

# NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION

erical Functional Analysis and  
Journal to your library for

NO POSTAGE  
NECESSARY  
IF MAILED  
IN THE  
UNITED STATES

AIL  
ORK NY  
E

illilllll

## al Analysis

### Publication

Newark

xamines the development and operator-theoretic methods in analysis, optimization, control, and interaction and unification of methods to provide insight and applications in the natural sciences. Numerical Functional Analysis and Optimization provides a comprehensive survey and synthesis of the major topics, book reviews, and topical topics.

Editor-in-Chief

M.Z. NASHED

Department of Mathematical Sciences  
University of Delaware  
Newark, Delaware 19716, U.S.A.

Editorial Board

- V. BARBU, *Universitatea "Al. I. Cuza", Seminarul Matematic, 6600 Iasi, Romania*  
A. BENSOUSSAN, *University of Paris IX-Dauphine, Place du Maréchal-de-Lattre-de-Tassigny, 75026 Paris, France*  
E. K. BLUM, *University of Southern California, Los Angeles, California 90007*  
J. H. BRAMBLE, *Texas A&M University, College Station, Texas 14853*  
B. H. BROSWOSKI, *Fachbereich Mathematik, Johann Wolfgang Goethe Universität Frankfurt/Main, 6 Frankfurt a. M., Germany*  
F. E. BROWDER, *Rutgers University, New Brunswick, New Jersey 08903*  
P. L. BUTZER, *Rhein.-Westf. Tech. Hochschule Aachen, Lehrstuhl A für Mathematik, Templergraben 55, D-51 Aachen, Germany*  
E. W. CHENEY, *The University of Texas, Austin, Texas 78712*  
Z. CIESIELSKI, *Instytut Matematyczny Polskiej Akademii Nauk, Oddział w. Gdańsk, ul. Abrahamia 18, 81-825 Sopot, Poland*  
V. F. DEMYANOV, *S. Petersburg State University, 10 Liniya dom 33, St. Petersburg 199178, Russia*  
H. FUJITA, *Meiji University, Tama-ku, Kawasaki-shi 214, Japan*  
R. P. GILBERT, *University of Delaware, Newark, Delaware 19716*  
R. GLOWINSKI, *University of Houston, Houston, Texas 77204*  
C. W. GROETSCH, *University of Cincinnati, Cincinnati, Ohio 45221*  
K.-H. HOFFMANN, *Lehrstuhl für Angewandte Mathematik, Technische Universität München, Dachauerstr. 9a, 80335 München, Germany*  
R. B. HOLMES, *M.I.T. Lincoln Laboratory, Lexington, Massachusetts 02173*  
H. B. KELLER, *California Institute of Technology, 101-50, Pasadena, California 91125*  
M. A. KRASNOSELSKII, *Institute of Control Sciences, Profsoyuznaya Street 6 Moscow 117342, Russia*  
I. MAREK, *Katedra Numerické Matematiky na Matematicko-Fyzikální Fakultě, Univerzity Karlovy, Malostranské Náměstí č 25, 11800 Praha 1, Czech Republic 91125*  
V. A. MOROZOV, *Moscow State University, 117234 Moscow, Russia*  
U. MOSCO, *Istituto Matematico, Università di Roma, Città Universitaria, 00100 Rome, Italy*  
J. T. ODEN, *The University of Texas, Austin, Texas 78712*

(continued)

NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION

Editorial Board (*continued*)

W. V. PETRYSHYN, *Rutgers University, New Brunswick, New Jersey 08903*

B. T. POLJAK, *Institute of Control Sciences, ul. Profsoyuznaya, 81 Moscow, GSP-1, Russia*

P. M. PRENTER, *Colorado State University, Fort Collins, Colorado 80523*

W. C. RHEINBOLDT, *Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*

Y. SAKAWA, *Kinki University, Uchita, Naga-gun, Wakayama 649-64, Japan*

I. SINGER, *INCREST and Institute of Mathematics, Bd Pačii 220, Sector 7, Bucharest 79622, Romania*

G. STRANG, *Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

F. STUMMEL, *Fachbereich Mathematik, Johann Wolfgang Goethe Universität, Robert-Mayer-Strasse 6-10, 6 Frankfurt a. M., Germany*

J. SZABADOS, *Mathematical Institute of the Hungarian Academy of Sciences, Reáltandoda u. 13-15, 1053 Budapest, Hungary*

R. TEMAM, *Mathématique, Université de Paris—Sud, Centre d'Orsay, 91405 Orsay Cedex, France*

G. M. VAINIKKO, *Helsinki University of Technology, Institute of Mathematics, 02150 Espoo, Finland*

E. ZEIDLER, *Sektion Mathematik, Universität Leipzig, Augustusplatz, 701 Leipzig, Germany*

Published and referenced as:

**On Consistent Regularities of Control and Value Functions,**  
Melnik, V.N., **Numer. Funct. Anal. and Optimiz.**, 18(3&4), 1997,  
401-426.

# ON CONSISTENT REGULARITIES OF CONTROL AND VALUE FUNCTIONS

(Roderick) V.Nick Melnik

Centre for Industrial & Applied Mathematics, School of Mathematics

University of South Australia, SA 5095, Australia

E-mail: matvnm@lv.levels.unisa.edu.au

## Abstract

In this paper we deal with nonsmooth optimal control problems in the case when the control is allowed to be a discontinuous function. We analyse smoothness assumptions on an adjoint process in deterministic and stochastic cases. Possibilities of steep generalized space-gradients of the adjoint function imply the necessity of an approximation of the Hamiltonian. The key question of such an approximation is a relationship between the control and the value function. Under quite general assumptions it is proved that the performance measure for the original process is determined by the control function with possible discontinuities.

**Key words:** Hausdorff topological spaces, probabilistic weight functions, discrete control, approximations of the Hamiltonian, coupling regularities.

## 1 Introduction: mathematical models in topological spaces and ramifications from initial conditions given approximately.

Many problems in applications of mathematics can be formulated in the following generic form. Let  $H(u, v)$  be a nonlinear operator defined on the product of the two topological spaces  $(\mathcal{T}, \tau)$  and  $(\mathcal{X}, h)$  (which may coincide) with values in a topological space  $(\mathcal{E}, \omega)$ , where  $\mathcal{T}$ ,  $\mathcal{X}$ ,  $\mathcal{E}$  are sets that form space-supports, whereas  $\tau$ ,  $h$ , and  $\omega$  are systems of their subsets that define topologies of the corresponding spaces. Let us denote the set of all ordered pairs  $\{(t, x) : t \in \mathcal{T}, x \in \mathcal{X}\}$  by  $\Omega_0^0 = \mathcal{T} \otimes \mathcal{X}$ . We assume that it is known that for a given (computed, estimated, observed, etc) value  $v = v_0 \in \Omega_0^0$  the equation  $H(u, v_0) = 0$  has a solution  $u_0^\epsilon \in \Omega_0^0$ . That is, for arbitrary small  $\epsilon > 0$  the following condition holds:

$$H(u_0^\epsilon, v_0) \rightarrow 0 \text{ when } \epsilon \rightarrow 0^+. \quad (1.1)$$

Assume that the limit of  $u_0^\epsilon$  exists and belongs to  $\Omega_0^0$  when  $\epsilon$  tends to zero from the right,

$$\lim_{\epsilon \rightarrow 0^+} u_0^\epsilon = u_0. \quad u_0 \in \Omega_0^0, \quad (1.2)$$

In this case, from the analysis of (1.1), (1.2), a natural question which arises is to find such solutions of the equations

$$H(u, v) = 0 \quad (1.3)$$

for values of the parameter  $v$  that are close to  $v_0$  in some sense and which “branch” from  $u_0^\epsilon$  when  $\epsilon \rightarrow 0^+$ . The solution of (1.3) depends on the choice of the non-linear operator  $H$  as well as on the parameter  $v$ , that may be a functional, a function, or a numerical parameter. From the definition of topology, it follows that  $\emptyset \subset \tau$ ,  $\emptyset \subset h$ , and  $\emptyset \subset \omega$ . If we further assume that  $u \in \mathcal{E}_1$ ,  $v \in \mathcal{E}_2$ , where  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are given topological spaces (not excluding the possibility  $\mathcal{E}_1 \equiv \mathcal{E}_2 \equiv \Omega_0^0$ ), this allows the possibility of a separation of the two elements  $u$  and  $v$  that in principle belong to the same set  $\Omega_0^0$ . Of course, as a partial case, it is possible that  $\mathcal{E}_1 \equiv \mathcal{T}$  and  $\mathcal{E}_2 \equiv \mathcal{X}$ . If it is a-priori given that at least one of these topological spaces is complete (typically this assumption is imposed on  $\mathcal{T}$ ), then the operator  $H$  can be specified in a more definite way using a generalization of the theorem on continuity of complex functions.

