ESAIM: COCV 25 (2019) 20
https://doi.org/10.1051/cocv/2018012

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

STRONG LAGRANGE DUALITY AND THE MAXIMUM PRINCIPLE FOR NONLINEAR DISCRETE TIME OPTIMAL CONTROL PROBLEMS

Eero V. Tamminen*

Abstract. We examine discrete-time optimal control problems with general, possibly non-linear or non-smooth dynamic equations, and state-control inequality and equality constraints. A new generalized convexity condition for the dynamics and constraints is defined, and it is proved that this property, together with a constraint qualification constitute sufficient conditions for the strong Lagrange duality result and saddle-point optimality conditions for the problem. The discrete maximum principle of Pontryagin is obtained in a straightforward manner from the strong Lagrange duality theorem, first in a new form in which the Lagrangian is minimized both with respect to the state and to the control variables. Assuming differentiability, the maximum principle is obtained in the usual form. It is shown that dynamic systems satisfying a global controllability condition with convex costs, have the required convexity property. This controllability condition is a natural extension of the customary directional convexity condition applied in the derivation of the discrete maximum principle for local optima in the literature.

Mathematics Subject Classification. 90C46, 93C55.

Received March 9, 2017. Accepted February 2, 2018.

1. Introduction

Optimality conditions based on Lagrange multipliers have a key position in the theory of optimization for constrained problems. In the classic form of the theory usually presented in textbooks, convexity properties of sets and functions have a central role in the development, and the Lagrangian is a linear functional defined on the image space of the objective and constraint functions of the problem. Lagrange multipliers are vectors in the dual space of the image space, and the existence and properties of the multipliers can be based on the separation of sets in the image space by hyperplanes defined by the multipliers; *cf.* [1, 22].

For both theoretical and computational reasons, the theory of optimality conditions has been deepened and extended beyond the classic linear form. Known results have been proven under weaker hypotheses and new results have been obtained. The classic theory has been extended by considering non-linear Lagrangian functions, separation of sets in the image space by curved manifolds, and theorems of the alternative for general systems. For original results and references see [8, 9, 17, 30].

Keywords and phrases: Optimal control, discrete time, Lagrange duality, Pontryagin discrete maximum principle, convexity condition, non-linear dynamics, controllability.

Department of Mathematics and Systems Analysis, Aalto University, 00076 Aalto, Helsinki, Finland.

* Corresponding author: eero.tamminen@pp2.inet.fi

The development of variational analysis has opened a new approach to conditions of optimality, not restricted by unnecessary requirements of convexity or differentiability. This line of research has led to profoundly new and more general results on all areas of optimization theory; *cf.* [6, 23, 31].

On the other hand, interest in the classic theory itself has not declined, because it has special concrete properties not always shared by deeper, often more abstract results. The optimality conditions based on the linear Lagrangian can be split into separate components, constraint by constraint and time-step by time-step, and the linear Lagrange duality theory is geometrically transparent and has profound symmetry properties. These features are important theoretically, computationally and in applications. In this work, we consider only classic linear Lagrangian functions.

Under appropriate assumptions, the optimality theorems establish the existence of a separating hyperplane in the image space, and a corresponding dual multiplier vector satisfying a set of conditions at a solution point of the problem. In the case of various problems, the existence of a separating hyperplane is a deeper issue, and the resulting explicit conditions of optimality are then more immediate. The mathematical theory of optimality conditions is extensive. In order to obtain a good overall picture of the theory, and to be able to place individual theorems in their proper context, it is helpful to organize the results in a suitable manner. Two main approaches to Lagrange multipliers for constrained optimization problems are traditionally distinguished in the literature – the local and the global, respectively [1, 22]. The local approach, *cf.* Chapter 4 in [1], is based on the differentiability of the objective and constraint functions of the problem, and leads to the Karush–Kuhn–Tucker necessary conditions for a local optimum. The global approach, the Lagrange duality theory, is based on convexity, and leads to the strong duality theorem and saddle point optimality conditions, *cf.* Chapter 6 in [1], which are necessary and sufficient for achieving a global optimum.

The strong Lagrange duality and saddle-point theorems for a constrained optimization problem are based on two basic assumptions about the problem: a convexity property in the image space of the constraints and a normality condition (constraint qualification) which rules out certain ill-behaved problems. The convexity property sufficient for duality is satisfied if the objective and the inequality constraint functions of the problem are convex, and the equality constraint functions are affine. These assumptions are transparent and easy to verify, and in the standard proofs of strong Lagrange duality for constrained optimization problems in the literature, convexity of the inequalities and affinity of the equalities are assumed, see [1] for example. These properties are not, however, necessary conditions for strong duality. The convexity property in the image space can be guaranteed by weaker assumptions. Alternative convexity properties sufficient for strong duality in different types of problems have been treated in Elster and Nehse [7], Hayashi and Komiya [13], Jeyakumar [20], Nehse [24], Simons [33], Tamminen [34], and Illés and Kassay [17].

The maximum principle of Pontryagin (PMP) gives necessary conditions for optimality for optimal control problems. Differentiability of the objective and constraint functions of the problem with respect to the state variables is assumed, and the principle establishes the existence of a dual vector (Lagrange multiplier), which together with the primal variables satisfies the optimality conditions at a solution point. In the theorem there are two central conditions: the Euler-Lagrange equation, and the maximum condition for the Hamiltonian function, respectively. The maximum principle gives necessary conditions for an optimum, which is local in the space of state variables, and global in the control set, *cf.* [18], and can be viewed as a counterpart of the Karush–Kuhn–Tucker theorem.

The principle was formulated and proved first for continuous-time problems by Pontryagin, Boltyanskii, Gamkrelidze and Mishchenko in [26]. Somewhat later the corresponding result for discrete-time problems, the discrete maximum principle (DMP), was proved by Halkin and Holzmann, and by Propoi, see [11, 12, 14, 15, 16, 27, 28]. However, the result for the discrete-time case is not an exact counterpart of the PMP. In the proof of the maximum condition for the Hamiltonian in the discrete maximum principle an additional condition of directional convexity for the dynamic equation, not needed in the proof of continuous maximum principle, is required.

The early development of the DMP was completed by Boltyanskii in [2]. He proved first that the Hamiltonian gradient is non-positive at every time point, which is a weaker form of the maximum principle. Assuming directional convexity, he then obtained the strong form of the DMP, with the optimality condition for the Hamiltonian.

