Journal of Mathematical Analysis and Applications **250,** 561–578 (2000)
doi:10.1006/jmaa.2000.6992, available online at http://www.idealibrary.com on IDEAL®

Hamiltonian Systems on Time Scales

Calvin D. Ahlbrandt

Department of Mathematics, University of Missouri, Columbia, Missouri 65211-0001
E-mail: calvin@math.missouri.edu

Martin Bohner

Department of Mathematics and Statistics, University of Missouri-Rolla, Rolla, Missouri 65409-0020
E-mail: bohner@umr.edu

and

Jerry Ridenhour[1]

Department of Mathematics and Statistics, Utah State University, Logan, Utah 84322-3900
E-mail: jride@math.usu.edu

Submitted by M. J. Balas

Received September 10, 1999

Linear and nonlinear Hamiltonian systems are studied on time scales $\mathbb{T}$. We unify symplectic flow properties of discrete and continuous Hamiltonian systems. A chain rule which unifies discrete and continuous settings is presented for our so-called *alpha derivatives* on generalized time scales. This chain rule allows transformation of linear Hamiltonian systems on time scales under simultaneous change of independent and dependent variables, thus extending the change of dependent variables recently obtained by Došlý and Hilscher. We also give the Legendre transformation for nonlinear Euler–Lagrange equations on time scales to Hamiltonian systems on time scales.

Key Words: Hamiltonian systems; time scales; Euler–Lagrange equations; delta derivatives; alpha derivatives; chain rule; symplectic flows.

[1] Visitor during 1998–1999 in the Department of Mathematics, University of Missouri, Columbia, MO 65211.

0022-247X/00 $35.00

1. LINEAR HAMILTONIAN SYSTEMS

During the past ten years, a calculus on time scales (or measure chains) has been developed by several authors with one goal being the unified treatment of differential equations (the continuous case) and difference equations (the discrete case). Nonempty closed subsets of the reals are considered to be time scales in, for example, [12–14]—or see [6, 11] for a discussion in the more general framework of measure chains. For the purposes of variable change and because solutions of differential equations are often defined on open intervals, we allow time scales that do not necessarily contain their finite supremum or finite infimum, if such numbers exist. That is, we define a *time scale* $\mathbb{T}$ as a nonempty subset of the real numbers with the property that every Cauchy sequence in $\mathbb{T}$ converges to a point of $\mathbb{T}$ with the possible exception of Cauchy sequences converging to a finite infimum or finite supremum of $\mathbb{T}$. Let $\mathbb{T}^i$ (previously denoted by $\mathbb{T}^\kappa$) denote the set of points of $\mathbb{T}$ except for a possible maximal isolated point (i.e., a "left scattered maximal point"). This notation is loosely thought of as the "interior" of $\mathbb{T}$. These concepts will be generalized further in the next section to allow a chain rule for differentiation on time scales.

Define the *right jump function* $\sigma : \mathbb{T} \to \mathbb{T}$ by $\sigma(t) = \inf\{s \in \mathbb{T} : \ s > t\}$ (supplemented by $\inf \emptyset = \sup \mathbb{T}$). The *graininess* (or *stepsize*) μ is then defined by $\mu(t) = \sigma(t) - t$ for each $t \in \mathbb{T}$. A point $t \in \mathbb{T}$ is called *right scattered* if $\mu(t) > 0$ while the terminology *right dense* is used in case of $\mu(t) = 0$. The $^\Delta$ derivative defined by Aulbach and Hilger [6, 11, 12] is the usual derivative if $\mathbb{T} = \mathbb{R}$ and the forward difference if $\mathbb{T} = \mathbb{Z}$. In order to define the $^\Delta$ derivative of a function, we say that a subset $\mathcal{U}$ of $\mathbb{T}$ is *open in* $\mathbb{T}$ if it is open in the relative topology [16, p. 51], i.e., if $\mathcal{U} = \mathcal{V} \cap \mathbb{T}$ for some open set $\mathcal{V}$ in $\mathbb{R}$. A *neighborhood* $\mathcal{U}$ of a point $t \in \mathbb{T}$ is a subset of $\mathbb{T}$ which is open in $\mathbb{T}$ and contains t.

Functions defined on a subset of a time scale are thought of as real or complex valued, vector valued, or as matrix valued.

A function f is said to be $^\Delta$ *differentiable* at a point $t \in \mathbb{T}^i$ if f is defined at $\sigma(t)$, f is defined in a neighborhood $\mathcal{U}$ of t, and there exists a quantity $f^\Delta(t)$, called the $^\Delta$ *derivative* of f at t, such that for each positive real number ε there exists a neighborhood $\mathcal{N}$ of t such that $\mathcal{N} \subseteq \mathcal{U}$ and

$$|f(\sigma(t)) - f(s) - (\sigma(t) - s)f^\Delta(t)| \leq \varepsilon|\sigma(t) - s| \tag{1}$$

for every $s \in \mathcal{N}$. We show uniqueness of $f^\Delta(t)$ in the next section as a special case of the so-called *alpha derivatives* introduced there. Since uniqueness does not apply at a point t which is maximal and isolated we must restrict our domains of $^\Delta$ differentiable functions in the discrete case.

We list some basic properties which will be used throughout this paper. Suppose that x is a function of t for t in a time scale $\mathbb{T}$.

• If x is defined on $\mathbb{R}$ and differentiable at a right dense point $t \in \mathbb{T}^i$, then x is $^{\Delta}$ differentiable at t with $x^{\Delta}(t) = x'(t)$.

• If x is continuous at a right scattered point $t \in \mathbb{T}$ and defined at $\sigma(t)$, then x is $^{\Delta}$ differentiable at t with $x^{\Delta}(t) = [x(\sigma(t)) - x(t)]/\mu(t)$.

• If x is $^{\Delta}$ differentiable at $t \in \mathbb{T}^i$, then

$$x^{\sigma} = x + \mu x^{\Delta} \tag{2}$$

holds at t, where we put $x^{\sigma} = x \circ \sigma$, i.e., $x^{\sigma}(t) = x(\sigma(t))$.

• If x and y are both $^{\Delta}$ differentiable at $t \in \mathbb{T}^i$ and the matrix product xy is defined, then xy is $^{\Delta}$ differentiable at t and at the point t we have

$$(xy)^{\Delta} = x^{\Delta} y + x^{\sigma} y^{\Delta}, \tag{3}$$

$$(xy)^{\Delta} = x^{\Delta} y^{\sigma} + x y^{\Delta}. \tag{4}$$

A function x on $\mathbb{T}$ is called right dense continuous [1, 11, p. 34], or just *rd-continuous*, provided x is continuous at all right dense points in $\mathbb{T}^i$, and at all left dense points $t \in \mathbb{T}$ the left hand limit $\lim_{s \to t^-} x(s)$ exists and is finite. It is known [11, Theorem 4.4] that rd-continuous functions have $^{\Delta}$ antiderivatives.

Let $\mathcal{J}$ be the skew-symmetric $2n \times 2n$ matrix $\mathcal{J} = \left[\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right]$. We say that a $2n \times 2n$ matrix function $\mathcal{H}$ is *Hamiltonian* at a point $t \in \mathbb{T}$ if

$$\mathcal{H}^*(t)\mathcal{J} + \mathcal{J}\mathcal{H}(t) + \mu(t)\mathcal{H}^*(t)\mathcal{J}\mathcal{H}(t) = 0. \tag{5}$$

This definition was motivated by recent work of Došlý and Hilscher [8] and extends the definition for continuous $\mathbb{T}$ in [7, p. 570, 18, p. 106].