**Theorem 1.1** [17] *If there exists a continuous mapping  $f : \mathcal{E}_1 \rightarrow \mathcal{E}_2$  and a continuous mapping  $\phi : \mathcal{E}_2 \rightarrow \mathcal{E}$  then the mapping  $F = \phi(f) : \mathcal{E}_1 \rightarrow \mathcal{E}$  is also continuous.*

Theorem 1.1 provides a technical convenience that allows us to assume the possible existence of a continuous mapping  $F : \mathcal{T} \otimes \mathcal{X} \rightarrow \mathcal{E}$ . Hence, the non-linear operator  $H$  in (1.3) may also, in principle, be defined by a continuous mapping. In this case, a constructive element in the determination of all solutions of (1.3) that branch from  $u_0^\epsilon$  in a range of  $v$  close to  $v_0$ , is introduced by a specification of topologies in the spaces  $\mathcal{E}_1$ ,  $\mathcal{E}_2$  and  $\mathcal{E}$ . The idea of such specifications in a rigorous mathematical sense came from the fundamental work of Newton [26] who considered the determination of branching solutions of (1.3) that tend to  $v_0$  when  $u \rightarrow u_0^\epsilon$  for the special case when

$$\lim_{\epsilon \rightarrow 0^+} \frac{\partial H}{\partial y}|_{(u_0^\epsilon, v_0)} = 0 \quad (1.4)$$

assuming that  $H$  can be expanded in a series of positive integral powers of  $(u - u_0^\epsilon)$  and  $(v - v_0)$  when  $\epsilon \rightarrow 0^+$ . Under these conditions Newton's technique consists of seeking the solution of (1.3) in the form

$$v = v_0 + \alpha_1(u - u_0^\epsilon)^{\epsilon_1} + \alpha_2(u - u_0^\epsilon)^{\epsilon_2} + \dots \quad (1.5)$$

where  $\epsilon_1, \epsilon_2, \dots$  is an increasing sequence of rational numbers. The basic idea of the Newton polygon method (which is often referred to as the parallelogram method) is intrinsically connected with the geometric interpretation of the problem. It is also known that fractional powers  $\epsilon_1, \epsilon_2, \dots$  have a finite common denominator and that the Newton polygon method can be used to find all continuous solutions such that  $v \rightarrow v_0$  provided  $u \rightarrow u_0^\epsilon$  and (1.4) holds [38]. From the topological point of view we implicitly assume that any point in  $\mathcal{E}_1$ ,  $\mathcal{E}_2$ , and  $\mathcal{E}$  is a closed set, and a finite set of such sets is again closed<sup>1</sup>. Then the problem (1.3) becomes reducible to the determination of possible values of

$$\alpha_1, \epsilon_1, \alpha_2, \epsilon_2, \dots \quad (1.6)$$

in (1.5). The principal difficulty in the solution of such a reduced problem in the general case resides in the fact that the terms in the sequence (1.6) are not independent of the parameter  $\epsilon$ . Thus, mathematically speaking we need a reinforcement of the first axiom of separability to be able to determine  $v$  from (1.5) with an a-priori given accuracy. To overcome this difficulty we may use the Hausdorff axiom of separability in the topological spaces  $\mathcal{E}_1$ ,  $\mathcal{E}_2$ ,  $\mathcal{E}$ . Since the Hausdorff property is a hereditary property<sup>2</sup>, it implies the possibility of the assumption

$$\lim_{\epsilon \rightarrow 0} H(u_0^\epsilon, v) = H(u_0, v_0), \quad (1.7)$$

where  $u_0 \in \mathcal{E}_1$ . Then by the standard change of variables

$$u = u_0 + x, \quad v = v_0 + y \quad (1.8)$$

the problem of “branching” from the point  $(u_0, v_0)$  is reduced to the problem of “branching” from the origin  $(0, 0)$ . Hence, in principle one can seek such solutions of the equation  $\tilde{H}(x, y) = 0$ , (here  $\tilde{H}(x, y) = H(u, v)$ ) that branch from the zero of  $\mathcal{E}_1$  for values of  $v$  that are close to the zero of  $\mathcal{E}_2$ .

<sup>1</sup>thus, in such topological spaces the first axiom of separability holds

<sup>2</sup>any subspace of a Hausdorff space is again Hausdorff

Also, due to (1.1), (1.2) it is usually assumed that  $\tilde{H}(0,0) = 0$ . The idea of such branching-from-zero solutions<sup>3</sup> is intrinsic to topological spaces equipped with the Hausdorff property. Indeed, in topological spaces with the Hausdorff separability axiom, we can claim simultaneous realization of two equalities, satisfied exactly, that in general connect three topological spaces. These equalities are introduced by (1.3) and by a relationship between a-priori given  $H$ ,  $u_0$ , and  $v_0$

$$H(u_0, v_0) = 0, u_0 \in \mathcal{E}_1, v_0 \in \mathcal{E}_2, H \in \mathcal{E}. \quad (1.9)$$

Although this approach has proved to be very useful in many applications, it has virtually inherited the classical mechanics idealization of temporal evolution as a motion of phase points along phase point trajectories. There are a number of situations where the validity of such an idealization is limited. On the whole, arbitrarily close initial conditions for models of the general type (1.3), (1.9) can give rise to exponentially diverging trajectories exhibiting qualitatively distinct physical behaviour [24]. In the general case, (1.3) may provide an approximation to sequences like (1.6), but it does not imply that a specified sequence may provide an appropriate approximation to the solution of the model (1.3). As a result, the quality of approximations depends decisively on a-priori information on  $\epsilon$  and the specification of topologies of the corresponding spaces.

In many applications a small numerical parameter  $\lambda$  plays the role of the variable  $v$  in the model (1.3). In these cases it is natural to attempt to represent *small solutions* of (1.3) in the form of series of integral (or fractional) powers of  $\lambda$ . However, if the uniqueness of the solution that branches from the origin is not part of the information given a-priori, this approach encounters serious mathematical difficulties. When the series is sought as integral powers of  $\lambda$  then the search for a majorant series is *problem specific*, whereas a fractional-power-series representation does not guarantee that all small solutions may be taken into consideration by such a form. Mathematically, these difficulties can be gradually removed by applying the Lyapunov-Schmidt theory in the topological spaces equipped with the Hausdorff axiom. Typically, such an approach is limited by either stationary models or by mathematical models, in which time appears as a pure deterministic category, being a parameter that labels states of the system [24]. Such mathematical formalization allows us to reduce the original mathematical model (1.1), (1.2) to a model with equalities satisfied exactly in (1.3) and (1.9) which in general are interpreted with the probability 1. This leads to a certain reversibility of mathematical models in time<sup>4</sup> that allows us to use a set-theoretic approach for the construction of mathematical models. The validity of the reduction of the mathematical model (1.1), (1.2) to the model (1.3), (1.9) is essentially based on a postulate proposed by J.W. Gibbs.

**Postulate 1.1** [10, 21] *An appropriate description of a macroscopic system in thermodynamic equilibrium may be given by certain probability measures on the phase space of the system.*

The main difficulty in a rigorous justification of this postulate lies with the fact that conditions for the measure stability are not independent from the specification of the operator  $H$  (which can be given only approximately) and the definition of topologies in the spaces  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}$ . At the same time, the possibility of translation (1.8) is defined by the conditions of measure stability. Indeed, (1.9) is a limiting case in which measure stability is implicitly assumed by equipping of the space topologies a-priori with the Hausdorff property. However, instead of (1.8), it is more realistic to consider a perturbed translation

$$u = u_0^\epsilon + x, \quad v = v_0 + y_\lambda, \quad (1.10)$$

where  $\lambda > 0$  is typically a small parameter. As a result, without some a-priori information on  $\epsilon > 0$  and  $\lambda > 0$ , the model (1.3), (1.9) cannot be justified rigorously. No matter how small the parameters  $\epsilon$  and  $\lambda$  are assumed to be, the translation (1.10) is always dependent on properties of the topology which is a-priori introduced in  $\Omega_0^0$ . If such a topology is assumed to be Hausdorff, then a transition from the point  $(0,0)$  to an arbitrary fixed point  $(u,v)$  in the space  $\Omega_0^0$  can not necessarily be described by a continuous governing operator  $H : \Omega_0^0 \rightarrow \mathcal{E}$ . On the other hand, if the governing operator  $H$  is continuous then the availability of the Hausdorff property in  $\Omega_0^0$  is questionable ([33], p.309). In the final analysis, simultaneous assumptions of continuity of the operator  $H$  and the Hausdorff property

<sup>3</sup>referred to as “small solutions” in the Lyapunov-Schmidt theory (see [38] and references therein)

<sup>4</sup>which physically speaking implies the invariance of densities along trajectories [20]

of the topological space on which it is defined cannot be justified rigorously for any mathematical model. The complexity of the possible justification was realized at the time when the groundwork in statistical mechanics was laid. In particular, J.W. Gibbs wrote that “we are rarely justified in excluding the considerations of the antecedent probability of the prior events” [9].

If we use the Hausdorff property as an a-priori given argument it is possible to assign the same probability, 1, to two non-identical events, that eventually allows us to assume that the governing operator  $H$  associated with these events may be continuous. In theory, such an extension of the continuity assumption from classical to statistical mechanics<sup>5</sup> still preserves time-reversibility of the constructed mathematical models. However it does not take into account the fact that the definition of the non-linear operator  $H$  in the model (1.3) is an inherently recursive approximation that is not independent of the definition of the topology. This mathematical idealization has turned out to be very useful in many practical applications including areas of statistical mechanics and classical irreversible thermodynamics. The success of applications of mathematical tools in statistical mechanics has been grounded on postulate 1.1, whereas the success of mathematical modelling in thermodynamics has been determined by the local-equilibrium hypothesis.