Both in the case of continuous-time, and discrete-time problems, the local approach to conditions for optimality was applied in the original derivation of the principle. Hager and Mitter [10], Rockafellar [29], and Zalmai [35] have applied the global approach for continuous-time optimal control problems with linear dynamics, and derived strong Lagrange duality, saddle point optimality, and maximum principle conditions for these problems.

Recently, Bourdin and Trélat have analyzed control problems on time scales (closed subsets of the time axis), and extended and proved the Pontryagin maximum principle for optimal control problems defined on arbitrary time scales in [3, 4]. Their result combines into one theorem the maximum principle results for continuous-time and discrete-time optimal control problems. Points on the time scale are divided into two mutually exclusive classes, right dense, and right scattered points, respectively. At right dense points time flows forward continuously, and the optimum condition for the Hamiltonian is satisfied without an additional convexity assumption, as in the continuous-time maximum principle. At right scattered points time takes a finite step forward, and the Hamiltonian satisfies a gradient condition. If the Hamiltonian is convex with respect to the control variable, a certain optimum condition can be derived at these points, too. The continuous, and discrete forms of the maximum principle are obtained as special cases.

After the original development, approaches to conditions of optimality in mathematical optimization, and in optimal control, were unified by many authors; *cf.* Luenberger [22], Canon *et al.* [5], and Ioffe and Tihomirov [18]. For a thorough presentation of the general (classic) theory of mathematical optimization in finite dimensional spaces see Bazaraa *et al.* [1], and the extensive literature cited there.

Optimal control problems with discrete time are mathematical optimization problems with a special structure, set up in finite-dimensional spaces. Control problems have two types of variables which are connected by the dynamic equation. State variables are elements of a finite dimensional vector space (the state space) and control variables belong to an abstract control set, which does not necessarily have a topological structure. From the viewpoint of mathematical optimization, the dynamic equation is an equality constraint within the problem. General results of optimization theory apply to control problems, and more specific results are obtained by utilizing the special structure. The discrete maximum principle, *cf.* [2, 5, 25, 32], is a key instance of such results.

Propoi proved in [28] the strong Lagrange duality result and saddle point optimality conditions for discrete optimal control problems with dynamics that are linear with respect to the state variables, and with convex inequality constraints. However, strong duality for control problems with nonlinear dynamic equations, leading to nonlinear equality constraints, has not been treated in the literature. In this work, we return to this problem and apply a general convexity condition, which is new in the context of discrete control problems and does not require linearity of the dynamic equations or the equality constraints of the problem, see [34].

Together with a constraint qualification, the convexity condition is sufficient for strong Lagrange duality and the existence of a saddle point for the problem. From the strong duality theorem, we first obtain the discrete maximum principle in a new form in which differentiability is not assumed, and the Lagrangian is minimized both with respect to the state variables and the control variables, then in sub-differential form and finally, assuming differentiability, in the customary form. The derivation of these results is transparent and, in a step-by-step manner, highlights the roles played by the different assumptions in establishing the results.

The control problem studied is defined in Section 2. For easy reference, the duality theorems needed are recalled without proofs in Section 3. Section 4 contains the results, the convexity condition is analysed in Section 5, and Section 6 concludes the paper with some remarks.

2. The optimization problem and definitions

We examine controlled dynamic processes in discrete time $t = 0, 1, \dots, T$. A process is a pair (x, u), where the trajectory $x = [x(t); t = 0, 1, \dots, T]$ is the sequence of the states $x(t) \in \mathbb{R}^n$ of the system at the time points t, where $\mathbb{R}$ is the set of real numbers, and the control of the system $u = [u(t); t = 0, 1, \dots, T-1]$ is the sequence of control elements $u(t) \in V_t$, where V_t is a general nonempty control set at time t. The processes (x, u) are elements of the set $W = X \times V$, where $x \in X = \mathbb{R}^{n(T+1)}$ and $u \in V = V_0 \times V_1 \times \cdots \times V_{(T-1)}$.

Definition 2.1 (The primal problem)**.** The following optimal control problem is called the primal problem:

$$\text{minimize } \left\{ C(x,u) = \sum_{t=0}^{T-1} c_t(x(t), u(t)) + c_T(x(T)) \right\} \text{ subject to the constraints} \tag{2.1}$$

$$x(t+1) = x(t) + f_t(x(t), u(t)), \quad t = 0, 1, \ldots, T-1, \tag{2.2}$$

$$x(0) = q, \tag{2.3}$$

$$g_t(x(t), u(t)) \leq 0, \quad t = 0, 1, \ldots, T-1, \tag{2.4}$$

$$h_t(x(t), u(t)) = 0, \quad t = 0, 1, \ldots, T-1, \text{ and} \tag{2.5}$$

$$u(t) \in V_t, \qquad t = 0, 1, \ldots, T-1. \tag{2.6}$$

The real function C on the set W in (2.1) is the objective to be minimized. The dynamic development of the system is governed by the vector difference equation (2.2), with the given initial state (2.3). The inequalities (2.4), and the equalities (2.5), define the state-control constraints of the problem. Constraints defined by functions of the state only, *i.e.* constraints defined by functions constant on the appropriate V_t are included among these. The relations (2.6) define the constraints on the control only.

The initial and terminal boundary conditions on the state are of the simplest possible type, with a given initial, and free final, state. This choice is made in order to concentrate on the essential, dynamic, primal-dual structure of the results, and to avoid unnecessary detail. Extension to other types of boundary conditions is straightforward. All constraint inequalities for vectors in real Euclidean spaces are defined in the natural manner, component for component.

In the primal problem, for each $t = 0, 1, \ldots, T-1$, the functions c_t, f_t, g_t, and h_t, are defined on the set $\mathbb{R}^n \times V_t$, into real Euclidean vector spaces of dimensions 1, n, r, and s, respectively. As there is no control vector at time T, the last term in the objective function C, the terminal cost c_T, is a real function on $\mathbb{R}^n$. We use the following short notation: the dynamic equation (2.2) with the initial Condition (2.3) is written $F(x, u) = 0$, the inequalities (2.4) are written $G(x, u) \leq 0$, and the equalities (2.5) $H(x, u) = 0$, where the functions $F\colon W \to X = \mathbb{R}^{n(T+1)}$, $G\colon W \to Y = \mathbb{R}^{rT}$, and $H\colon W \to Z = \mathbb{R}^{sT}$, are defined by:

$$F(x,u) = [-x(0) + q, -x(t+1) + x(t) + f_t(x(t), u(t)); t = 0, 1, \ldots, T-1], \tag{2.7}$$

$$G(x,u) = [g_t(x(t), u(t)); t = 0, 1, \ldots, T-1], \text{ and} \tag{2.8}$$

$$H(x,u) = [h_t(x(t), u(t)); t = 0, 1, \ldots, T-1]. \tag{2.9}$$

Using this notation, the primal problem (2.1)–(2.6) is written in the following short form:

$$\text{minimize } C(x,u), \text{ subject to:} \tag{2.10}$$

$$F(x,u) = 0, \tag{2.11}$$

$$G(x,u) \leq 0, \tag{2.12}$$

$$H(x,u) = 0, \text{ and} \tag{2.13}$$

$$(x,u) \in W = X \times V. \tag{2.14}$$

All of the inequality constraints are combined into (2.12), whereas the different roles of the state equation (2.11) and the constraint equalities (2.13) within the maximum principle provide the reason for writing them as separate equations in the problem. No special assumptions are made about the structure of control set V. Throughout the paper, the primal problem is assumed to have a finite optimal solution.