A $2n \times 2n$ matrix valued function S is said to be *symplectic* [4, p. 74] at a point $t \in \mathbb{T}$ if $S^*(t)\mathcal{J}S(t) = \mathcal{J}$. From the relation

$$[I + \mu\mathcal{H}]^*\mathcal{J}[I + \mu\mathcal{H}] = \mathcal{J} + \mu[\mathcal{H}^*\mathcal{J} + \mathcal{J}\mathcal{H} + \mu\mathcal{H}^*\mathcal{J}\mathcal{H}] \tag{6}$$

it can be seen that $S := I + \mu\mathcal{H}$ is symplectic at t provided $\mathcal{H}$ is Hamiltonian at t.

This section is concerned with *linear Hamiltonian systems*

$$x^{\Delta} = \mathcal{H}(t)x \qquad \text{with } \mathcal{H}(t) \text{ Hamiltonian for } t \in \mathbb{T}^i. \tag{7}$$

Došlý and Hilscher [8] call such a system a "symplectic dynamic system." Formula (2) implies that a solution x of (7) satisfies $x^{\sigma} = x + \mu x^{\Delta} = x + \mu\mathcal{H}x = [I + \mu\mathcal{H}]x$; thus if $S := I_{2n \times 2n} + \mu\mathcal{H}$, then any solution x of (7) is also a solution of the system

$$x^{\sigma} = S(t)x. \tag{8}$$

Whereas systems of this latter form are important for discrete problems [4], information about $\mathcal{H}(t)$ is lost in system (8) at right dense points t.

PROPOSITION 1.1 (Linear Canonical Equations are Hamiltonian). *Let A, B, and C be $n \times n$ real or complex matrix valued functions defined on $\mathbb{T}^i$ with B and C Hermitian on $\mathbb{T}^i$. Assume that the matrix $D := (I - \mu A)$ is nonsingular on $\mathbb{T}^i$ and let $E := (I - \mu A)^{-1}$ on $\mathbb{T}^i$. If y and z are defined on $\mathbb{T}$, satisfy the linear canonical system* [1, 8]

$$y^\Delta = Ay^\sigma + Bz, \qquad z^\Delta = Cy^\sigma - A^* z, \quad on \quad \mathbb{T}^i, \tag{9}$$

and $\mathcal{H}$ and x are defined by

$$\mathcal{H} := \begin{bmatrix} AE & EB \\ CE & \mu CEB - A^* \end{bmatrix} \quad on \quad \mathbb{T}^i \quad and \quad x := \begin{bmatrix} y \\ z \end{bmatrix} \quad on \quad \mathbb{T}, \tag{10}$$

then $\mathcal{H}$ is Hamiltonian on $\mathbb{T}^i$ and x satisfies $x^\Delta = \mathcal{H}x$ on $\mathbb{T}^i$.

Proof. Assume that y, z is a solution of the linear canonical system (9). Since $y^\sigma = y + \mu y^\Delta = y + \mu(Ay^\sigma + Bz)$, we have $Dy^\sigma = y + \mu Bz$ and $y^\sigma = Ey + \mu EBz$. Thus $y^\Delta = AEy + (\mu AE + I)Bz = AEy + EBz$ since $\mu AE + I = \mu AE + DE = (\mu A + D)E = I \cdot E = E$. Similarly, $z^\Delta = CEy + (\mu CEB - A^*)z$. Thus x is a solution of the linear system $x^\Delta = \mathcal{H}x$ on $\mathbb{T}^i$. To verify condition (5), use the identity $\mu AE = E - I$ and its complex conjugate $\mu E^* A^* = E^* - I$ (except in the A^2 and $(A^*)^2$ terms) to reduce the problem of showing that $\mathcal{H}$ is Hamiltonian on $\mathbb{T}^i$ to verifying the identity $AE - A = \mu A^2 E$, which follows from $A(E - I) = A(\mu AE)$. A useful error check for the calculations is to note that the left side of Eq. (5) is skew-Hermitian. ∎

The phrase "linear Hamiltonian systems" for systems (7) whose coefficient matrix $\mathcal{H}$ satisfies the condition (5) extends previous terminology. Indeed, these systems include as special cases the continuous linear [20, p. 303] and discrete linear cases [3, 10], including the variable step case [4, p. 187]. But, more importantly, for $A(t)$, $B(t)$, and $C(t)$ $n \times n$ with real entries and $B(t)$ and $C(t)$ real symmetric, Eq. (9) is the special case of a Hamiltonian system of equations

$$y^\Delta = H_z(t, y^\sigma(t), z(t)), \qquad z^\Delta = -H_y(t, y^\sigma(t), z(t)) \tag{11}$$

(see Section 4), for $H(t, y, z)$ the quadratic Hamiltonian

$$H(t, y, z) = \frac{1}{2} [\, y^T \quad z^T \,] \begin{bmatrix} C & A^T \\ A & B \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}. \tag{12}$$

The partial derivative notation H_y denotes the column vector (H_{y_i}). In addition, linear Hamiltonian systems (7) arise in Section 4 as variational equations for general nonlinear Hamiltonian systems on time scales.

Remark 1.2 (Convergence to Continuous). Suppose that $\{\mathbb{T}_m\}$, for $m = 1,2,\ldots$, is a family of time scales containing the point t with $\mu_m(t) > 0$ and $\mu_m(t) \to 0$ as $m \to \infty$. Suppose that $B(t)$ and $C(t)$ are Hermitian and for each m the matrix $(I - \mu_m(t)A(t))$ is nonsingular. Let $H_m(t)$ be defined by (10) with $\mu(t) = \mu_m(t)$ and $E(t) = E_m(t) := (I - \mu_m(t)A(t))^{-1}$ for each m. Then $E_m(t) \to I$ as $m \to \infty$ and $\mathcal{H}_m(t)$ converges to the $\mathcal{H}(t)$ corresponding to the case where $\mu(t) = 0$, namely,

$$\mathcal{H}(t) = \begin{bmatrix} A(t) & B(t) \\ C(t) & -A^*(t) \end{bmatrix}. \tag{13}$$

We now show that as for previously studied linear Hamiltonian systems the transition matrix $X(t,s)$, which carries a solution at time s to a time t, is symplectic. For differential equations this result is due to Radon [19], and for discrete symplectic systems (8) it is a consequence of the group property of symplectic matrices [4, p. 74].

THEOREM 1.3 (Symplectic Flows). *If $\mathcal{H}$ is Hamiltonian on $\mathbb{T}^i$, s is fixed in $\mathbb{T}^i$, and $X(t) := X(t,s)$ defined for $t \in \mathbb{T}$ satisfies $X^\Delta = \mathcal{H}(t)X$ on $\mathbb{T}^i$ with initial condition $X(s,s) = I_{2n}$, then $X(t,s)$ is symplectic on $\mathbb{T}$.*

Proof. If we show that for s fixed, $[(X(t,s))^*\mathcal{J}X(t,s)]^\Delta \equiv 0$ on $\mathbb{T}^i$, then (by [15, Corollary 1.3.1]) the function of t defined by $(X(t,s))^*\mathcal{J}X(t,s)$ has constant value on $\mathbb{T}$ of $(X(s,s))^*\mathcal{J}X(s,s) = \mathcal{J}$ and $X(t,s)$ is symplectic on $\mathbb{T}$. For the proof we abbreviate $X(t,s)$ by X, write X^σ for $X(\sigma(t),s)$, and denote the delta derivative of $X(t,s)$ with respect to t by X^Δ. We use formulas (4) and (2) to obtain

$$\begin{aligned} [X^*\mathcal{J}X]^\Delta &= [X^*]^\Delta\mathcal{J}X^\sigma + X^*\mathcal{J}X^\Delta \\ &= [X^*]^\Delta\mathcal{J}[X + \mu X^\Delta] + X^*\mathcal{J}X^\Delta \\ &= X^*\mathcal{H}^*\mathcal{J}[X + \mu\mathcal{H}X] + X^*\mathcal{J}\mathcal{H}X \\ &= X^*[\mathcal{H}^*\mathcal{J} + \mathcal{J}\mathcal{H} + \mu\mathcal{H}^*\mathcal{J}\mathcal{H}]X = 0. \end{aligned}$$

