|_x} \in int(P), |\mu| < \epsilon.$$

Given $x \in int(P)$, consider iterations $x, \nu(x), \nu^2(x), \ldots,$. This procedure is known as Dikin's algorithm [2] for the case where all constraints are linear. It is known to be an efficient tool for solving linear programming problems. To the best of our knowledge, no convergence results have been reported for the case of nonlinear constraints. It is clear, however, that convergence to the optimal solution should take place (at least for small μ) due to the global convergence of all trajectories of the vector field V_c to the optimal solution. Observe that conditions 4.1 are satisfied provided the functions $\ln f_i, i = 1, \ldots m$ are self-concordant functions (this is always the case when f_i are linear or quadratic [10]). The properties of this type provide a special status for logarithmic barrier functions at least for now.

Our analysis can be generalized to certain infinite-dimensional situations with interesting applications to control problems like linear-quadratic regulators with quadratic constraints [11]. We plan to consider these applications separately.

References

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

[2] I. Dikin. Iterative solution of problems of linear and quadratic programming, *Soviet. Math. Dokl.* **8** (1967), 674-675.

[3] L. Faybusovich. Hamiltonian structure of dynamical systems which solve linear programming problems, *Physica* **D53** (1991), 217-232.

[4] L. Faybusovich. Dynamical systems which solve optimization problems with linear constraints, *IMA Journ. of Mathematical Control and Information* **8** (1991), 135-149.

[5] L. Faybusovich. Gibbs variational principal,Interior-point methods and gradient flows, (to appear)

[6] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming:Sequential Unconstrained Minimization Techniques.* New York: John Wiley, 1968.

[7] C. Gonzaga. Path-following algorithms for linear programming, *SIAM Review* **34** (1992), 167-224.

[8] F. Jarre. Interior-point methods for convex programming, *Applied Mathematics and Optimization* **26** (1992), 287-311.

[9] J. Marsden. *Lectures on Mechanics.* Cambridge: Cambridge University Press, 1992.

[10] Yu.E. Nesterov and A.S. Nemirovsky. Self-concordant functions and polynomial-time methods in convex programming. Technical report, Centr. Econ. Inst.,Moscow, USSR,1989.

[11] V.A. Yakubovich. Nonconvex optimization problem: The infinite-horizon linear-quadratic control problem with quadratic constraints, *System and Control Letters* **19** (1992), 13-22.

Department of Mathematics, University of Notre Dame, Notre Dame, Indiana, USA 64556-5683

Communicated by Clyde F. Martin