3. The Lagrangian function and duality

For Lagrange multipliers, we use the notation of Luenberger [22]. Continuous linear functionals on a normed vector space Q form its topological dual vector space, denoted as Q^*, and the value of a functional $\rho \in Q^*$ at the point $r \in Q$ is written as $\langle r, \rho \rangle$. Lagrange multipliers are elements of the dual vector space of the image space of the constraint functions. The image spaces X, Y, and Z of the constraints in the definition of the primal problem are finite dimensional Euclidean vector spaces, and consequently, $X^* = X = \mathbb{R}^{n(T+1)}$, $Y^* = Y = \mathbb{R}^{rT}$, and $Z^* = Z = \mathbb{R}^{sT}$.

Definition 3.1 (The Lagrangian function)**.** The Lagrangian function $L\colon W \times X^* \times Y^* \times Z^* \to \mathbb{R}$ for the primal problem is defined for $(x,u) \in W$, and $(\mathit{\Psi}, \mathit{\Lambda}, \mathit{\Gamma}) \in X^* \times Y^* \times Z^*$ by:

$$L(x,u;\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma}) = C(x,u) + \langle F(x,u), \mathit{\Psi}\rangle + \langle G(x,u), \mathit{\Lambda}\rangle + \langle H(x,u), \mathit{\Gamma}\rangle. \tag{3.1}$$

The Lagrange multiplier $(\mathit{\Psi}, \mathit{\Lambda}, \mathit{\Gamma})$ has the components $\mathit{\Psi} \in X^*$, $\mathit{\Lambda} \in Y^*$, and $\mathit{\Gamma} \in Z^*$, and these are further partitioned into vector components corresponding to the structure (2.7)–(2.9) of the functions F, G, and H, respectively, and indexed as follows: $\mathit{\Psi} = [\psi(t) \in \mathbb{R}^n; t = 0,1,\ldots,T]$, $\mathit{\Lambda} = [\lambda(t) \in \mathbb{R}^r; t = 0,1,\ldots,T-1]$, and $\mathit{\Gamma} = [\gamma(t) \in \mathbb{R}^s; t = 0,1,\ldots,T-1]$. Using this notation, the terms of the Lagrangian function (3.1), associated with the constraints, have the following detailed expressions:

$$\langle F(x,u), \mathit{\Psi}\rangle = \langle -x(0) + q, \psi(0)\rangle + \sum_{t=0}^{T-1} \big\langle -x(t+1) + x(t) + f_t(x(t),u(t)), \psi(t+1)\big\rangle, \tag{3.2}$$

$$\langle G(x,u), \mathit{\Lambda}\rangle = \sum_{t=0}^{T-1} \big\langle g_t(x(t),u(t)), \lambda(t)\big\rangle, \text{ and} \tag{3.3}$$

$$\langle H(x,u), \mathit{\Gamma}\rangle = \sum_{t=0}^{T-1} \big\langle h_t(x(t),u(t)), \gamma(t)\big\rangle. \tag{3.4}$$

In (3.2), it is natural to name the first (vector) component of $\mathit{\Psi}$ as $\psi(0)$, and the component multiplying the term $[-x(t+1) + x(t) + f_t(x(t),u(t))]$ as $\psi(t+1)$, for $t = 0,1,\ldots,T-1$. For the primal problem, with the Lagrangian function (3.1), the corresponding dual function and the Lagrange dual optimization problem are defined as follows:

Definition 3.2 (The dual function)**.** The dual function $\mathit{\Phi}\colon X^* \times Y^* \times Z^* \to \mathbb{R} \cup \{-\infty\}$ is defined by

$$\mathit{\Phi}(\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma}) = \inf\{L(x,u;\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma}) : (x,u) \in W\}. \tag{3.5}$$

Definition 3.3 (The Lagrange dual optimization problem)**.** The Lagrange dual optimization problem is defined as:

$$\max\{\mathit{\Phi}(\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma}) : (\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma}) \in X \times Y \times Z\}, \text{ subject to} \tag{3.6}$$

$$\mathit{\Lambda} \geq 0. \tag{3.7}$$

The set $P^* = \{\mathit{\Lambda} \in Y^* : \mathit{\Lambda} \geq 0\}$ is defined as the positive cone in Y^*. Any (x,u) satisfying the constraints (2.11)–(2.14) of the primal problem is called primal feasible, and any $(\mathit{\Psi},\mathit{\Lambda},\mathit{\Gamma})$ satisfying the constraint (3.7) of the dual problem is dual feasible. We recall the following fundamental relations between solutions to the primal, dual and Lagrangian minimization problems required in the sequel. Theorems 3.4 and 3.5 are algebraic consequences of the definitions made above, for proofs *cf.* [1, 34], the details are omitted here.

Theorem 3.4 (Weak and strong duality)**.** *If $(\hat{x}, \hat{u}) \in W$ is primal feasible, and $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times Y^* \times Z^*$ is dual feasible, then $\Phi(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \leq C(\hat{x}, \hat{u})$. Further, if $\Phi(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) = C(\hat{x}, \hat{u})$, then $(\hat{x}, \hat{u})$ and $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ are optimal solutions to the primal and dual problems, respectively.*

Theorem 3.5 (Conditions equivalent to strong duality)**.** *Let $(\hat{x}, \hat{u}) \in W$ be primal feasible and $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times Y^* \times Z^*$ dual feasible. Then the strong duality condition $\Phi(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) = C(\hat{x}, \hat{u})$ is satisfied if and only if the following conditions are satisfied:* $\min\{L(x, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : (x, u) \in W\}$ *is obtained at the point $(\hat{x}, \hat{u})$, and $\langle G(\hat{x}, \hat{u}), \hat{\Lambda}\rangle = 0$.*

According to duality Theorem 3.4, any dual feasible vector provides a lower bound for the primal optimum, and any primal feasible vector an upper bound for the dual optimum. The strong duality condition, $\Phi(\Psi, \Lambda, \Gamma) = C(x, u)$, is sufficient for the primal and dual optimality of the respective vectors. Theorem 3.5 states that, for the feasible primal and dual vectors $(\hat{x}, \hat{u})$ and $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$, minimization of the Lagrangian $L(x, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ over W at the point $(\hat{x}, \hat{u})$, and the condition of complementary slackness, $\langle G(\hat{x}, \hat{u}), \hat{\Lambda}\rangle = 0$, are necessary and sufficient for the strong duality and, consequently, for the primal and dual optimality of the respective vectors. The point $(\hat{x}, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ is also a saddle point of the Lagrangian function on the set $W \times X^* \times P^* \times Z^*$.