∎

In the final section of this paper we use this result to establish a generalization to the nonlinear case. We will also be considering a Lagrangian $L(t,y,r)$ for $t \in \mathbb{T}$, $y \in R^n$ and $r \in R^n$. The Euler–Lagrange equation on a time scale $\mathbb{T}$ is

$$L_y(t, y^\sigma(t), y^\Delta(t)) = \left[L_r(t, y^\sigma(t), y^\Delta(t))\right]^\Delta. \tag{14}$$

EXAMPLE 1.4 (Jacobi Equations). *Let L be the quadratic Lagrangian*

$$L(t,y,r) = \frac{1}{2}\begin{bmatrix} y^T & r^T \end{bmatrix}\begin{bmatrix} P(t) & Q(t) \\ Q^T(t) & R(t) \end{bmatrix}\begin{bmatrix} y \\ r \end{bmatrix} \tag{15}$$

for $t \in \mathbb{T}$ with real $n \times n$ matrices $P(t)$, $Q(t)$, and $R(t)$ with $P(t)$ and $R(t)$ symmetric. We also assume that $R(t)$ is nonsingular. Then the Euler–Lagrange equation (14) *is the Jacobi equation*

$$Py^{\sigma} + Qy^{\Delta} = (Q^T y^{\sigma} + Ry^{\Delta})^{\Delta}. \tag{16}$$

Introduce a generalized momentum coordinate $z(t)$ by (and if $\mathbb{T}^i \neq \mathbb{T}$, also by Section 4, *Eq.* (38))

$$z = Q^T y^{\sigma} + Ry^{\Delta}. \tag{17}$$

Then the Jacobi equation (16) *is a special case of the linear canonical system* (9) *with coefficients*

$$A = -R^{-1}Q^T, \qquad B = R^{-1}, \qquad C = P - QR^{-1}Q^T. \tag{18}$$

2. ALPHA DERIVATIVES AND THE CHAIN RULE

As presented thus far a time scale is a set $\mathbb{T}$, as defined in Section 1, together with the induced right jump function σ. For such a time scale $\mathbb{T}$ suppose that g is a real valued function defined on $\mathbb{T}$. We are interested in conditions such that the set $\mathbb{X} = g(\mathbb{T})$ is also a time scale. We have allowed finite extrema of the time scale to be omitted because examples such as $\mathbb{T} = \mathbb{R}$ under $x = g(t) = \text{Arctan}(t)$ or $\mathbb{T} = \mathbb{Z}^+$ under $x = g(t) = 1/t$ take a closed set to a nonclosed set. Because g could be decreasing, we need to generalize the concept of a time scale and $^{\Delta}$ differentiablity to take orientation into account. Our motivation comes from the need for a chain rule in order to transform $^{\Delta}$ equations. We apply our chain rule to that problem in the next section. Furthermore, the chain rule of ordinary calculus has no monotonicity assumptions, so we make definitions which do not require strictly monotone changes of variable.

For purposes of this section, we generalize our previous definition of a *time scale* as follows: A *time scale* is a pair $\{\mathbb{T}, \alpha\}$ such that

(1) $\mathbb{T}$ *is a nonempty subset of the real numbers such that every Cauchy sequence in $\mathbb{T}$ converges to a point of $\mathbb{T}$ with the possible exception of Cauchy sequences which converge to a finite infimum or finite supremum of $\mathbb{T}$*;

(2) α *maps $\mathbb{T}$ into $\mathbb{T}$*.

We think of α as a (*generalized*) *jump function*. If $\alpha = \sigma$, then this is consistent with our previous definition.

Let us define the *interior of $\mathbb{T}$ relative to α* to be the set

$$\mathbb{T}^i = \{t \mid t \in \mathbb{T} \text{ such that either } \alpha(t) \neq t \text{ or } \alpha(t) = t \text{ and } t \text{ is not isolated}\}.$$

DEFINITION 2.1 (The Alpha Derivative). Let $\{\mathbb{T}, \alpha\}$ be a time scale. A real valued function g is said to be α-differentiable at $t \in \mathbb{T}^i$ if

(i) g is defined in a neighborhood $\mathcal{U}$ of t;

(ii) g is defined at $\alpha(t)$;

(iii) there exists a unique real number $g_\alpha(t)$, called the α-derivative of g at t, such that for each positive real number ε, there exists a neighborhood $\mathcal{N}$ of t with $\mathcal{N} \subseteq \mathcal{U}$ and

$$|g(\alpha(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)| \leq \varepsilon|\alpha(t) - s| \tag{19}$$

for every $s \in \mathcal{N}$.

We also use the notation $dg(t)/d\alpha(t)$ for $g_\alpha(t)$. This agrees with standard notation for derivatives if $\mathbb{T}$ is $\mathbb{R}$ and $\alpha(t) = t$. If $\alpha = \sigma$, then the σ-derivative is the $^\Delta$ derivative. If $\alpha = \rho$, Hilger's left jump function, then the ρ-derivative is the left-hand difference quotient at left scattered points. In the example of $\mathbb{T} = \{-1\} \cup [0, 1] \cup \{2\} \cup \{3\}$, the point $t = 3$ is not in the interior of $\mathbb{T}$ relative to σ whereas -1 is not in the interior of $\mathbb{T}$ relative to ρ.

The set $\mathbb{T} = \mathbb{R}$ with two jump functions $\alpha(t) = t$ and $\beta(t) = t + 1$ gives rise to *differential-difference operators* $(f_\alpha)_\beta$.

Note 2.2. If t is an isolated point of $\mathbb{T}$ with $\alpha(t) = t$, then the singleton set $\mathcal{N} = \{t\}$ is a nbhd. of t and for any real L and any $\varepsilon > 0$, the condition

$$|g(\alpha(t)) - g(s) - (\alpha(t) - s)L| = \varepsilon|\alpha(t) - s| = 0$$

holds for every $s \in \mathcal{N}$, i.e., for $s = t$. Since we demand uniqueness of $g_\alpha(t)$, we conclude that no function can be α-differentiable at a point $t \in \mathbb{T} \setminus \mathbb{T}^i$.

Note 2.3. If $t \in \mathbb{T}^i$, g is defined in a nbhd. $\mathcal{U}$ of t, $g(\alpha(t))$ is defined, and there exist numbers L_1 and L_2 such that for each $\varepsilon > 0$, there exist neighborhoods $\mathcal{N}_1$ and $\mathcal{N}_2$ of t such that

$$|g(\alpha(t)) - g(s) - (\alpha(t) - s)L_i| \leq \varepsilon|\alpha(t) - s|$$

for every $s \in \mathcal{N}_i$, then $L_1 = L_2$ and $g_\alpha(t)$ exists.