**Hypothesis 1.1** *Any system out of equilibrium is assumed to depend locally on the same set of variables as when it is in equilibrium.*

During recent years experimental and theoretical physicists have shown a growing interest in going beyond the limits of this hypothesis (see [25,13] and references therein). Associated mathematical challenges require relaxation of some mathematical assumptions typically made when classical approaches are applied even in the non-relativistic case. This paper deals with such relaxations, bearing in mind the problems of optimal control theory.

The remainder of the paper is organized as follows:

- In Section 2 we analyse mathematical definitions of controlled dynamic rules in deterministic and probabilistic cases when the topology of the underlying spaces is equipped with the Hausdorff property. We argue that if the control is allowed to be a discontinuous function, then a *sequential regularization procedure* is necessary to ensure system stability.
- In Section 3 we show why the stability of the system may be violated when passing from an optimal control problem to the associated Hamilton-Jacobi-Bellman equation. It is emphasized that an appropriate approximation of the Hamiltonian of a controlled dynamic system, considered in a Hausdorff topological space, is not necessarily provided by a continuous mapping.
- Section 4 deals with the case when both control and the value function are allowed to come from the non-Euclidean Banach space  $L^1(Q_0^0)$  where  $Q_0^0$  is an approximation of  $\Omega_0^0$ . Under quite general assumptions it is shown that, in this case, control can always be chosen arbitrarily close to a specified time-averaged performance measure. The main result has been derived without reference to a possible continuity of the value function. The proposed proof is a properly constructed algorithm which ensures the measure stability in  $Q_0^0$ .
- In Section 5 we discuss the validity of assumptions made for the derivation of theorem 4.1 and logical issues related to this.
- Conclusions and future directions are addressed in Section 6.

## 2 Deterministic and stochastic models of optimal control in Hausdorff spaces.

As above, let  $\Omega_0^0$  be a set defined as  $\mathcal{T} \otimes \mathcal{X}$ . We assume that the evolution of a dynamic system takes place in  $\Omega_0^0$  and such evolution can be controlled to achieve a certain goal. Let us denote a space-time domain of definition for admissible controls as  $\Sigma$ , and the set of all admissible controls as  $\mathcal{U}$ . We assume that the state dynamics of the system can be effectively defined by the two coupled mappings:

$$f : \Omega_0^0 \otimes \mathcal{U} \rightarrow \mathcal{B}_1 \quad (2.1)$$

---

<sup>5</sup>according to J.W.Gibbs “conservation of extension in phase”

and

$$u : \Sigma \rightarrow \mathcal{U} \subseteq \mathcal{B}_2, \quad (2.2)$$

where  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are assumed to be Hausdorff spaces. The topological space  $\mathcal{B}_1$  plays the role of the space  $\mathcal{E}$  introduced in Section 1.

If  $\mathcal{B}_1$  is a Hilbert topological space, then it is possible to introduce an energetic inner product in  $\mathcal{B}_1$  with respect to a linear, symmetric, and strongly monotone operator  $L : D(L) \subseteq \mathcal{B}_1 \rightarrow \mathcal{B}_1$  as  $(Lu|v) \forall u, v \in D(L)$ . Hence, we conceptually assume that  $\mathcal{B}_1$  is such that there exists an admissible sequence  $u_k \in D(L)$  for all  $u \in \mathcal{B}_1$ , and the limit  $\lim_{k \rightarrow \infty} (Lu_k|v_k)$  does not depend on the chosen admissible sequence [40].

The topological space  $\mathcal{B}_1$  from (2.1) can be viewed as a generalization of such energetic spaces on spaces equipped with the Hausdorff property. If  $\Omega_0^0$  is a topological space with a topology  $\pi$ , then in order to specify the model (2.1), (2.2) it is important to have some a-priori information on the possibility of a closure of  $\Omega_0^0$  or its subset in some rigorous mathematical sense. If  $\Omega_0^0$  is a compact topological space, then equipping  $\Omega_0^0$  with the Hausdorff property provides a rigorous mathematical basis for a closure procedure on the basis of the Heine-Borel lemma arguments. In this case  $\Omega_0^0$  can be imbedded into a “richer” topological space with the Hausdorff property. Since compactness is an inner and integral property of Hausdorff spaces, it can always be used as a mathematical argument for a closure procedure, no matter how “rich” the actual topology  $\pi$  is, by requiring appropriate regularity assumptions on the mappings (2.1), (2.2). Such assumptions are typically based on the continuity arguments which in many cases allow us to ensure homeomorphic properties of the constructing mappings:

**Theorem 2.1** *If there exists a continuous one-to-one mapping of a compact set  $X$  into a Hausdorff space  $Y$  then such a mapping is a homeomorphism.*

If  $\Omega_0^0$  is not a compact set then, in addition to the continuity arguments, one needs an exact specification of a point or a surface in a topological space to ensure that the mathematical model is well-posed. This leads to the main difficulty in gathering a-priori information about the problem itself, for in the general case such a specification is not topology-independent and is an inherently approximating procedure. Therefore, an a-priori equipping of the topology with the Hausdorff property, followed by the imbedding of the obtained Hausdorff space into a “richer” Hausdorff space, cannot ensure that the model with an a-priori given continuous mapping between the two spaces is well-posed. Coupling of the mappings (2.1) and (2.2) precludes a specification of topological properties of the sets  $\Sigma$  and  $\Omega_0^0$  with deterministic certainty, as well as with the probability exactly 1. The reverse statement is also true, and unknown topological properties of  $\Sigma$  and  $\Omega_0^0$  cause difficulties in the definition of the mappings  $f$  and  $u$  which are approximate by their nature.

However, the possibility of either of the following two cases always exists

$$\Sigma \subseteq \Omega_0^0 \text{ or } \Omega_0^0 \subseteq \Sigma, \quad (2.3)$$

and this allows an effective construction of models in mathematical applications.

If  $\varpi(\Sigma)$  and  $\varpi(\Omega_0^0)$  are powers of these sets, then we have four logical choices to be considered:

- $\varpi(\Sigma) = \varpi(\Omega_0^0)$ ;
- $\varpi(\Sigma) \succ \varpi(\Omega_0^0)$ ;
- $\varpi(\Sigma) \prec \varpi(\Omega_0^0)$ ;
- In neither of these sets there is a part that is equivalent to the other set.

The first case is a consequence of the Cantor-Bernstein theorem, which requires a one-to-one mapping of  $\Sigma$  into a subset of  $\Omega_0^0$ , and a one-to-one mapping of  $\Omega_0^0$  into a subset of  $\Sigma$ . The last case is usually excluded in the set theory by Zermelo-type arguments, whereas the two remaining cases are typical in the majority of applications. Each specific logical choice provides some a-priori constraints on the resulting mathematical models. Such constraints define the limitations of associated computational models from which particular algorithms can be derived (see also [37]).

Rigorously speaking, mappings (2.1) and (2.2) cannot be defined independently of each other for any a-priori choice of  $\mathcal{B}_1$ ,  $\mathcal{B}_2$ , and the set inclusions (2.3) are responsible for the dependency between their topologies. As the first example, we consider a class of nonsmooth deterministic optimal control problem.

## 2.1. Deterministic control with a-priori given topology and initial conditions.

We restrict the theoretical generality of the problem, by bounded, Lebesgue measurable in time  $U$ -valued control functions. We consider the set of all controls that can be defined as  $u_{\beta_1}^{\beta_2}(\cdot, \cdot) \in \mathcal{B}_2 \equiv L^\infty([t, T]; U)$ , where  $t_0 \leq t \leq T$  (with the possibility  $T \rightarrow \infty$ ), and  $U$  is a subset in  $\mathcal{B}_2$  defined as the control space, whereas  $\beta_1, \beta_2$  are a-priori specified topologies in  $\mathcal{B}_1, \mathcal{B}_2$  respectively. The dot-arguments reflect the dependency of the control on both time and state-space. As a partial case the topological spaces  $\mathcal{B}_1$  and  $\mathcal{B}_2$  may have the arithmetic Euclidean structure, that is

$$\mathcal{B}_1 \equiv R^n, \quad \mathcal{B}_2 \equiv R^m. \quad (2.4)$$

The set of admissible controls as a subset of  $\mathcal{B}_1$  depends on the type of required constraints, and as soon as such constraints are specified we denote the set of all admissible controls by  $\mathcal{U}$ .

Given the above assumptions and notations, the classical deterministic problem of optimal control can be formulated in the following way. We assume that the process of evolution of a controlled dynamic system in  $\Omega_0^0$  can be appropriately described by a differential equation in terms of a space-time function  $x(\cdot)$ , the initial conditions and state constraints, that is

$$\frac{dx(t)}{dt} = \tilde{f}(t, x, u_{\beta_1}^{\beta_2}) \text{ a.e. in } \mathcal{T} = (t_0, T), \quad \text{for } x \in \mathcal{X} \subseteq \mathcal{B}_1 \quad (2.5)$$

$$x(t_0) = x_0 \in \mathcal{X}, \quad u \in \mathcal{U}. \quad (2.6)$$

Our goal is to minimize the functional

$$J(u_{\beta_1}^{\beta_2}) = \int_{t_0}^T f_0(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, x(\tau))) d\tau + g(T, x(T)) \rightarrow \min \quad (2.7)$$

on the set  $\mathcal{U}$  of all admissible controls, where  $f_0$  and  $g$  are given functions (the running and terminal costs respectively).