For any constrained optimization problem, the question of the existence of a Lagrange multiplier vector which, together with a primal optimal solution vector, satisfies the strong duality condition is of great interest. If such a multiplier exists, then the strong duality condition and the equivalent conditions of Theorem 3.5 provide a starting point for the structural analysis and solution of the problem.

4. Strong Lagrange duality theorem and the discrete maximum principle for the primal problem

The existence of a Lagrange multiplier vector satisfying the strong duality condition can be based on convexity and normality properties associated with the problem. Of these, the convexity requirement concerns the structure of the problem. Here, we apply the general convexity property of [34] which, when stated in relation to the primal problem, takes the form of Condition 4.1. This condition is analysed in detail in the next section. For some other applications of the convexity property of [34] and other related generalized convexity conditions in optimization see [17, 21]. The normality Condition 4.2 is required in order to guarantee that the Lagrange multiplier vector constructed in Theorem 4.3 is normal, i.e. that the scalar multiplier of the objective function in the Lagrangian is strictly positive. The Lagrangian is linear in the multiplier vector, and it is then possible to take the component multiplying the objective equal to one, as is already done in the Definition 3.1 of the Lagrangian. A number of different normality conditions for constrained optimization problems have been treated in the literature, see *e.g.* [1]. Condition 4.2 is simple in form, and all the constraints are treated on a formally equal basis. It is easy to prove, that the normality condition for general optimization problems with a similar structure in [18] implies Condition 4.2.

Condition 4.1 (Generalized convexity)**.** For every $(x_1, u_1) \in W$, $(x_2, u_2) \in W$, and $\kappa \in [0, 1]$, there is an element $(x_3, u_3) \in W$, such that:

$$C(x_3, u_3) \leq \kappa C(x_1, u_1) + (1 - \kappa)C(x_2, u_2), \tag{4.1}$$

$$F(x_3, u_3) = \kappa F(x_1, u_1) + (1 - \kappa)F(x_2, u_2), \tag{4.2}$$

$$G(x_3, u_3) \leq \kappa G(x_1, u_1) + (1 - \kappa)G(x_2, u_2), \text{ and} \tag{4.3}$$

$$H(x_3, u_3) = \kappa H(x_1, u_1) + (1 - \kappa)H(x_2, u_2). \tag{4.4}$$

Condition 4.2 (Normality (constraint qualification))**.** For every nonzero $(\Psi, \Lambda, \Gamma) \in X^* \times P^* \times Z^*$, there is $(x, u) \in W$ such that:

$$\langle F(x, u), \Psi\rangle + \langle G(x, u), \Lambda\rangle + \langle H(x, u), \Gamma\rangle < 0. \tag{4.5}$$

Note that the inequality (4.5) must be satisfied only for multiplier vectors (Ψ, Λ, Γ) satisfying $\Lambda \in P^* = \{\Lambda \in Y^* : \Lambda \geq 0\}$. By application of Theorem 5.2 of [34], we obtain the Lagrange duality Theorem 4.3 for the primal problem.

Theorem 4.3 (Lagrange duality theorem)**.** *Assume that the conditions of convexity and normality 4.1 and 4.2 are satisfied in the primal problem. A feasible process $(\hat{x}, \hat{u}) \in W$ is an optimal solution to the primal problem if and only if there exists a dual feasible Lagrange multiplier vector $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times P^* \times Z^*$, such that the strong duality condition $\Phi(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) = C(\hat{x}, \hat{u})$ is satisfied. It follows that the vector $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ is an optimal solution to the dual problem.*

On the strength of Theorem 3.5, the Lagrange duality theorem can also be stated in the following, alternative form.

Theorem 4.4 (Necessary and sufficient conditions for optimality in the case of strong Lagrange duality)**.** *Assume convexity and normality 4.1 and 4.2. A primal feasible process $(\hat{x}, \hat{u}) \in W$ is optimal if and only if there exists a Lagrange multiplier vector $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times Y^* \times Z^*$, such that the following conditions are satisfied:*

$$\min\{L(x, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : (x, u) \in W\} \text{ is attained at the point } (\hat{x}, \hat{u}), \tag{4.6}$$

$$\hat{\Lambda} \geq 0, \text{ and} \tag{4.7}$$

$$\langle G(\hat{x}, \hat{u}), \hat{\Lambda} \rangle = 0. \tag{4.8}$$

The discrete maximum principle of Pontryagin is a direct consequence of the theorem. The convexity and normality assumptions are only needed in establishing the existence of a feasible dual vector satisfying the strong duality condition. All subsequent results follow from this condition only. In other words, for the control problem studied, convexity and normality are sufficient conditions for strong duality, which is in turn sufficient for the maximum principle. In order to obtain the conditions of the principle in detail, we derive two separate conditions from the minimum Condition (4.6) for the Lagrangian function, one for the state and one for the control, respectively, see Proposition 4.5 below.

Proposition 4.5 (Elementary necessary conditions of optimality for the Lagrangian function)**.** *Let$(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times Y^* \times Z^*$. If $\min\{L(x, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : (x, u) \in W = X \times V\}$ is obtained at the point $(\hat{x}, \hat{u})$, then*

$$\min\{L(\hat{x}, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : u \in V\} \text{ is obtained at the point } \hat{u}, \text{ and} \tag{4.9}$$

$$\min\{L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : x \in X\} \text{ is obtained at the point } \hat{x}. \tag{4.10}$$

The set X is a vector space and the minimum Condition (4.10) is satisfied, if and only if, the null vector θ is a subgradient of the Lagrangian function $L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ at $\hat{x}$:

$$\theta \in \partial_x L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}); x = \hat{x}. \tag{4.11}$$

Further, if the Lagrangian function $L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ is differentiable with respect to x at $\hat{x}$, then the gradient of the Lagrangian function at the minimum point is zero:

$$\theta = L_x(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}); x = \hat{x}. \tag{4.12}$$

In the proposition the subgradient, *cf.* [1, 18, 22], is defined as in convex analysis: For a (not necessarily convex) function $B \colon X \to \mathbb{R}$ on the vector space X, the vector $d \in X^*$ is a subgradient of B at $\hat{x}$ if $B(\hat{x}) + \langle (x - \hat{x}), d \rangle \leq B(x)$ for every $x \in X$. The set of subgradients of B at the point $\hat{x}$ is the subdifferential $\partial_x B(x); x = \hat{x}$, and it may be an empty set for a nonconvex function.