Proof. Assume $L_1 \neq L_2$ and set $\varepsilon = |L_1 - L_2|/4$. Then for $s \in \mathcal{N}_1 \cap \mathcal{N}_2$ we have the inequalities

$$-\varepsilon|\alpha(t) - s| \leq g(\alpha(t)) - g(s) - (\alpha(t) - s)L_1 \leq \varepsilon|\alpha(t) - s|,$$
$$-\varepsilon|\alpha(t) - s| \leq -g(\alpha(t)) + g(s) + (\alpha(t) - s)L_2 \leq \varepsilon|\alpha(t) - s|.$$

Add these inequalitites and use $2\varepsilon = |L_1 - L_2|/2$ for the conclusion that

$$|\alpha(t) - s| \cdot |L_1 - L_2| \leq (1/2)|L_1 - L_2| \cdot |\alpha(t) - s|.$$

If $\alpha(t) \neq t$, choose $s = t$ for a contradiction. If $\alpha(t) = t$ and t is not isolated, then there exists a point $s \in \mathcal{N}_1 \cap \mathcal{N}_2$ other than t which also gives a contradiction. Thus if $t \in \mathbb{T}^i$ and one such L exists, then $g_\alpha(t)$ exists. ∎

THEOREM 2.4. *If f is α-differentiable at t, then it is continuous at t.*

Proof. Let $\varepsilon > 0$. Define $\varepsilon^* = \varepsilon[1 + |f_\alpha(t)| + 2|\alpha(t) - t|]^{-1}$, where $\varepsilon^* \in (0, 1)$ wlog. Hence there exists a neighborhood $\mathcal{N}$ of t (wlog diam$\mathcal{N} < \varepsilon^*$) such that

$$|f(\alpha(t)) - f(s) - (\alpha(t) - s)f_\alpha(t)| \leq \varepsilon^*|\alpha(t) - s| \quad \text{for all} \quad s \in \mathcal{N}.$$

Let $s \in \mathcal{N}$. Then

$$\begin{aligned}|f(t) - f(s)| &= \big|\{f(\alpha(t)) - f(s) - (\alpha(t) - s)f_\alpha(t)\} \\ &\qquad -\{f(\alpha(t)) - f(t) - (\alpha(t) - t)f_\alpha(t)\} + (t - s)f_\alpha(t)\big| \\ &\leq \varepsilon^*|\alpha(t) - s| + \varepsilon^*|\alpha(t) - t| + |t - s||f_\alpha(t)| \\ &\leq \varepsilon^*[|\alpha(t) - t| + |t - s| + |\alpha(t) - t| + |f_\alpha(t)|] \\ &< \varepsilon^*[1 + |f_\alpha(t)| + 2|\alpha(t) - t|] = \varepsilon. \quad \blacksquare\end{aligned}$$

Since the choice of $\mathbb{T} = \mathbb{R}$ and $\alpha(t) = t$ gives the usual calculus, we know that there exist examples of functions which are continuous at a point but not α-differentiable there.

Note 2.5. If $\{\mathbb{T}, \alpha\}$ is a time scale and $t \in \mathbb{T}^i$ has $\alpha(t) \neq t$, f is continuous at t, and $f(\alpha(t))$ is defined, then f is α-differentiable at t with

$$f_\alpha(t) = \frac{df(t)}{d\alpha(t)} = \frac{f(\alpha(t)) - f(t)}{\alpha(t) - t}. \tag{20}$$

Proof. Note that if $f_\alpha(t)$ exists, then this form of $f_\alpha(t)$ follows from Eq. (19) (with $g = f$) since every neighborhood of t must contain $s = t$ and $t \in \mathbb{T}^i$ implies that the derivative is unique. To see that f is α-differentiable at t, let $\varepsilon > 0$. Since $\alpha(t) - t \neq 0$, the function g defined for $s \in \mathbb{T}$ where $s \neq \alpha(t)$ by $g(s) = [f(\alpha(t)) - f(s)]/[\alpha(t) - s]$ is continuous at t. Thus there exists a neighborhood $\mathcal{N}$ of t such that for $f_\alpha(t)$ given by (20) and $s \in \mathcal{N}$

$$|f(\alpha(t)) - f(s) - (\alpha(t) - s)f_\alpha(t)| = |g(s) - g(t)| \cdot |\alpha(t) - s| \leq \varepsilon|\alpha(t) - s|.$$

$\blacksquare$

This suggests the useful formula

$$f(\alpha(t)) = f(t) + h(t)f_\alpha(t), \quad \text{with } \textit{stepsize} \;\; h(t) := \alpha(t) - t, \tag{21}$$

which holds if $f_\alpha(t)$ exists. As in the notation f^σ, we write $f^\alpha(t) := f(\alpha(t))$.

THEOREM 2.6 (Product Rules). *If f and g are real valued and α-differentiable at $t \in \mathbb{T}^i$, then fg is α-differentiable at t and the formulas $(fg)_\alpha = f_\alpha g^\alpha + f g_\alpha$ and $(fg)_\alpha = f_\alpha g + f^\alpha g_\alpha$ hold at t.*

Proof. Let $\varepsilon > 0$. Define $\varepsilon^* = \varepsilon[1 + |f(t)| + |g(\alpha(t)| + |g_\alpha(t)|]^{-1}$, where $\varepsilon^* \in (0, 1)$ wlog. Hence there exist neighborhoods $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$ of t with

$$|f(\alpha(t)) - f(s) - (\alpha(t) - s)f_\alpha(t)| \leq \varepsilon^*|\alpha(t) - s| \quad \text{for all} \quad s \in \mathcal{N}_1,$$

$$|g(\alpha(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)| \leq \varepsilon^*|\alpha(t) - s| \quad \text{for all} \quad s \in \mathcal{N}_2,$$

and (from Theorem 2.4)

$$|f(t) - f(s)| \leq \varepsilon^* \quad \text{for all} \quad s \in \mathcal{N}_3.$$

Put $\mathcal{N} = \mathcal{N}_1 \cap \mathcal{N}_2 \cap \mathcal{N}_3$ and let $s \in \mathcal{N}$. Then

$$\begin{aligned}
&\big|(fg)(\alpha(t)) - (fg)(s) - (\alpha(t) - s)[f_\alpha(t)g(\alpha(t)) + f(t)g_\alpha(t)]\big| \\
&\quad = \Big|[f(\alpha(t)) - f(s) - (\alpha(t) - s)f_\alpha(t)]g(\alpha(t)) \\
&\qquad + [g(\alpha(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)]f(t) \\
&\qquad + [g(\alpha(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)][f(s) - f(t)] \\
&\qquad\qquad + (\alpha(t) - s)g_\alpha(t)[f(s) - f(t)]\Big| \\
&\quad \leq \varepsilon^*|\alpha(t) - s| \cdot |g(\alpha(t))| + \varepsilon^*|\alpha(t) - s||f(t)| \\
&\qquad + \varepsilon^*\varepsilon^*|\alpha(t) - s| + \varepsilon^*|\alpha(t) - s||g_\alpha(t)| \\
&\quad = \varepsilon^*|\alpha(t) - s| \cdot [\,|g(\alpha(t))| + |f(t)| + \varepsilon^* + |g_\alpha(t)|\,] \\
&\quad \leq \varepsilon^*|\alpha(t) - s| \cdot [\,1 + |f(t)| + |g(\alpha(t))| + |g_\alpha(t)|\,] \\
&\quad = \varepsilon|\alpha(t) - s|.
\end{aligned}$$

By Note 2.3 existence of one derivative at a point $t \in \mathbb{T}^i$ implies uniqueness. Thus $(fg)_\alpha = f_\alpha g^\alpha + fg_\alpha$. The second formula follows from the first since products commute. ∎

Matrix case product formulas are valid if we also require that the matrix product of the functions is defined. The definition of the α-derivative of a matrix follows from the choice of 2-norm in the definition of α-derivatives. Then use transposes twice: $(FG)_\alpha = [(G^T F^T)_\alpha]^T$ in order to derive the second formula from the first. Since $||A^T||_2 = ||A||_2$, [22, Theorem 2.10, p. 180], the choice of 2-norm assures that the operations of transpose and α-differentiation commute.