The model (2.5), (2.6), (2.7) may provide an approximation to the coupled mapping (2.1), (2.2). However, additional constraints on the set of admissible controls  $\mathcal{U}(\Sigma)$  are required to construct such an approximation. For a given set of initial data  $(t, x)$ , these constraints imply some additional a-priori assumptions on the mapping  $f$ , and typically require the a-priori equipping of the topological product  $\Omega_0^0 \otimes \mathcal{U}$  with the Hausdorff property. Then, the chain of logical steps for the model construction is typically based on the assumption of the possibility of the precise definition of initial conditions (2.6) and the existence of a continuous mapping between two subsets in  $\Omega_0^0 \otimes \mathcal{U}$ , one of which is a compact set allowing the use of the arguments on theorem 2.1.

In many applications these assumptions appear naturally, especially in those problems when the dynamic motion can be relatively easily represented by continuous phase trajectories. For example, this is usually the case when the topological spaces  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are finite dimensional Euclidean spaces defined by (2.4), the set  $\mathcal{X}$  is an open set in  $R^n$ , and  $\mathcal{T} \equiv [t_0, t_1]$ ,  $t_0 < t_1 \leq T$ . In this case the topological product  $\Omega_0^0$  can be defined as  $[t_0, t_1] \times \mathcal{X}$ , and for finite time problems the closure of  $\Omega_0^0$  can be imbedded in a wider set  $\bar{\Omega}_0^0 = [t_0, t_1] \times R^n$ .

Now an approximation can be specified for the mapping (2.1), (2.2) using the a-priori assumption of its continuity

$$\bar{f} : \bar{\Omega}_0^0 \times U \rightarrow \mathcal{B}_1, \quad (2.8)$$

(for the infinite time horizon problems with  $T \rightarrow \infty$  we can use the assumption  $\bar{f} \in C(R^n \times U)$ ). Typically this consideration requires that for all  $t \in [t_0, T]$  (or  $[t_0, \infty)$ ) and  $x_1, x_2 \in \mathcal{X} \subset R^n$  (or  $x_1, x_2 \in R^n$ ) the following inequality holds,

$$|\bar{f}(t, x_1, u_{\beta_1}^{\beta_2}) - \bar{f}(t, x_2, u_{\beta_1}^{\beta_2})| < \epsilon(\delta)|x_1 - x_2|, \quad (2.9)$$

provided

$$u_{\beta_1}^{\beta_2} \in \mathcal{U} \text{ and } |u_{\beta_1}^{\beta_2}| < \delta. \quad (2.10)$$

Whenever  $\mathcal{U}$  is a compact set, (2.9) becomes the Lipschitz-type condition with constant  $\epsilon$ . The assumption (2.9) is quite restrictive implying a-priori uniqueness for the solution of the following mathematical model:

$$\frac{dx(\tau)}{d\tau} = \bar{f}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot)) \quad (2.11)$$

for any given control  $u_{\beta_1}^{\beta_2}(\tau, \cdot)$ , all  $\tau \in [t, T]$ ,  $t \in \bar{T}$  and the initial data  $x(t) = x$ , where the dot argument reflects the dependency of the control on the neighborhood states when time is fixed.

There are at least two practically important cases when such a consideration is inappropriate:

- a set  $\mathcal{U}$  is given approximately and no *a-priori* information is provided as for its compactness;
- initial conditions (2.6) in the governing equation (2.5) are given approximately.

Both situations are typical in many applications which makes the assumption of continuity of  $\bar{f}$  (as well as  $\tilde{f}$ ) inappropriate in general. The current field of optimal control theory essentially relies on the assumption that an admissible control will always remain such under the flow of temporal evolution. This mathematical assumption was induced by the understanding of an optimizer as an error-nulling device and represents a relatively narrow view of the concept of cybernetics [32]. Of course, such a simplified view has allowed the achievement of a number of successes in modelling some aspects of automobile driving, aircraft piloting, aerospace engineering, economics and in some other areas of application. However, the limitations of this view are being reached as the topic of supervisory control is being pursued [16].

Even if we assume the existence of an admissible sequence of controls in  $\mathcal{B}_1$ , it does not necessarily mean that such a sequence is feasible in  $\mathcal{B}_2$ . On the other hand, a possible continuity of  $x(\cdot)$  does not imply continuity of  $\tilde{f}(t, x, u_{\beta_1}^{\beta_2})$  (and as a partial case the continuity of  $\bar{f}$  defined by (2.8)-(2.11)) and vice versa.

This leads to the major difficulty which is the *construction* of a mapping between  $x$  and  $\tilde{f}$  ( $\bar{f}$ ) that provides an appropriate approximation to the coupled mapping (2.1), (2.2). This difficulty is twofold. On the one hand, the well-posedness of the Cauchy problem (2.5), (2.6) is a highly desirable feature, not only from the mathematical point of view, but also from the point of view of physics [25]. On the other hand, in constructing mathematical models it is important to take into account that the assumptions of continuity together with the possibility of equipping of the topological space  $\mathcal{B}_1$  with the Hausdorff property are not independent of the result of such a construction.

A consequence of an approximation of the coupled mapping (2.1), (2.2) by mappings  $\tilde{f}$  ( $\bar{f}$ ) and  $u_{\beta_1}^{\beta_2}$  is that regularities of these mappings become coupled through the model stability. Certain *a-priori* regularities on  $x(\cdot)$  imply a stabilizing trade-off between regularities of  $\tilde{f}$  ( $\bar{f}$ ) and  $u_{\beta_1}^{\beta_2}$ . Conversely, *a-priori* regularities on  $\tilde{f}$  ( $\bar{f}$ ) require a trade-off between the smoothness of  $x$  and  $u_{\beta_1}^{\beta_2}$ . This reflects the recursive structure of dynamic mappings that eventually leads to the question on the singular nature of the problem of initial conditions [11]. Mathematically, this can be formalized by the choice of one of the two possibilities in (2.3), and such a formalization allows the mathematical closure of either  $\Omega_0^0$  or  $\Sigma$  (or both). If the topology of  $\mathcal{X}$  is *a-priori* chosen, then the quality of approximation of the coupled mapping (2.1), (2.2) by the deterministic model (2.5), (2.6), (2.7) is decisively dependent on the choice of the sets  $\mathcal{U}$  and  $\mathcal{T}$ .

Let us assume the Lebesgue measurability of the approximating functions  $u_{\beta_1}^{\beta_2} \in L^1(\Sigma)$  and  $\tilde{f} \in L^1(\Omega_0^0 \times \mathcal{U})$  in the deterministic model (2.1), (2.2), (2.3). Then the governing equation (2.5) can be interpreted as an integral equation using a consequence of the Chebyshev inequality [17]

$$x(t) - x(t_0) = \int_{t_0}^t \tilde{f}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot)) d\tau + \int_{t_0}^t |\bar{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))| d\tau, \quad (2.12)$$

where

$$\int_{t_0}^T |\bar{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))| d\tau = 0. \quad (2.13)$$

The key *a-priori* assumption in the approximation of (2.1) and (2.2) by (2.12) and (2.13) is the possibility of equipping of the topological space  $\mathcal{X}$  with the Hausdorff property<sup>6</sup>. Since  $\mathcal{T}$  is also considered a Hausdorff topological space, the topological product  $\Omega_0^0$  of  $\mathcal{T}$  and  $\mathcal{X}$  inherits the Hausdorff property.

Under these assumptions let us consider two sets  $X_1^{\mathcal{T}}$  and  $X_2^{\mathcal{T}}$  in  $\Omega_0^0$ . Denote  $\Upsilon_\epsilon(X_i^{\mathcal{T}})$  as an  $\epsilon$ -neighborhood of  $X_i^{\mathcal{T}}$ ,  $i = 1, 2$ . Then the Hausdorff separation of the set  $X_1^{\mathcal{T}} \subset \Omega_0^0$  from  $X_2^{\mathcal{T}} \subset \Omega_0^0$  is

<sup>6</sup>in this case we will write  $\Omega_0^0 \otimes \mathcal{U} \equiv \Omega_0^0 \times \mathcal{U}$

defined as [3]

$$Sep(X_1^T, X_2^T) = \inf\{\epsilon : X_1^T \subseteq \Upsilon_\epsilon(X_2^T)\}.$$

If  $\Upsilon_i(\tau_i) \subset \Omega_0^0$ ,  $i = 1, 2$  are open neighborhood sets of all admissible states under control actions  $u_{\beta_1}^{\beta_2}(\tau_i, \cdot) \in \mathcal{U}$ ,  $i = 1, 2$ , then it is assumed that  $\forall \epsilon > 0$  in the state-space of the system we have

$$Sep(\Upsilon_1(\tau_1), \Upsilon_2(\tau_2)) < \epsilon \quad (2.14)$$

provided there exists  $\delta(\epsilon, \Upsilon_1, \Upsilon_2) > 0$  such that

$$|\tau_1 - \tau_2| < \delta. \quad (2.15)$$

The conditions (2.14), (2.15) store the connection between the sets  $\mathcal{X}$  and  $\mathcal{T}$  through the intermediate influence of the set  $\mathcal{U}$ . This may not be explicitly mentioned when the Heine-Borel lemma arguments are applied to the set defined by  $\delta$  in (2.15). Indeed, such arguments allow us to claim the existence of  $\delta$  for any arbitrary given  $\epsilon$ . However, they cannot guarantee stability of the resulting mathematical model for the described dynamic system. Stability conditions express a local dependency between  $\Upsilon_i$ ,  $i = 1, 2$  and  $\epsilon$  on the global scale of  $\Omega_0^0$ . Whenever the state-space of a dynamic system is equipped with the Hausdorff property such conditions are “frozen” in  $\delta$  mathematically defined by (2.15). After the model has been constructed the question of its stability can be approached in the general case through the investigation of the dependency between  $\beta_1$ ,  $\beta_2$  and  $t_0$ . The complexity of this question has led to the attempts of relaxation of topological constraints on the initial data. To a large degree these attempts are connected to the development of stochastic mathematical models.