The function $L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ on X is not assumed to be convex, neither is convexity required for the equivalence of the conditions (4.10) and (4.11). Condition (4.11) states that the subdifferential of the Lagrangian function $L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ at the optimal point $\hat{x}$ contains the null vector θ, and is thus not empty. For the differentiability of the Lagrangian function with respect to x at $\hat{x}$, it is sufficient, that all the functions in the definition of the primal problem are differentiable with respect to x at the optimum point $(\hat{x}, \hat{u})$. The proof of the proposition is standard, and omitted here. The Condition (4.9) is the optimality condition with respect to the control in the maximum principle, and the different variants (4.10)–(4.12) of the optimality condition with respect to the state correspond to the Euler–Lagrange equation. We then substitute the Condition (4.9), and one of the alternatives (4.10)–(4.12), for the Condition (4.6) in Theorem 4.4, and obtain the maximum principle as a consequence.

Theorem 4.6 (The discrete maximum principle of Pontryagin without, and with, differentiability, in short notation)**.** *Assume that convexity Condition 4.1 and normality Condition 4.2 are satisfied. If $(\hat{x}, \hat{u})$ is an optimal solution to the primal problem, then a Lagrange multiplier vector $(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) \in X^* \times Y^* \times Z^*$ exists such that*

$$\min\{L(\hat{x}, u; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : u \in V\} \textit{ is obtained at point } \hat{u}, \textit{ and} \tag{4.13}$$

$$\min\{L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) : x \in X\} \textit{ is obtained at point } \hat{x}, \tag{4.14a}$$

equivalent to the condition:

$$\theta \in \partial_x L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}); x = \hat{x}, \tag{4.14b}$$

and the conditions of dual feasibility and complementary slackness are satisfied:

$$\hat{\Lambda} \geq 0, \textit{ and} \tag{4.15}$$

$$\langle G(\hat{x}, \hat{u}), \hat{\Lambda} \rangle = 0. \tag{4.16}$$

Further, if the Lagrangian function $L(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma})$ is differentiable with respect to the state x at $\hat{x}$, then the gradient of the Lagrangian function at the optimal point is zero:

$$\theta = L_x(x, \hat{u}; \hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}); x = \hat{x}. \tag{4.14c}$$

Theorem 4.6 provides necessary conditions for optimality for the primal problem written in the short form (2.10)–(2.14). The conditions of the theorem take the familiar form of the discrete maximum principle of Pontryagin, when the detailed structure and notation (2.1)–(2.6) and (3.2)–(3.4) are taken into account, and the conditions are rewritten accordingly. We assume differentiability and use the minimum condition with respect to the state variable x in the form (4.14c), obtaining Theorem 4.7.

Theorem 4.7 (The discrete maximum principle of Pontryagin)**.** *Assume that convexity Condition 4.1 and normality Condition 4.2 are satisfied, and that for every control $u \in V$ all functions of the state x in the definition of the primal problem are differentiable with respect to x. If $(\hat{x}, \hat{u})$ is an optimal solution to the primal problem, then there exist Lagrange multiplier vectors $\hat{\psi}(t) \in \mathbb{R}^n; t = 0, 1, \dots, T$, $\hat{\lambda}(t) \in \mathbb{R}^r; t = 0, 1, \dots, T-1$, and $\hat{\gamma}(t) \in \mathbb{R}^s; t = 0, 1, \dots, T-1$ which satisfy the following propositions (i)–(vi):*

(i) The adjoined state $\psi(t)$ satisfies the dynamic equation with the terminal condition: For every $t = 0, 1, \dots, T-1$,

$$\begin{aligned}\hat{\psi}(t) = \hat{\psi}(t+1) &+ \left[f_{tx}(\hat{x}(t), \hat{u}(t))\right]^\top \hat{\psi}(t+1) \\ &+ c_{tx}(\hat{x}(t), \hat{u}(t)) + \left[g_{tx}(\hat{x}(t), \hat{u}(t))\right]^\top \hat{\lambda}(t) + \left[h_{tx}(\hat{x}(t), \hat{u}(t))\right]^\top \hat{\gamma}(t), \textit{ and}\end{aligned} \tag{4.17}$$

$$\hat{\psi}(T) = c_{Tx}(\hat{x}(T)). \tag{4.18}$$

(ii) *The minimum condition for the local (time-point-wise) Lagrangian is satisfied: For every* $t = 0, 1, \ldots, T-1$,

$$\min\Big\{c_t(\hat{x}(t), v) + \big\langle f_t(\hat{x}(t), v), \hat{\psi}(t+1)\big\rangle + \big\langle g_t(\hat{x}(t), v), \hat{\lambda}(t)\big\rangle + \big\langle h_t(\hat{x}(t), v), \hat{\gamma}(t)\big\rangle : v \in V_t\Big\} \tag{4.19}$$

is obtained at the point $\hat{u}(t)$.

(iii) *Conditions of non-negativity and complementary slackness for the Lagrange multipliers associated with the state-control inequality constraints are satisfied: For every* $t = 0, 1, \ldots, T-1$,

$$\hat{\lambda}(t) \geq 0, \ \textit{and} \tag{4.20}$$

$$\big\langle g_t(\hat{x}(t), \hat{u}(t)), \quad \hat{\lambda}(t)\big\rangle = 0. \tag{4.21}$$

Propositions (ii) and (iii) are equivalent to the following, strong duality conditions (iv)–(vi) for the constrained minimization of the Hamiltonian function $c_t(\hat{x}(t), v) + \langle f_t(\hat{x}(t), v), \hat{\psi}(t+1)\rangle$ *of the problem, and it's Lagrange dual optimization problem. The dual functional is given below, see Definition 4.8.*

(iv) *The time-point-wise primal minimum condition for the Hamiltonian: For every* $t = 0, 1, \ldots, T-1$,

$$\min\Big\{c_t(\hat{x}(t), v) + \big\langle f_t(\hat{x}(t), v), \hat{\psi}(t+1)\big\rangle\Big\}, \ \textit{subject to:} \tag{4.22}$$

$$g_t(\hat{x}(t), v) \leq 0, \tag{4.23}$$

$$h_t(\hat{x}(t), v) = 0, \ \textit{and} \tag{4.24}$$

$$v \in V_t, \tag{4.25}$$

is obtained at the point $\hat{u}(t)$.