THEOREM 2.7 (Chain Rule). *Let $\{\mathbb{T}, \alpha\}$ and $\{\mathbb{X}, \beta\}$ be time scales related by a function $g : \mathbb{T} \to \mathbb{X}$. Let $w : \mathbb{X} \to \mathbb{R}$ and let $z = w \circ g$. Suppose that t is a point of $\mathbb{T}^i$ such that g has the property $g(\alpha(t)) = \beta(g(t))$. If $g_\alpha(t)$ and $w_\beta(g(t))$ exist, then $z_\alpha(t)$ exists and satisfies the chain rule*

$$z_\alpha = (w \circ g)_\alpha = (w_\beta \circ g)g_\alpha \quad \text{at } t.$$

Proof. Let $\varepsilon > 0$. Define $\varepsilon^* = \varepsilon[1 + |g_\alpha(t)| + |w_\beta(g(t))|]^{-1}$, where $\varepsilon^* \in (0, 1)$ wlog. According to the assumptions, there exist neighborhoods $\mathcal{N}$ of t and $\mathcal{V}$ of $g(t)$ on which, respectively,

$$|g(\alpha(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)| \le \varepsilon^*|\alpha(t) - s|, \quad s \in \mathcal{N},$$
$$|w(\beta(g(t))) - w(r) - (\beta(g(t)) - r)w_\beta(g(t))| \le \varepsilon^*|\beta(g(t)) - r|, \quad r \in \mathcal{V}.$$

Since g is α-differentiable at $t \in \mathbb{T}^i$, it is continuous at t by Theorem 2.4 and there exists a neighborhood $\mathcal{U}$ of t such that $s \in \mathcal{U}$ implies $g(s) \in \mathcal{V}$. Put $\mathcal{N}_1 = \mathcal{N} \cap \mathcal{U}$ and let $s \in \mathcal{N}_1$. Then $s \in \mathcal{N}$, $g(s) \in \mathcal{V}$, and

$$\begin{aligned}
&\big|w(g(\alpha(t))) - w(g(s)) - (\alpha(t) - s)[w_\beta(g(t))g_\alpha(t)]\big| \\
&\quad = \big|w(g(\alpha(t))) - w(g(s)) - (\beta(g(t)) - g(s))w_\beta(g(t)) \\
&\qquad\quad + [\beta(g(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)]w_\beta(g(t))\big| \\
&\quad \le \varepsilon^*|\beta(g(t)) - g(s)| + \varepsilon^*|\alpha(t) - s||w_\beta(g(t))| \\
&\quad \le \varepsilon^*\big\{ |\beta(g(t)) - g(s) - (\alpha(t) - s)g_\alpha(t)| + |\alpha(t) - s||g_\alpha(t)| \\
&\qquad\quad + |\alpha(t) - s||w_\beta(g(t))| \big\} \\
&\quad \le \varepsilon^*\big\{ \varepsilon^*|\alpha(t) - s| + |\alpha(t) - s||g_\alpha(t)| + |\alpha(t) - s||w_\beta(g(t))| \big\} \\
&\quad = \varepsilon^*|\alpha(t) - s|\big\{ \varepsilon^* + |g_\alpha(t)| + |w_\beta(g(t))| \big\} \\
&\quad \le \varepsilon^*\big\{ 1 + |g_\alpha(t)| + |w_\beta(g(t))| \big\}|\alpha(t) - s| \\
&\quad = \varepsilon|\alpha(t) - s|.
\end{aligned}$$

This establishes our chain rule. ∎

The hypothesis $g(\alpha(t)) = \beta(g(t))$ holds in the usual chain rule of calculus because in that setting $\alpha(t) = t$ and $\beta(x) = x$.

EXAMPLE 2.8 (The Discrete Chain Rule). *Suppose that $\mathbb{T} = \mathbb{Z}^+$ and $\alpha = \sigma$, i.e., $\alpha(t) = t + 1$ for $t \in \mathbb{Z}^+$. Let $g(t) := 1/t$ on $\mathbb{Z}^+$, let $\mathbb{X} = \{1/t \mid t \in \mathbb{Z}^+\}$, and consider $w : \mathbb{X} \to \mathbb{R}$ defined by $w(x) = x^2$. Then $z(t) = w(g(t)) = 1/t^2$ and the induced jump function β on $\mathbb{X}$ such that $\beta(x) = g(\alpha(t))$ for $x = g(t)$ is $\beta(x) = x/(x + 1)$. Thus $w_\beta(x) = [w(\beta(x)) - w(x)]/[\beta(x) - x]$, $x/(x + 1) = 1/(t + 1)$, and the chain rule reads*

$$z_\alpha(t) = \frac{1}{(t+1)^2} - \frac{1}{t^2} = \frac{1/(t+1)^2 - 1/t^2}{1/(t+1) - 1/t} \cdot \left[\frac{1}{t+1} - \frac{1}{t}\right] = w_\beta\left(\frac{1}{t}\right) \cdot g_\alpha(t).$$

It has been pointed out to the authors that the chain rule presented by B. Kaymakçalan *et al.* in their book [15, Theorem 1.2.3 (iv), pp. 17–18] seems to be in error. They have a chain rule for a function of two variables $V(t, x)$. In the usual calculus we had

$$\frac{d}{dt}[V(t, x(t))] = V_t(t, x(t)) + V_x(t, x(t))x'(t).$$

Their generalization to time scales is given in [15, p. 18] as

$$[V(t, x(t))]_t^{\Delta} = V_t^{\Delta}(t, x(t)) + V_x(\sigma(t), x(t))x_t^{\Delta}.$$

If one takes their function V to be a function of x alone, then their result would read $[V(x(t))]_t^{\Delta} = V_x(x(t))x_t^{\Delta}$. They previously used the notation [15, Eq. (1.2.7), p. 14] of f_t^{Δ} for the $^{\Delta}$ derivative of f at the point t. Now the symbol V_x on right hand side apparently denotes the partial of $V(t, x)$ with respect to x. If this is what they intend, then the example of $\mathbb{T} = \mathbb{Z}^+$, $x(t) = t$, and $V(x) = x^2$ would give the erroneous conclusion that $[t^2]^{\Delta} = 2x(t) \cdot 1 = 2t$, i.e., the contradiction $[t^2]^{\Delta} = (t+1)^2 - t^2 = 2t + 1 = 2t$. Their chain rule ignores the induced jump function β on the time scale for the x variable and the concept of a β-derivative on that induced time scale.

Remark 2.9 (Differentiation of the Inverse Function). Let $\{\mathbb{T}, \alpha\}$ be a time scale. Suppose that g is a strictly monotone continuous real valued function on $\mathbb{T}$ and $\{\mathbb{X}, \beta\}$ is the induced time scale defined by $\mathbb{X} := g(\mathbb{T})$ with jump function $\beta(x) := g(\alpha(t))$ for $x = g(t)$. Let f be the inverse function of g. Then $g \circ \alpha = \beta \circ g$ on $\mathbb{T}$. Assume that t is a point of $\mathbb{T}^i$ such that $g_{\alpha}(t)$ exists and f_{β} exists at $x = g(t)$. Then $g_{\alpha}(t)$ is nonzero and

$$\frac{1}{g_{\alpha}(t)} = f_{\beta}(x). \tag{22}$$

With these developments of time scales $\{\mathbb{T}, \alpha\}$ and their properties one could readily convert most of Section 1 to this more general setting by replacing $^{\Delta}$ derivatives by α-derivatives. One gap in our development has been in generalizing the result that an *rd*-continuous function has a $^{\Delta}$ antiderivative. Perhaps some type of a class of α-continuous functions would have the analogous property of having α antiderivatives. Because we have not resolved that question we now revert to time scales as defined in Section 1.