## 2.2. Hamiltonian approximations in deterministic and stochastic models of optimal control.

During the last decade a general framework based on an intrinsic connection between nonlinear partial differential equations in infinite dimensional spaces and optimal control in Hilbert spaces has been extensively developed (see, for example, [27] and references therein). The development of this new approach is essentially being motivated by a desire to develop a rigorous mathematical basis for applications of the dynamic programming method (DPM) in optimal control problems. The main difficulty in practical applications of DPM stem from *a-priori* regularity assumptions on the value function<sup>7</sup> that is not known from the control problem itself. On the other hand, the other fundamental approach in optimal control, the maximum principle method (MPM), encounters serious difficulties when steep space-time gradients occur. This produces a challenging problem on a connection between MDP and MPM in the general non-smooth case.

Let us recall the main steps in the application of DPM to deterministic models (2.5)-(2.7). We restrict terminal behaviour of a system by a target set  $\mathcal{K}$  assuming that  $x(T) \in \mathcal{K}$  and at this stage limiting our consideration to arithmetic Euclidean spaces (2.4). When the initial conditions  $(t, x)$  and the target set are specified we define the performance measure as a time-averaging functional [15]

$$J(t, x; u_{\beta_1}^{\beta_2}) = \tilde{g}(\tau, x(\tau)) + \int_t^\tau f_0(s, x(s), u_{\beta_1}^{\beta_2}(s, \cdot))ds \quad (2.16)$$

where  $x$  is the solution of (2.5) for the specified initial conditions  $(t, x)$ ,  $x \in \mathcal{K}$ , and  $\tau$  is the “exit” time of  $(s, x(s))$  from a closure of  $\Omega_0^0$ . The function  $\tilde{g}$  is defined as

$$\tilde{g}(t, x) = \begin{cases} g(t, x) & \text{when } (t, x) \in [t_0, T) \times \mathcal{X} \\ \hat{g}(x) & \text{when } (t, x) \in T \times \mathcal{X} \end{cases}$$

(we can formally set  $\tilde{g} = \infty$  whenever  $x$  does not belong  $\mathcal{K}$  [8]). This allows us to consider a family of optimization problems with different initial conditions  $(t, x)$  if we define the value function as the greatest lower bound of the functional (2.16) on the set of all admissible controls

$$V(t, x) = \inf_{u_{\beta_1}^{\beta_2} \in \mathcal{U}} J(t, x; u_{\beta_1}^{\beta_2}) \quad (2.17)$$

<sup>7</sup>even in linear cases this function is not necessarily differentiable

for all  $(t, x) \in \bar{\Omega}_0^0$ . If it is a-priori assumed that  $V(t, x) > -\infty$  [7] and that the set  $\mathcal{U}$  is nonempty, then the basis for applying a DP approach is provided by the Bellman dynamic programming principle.

**Theorem 2.2** [7] *For arbitrary initial conditions  $(t, x) \in \bar{\Omega}_0^0$  and  $r \in [t, T]$  the following formula holds*

$$V(t, x) = \inf_{u_{\beta_1}^{\beta_2} \in \mathcal{U}} \left[ \int_t^{\min\{r, \tau\}} f_0(s, x(s), u_{\beta_1}^{\beta_2}) ds + \tilde{g}(\tau, x(\tau)) \chi(\tau; r) + V(r, x(r)) \chi(r; \tau) \right], \quad (2.18)$$

where the set-characteristic function  $\chi$  is defined as

$$\chi(x_1; x_2) = \begin{cases} 1 & \text{if } x_1 < x_2 \\ 0 & \text{otherwise} \end{cases}$$

Since the topological space  $\mathcal{B}_1$  is a partially ordered set let us choose a chain<sup>8</sup>  $\mathcal{M}$  in it. This chain is contained in a maximal chain of  $\mathcal{B}_1$  (due to the Hausdorff theorem). In the general case it is not required that  $\mathcal{M}$  has to have the least upper bound unless  $\mathcal{B}_1$  is a structure<sup>9</sup>) [17]. Hence in principle, any mapping defined with respect to the chain  $\mathcal{M}$  may only provide an approximation of a mapping defined with respect to the whole topological space  $\mathcal{B}_1$ . If we assume that  $\mathcal{B}_1$  is a lattice, then the quality of such an approximation influences the stability properties of the lattice as a whole, and this gives rise to difficulties in the definition of the least upper bound for the chain  $\mathcal{M}$  provided it exists.

In some cases such difficulties can be removed using the classical mechanical analogy related to the quantity which is a constant along the phase trajectory  $(x, p)$  of a system, where variables  $x$  and  $p$  denote space positions and generalized impulses of the system respectively. If we assume that such a quantity does not depend on time then it can be defined as a first integral of the Hamiltonian canonical system

$$\frac{\partial x}{\partial t} = \frac{\partial H}{\partial p}, \quad \frac{\partial p}{\partial t} = -\frac{\partial H}{\partial x}, \quad (2.19)$$

where  $H(x, p)$  is the system Hamiltonian which provides a mathematical generalization of the mapping defined by (1.3). The other quantity of interest is the system Lagrangian  $L$  that allows us to relate time-averaging functionals to a canonical system (2.19). In the classical case this relationship is provided by the formula  $p = \partial L / (\partial x \partial t)$ .

In general, both quantities  $H$  and  $L$  are time dependent. When the motion of the system can be idealized as a motion of a material point along a phase space trajectory, the dynamics of the system can be described by the Hamiltonian principle of the least motion, allowing us to apply techniques of the calculus of variations. In classical mechanical applications, the definition of the quantity  $H + L$  is straightforward. For example, for the motion in  $R^n$  this quantity is defined by the inner product between velocities and impulses

$$H + L = (\dot{x}|p) = \sum_{i=1}^n \dot{x}_i p_i, \quad (2.20)$$

and is easily generalizable on Hilbert spaces. This definition provides an appropriate approximation to the total energy of the system in many different applications of dynamic systems.

In attempts to generalize (2.20) on a wider class of dynamic systems one encounters two difficulties. Firstly, in the general case it may be inappropriate to define the Lagrangian as a function of only  $(t, x(t), \dot{x}(t))$ . Secondly, if the Lagrangian is taken as an integrand in the optimizing functional of an optimal control problem, then one needs to specify the topology of a set that may be embedded in a lattice produced by  $\mathcal{B}_1$ . This set may play the role of the value domain of the system Hamiltonian.

Since this set is not required to be countable, the usual induction arguments may fail. This requires appealing to the Zorn lemma [40] under the a-priori assumption that such set can be partially

<sup>8</sup>a subset in which any two elements are comparable in the sense of an a-priori introduced partial ordering

<sup>9</sup>often referred to as a lattice, i.e. a partially ordered set for which any of its finite subsets has the greatest lower and the least upper bounds

ordered. Of course, the possible existence of the least upper bound of the set does not follow from such a consideration. However, if we assume that the value domain of the Hamiltonian is a Hausdorff topological space, it allows us to approximate the coupled mapping (2.1), (2.2) by the pair of “weakly” coupled mappings that define the Hamiltonian and the value function respectively:

$$H : \tilde{\Omega}_0^{0,1} \rightarrow \tilde{\mathcal{B}}_1 \quad (2.21)$$

and

$$V : \tilde{\Sigma} \rightarrow \tilde{\mathcal{B}}_2. \quad (2.22)$$

The quality of such an approximation decisively depends on an appropriate choice of the Hamiltonian and the specification of  $V$  on a subset of  $\Omega_0^0 \times \mathcal{U}$ , and eventually on the quality of the approximation of the initial conditions for the governing equation for the original optimal control problem. The point  $(t_0, x_0)$ , defined by the initial conditions of the model, may be a unique point of tangency of the sets  $\tilde{\Sigma}$  and  $\tilde{\Omega}_0^{0,1}$ . In this case the *oscillation of the value function on the set  $\mathcal{U}$*  does not have to be defined explicitly<sup>10</sup>, and one can assume that the function  $V$  is merely a semi-continuous function assuming its equality to  $+\infty$  at all points from which the target  $\mathcal{K}$  is unreachable [8,36].