(v) *The time-point-wise dual maximum condition: For every* $t = 0, 1, \ldots, T-1$,

$$\max\Big\{\varphi_t\big(\hat{x}(t), \hat{\psi}(t+1); \lambda, \gamma\big) : (\lambda, \gamma) \in \mathbb{R}^r \times \mathbb{R}^s\Big\}, \ \textit{subject to:} \tag{4.26}$$

$$\lambda \geq 0, \tag{4.27}$$

is obtained at the point $(\hat{\lambda}(t), \hat{\gamma}(t))$.

(vi) *The (local) strong duality condition: For every* $t = 0, 1, \ldots, T-1$,

$$c_t(\hat{x}(t), \hat{u}(t)) + \big\langle f_t(\hat{x}(t), \hat{u}(t)), \hat{\psi}(t+1)\big\rangle = \varphi_t\big(\hat{x}(t), \hat{\psi}(t+1); \hat{\lambda}(t), \hat{\gamma}(t)\big). \tag{4.28}$$

This completes the theorem.

Definition 4.8 (The dual functional for the time-point-wise Hamiltonian minimization problem)**.** For every $t = 0, 1, \ldots, T-1$, $\xi \in \mathbb{R}^n$, and $\Sigma \in \mathbb{R}^n$, and for every $(\lambda, \gamma) \in \mathbb{R}^r \times \mathbb{R}^s$, the local dual functional $\varphi_t(\xi, \Sigma; \cdot, \cdot) \colon \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R} \cup \{-\infty\}$ is defined by:

$$\varphi_t\big(\xi, \Sigma; \lambda, \gamma\big) = \inf\Big\{c_t(\xi, v) + \big\langle f_t(\xi, v), \Sigma\big\rangle + \big\langle g_t(\xi, v), \lambda\big\rangle + \big\langle h_t(\xi, v), \gamma\big\rangle : v \in V_t\Big\}. \tag{4.29}$$

The set V_t is never empty, and it follows that either a finite infimum in (4.29) exists, or the infimum equals $-\infty$. We have assumed that the primal problem has a finite optimum solution, and the Lagrangian minimum condition (ii) of Theorem 4.7, satisfied by the optimal solution, implies that for every t, for $\xi = \hat{x}(t)$, and $\Sigma = \hat{\psi}(t+1)$, the set of points $(\lambda, \gamma) \in \mathbb{R}^r \times \mathbb{R}^s$, $\lambda \geq 0$, where the infimum (4.29) is finite, is nonempty and the dual maximization problem (4.26) and (4.27) is properly defined.

Proof of Theorem 4.7. The detailed conditions of Theorem 4.7 are obtained from Theorem 4.6 by inserting the expressions for the cost and constraint functions of the problem. The Euler–Lagrange equation (4.14c) gives the conditions (i). The minimum Condition (4.13) takes the form (ii), and (4.15) and (4.16) give the conditions (iii). Note that as $G(\hat{x}, \hat{u}) \leq 0$, and $\hat{\Lambda} \geq 0$, it follows that each additive term in $\langle G(\hat{x}, \hat{u}), \hat{\Lambda} \rangle$ is non-positive, and the complementary slackness Condition (4.16) implies that each term equals zero. The equivalence of propositions (ii) to (iii) and propositions (iv)–(vi) is an application of Theorems 3.4 and 3.5 to the time-point-wise Hamiltonian minimization problems. The details of the proof are omitted here. □

Propositions (i)–(iii) of Theorem 4.7 give the usual form of the necessary conditions for optimality of the discrete maximum principle of Pontryagin in the literature, see [5]. Propositions (ii) to (vi) in Theorem 4.7 are not affected by the choice between the alternative forms (4.14a)–(4.14c) of the minimum condition for the Lagrangian with respect to the state variables. Finally we note that in addition to the necessary conditions for optimality of the Theorems 4.6 and 4.7, under the assumptions made also the strong global duality condition, $\Phi(\hat{\Psi}, \hat{\Lambda}, \hat{\Gamma}) = C(\hat{x}, \hat{u})$, which is sufficient for primal and dual optimality, is also satisfied.

5. Examples of problems with the required convexity property

Condition 4.1 is an application of the convexity condition for general optimization problems of [34] to a discrete-time optimal control problem. By direct substitution, it can be seen that Condition 4.1 is satisfied in a primal problem with the following linear structure: (i) All the functions of state and control in the problem are additively separable to state-dependent and control-dependent terms. (ii) The state-dependent terms are linear. (iii) The control-dependent terms satisfy a straightforward derived convexity condition. This derived condition is always satisfied, if the control sets are convex subsets of vector spaces, and if also all the functions of control in the problem are linear.

Next, we examine the possibility of satisfying Condition 4.1 if we choose $x_3 = \kappa x_1 + (1-\kappa)x_2$ within it, *i.e.* we assume that for any $(x_1, u_1) \in W$, any $(x_2, u_2) \in W$, and for any $\kappa \in [0,1]$ a $u_3 \in V$ exists, such that Condition 4.1 is satisfied by the element $(x_3, u_3) \in W$, where $x_3 = \kappa x_1 + (1-\kappa)x_2$. We also assume that the Condition 4.1 for the objective function is satisfied by each additive term c_t separately. Inserting these assumptions into Condition 4.1 leads to the following time-step-wise Condition 5.1, implying Condition 4.1:

Condition 5.1 (Sufficient conditions for convexity property 4.1)**.** For every $t = 0, 1, \ldots, T-1$, for every $\xi_1, \xi_2 \in \mathbb{R}^n$, for every $\nu_1, \nu_2 \in V_t$, and for every $\kappa \in [0,1]$, there is a $\nu_3 \in V_t$ such that:

$$c_t(\xi, \nu_3) \leq \kappa c_t(\xi_1, \nu_1) + (1-\kappa) c_t(\xi_2, \nu_2), \tag{5.1}$$

$$f_t(\xi, \nu_3) = \kappa f_t(\xi_1, \nu_1) + (1-\kappa) f_t(\xi_2, \nu_2), \tag{5.2}$$

$$g_t(\xi, \nu_3) \leq \kappa g_t(\xi_1, \nu_1) + (1-\kappa) g_t(\xi_2, \nu_2), \text{ and} \tag{5.3}$$

$$h_t(\xi, \nu_3) = \kappa h_t(\xi_1, \nu_1) + (1-\kappa) h_t(\xi_2, \nu_2), \tag{5.4}$$

where $\xi = \kappa\xi_1 + (1-\kappa)\xi_2$, and for $t = T$, the final cost c_T is a convex function on $\mathbb{R}^n$.