THEOREM 2.10 (Substitution). *Let $\mathbb{T}$ be a time scale as in Section* 1 *such that $\mathbb{T}$ is compact and let $g : \mathbb{T} \to \mathbb{R}$ be a strictly increasing function such that $\tilde{\mathbb{T}} = g(\mathbb{T})$ is also a time scale as in Section* 1. *Then $g \circ \sigma = \tilde{\sigma} \circ g$. Let $w : \tilde{\mathbb{T}} \to \mathbb{R}$. If $f : \mathbb{T} \to \mathbb{R}$ is an rd-continuous function and g is $^{\Delta}$ differentiable with rd-continuous $^{\Delta}$ derivative, then*

$$\int_{\mathbb{T}} f(t)g^{\Delta}(t)\Delta t = \int_{\tilde{\mathbb{T}}} (f \circ g^{-1})(s)\tilde{\Delta}s.$$

Proof. Let $a = \min \mathbb{T}$ and $b = \max \mathbb{T}$. Since fg^{Δ} is an rd-continuous function, it possesses a $^{\Delta}$ antiderivative F, i.e., $F^{\Delta} = fg^{\Delta}$, and

$$\begin{aligned}
\int_{\mathbb{T}} f(t)g^{\Delta}(t)\Delta t &= \int_{\mathbb{T}} F^{\Delta}(t)\Delta t \\
&= F(b) - F(a) \\
&= (F \circ g^{-1})(g(b)) - (F \circ g^{-1})(g(a)) \\
&= \int_{\tilde{\mathbb{T}}} (F \circ g^{-1})^{\tilde{\Delta}}(s)\tilde{\Delta} s \\
&= \int_{\tilde{\mathbb{T}}} (F^{\Delta} \circ g^{-1})(s)(g^{-1})^{\tilde{\Delta}}(s)\tilde{\Delta} s \\
&= \int_{\tilde{\mathbb{T}}} ((fg^{\Delta}) \circ g^{-1})(s)(g^{-1})^{\tilde{\Delta}}(s)\tilde{\Delta} s \\
&= \int_{\tilde{\mathbb{T}}} (f \circ g^{-1})(s)[(g^{\Delta} \circ g^{-1})(g^{-1})^{\tilde{\Delta}}](s)\tilde{\Delta} s \\
&= \int_{\tilde{\mathbb{T}}} (f \circ g^{-1})(s)\tilde{\Delta} s,
\end{aligned}$$

where we have used the chain rule and Remark 2.9. ∎

3. TRANSFORMATION THEORY

To simplify the notation we will start with a time scale $\{\mathbb{X}, \sigma\}$ with $\sigma(x)$ being Hilger's right jump function defined on $\mathbb{X}$ as in Section 1. Let the graininess h (i.e., the stepsize) of $\mathbb{X}$ be defined on $\mathbb{X}^i$ by $h(x) = \sigma(x) - x$. Note that $f^{\Delta}(x) = f_{\sigma}(x)$. Suppose that f is a real valued C^1 function defined on the convex hull $[X]$ (i.e., on the smallest interval containing X), such that $f'(x)$ is either always positive on $[X]$ or always negative on $[X]$. Let $\mathbb{T} = f(\mathbb{X})$ and let g be the inverse function of f which carries $\mathbb{T}$ back onto $\mathbb{X}$.

Let $\{\mathbb{T}, \tau\}$ be the time scale $\mathbb{T} = f(\mathbb{X})$ with jump function $\tau(t) = f(\sigma(x))$ and stepsize $k(t) = \tau(t) - t$ on $\mathbb{T}^i = f(\mathbb{X}^i)$. Note that if $f'(x) < 0$ on $[\mathbb{X}]$, then $\mathbb{T}$ with jump function τ is oriented in reverse direction to $\mathbb{X}$ with jump function σ. Also, $t \in \mathbb{T}$ implies there exists a unique $x \in \mathbb{X}$ such that $t = f(x)$. We use the notation $z^o(t)$ for the τ-derivative of a function on the time scale $\mathbb{T}$.

For time scales $\mathbb{X}$ and $\mathbb{T}$ related as above with $[\mathbb{X}]$ and $[\mathbb{T}]$ related by strictly monotone C^1 functions f and g, suppose w is $^{\Delta}$ differentiable at a point $x \in \mathbb{X}^i$ and $z := w \circ g$ is defined in a neighborhood of $t = f(x)$ and at $\tau(t) = f(\sigma(x))$. [Note that $z(t) = w(x)$ and $z(\tau(t)) = w(\sigma(x))$.] Then the chain rule tells us that z is ^{o} differentiable at t with ^{o} derivative at t

given by (here w^{Δ} replaces w' and z^{o} replaces $\dot{z}$ of usual calculus derivative notation)

$$z^{o}(t) = w^{\Delta}(x)g^{o}(t), \qquad \text{where} \quad x = g(t), \quad \text{i.e.,} \tag{23}$$

$$\frac{dz(t)}{d\tau(t)} = \frac{dw(x)}{d\sigma(x)} \cdot \frac{dg(t)}{d\tau(t)} = \frac{dw(x)}{d\sigma(x)} \cdot \frac{dx(t)}{d\tau(t)}. \tag{24}$$

Note that $g^{o}(t)$ is for the function g restricted to the time scale $\mathbb{T}$, even though g is defined on the convex hull $[\mathbb{T}]$.

The following result generalizes Theorem 6.2 of [2], Section 3.18 of [4, p. 144], and Theorem 1 of Došlý and Hilscher [8].

THEOREM 3.1 (Transformations of Systems). *Let time scales $\{\mathbb{X}, \sigma\}$ and $\{\mathbb{T}, \tau\}$ be related as above by a strictly monotone class C^1 function f. Suppose that M is a square matrix valued function on $\mathbb{X}^i$ and y is a vector valued function on $\mathbb{X}$ which is $^{\Delta}$ differentiable on $\mathbb{X}^i$. For such y consider the operator L defined by*

$$L[y](x) := y^{\Delta}(x) - M(x)y(x), \qquad x \in \mathbb{X}^i. \tag{25}$$

Suppose that $N(x)$ is nonsingular on $\mathbb{X}$ such that N^{Δ} exists on $\mathbb{X}^i$ and a vector function $z(t)$ is defined on $\mathbb{T}$ by

$$y(x) = N(x)z(t), \qquad t = f(x), \qquad x \in \mathbb{X}. \tag{26}$$

Then z is ^{o} differentiable on $\mathbb{T}^i = f(\mathbb{X}^i)$ and the operator L_0 defined by

$$L_0[z](t) := z^{o}(t) - Q(t)z(t), \qquad t \in \mathbb{T}^i \tag{27}$$

for

$$Q(t) = \frac{1}{f^{\Delta}(x)}\left[N^{\sigma}(x)\right]^{-1}\left[M(x)N(x) - N^{\Delta}(x)\right] \qquad \text{with } t = f(x), \tag{28}$$

is related to L by

$$L_0[z](t) = ([f^{\Delta}N^{\sigma}]^{-1}L[y])(x), \qquad \text{with } t = f(x). \tag{29}$$

If M is Hamiltonian on $\mathbb{X}^i$ and N is symplectic on $\mathbb{X}$, then Q is Hamiltonian on $\mathbb{T}^i$.

Note that if N is constant on $\mathbb{X}$ and $f(x) \equiv x$, then $Q(x)$ is similar to $M(x)$ for $x \in \mathbb{X}^i$.