In principle, the application of this idea is virtually unlimited [36], provided the Hamiltonian of the system is defined. However, the dilemma of this approach consists in the fact that any *particular* specification of the Hamiltonian automatically imposes an *upper bound on the oscillation* of the function  $V$  on the set  $\mathcal{U}$ . Of course, in many practical applications this dilemma can be easily resolved when the evolution of a dynamic system can be idealized by the analogy with classical mechanics. As a result, this approach leads to the construction of mathematical models that give appropriate approximations of the system. For example, under quite general topological assumptions we can introduce the quantity

$$H(t, x, p) = \sup_{u_{\beta_1}^{\beta_2} \in \mathcal{U}} \{-p \cdot \tilde{f}(t, x, u_{\beta_1}^{\beta_2}) - f_0(t, x, u_{\beta_1}^{\beta_2})\} \quad (2.23)$$

that is called the Hamiltonian [7] using classical mechanics similarities. This allows us to use Theorem 2.1 to construct the model

$$\frac{\partial V}{\partial t} = H(t, x, D_x V), \quad V(T, \cdot) = \tilde{g}(\cdot) \quad (2.24)$$

known as the deterministic Hamilton-Jacobi-Bellman (HJB) equation, where in the general case  $D_x V$  denotes the generalized gradient of the function  $V$  defined by (2.17). As in the original problem of optimal control the model (2.24) provides an approximation to the coupled mapping (2.1), (2.2) on the basis of a *competition* between the topological completeness of the domain of the definition of  $H$  and a-priori regularity assumptions on  $V$ . If the domain of definition of  $H$  is a-priori equipped with the Hausdorff property, then the quality of such an approximation depends decisively on the quality of the *approximation* of the function  $V$  on a subset of  $\Omega_0^0 \times \mathcal{U}$ . One of the possibilities to improve such an approximation is connected with stochastic mathematical models in which *careful examination of the neighborhood states of motion* is required. This requirement always competes with the requirement of conservation extension in phase postulated in statistical mechanics.

In the case when dynamics can be reasonably well described by the model (2.5)-(2.7), the main difficulty stems from a connection between  $u_{\beta_1}^{\beta_2}$  and  $T$  expressed by the identity (2.13). Since a function  $\bar{\sigma}$  is unknown, many important results in the deterministic theory of optimal control have been obtained for the limiting case of  $T \rightarrow \infty$ . The other possible approach to this problem is based on Tichonov type models with singular perturbations, which require the investigation of dynamics in the limit of vanishing perturbations,  $\epsilon \rightarrow 0^+$ . The models, resulting from both approaches, are natural developments of the model (1.1), (1.2), which in the view of optimal control theory require information on a connection between  $u_{\beta_1}^{\beta_2}$  and  $T$ .

Attempts to capture this connection in its generality lead to the limiting mathematical models for  $T \rightarrow \infty$  simultaneously with  $\epsilon \rightarrow 0^+$ . However, each model of this type implies a-priori knowledge on a connection between topological properties of  $\mathcal{B}_1$  and  $\mathcal{B}_2$  (or  $\tilde{\mathcal{B}}_1$  and  $\tilde{\mathcal{B}}_2$ ), their subsets, or their approximations by other topological spaces. If such information is available, then the results of the

<sup>10</sup>If a function  $\tilde{f}$  has infimum  $m_0$  and supremum  $M_0$  on a set  $X$ , then the difference  $M_0 - m_0$  is called the oscillation of  $\tilde{f}$  on  $X$

Ljusternik-Schnirelman theory can be used to draw a conclusion on sets of critical points of functionals that define the mathematical model itself (see, for example, [36] and references therein). In the general case, the problem lies in the fact that the definition of topologies in  $\mathcal{B}_1$  and  $\mathcal{B}_2$  is not independent of the definition of the functionals that define the model itself. An approximate character of their connection is captured in evolutionary mathematical models that are obtained from the formal definition of the coupled mapping (2.1) and (2.2).

Which of the approximate models should be used is determined by the specific area of model application. For example, although neither  $\bar{\sigma}$  nor  $x(\cdot)$  in (2.12) are known a-priori, and in the general case equality (2.13) is one of our a-priori assumptions, these assumptions can be reasonably well justified for a large number of dynamic systems. The development of physics beyond the scope of classical mechanics led to the necessity to analyze the functional  $\int_{t_0}^t \bar{\sigma}$  in (2.12) and consequently to relax the requirement (2.13). A fundamental idea for such a relaxation that allows us to approximate (2.1), (2.2) is connected with the following stochastic model with random coefficients

$$x(t) - x(t_0) = \int_{t_0}^t \bar{f}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot)) d\tau + \int_{t_0}^t \tilde{\sigma}(\tau, x, u_{\beta_1}^{\beta_2}(\tau, \cdot)) d\omega(\tau). \quad (2.25)$$

In spite of the general nature of this approximation, its quality in applications is determined by the regularity assumptions on  $x(\cdot)$  and  $\omega(\cdot)$ . For example, we may assume that  $x(\cdot)$  is a continuous sample path  $R^n$ -valued process and  $\omega(\cdot)$  is a  $\mathcal{F}_t$ -adapted  $R^k$ -valued Wiener process on a probability space  $(\Omega, \mathcal{F}, P)$  (the family of  $\sigma$ -algebra  $\mathcal{F}, t \geq 0$  defines a filtration on this space). Then the problem of stochastic optimal control of the dynamics described by the governing equation (2.25) is to minimize the conditional mathematical expectation

$$E_{x,t}^{u_{\beta_1}^{\beta_2}}[J(u_{\beta_1}^{\beta_2})] \rightarrow \min, \quad (2.26)$$

where the functional  $J(u_{\beta_1}^{\beta_2})$  is defined by (2.16).

As for the deterministic models, we require information on the connection between  $u_{\beta_1}^{\beta_2}$  and  $T$ , which always comes implicitly from the a-priori regularity assumptions on the mappings which define the model. In principle, the situation is the same, when we extend the model (2.26) to a model with jumps by adding a “jump-term” to (2.25) defined by  $\mathcal{F}_t$ -Poisson random measures (see, for example, [19] and references therein).

If the process described by (2.25) and (2.26) is assumed to be a controlled (Markov) diffusion, then its dynamic programming equation (2.18) is reducible to the HJB equation, and this takes the form of the second order PDE

$$\frac{\partial V}{\partial x} = H(t, x, D_x V, D_x^2 V), \quad V(T, \cdot) = \tilde{g}(\cdot) \quad (2.27)$$

under appropriate regularity assumptions (see [7] for details), where

$$H(t, x, p, Q) = \sup_{u_{\beta_1}^{\beta_2} \in \mathcal{U}} \left[ -p \cdot \bar{f}(t, x, u_{\beta_1}^{\beta_2}) - \frac{1}{2} \sum_{i,j=1}^n (\tilde{\sigma} \tilde{\sigma}')_{ij} q_{ij} - f_0(t, x, u_{\beta_1}^{\beta_2}) \right], \quad (2.28)$$

and  $Q = (q_{ij})$  is a symmetric, non-negative definite matrix. A procedure for the reduction of (2.25) and (2.26) to (2.27) and (2.28) preserves a connection between  $\bar{f}$ ,  $\tilde{\sigma}$ , and  $\omega$  through topological properties of their functional space definition. From the statistical point of view, such a procedure is a construction of a canonical averaging ensemble which requires consistency with postulates of statistical mechanics. In general, such procedures cannot guarantee continuity of  $x(\cdot)$ , and thus, require the consideration of *singular stochastic control* models. Moreover, even if the original process is Markovian, it does not necessarily imply that it is a continuous process with transition probabilities which can be approximated arbitrarily well by a diffusion processes. However, in many practical applications such an approximation allows us to catch some important aspects of dynamic systems.

From a physical point of view, the necessity of relaxing requirement (2.13) has been strongly motivated by Eddington’s idea of the “time arrow” [20]. In addition, such a relaxation has mathematical and technical convenience in its favour.

We consider a class of relaxed systems by considering the set  $\Lambda$  of all Borel measures on  $\mathcal{T} \times \mathcal{X} \times \mathcal{U} \subset \Omega_0^0 \otimes \mathcal{U}$  such that

$$\lambda([0, s] \times \mathcal{X} \times \mathcal{U}) = s, \quad 0 \leq s < T \quad (2.29)$$

under the assumption that  $\mathcal{T} \equiv [0, T]$  (or  $\mathcal{T} = [0, \infty)$ ). The property (2.29) (the “fixed-time control-iteration”) allows us to construct mathematical models which, in principle, are reversible in time. This property subtends the Hausdorff property of an approximation to  $\Omega_0^0 \otimes \mathcal{U}$ . Then applying the Prochorov metric one can show that  $\Lambda$  is a compact metric space [27]. If we consider the Borel field  $\mathcal{B}$  on  $\mathcal{U}$ , and the  $\sigma$ -field  $\sigma_t(\Lambda)$  generated by  $\{\lambda([0, s] \times \mathcal{X} \times \mathcal{U}_0), s \leq t, \mathcal{U}_0 \in \mathcal{B}\}$  and finally introduce the  $\sigma$ -field generated by  $\sigma_t(\Lambda)$ ,  $t > 0$ , then the space  $\mathcal{P} = \mathcal{P}(\Lambda)$  of all probabilities on  $(\Lambda, \sigma(\Lambda))$  (with weak convergence topology) is also a compact metric space due the Prochorov theorem. Now if  $\omega(0) = 0$ ,  $\mu$  is an  $\mathcal{F}_t$ -adapted  $\Lambda$ -valued random variable (relaxed control)<sup>11</sup> and  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathcal{P})$  is a probability space, then any system of the form  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathcal{P}, \omega, \mu)$  is a relaxed system. Since conditional mathematical expectation can be defined using progressively measurable Radon-Nikodim derivatives, appealing to the Radom-Nikodim theorem provides a way to a justification of the mathematical model (2.25), (2.26) for relaxed controlled systems.