Condition 5.1 for time $t = 0, 1, \ldots, T-1$ can be illustrated as follows. Suppose first that the problem has no inequality or equality constraints, and the convexity property consequently has no Conditions (5.3) and (5.4). We examine the system at time t, and Conditions (5.1) and (5.2), and choose $x(t) = \xi_1$, and $u(t) = \nu_1$. Then the state of the system takes step $f_t(\xi_1, \nu_1)$ from the point $x(t) = \xi_1$ to the point $x(t+1) = \xi_1 + f_t(\xi_1, \nu_1)$ at time $t+1$, and the contribution to the cost function at step t is $c_t(\xi_1, \nu_1)$, and correspondingly for $x(t) = \xi_2$, and $u(t) = \nu_2$. Condition 5.1 for $t = 0, 1, \ldots, T-1$ is twofold: (i) The equation (5.2) is a controllability requirement which states that, for any two initial points ξ_1 and ξ_2, for any $\kappa \in [0,1]$, and for any two controls $\nu_1, \nu_2 \in V_t$, control $\nu_3 \in V_t$ exists, which from the convex combination $x(t) = \xi = \kappa\xi_1 + (1-\kappa)\xi_2$ of initial points ξ_1 and ξ_2 gives the convex combination transfer step $f_t(\xi, \nu_3) = \kappa f_t(\xi_1, \nu_1) + (1-\kappa) f_t(\xi_2, \nu_2)$, and consequently at time $t+1$ leads to the corresponding convex combination of the end points. (ii) The term in the objective function satisfies the convexity requirement (5.1). A sufficient condition for this is that the cost term is of the following type:

$$c_t(\xi,\nu) = c_t^1[f_t(\xi,\nu)] + c_t^2[\xi + f_t(\xi,\nu)], \tag{5.5}$$

where c_t^1 and c_t^2 are any convex functions on $\mathbb{R}^n$.

For problems with state-control constraints Condition 5.1 has inequalities (5.3) and equalities (5.4). In this case, we can repeat the same argument as above, with the additional requirement that constraints (5.3) and (5.4) are satisfied by the control $\nu_3 \in V_t$.

For the primal problem, for each $t = 0, \ldots, T-1$, the attainability set $A_{t+1} \in \mathbb{R}^n$ is defined as the set of states $x(t+1)$, reachable by trajectories satisfying all the constraints of the problem for all time steps $s = 0, 1, \ldots, t$. Condition 5.1 implies that all the attainability sets are convex. Condition 5.1 also implies that, if the trajectories $x_1 = [x_1(t); t = 0, 1, \ldots, T]$, and $x_2 = [x_2(t); t = 0, 1, \ldots, T]$ are feasible, then the convex combination trajectory $x_3 = \kappa x_1 + (1-\kappa)x_2 = [\kappa x_1(t) + (1-\kappa)x_2(t); t = 0, 1, \ldots, T]$ is feasible too, for any $\kappa \in [0,1]$, and satisfies the convexity requirement for the objective function.

Condition 5.1 can be viewed as a natural generalization to a global optimum of the customary assumption of directional convexity made in the derivation of the discrete maximum principle for a local optimum in the literature, see Theorem 4.2 of [5]. Following the treatment of the reference, and applying the notation of this paper, we recall the directional convexity condition for problems without state constraints.

Condition 5.2 (Convexity for an optimum, local in the state space)**.** For every $t = 0, 1, \ldots, T-1$, for every $\xi \in \mathbb{R}^n$, for every $\nu_1, \nu_2 \in V_t$, and for every $\kappa \in [0,1]$, there is a $\nu_3 \in V_t$ such that:

$$c_t(\xi,\nu_3) \le \kappa c_t(\xi,\nu_1) + (1-\kappa)c_t(\xi,\nu_2), \text{ and} \tag{5.6}$$

$$f_t(\xi,\nu_3) = \kappa f_t(\xi,\nu_1) + (1-\kappa)f_t(\xi,\nu_2). \tag{5.7}$$

By choosing $\xi_1 = \xi_2 = \xi$ in Condition 5.1 we see that it implies Condition 5.2.

6. Concluding remarks

In this article we examine discrete time optimal control problems by applying the global approach via Lagrange duality theory. The main contribution is the generalized convexity Condition 4.1, which, together with a constraint qualification, is sufficient for strong Lagrange duality, and saddle point conditions for optimality for these problems. The conditions are sufficient for global optimality, and state that there exists for the problem a normal Lagrange multiplier, the Lagrangian is optimized at the primal optimal solution, the multiplier is an optimal solution to the dual problem, and the complementary slackness conditions are satisfied.

The discrete maximum principle of Pontryagin is derived in a straight-forward way from this result. The maximum principle is a necessary, but not sufficient, condition for optimality, and consists of three parts: the Euler-Lagrange equation, the Hamiltonian maximization condition and the conditions of complementary slackness. The optimum condition for the global Lagrangian in the duality result gives the two first of these, and the conditions of complementary slackness are the same in both theorems. The Euler-Lagrange equation follows from the fact that the global Lagrangian attains an optimum with respect to the state variables at the primal optimal point. It is obtained first as an optimum condition, and then, assuming differentiability, as a difference equation. The condition for the Hamiltonian follows from the optimum condition for the global Lagrangian with respect to the control variables.

In the literature the DMP is derived by a local approach. Both the development in [11, 12, 14, 15, 16, 25, 27, 28], the detailed analysis in [2], and the derivation of the principle from general results covering wide classes of mathematical programming problems in [5, 18], apply a local approach. Also the totally different general development of the maximum principle for arbitrary time scales, including discrete time, in [3, 4] follows a local approach. Differentiability with respect to the state is assumed from the beginning, and this is sufficient for the weak form of the principle, with Euler-Lagrange equation, conditions of complementary slackness, and a gradient condition for the Hamiltonian. By making a proper assumption of convexity, directional convexity *e.g.*, the optimum condition for the Hamiltonian is obtained.