Proof. Set

$$y(x) = N(x)w(x) \quad \text{and} \quad w(x) = z(t) \quad \text{for} \quad t = f(x), \quad x \in \mathbb{X}.$$

Then formula (3) gives

$$L[y] = (Nw)^\Delta - MNw = N^\Delta w + N^\sigma w^\Delta - MNw = N^\sigma w^\Delta - (MN - N^\Delta)w.$$

Since y is $^\Delta$ differentiable on $\mathbb{X}^i$, N is nonsingular on $\mathbb{X}$, and N^Δ exists on $\mathbb{X}$, we know that $w(x)$ is defined on $\mathbb{X}$ and $^\Delta$ differentiable on $\mathbb{X}^i$. The chain rule implies that z is ^{o} differentiable on $\mathbb{T}^i$ with $z^o(t) = g^o(t)w^\Delta(x)$. Remark 2.9 implies that $g^o(t) = (f^\Delta(x))^{-1}$ which gives $w^\Delta(x) = f^\Delta(x)z^o(t)$ and $((f^\Delta N^\sigma)^{-1}L[y])(x) = L_0[z](t)$ for L_0 defined as in (27) with Q given by (28). To show that M Hamiltonian and N symplectic implies that Q is Hamiltonian, first show that $P := (N^\sigma)^{-1}[MN - N^\Delta]$ is Hamiltonian on $\mathbb{X}^i$. Introduce the notation

$$F(M,h) := M^*\mathcal{J} + \mathcal{J}M + hM^*\mathcal{J}M, \qquad F(P,h) := P^*\mathcal{J} + \mathcal{J}P + hP^*\mathcal{J}P.$$

Then M Hamiltonian on $\mathbb{X}^i$ implies $F(M,h) = 0$ on $\mathbb{X}^i$. We wish to show that $F(P,h) = 0$ on $\mathbb{X}^i$. Rephrase the proof of Došlý and Hilscher [8], by using the identities

$$N^*\mathcal{J}N = \mathcal{J};\ (N^*)^{-1}\mathcal{J}N^{-1} = \mathcal{J} \text{ on } \mathbb{X};\ (N^\sigma)^*\mathcal{J}N^\sigma = \mathcal{J} \text{ on } \mathbb{X}^i; \tag{30}$$

also, the second product formula for $^\Delta$ derivatives (4) applied to $N^*\mathcal{J}N = \mathcal{J}$ (and the formula $N^\sigma = N + hN^\Delta$) gives, in the notation $G(N,h) := (N^*\mathcal{J}N)^\Delta = \mathcal{J}^\Delta = 0$ on $\mathbb{X}^i$, the resulting identity

$$G(N,h) = (N^\Delta)^*\mathcal{J}N + (N^\Delta)^*\mathcal{J}hN^\Delta + N^*\mathcal{J}N^\Delta \equiv 0 \qquad \text{on } \mathbb{X}^i.$$

We compute $F(P,h)$ by noting from (30) that

$$P^*\mathcal{J} = [MN - N^\Delta]^*\mathcal{J}N^\sigma$$
$$\mathcal{J}P = (N^\sigma)^*\mathcal{J}[MN - N^\Delta]$$
$$hP^*\mathcal{J}P = h[MN - N^\Delta]^*\mathcal{J}[MN - N^\Delta].$$

Use $N^\sigma = N + hN^\Delta$ in $P^*\mathcal{J}$ and $\mathcal{J}P$. Then expand $F(P,h)$ to get 12 terms, 6 of which cancel, for the identity

$$F(P,h) = N^*F(M,h)N - G(N,h) \equiv 0 \quad \text{on} \quad \mathbb{X}^i. \tag{31}$$

Thus P is Hamiltonian on $\mathbb{X}^i$. Since $P(x)$ and $Q(t)$ are related by $P(x) = f^\Delta(x)Q(t)$ and by formula (2) we have

$$h(x)f^\Delta(x) = f^\sigma(x) - f(x) = \tau(t) - t = k(t)$$

and $F(P,h) = f^\Delta F(Q,k) \equiv 0$. Since f^Δ is never 0, we conclude that

$$F(Q,k) := Q^*\mathcal{J} + \mathcal{J}Q + kQ^*\mathcal{J}Q \equiv 0 \quad \text{on} \quad \mathbb{T}^i$$

and Q is Hamiltonian on $\mathbb{T}^i$. ∎

4. NONLINEAR HAMILTONIAN SYSTEMS

Assume throughout this section that $\mathbb{T}$ is a time scale as defined in Section 1 with right jump function σ. Consider a Lagrangian $L(t, y, r)$ defined on $\mathbb{T} \times \mathbb{R}^n \times \mathbb{R}^n$ and a fixed endpoint variational problem

$$J[y] = \int_a^b L(t, y^\sigma(t), y^\Delta(t))\,\Delta t, \qquad y(a) = \alpha, \quad y(b) = \beta, \tag{32}$$

where the notation $F(t) = \int_a^t f(s)\Delta s$ means that F is the $^\Delta$ antiderivative of f, i.e., $F^\Delta(t) = f(t)$ on $\mathbb{T}$, with $F(a) = 0$. Assume that $L(t, y, r)$ is continuous at each right dense point t and has continuous partials of the first two orders with respect to the components of y and r. The Euler–Lagrange equation associated with J is

$$L_y(t, y^\sigma(t), y^\Delta(t)) = \left[L_r(t, y^\sigma(t), y^\Delta(t))\right]^\Delta. \tag{33}$$

We now develop the Legendre transformation. This proof closely follows the discrete proof given in [4, pp. 165 and 168]. Also, for comparisons between the discrete and continuous proofs, see [3, 17, Section 2 of Chap. VI, 18, p. 117].

Assume there exists a vector function g such that

$$L_r(t, u, v) = w \quad \Longleftrightarrow \quad v = g(t, u, w). \tag{34}$$

Define a Hamiltonian $H(t, u, w)$ by

$$H(t, u, w) = \langle w, g(t, u, w)\rangle - L(t, u, g(t, u, w)). \tag{35}$$

Define a generalized momentum variable z corresponding to a solution y of the Euler–Lagrange equation by

$$z(t) := L_r(t, y^\sigma(t), y^\Delta(t)). \tag{36}$$

From the definition of g we have

$$L_r(t, y^\sigma(t), y^\Delta(t)) = z(t) \iff y^\Delta(t) = g(t, y^\sigma(t), z(t)).$$

Thus we have the system

$$y^\Delta(t) = g(t, y^\sigma(t), z(t)), \qquad z^\Delta(t) = L_y(t, y^\sigma(t), g(t, y^\sigma(t), z(t)). \tag{37}$$

If y is defined on a time scale $\mathbb{T}$ which contains at least three points, then $\mathbb{T}^i$ and $(\mathbb{T}^i)^i$ are nonempty. Assume that y satisfies the Euler–Lagrange equation (33) on $(\mathbb{T}^i)^i$. Then z is defined by (36) only for $t \in \mathbb{T}^i$ and $z^\Delta(t)$ only has meaning on $(\mathbb{T}^i)^i$. To relax this domain restriction when $\mathbb{T}^i \neq \mathbb{T}$, define z at a maximal isolated (i.e., left scattered) point $T \in \mathbb{T}$ by letting $\tau := \rho(T)$, i.e., τ satisfies $\sigma(\tau) = T$, and define z at T by

$$z(T) := z(\tau) + \mu(\tau)L_y(\tau, y^\sigma(\tau), y^\Delta(\tau)). \tag{38}$$

Then the phase space coordinates $(y(t), z(t))$ are defined for $t \in \mathbb{T}$ and the system (37) is satisfied on $\mathbb{T}^i$.

THEOREM 4.1 (Legendre Transformation). *Assume that there exists a function g as in* (34), *the Hamiltonian H is defined by* (35), *y is defined on* $\mathbb{T}$, *the momentum variable z is defined by* (36), *and if* $\mathbb{T}^i \neq \mathbb{T}$, *also by* (38). *If y satisfies the Euler–Lagrange equation* (33) *on* $(\mathbb{T}^i)^i$, *then* (y, z) *is defined on* $\mathbb{T}$ *and satisfies the Hamiltonian system*

$$y^\Delta(t) = H_w(t, y^\sigma(t), z(t)), \qquad z^\Delta(t) = -H_u(t, y^\sigma(t), z(t)) \text{ for } t \in \mathbb{T}^i. \tag{39}$$