Such models are less sensitive to a topological connection between the spaces  $\mathcal{B}_1$  and  $\mathcal{B}_2$  because for both spaces, the possibility of embedding into a Hausdorff topological space with the property (2.29) is assumed a-priori. This connection is still preserved through the definition of the process  $\omega(\cdot)$  and the specification of the initial conditions of the problem. If the process  $\omega(\cdot)$  is assumed to be continuous, such specification provides an *important class of limiting evolutionary models* with vanishing drift and normalized diffusion. This possibility follows from the second case in the comparison of the sets  $\Sigma$  and  $\Omega_0^0$  in (2.3).

The other important limiting case of a mathematical model for evolution is provided by the classical dynamical systems where we assume the existence of an evolutionary process with vanishing “diffusion” and a normalized “drift”

$$\dot{x} = f(t, x), \quad (2.30)$$

which is formally reducible to the autonomous system  $\dot{x} = f(x^{n+1}, x)$ ,  $\dot{x}^{n+1} = 1$  by a change of variables  $x^{n+1} = t$  [30], the analogue of which for stochastic mathematical models, takes the form (2.29) [27]. The possibility of approximations based on the models (2.30) follows from the third case in comparison of the sets  $\Sigma$  and  $\Omega_0^0$  in (2.3). The first case in (2.3) corresponds to the situation when both processes  $x(\cdot)$  and  $\omega(\cdot)$  are continuous.

### 2.3. Coupling modes in mathematical models for evolution.

All three types of models considered in this section<sup>12</sup> assume the a-priori possibility of the regularization of mathematical models when the topological space (that has to approximate space-time of the dynamic system evolution) is equipped with the Hausdorff property. This consideration relies on the possibility of knowing the initial conditions for evolutionary mathematical models exactly. In turn, this transfers the complexity issues related to the definition of a balance between “drift” and “diffusion” components in the governing dynamic equation to the computational level where these issues are typically addressed.

An approximating mapping (possibly expressed on a finite lattice) between  $\Omega_0^0$  and  $\Sigma$ , which approximates the problem (2.1) and (2.2), implies that the control depends on both time and neighborhood states. Hence, whenever we assume the Hausdorff property of state-space and allow the control to be a discontinuous function a *sequential regularization procedure* is necessary to ensure the model stability. The complexity of this regularization has two sides.

- On the one hand, stochastic processes are not necessarily diffusion-dominated, and the problem of dynamic system control described by stochastic mathematical models may not necessarily be reduced to either elliptic or parabolic (possibly degenerate) PDEs. In the general case, the method of diffusion approximation may not provide an appropriate framework for the approximation of the original problem. This is shown by the growing number of physical examples (see [25,13] and references therein). Allowing the process  $x(\cdot)$  to be a discontinuous function in stochastic singular control problems and assuming the control function to be from the class of functions with bounded variation, the regularity

<sup>11</sup>for  $\mathcal{F}_t$  measurability of  $\mu(X \times Y)$  it is required only that  $X \in \mathcal{B}([0, t] \times \mathcal{X})$  and  $Y \in \mathcal{B}(\mathcal{U})$

<sup>12</sup>based on the first three logical choices in comparison of  $\Sigma$  and  $\Omega_0^0$

of the value function is virtually determined by the regularity of the interface, which is unknown from the control problem itself. This coupled regularity problem which in some cases can be resolved through different versions of the principle of smooth fit [2,34,7] or by the compactness arguments [18] in general, is problem specific. This leads to the situation when the quality of the mathematical model is completely defined by its consistency to the real-world phenomenon.

- On the other hand, deterministic processes may not necessarily be convection-dominated or purely autonomous (or such that they exhibit a strictly periodic behavior), and the problem of dynamic system control described by deterministic mathematical models may not necessarily be reduced to purely hyperbolic PDEs. Taking the limit of small diffusivity in the reaction-diffusion equation, exponential ill-conditioning may manifest itself. In this case, a numerical treatment of possible discontinuities in the density function is necessary [31].

Such competition between deterministic and stochastic features of mathematical models is essentially part of their effectiveness in solving real-world problems. This requires the study of phenomena connected with possible discontinuities of the density function in deterministic mathematical models as well as possible discontinuities in the function  $x(\cdot)$  in stochastic mathematical models. An efficient way to study such phenomena can be provided by the analysis of the value function for the processes described by such mathematical models. The regularity of this function can be implemented in the model by the problem-specific information rather than by a-priori arguments of continuity.

### 3 Coupling regularities of adjoint function and control.

On the whole, the process of validation of mathematical models is defined by the possible existence of a separation principle between control issues of dynamics and statistical estimation aspects related to this dynamics. Such validation is based on some a-priori smoothness assumptions on functions that are not known from the mathematical model itself. Of course, in many important cases these assumptions can be justified making the model consistent with a certain (possibly very large) range of applications. Moreover, it is often possible to establish a scale of a-priori estimates that control effectiveness of the model applicability in practical problems (see, for example, [22] and references therein).

#### 3.1. Control, boundary and mesh refinement.

However there are many other situations when regularities of unknown functions can not easily be predicted, and a natural experiment can not be easily undertaken. A control problem arises naturally from such situations, and yet the importance in the real application of controlling the boundary to obtain the physically stable situation results in the main difficulty in a rigorous mathematical formalization of such a problem. This gives rise to two main approaches to the control problems.

- In practice such control can be conducted effectively by minimizing the difference between the computed and partially-observed state variables that allow us in many cases to estimate values of those variables at other points as well as to approximately determine physical parameters that may not be distinct in observations [14].

- Alternatively, we can assume that in theory any region of interest in a specific problem can be arbitrary densely discretized. However, for a specific mathematical model such fine properties of the mesh may result in a discretized model which may not only be computationally inefficient but may also be practically infeasible.

Hence, using either of these approaches, it is crucial for the algorithm performance to predict the locations of the regions with large space-gradient. This requires some reasonable a-priori assumptions about the smoothness of unknown functions that eventually leads to the validation of the mathematical model on a class of real-world problems. Although it is often the case that such assumptions can be relatively easy justified for stationary models, in the time dependent problems the possible change of shapes and location of large-gradient regions requires an increase attention to the validation procedure for the mathematical model. The topology of the space becomes dependent on a-posteriori error estimations [12]. This in turn requires an adaptive mesh refinement using some a-priori defined indicators [28].

The definition of such indicators is a necessary step in the construction of the mathematical model itself and this gives rise to difficulties in approximating the system Hamiltonian. When the model

has been constructed, an approximate character of the Hamiltonian is still reflected in the coupling regularities of control and value functions. In other words, we approximate the coupled mapping (2.1), (2.2) by a mathematical model that couples three main functions - control, value function, and the system Hamiltonian. If the mathematical formalization of the Hamiltonian has been made, the key problem to be solved is the problem of coupling regularities of the control and value functions.

### 3.2. The Pontryagin maximum principle and normalizations of the Hamiltonian.

In the deterministic theory of optimal control the Pontryagin maximum principle (PMP) allows us to solve a wide range of important practical problems [30]. However, there are difficulties in the application of this principle when the value function has steep space-gradients, especially when their locations are not known a-priori. These difficulties can be overcome if we assume continuity (or at least semi-continuity) of the value function. This assumption allows us to reduce the original control problem to the problem of solving the associated HJB equation which follows from the Dynamic Programming Principle (DPP). When the value function has classical smoothness  $V(t, x) \in C^{1,2}(\Omega_0^0)$  both approaches are connected by the method of characteristics which implies that

$$\psi(t) = \frac{\partial V}{\partial x}(t, x^*(t)), \quad (3.1)$$

where  $\psi$  is the adjoint function and  $x^*$  is the optimal trajectory from the optimal solution pair  $(x^*, \{u_{\beta_1}^{\beta_2}\}^*)$ .

If the model is stochastic, then for a diffusion approximation the dynamic programming formalism can still be put on a rigorous mathematical basis using viscosity solution theory. A non-smooth analogue of a connection between the PMP and DPP (3.1) involving the adjoint and value functions respectively, can be expressed in the form of embedding theorems.

**Theorem 3.1** [41,42] *If  $(x^*, \{u_{\beta_1}^{\beta_2}\}^*)$  is an optimal process with the Pontryagin adjoint function  $\psi$ , and  $V(t, x)$  is a viscosity solution of the problem (2.24) (or (2.27) when the function  $\tilde{\sigma}$  in (2.25) does not depend on the control explicitly and  $\omega(\cdot)$  is a Brownian motion<sup>13</sup>, then the inclusion*

$$D_x^- V(t, x^*(t)) \subset \{\psi(t, \cdot)\} \subset D_x^+ V(t, x^*(t)), \quad (3.2)$$

holds for any  $t \in [s, T]$ ,  $t_0 < s < T$ , where  $D_x^+ V$  and  $D_x^- V$  denote the superdifferential and subdifferential of  $V$  respectively.