Also in [28], the strong Lagrange duality result for a discrete-time optimal control problem is derived from the conditions of DMP, obtained trough a local approach, by assuming that the dynamic equation, and the state equality constraints are linear, and the inequality constraints are convex. This case is also covered by general Lagrange duality results for mathematical programming problems in the literature, see [1].

In comparison with the local approaches to the maximum principle, the global approach applied here is based on a very restrictive convexity condition, does not require differentiability, and consequently can be applied to problems with non-smooth data. In addition to the conditions of the DMP it furnishes the optimum condition for the overall Lagrangian, characterizes the Lagrange multiplier as an optimal dual solution, and gives an exact convergence criterion for primal-dual solution methods. In the local approaches only differentiability is assumed in order to obtain the principle in the weak form, and a suitable assumption of convexity gives the strong form. The different local approaches are thus much more widely applicable. The generalized convexity Condition 4.1 is satisfied in problems with a strong controllability property, with a convex objective function. Linear structure is also sufficient. The condition is, however, not needed in the derivation of the maximum principle from the strong duality result, i.e. strong duality alone is sufficient for the maximum principle. Finally, we note that the approach of this paper is very similar to the approch to continuous-time optimal control problems with linear dynamics applied in [10].

The statement of the primal problem, and the conditions of the maximum principle, Theorem 4.7, have a beautiful symmetry between the primal and dual forms of the problem. For linear problems this symmetry is known as the duality relations of dynamic linear programming, see Ivanilov and Propoi [19].

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, Nonlinear Programming: Theory and Algorithms, John Wiley & Sons, New York (2006).

[2] V.G. Boltyanskii, Optimal Control of Discrete Systems, John Wiley & Sons, New York, Toronto, Ontario (1978).

[3] L. Bourdin and E. Trélat, Pontryagin maximum principle for finite dimensional nonlinear optimal control problems on time scales. *SIAM J. Control Optim.* **51** (2013) 3781–3813.

[4] L. Bourdin and E. Trélat, Optimal sampled-data control, and generalizations on time-scales. *Math. Control Related Fields* **6** (2016) 53–94.

[5] M.D. Canon, C. Cullum and E. Polak, Theory of Optimal Control and Mathematical Programming, McGraw-Hill, New York (1970).

[6] F.C. Clarke, Optimization and Nonsmooth Analysis, John Wiley & Sons, New York (1983). Republished by Université de Montréal, Montréal (1990).

[7] K.H. Elster and R. Nehse, Optimality conditions for some nonconvex problems, in Optimization Techniques, Part 2, edited by K. Iracki, K. Malanowski, S. Walukiewicz, Springer, Berlin (1980) 1–9.

[8] F. Giannessi, Theorems of the alternative and optimality conditions. *J. Optim. Theory Appl.* **42** (1984) 331–365.

[9] F. Giannessi, Theorems of the alternative for multifunctions with applications to optimization: general results. *J. Optim. Theory Appl.* **55** (1987) 233–256.

[10] W.W. Hager and S.K. Mitter, Lagrange duality theory for convex control problrms. *SIAM J. Control Optim.* **14** (1976) 843–856.

[11] H. Halkin, Optimal control for systems described by difference equations, in Advances in Control Systems: Theory and Applications, Academic Press, New York (1964) 173–196.

[12] H. Halkin, A Maximum principle of the Pontryagin type for systems described by nonlinear difference equations. *J. SIAM Control* **4** (1966) 90–111.

[13] M. Hayashi and H. Komiya, Perfect duality for convexlike programs. *J. Optim. Theory Appl.* **38** (1982) 179–189.

[14] J.M. Holtzman, Convexity and the maximum principle for discrete systems. *IEEE Trans. Autom. Control* **11** (1966) 30–35.

[15] J.M. Holtzman, On the maximum principle for nonlinear discrete-time systems. *IEEE Trans. Autom. Control* **11** (1966) 273–274.

[16] J.M. Holtzman and H. Halkin, Directional convexity and the maximum principle for discrete systems. *J. SIAM Control* **4** (1966) 263–275.

[17] T. Illés and G. Kassay, Theorems of the alternative and optimality conditions for convexlike and general convexlike programming. *J. Optim. Theory Appl.* **101** (1999) 243-257.

[18] A.D. Ioffe and V.M. Tihomirov. Theory of Extremal Problems, North-Holland, Amsterdam (1979).

[19] Y.P. Ivanilov and A.I. Propoi, Duality relations in dynamic linear programming. *Autom. Remote Control* **34** (1973) 1945–1952.
[20] V. Jeyakumar, Convexlike alternative theorems and mathematical programming. *Optimization* **16** (1985) 643–652.
[21] Z. Li and G. Chen, Lagrangian multipliers, saddle points, and duality in vector optimization of set-valued maps. *J. Math. Anal. Appl.* **215** (1997) 297–316.
[22] D.G. Luenberger, Optimization by Vector Space Methods. John Wiley & Sons, New York (1969).
[23] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation. Springer-Verlag, Berlin (2006).
[24] R. Nehse, Some general separation theorems. *Math. Nachr.* **84** (1978) 319–327.
[25] J.A. Ortega and J.R. Leake, Discrete maximum principle with state constrained control. *J. SIAM Control* **15** (1977) 984–990.
[26] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko, The Mathematical Theory of Optimal Processes. John Wiley & Sons, New York (1962).
[27] A.I. Propoi, The maximum principle for discrete systems. *Autom. Remote Control* **26** (1965) 1169–1177.
[28] A.I. Propoi, Discrete control problems with phase constraints. *Zh. Vychisl. Mat. Mat. Fiz.* **12** (1972) 1128–1144.
[29] R.T. Rockafellar, Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs. *SIAM J. Control Optim.* **27** (1989) 1007–1025.
[30] R.T. Rockafellar, Lagrange multipliers and optimality. *SIAM Rev.* **35** (1993) 183–238.
[31] R.T. Rockafellar and R.J.-B. Wets. Variational Analysis. Springer-Verlag, Berlin (1998).
[32] S.P. Sethi and G.L. Thompson. Optimal Control Theory, 2nd edn. Kluwer Academic Publishers, Boston, MA (2000).
[33] S. Simons, Abstract Kuhn-Tucker theorems. *J. Optim. Theory Appl.* **58** (1988) 147–152.
[34] E.V. Tamminen, Sufficient conditions for the existence of multipliers and Lagrangian duality in abstract optimization problems. *J. Optim. Theory Appl.* **82** (1994) 93–104.
[35] G.J. Zalmai, Saddle-poit-type optimality conditions and Lagrangian-type duality for a class of constrained generalized fractional optimal control problems. *Optimization* **44** (1998) 351–372.