Assume henceforth that $H(t, u, w)$ is a continuous function on $\mathbb{T}^i \times \mathbb{R}^n \times \mathbb{R}^n$ which has continuous partials of the first two orders in the components of u and w. Suppose that for s a given point of $\mathbb{T}^i$, that Ω is a nonempty open set in $\mathbb{R}^n \times \mathbb{R}^n$ such that for $(p, q) \in \Omega$ the solution $y(t, p, q)$, $z(t, p, q)$ of the initial value problem

$$\begin{aligned} y^\Delta(t, p, q) &= H_w(t, y^\sigma(t, p, q), z(t, p, q)), \quad y(s, p, q) = p, \\ z^\Delta(t, p, q) &= -H_u(t, y^\sigma(t, p, q), z(t, p, q)), \quad z(s, p, q) = q \end{aligned} \tag{40}$$

exists and is unique on $\mathbb{T}$. Also assume that $y(t, p, q)$ and $z(t, p, q)$ have continuous first partials with respect to each of the components of p and q and the $^\Delta$ derivatives and partial derivatives of y and z (with respect to p or q) interchange. Since y is a column n-vector, we use the notation y_p to denote the $n \times n$ matrix $(\partial y_i/\partial p_j)$. Use the notation H_{wu} to denote the matrix $(H_{w_i u_j})$. Then if we take the partial of the ith component of y^Δ with respect to p_j we will have

$$y^\Delta_{ip_j}(t, p, q) = \sum_{k=1}^n (H_{w_i u_k})(y_k^\sigma)_{p_j} + \sum_{k=1}^n (H_{w_i w_k})(z_k)_{p_j},$$

where the partials of H are evaluated at $(t, y^\sigma(t, p, q), z(t, p, q))$. Matrix notation and similar calculations give the system (the variational equations)

$$\begin{aligned} y_p^\Delta &= H_{wu} y_p^\sigma + H_{ww} z_p, \qquad y_q^\Delta = H_{wu} y_q^\sigma + H_{ww} z_q, \\ z_p^\Delta &= -H_{uu} y_p^\sigma - H_{uw} z_p, \quad z_q^\Delta = -H_{uu} y_q^\sigma - H_{uw} z_q, \\ y_p(s, p, q) &= I, \quad y_q(s, p, q) = 0, \quad z_p(s, p, q) = 0, \quad z_q(s, p, q) = I. \end{aligned} \tag{41}$$

For fixed p and q set $A(t) = H_{wu}$, $B(t) = H_{ww}$, $C(t) = -H_{uu}$ where the partials of H are evaluated at $(t, y^\sigma(t, p, q), z(t, p, q))$. Then $-H_{uw} = -(H_{wu})^T = -A^T(t)$ and if we assume that the matrix

$$\left[I - \mu(t) H_{wu}(t, y^\sigma(t, p, q), z(t, p, q))\right] \text{ is nonsingular on } \mathbb{T}^i, \tag{42}$$

then a solution of the linear canonical system

$$u^\Delta = A u^\sigma + B v, \qquad v^\Delta = C u^\sigma - A^T v \tag{43}$$

is a solution of a linear Hamiltonian system by Proposition 1.1.

For the continuous case, an excellent discussion of the importance of symplectic mappings can be found in [21, Chap. 2]. For methods based on differential forms, they refer to [5, Chap. 7 and Sect. 44] for a proof of the continuous version of the following theorem. For the discrete case, see [4, Theorem 4.45, p. 190].

THEOREM 4.2 (Symplectic Flows). *Let s be given in $\mathbb{T}^i$. Under the above assumptions and for each fixed $t \in \mathbb{T}$, let $\Psi(p, q)$ be the mapping which carries the solution y, z from initial conditions (p, q) at time s to values $\phi(p, q) := y(t, p, q)$, $\theta(p, q) := z(t, p, q)$ at time t. Then the matrix Ψ' defined by*

$$\Psi' = \begin{bmatrix} \phi_p & \phi_q \\ \theta_p & \theta_q \end{bmatrix} \tag{44}$$

is symplectic on Ω.

Proof. Let $U_1(t) = y_p(t, p, q)$, $V_1(t) = z_p(t, p, q)$, $U_2(t) = y_q(t, p, q)$, and $V_2(t) = z_q(t, p, q)$. Then the matrix $X(t, s) = \begin{bmatrix} U_1(t) & U_2(t) \\ V_1(t) & V_2(t) \end{bmatrix}$ is the fundamental matrix solution of a linear Hamiltonian system with $X(s, s) = I_{2n}$. By Theorem 1.3 the matrix $X(t, s)$ is symplectic and hence $\Psi' = X(t, s)$ is symplectic. ∎

REFERENCES

1. R. P. Agarwal, C. Ahlbrandt, M. Bohner, and A. Peterson, Discrete linear Hamiltonian systems: A survey, *Dynam. Systems Appl.*, **8**, (1999), 307–333.
2. C. Ahlbrandt, D. Hinton, and R. Lewis, The effect of variable change on oscillation and disconjugacy criteria with applications to spectral theory and asymptotic theory, *J. Math. Anal. Appl.* **81** (1981), 234–277.
3. C. Ahlbrandt, Equivalence of discrete Euler equations and discrete Hamiltonian systems, *J. Math. Anal. Appl.* **180** (1993), 498–517.
4. C. Ahlbrandt and A. Peterson, "Discrete Hamiltonian Systems: Difference Equations, Continued Fractions, and Riccati Equations," Kluwer Academic, Boston, 1996.
5. V. Arnold, "Mathematical Methods of Classical Mechanics," 2nd Ed., Graduate Texts in Mathematics, Vol. 60, Springer-Verlag, New York, 1989.
6. B. Aulbach and S. Hilger, Linear dynamic processes with inhomogeneous time scale, in "Nonlinear Dynamics and Quantum Dynamical Systems, (Gaussig, 1990)," *Math. Res.*, Vol. 59, pp. 9–20, Akademie Verlag, Berlin, 1990.
7. L. Dieci, B. Morini, and A. Papini, Computational techniques for real logarithms of matrices, *SIAM J. Matrix Anal. Appl.* **17** (1996), 570–593.
8. O. Došlý and R. Hilscher, Disconjugacy, transformations and quadratic functionals for symplectic dynamic systems on time scales, *J. Differ. Equations Appl.,* to appear.
9. L. Erbe and S. Hilger, Sturmian theory on measure chains, *Differential Equations Dynam. Systems* **1** (1993), 223–246.
10. L. Erbe and P. Yan, Disconjugacy for linear Hamiltonian difference systems, *J. Math. Anal. Appl.* **167** (1992), 355–367.

11. S. Hilger, Analysis on measure chains—A unified approach to continuous and discrete calculus, *Results Math.* **18** (1990), 18–56.
12. S. Hilger, Special functions, Laplace and Fourier transforms on measure chains, *Dynam. Systems Appl.*, **8** (1999), 471–488.
13. R. Hilscher, Linear Hamiltonian systems on time scales: Transformations, *Dynam. Systems Appl.*, **8** (1999), 489–501.
14. R. Hilscher, Linear Hamiltonian systems on time scales: positivity of quadratic functionals, *Math. Comput. Modelling,* to appear.
15. B. Kaymakçalan, V. Lakshmikantham, and S. Sivasundaram, "Dynamical Systems on Measure Chains," Kluwer Academic, Boston, 1996.
16. J. Kelley, "General Topology," Amer. Book Co., New York, 1969.
17. C. Lanczos, "The Variational Principles of Mechanics," 4th Ed., Univ. Toronto Press, Toronto, 1970.
18. L. Michelotti, "Intermediate Classical Dynamics with Applications to Beam Physics," Wiley, New York, 1995.
19. J. Radon, "Zum Problem von Lagrange: Vier Vorträge gehalten im Mathematischen Seminar der Hamburgischen Universität," Hamburger Mathematische Einzelschriften, Heft **6**, Teubner, Leipzig, 1928.
20. W. T. Reid, "Ordinary Differential Equations," Wiley, New York, 1971.
21. J. M. Sanz-Serna and M. P. Calvo, "Numerical Hamiltonian Problems," Chapman & Hall, London, 1994.
22. G. W. Stewart, "Introduction to Matrix Computations," Academic Press, New York, 1973.