Since the Hamiltonian is dependent on the adjoint process it is important to know how the adjoint function, the Hamiltonian, and the value function are connected. In the deterministic case such a connection is expressed by the inclusion [41]

$$D_{t,x}^{1,-} V(t, x^*(t)) \subset \{H(t, x^*, \{u_{\beta_1}^{\beta_2}\}^*(t, \cdot), \psi(t, \cdot)), \psi(t, \cdot)\} \subset D_{t,x}^{1,+} V(t, x^*(t)) \quad (3.3)$$

which holds a.e. for  $t \in [s, T]$ .

In the stochastic case the embedding (3.3) may be violated, so this requires more careful examination of the dependency between the Hamiltonian and the adjoint function. The importance of such an examination lies with the fact that in using the DPP approach and specifying the form of the system Hamiltonian a-priori, the decisive role in the justification of resulting mathematical models belongs to the algorithm for the solution of the associated dynamic equation of the Hamilton-Jacobi-Bellman type [5,6,23,29]. The choice of an approximation to the Hamiltonian eventually defines the choice of the adjoint function and vice versa, and thus the regularities of the value function become a reflection of a connection between the Hamiltonian approximation and the adjoint process.

Let us analyze this connection with the example of deterministic models of optimal control (2.5)-(2.7) where in the classical smooth case the connection between PMP and DPP may be given by the formula (3.1). In the classical deterministic case the adjoint function  $\psi(t, \cdot)$  is defined by the differential equation [30]

$$\frac{\partial \psi}{\partial t} = -\frac{\partial f_0}{\partial x} - \frac{\partial \tilde{f}}{\partial x} \psi \quad (3.4)$$

---

<sup>13</sup>in this case (3.2) holds with the probability 1

with the Cauchy type terminal condition

$$\psi(T, \cdot) = \tilde{g}(\cdot), \quad (3.5)$$

where  $\tilde{g}(\cdot) = 0$  or  $\psi(T, \cdot) = \partial\tilde{g}(T, x(T))/\partial x$  depending on the type of the problem we consider. In the general case, similar to the control function the adjoint function depends not only on time but on the topology of the state-space as well. Having defined the adjoint function, the definition of the Hamiltonian is typically assumed to be of the form [35,15]

$$H = f_0 + \tilde{f}\psi. \quad (3.6)$$

From the mathematical theory of optimal processes developed on the basis of the PMP [30], it follows that, for the process  $(x(t), u_{\beta_1}^{\beta_2}(t, \cdot))$  to be optimal, it is necessary for an adjoint function (which is not identically zero) to exist such that the relationship

$$\min_{u_{\beta_1}^{\beta_2} \in U} H(t, x(t), u_{\beta_1}^{\beta_2}(t, \cdot), \psi(t, \cdot)) = H^*(t, x(t), \{u_{\beta_1}^{\beta_2}\}^*(t, \cdot), \psi(t, \cdot)) \quad (3.7)$$

holds a.e. in  $[t_0, T]$ .

On the other hand we can use the optimality principle of the Dynamic programming approach based on theorem 2.2 to obtain

$$V(t, x(t)) = \min_{u_{\beta_1}^{\beta_2}(\tau, \cdot): t \leq \tau \leq t + \Delta t} \left\{ \int_t^{t+\Delta t} f_0[\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot)] d\tau + V(t + \Delta t, x(t + \Delta t)) \right\}. \quad (3.8)$$

In (3.8) the minimum-cost value function  $V$  is defined similarly to (2.17), that is as the minimum of the performance measure

$$V(t, x(t)) = \min_{u(\tau, \cdot): t \leq \tau \leq T} J(t, x; u_{\beta_1}^{\beta_2})$$

where  $J(t, x; u_{\beta_1}^{\beta_2})$  in its turn is defined as the performance measure (2.16) changing  $s$  for  $\tau$  and  $\tau$  for  $T$

$$J(t, x; u_{\beta_1}^{\beta_2}) = \tilde{g}(T, x(T)) + \int_t^T f_0(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot)) d\tau.$$

Then by the standard procedure from (3.8) we derive the HJB equation in the classical form (2.24), i.e. we have

$$0 = \frac{\partial V}{\partial t}(t, x(t)) + H^*\left(t, x(t), \{u_{\beta_1}^{\beta_2}\}^*, \frac{\partial V}{\partial x}\right), \quad (3.9)$$

where

$$V(T, x(T)) = \tilde{g}(T, x(T)), \quad (3.10)$$

and

$$H^*\left(t, x(t), \{u_{\beta_1}^{\beta_2}\}^*, \frac{\partial V}{\partial x}\right) = \min_{u_{\beta_1}^{\beta_2}(t, \cdot) \in U} H\left(t, x(t), u_{\beta_1}^{\beta_2}(t, \cdot), \frac{\partial V}{\partial x}\right) \quad (3.11)$$

whereas the Hamiltonian is defined as

$$H\left(t, x(t), u_{\beta_1}^{\beta_2}(t, \cdot), \frac{\partial V}{\partial x}\right) = f_0(t, x(t), u_{\beta_1}^{\beta_2}(t, \cdot)) + f(t, x(t), u_{\beta_1}^{\beta_2}(t, \cdot)) \frac{\partial V}{\partial x}. \quad (3.12)$$

Since PMP (3.7) gives only necessary conditions for optimality we cannot guarantee that the adjoint function  $\psi$  that satisfies (3.7) and is defined by the model (3.4), (3.5) is unique. This implies that in the general case the validity of passage from (3.4), (3.5) to (3.9)-(3.12) depends decisively on the definition of the Hamiltonian. However, there is a striking difference between the definition of the Hamiltonian by the formulas (3.6), (3.12) and the definition (2.20) in classical mechanics. If the function  $f_0$  defines the system Lagrangian<sup>14</sup>, then the formulas (3.6), (3.12) can be seen as those

<sup>14</sup>recall that in classical mechanics the quantity  $\int_{t_0}^T L dt$  is called the action functional

that “normalize” the quantity  $H - L$  whereas the formula (2.20) “normalizes” the quantity  $H + L$ . Strictly speaking, it is not clear from *a-priori* reasoning that the coefficient near the function  $f_0$  given *a-priori* in (3.6) and (3.12) should be equal to 1. It is this assumption that allows us to prove the uniqueness results for the value function in the types of models (2.24), (2.27) under relaxed smoothness assumptions (see [7] and references therein).

To explain this phenomenon let us substitute the Hamiltonian (3.6) into the second equation of the set of canonical equations (2.19) which in the theory of optimal control are known as the Euler-Hamilton canonical equations [30]. We assume that the initial condition  $x(t_0) = x_0$  may be given precisely for the model

$$\frac{\partial x}{\partial t} = \frac{\partial H}{\partial \psi}, \quad (3.13)$$

and that the terminal condition (3.5) for the adjoint model

$$\frac{\partial \psi}{\partial t} = -\frac{\partial H}{\partial x} \quad (3.14)$$

may also be given precisely. Then we have the following relationship between  $f_0$ ,  $\psi$  and  $\tilde{f}$

$$\frac{\partial \psi}{\partial t} = -\frac{\partial f_0}{\partial x} - \psi \frac{\partial \tilde{f}}{\partial x} - \tilde{f} \frac{\partial \psi}{\partial x}. \quad (3.15)$$

From the comparison (3.4) and (3.15) we conclude that if the Hamiltonian is defined with the normalizing factor 1 near the Lagrangian<sup>15</sup> then the adjoint function must have the property

$$\tilde{f} \frac{\partial \psi}{\partial x} = 0. \quad (3.16)$$

This property has a natural physical interpretation, that is if the average velocity of a controlled process is high, the adjoint process should be characterized by small space gradients, and conversely steep gradients in the adjoint process necessarily should imply a slow controlled dynamics.

This mathematical formalization of a connection between slow and fast motions may be inappropriate for complex dynamic systems that are often characterized by steep space-time gradients of functions unknown *a-priori*. It leads to the situation when an adequate approximation of the Hamiltonian can be made in general only on the basis of the problem-specific information.

A level of unification in the mathematical theory of optimal control can be obtained when the Hausdorff assumption is imposed on the topological space of dynamic system state-space. In some cases this allows the dynamic programming equation to be reduced to a partial differential equation of the HJB type. This assumption is natural when it is applied to deterministic finite dimensional systems for which the continuity of the function  $x(\cdot)$  may be taken for granted. In such deterministic situations the forcing term  $\tilde{f}$  of the system dynamics, defined by the set of *statistical equalities*

$$x(t + \Delta t) - x(t) = \tilde{f}(t, x(t), u_{\beta_1}^{\beta_2}(t, x))\Delta t, \quad (3.17)$$

can be reasonably well approximated by a differential equation (2.5) provided that the point  $x(t_0)$  is specified. To generalize approximations of (3.17) obtainable from the model (2.5), we can use stochastic mathematical models or, alternatively, some averaging procedures for deterministic models. In both situations we eventually split the forcing term  $\tilde{f}$  into several parts using *problem-specific information*. A wide class of mathematical models obtained from such splitting is provided by equation (2.25). Let us assume that  $t_0 \leq t < t + \Delta t \leq T$ . Then from (2.25) we have

$$x(t + \Delta t) = x(t) + \int_t^{t+\Delta t} \bar{f}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))d\tau + \int_t^{t+\Delta t} \tilde{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))d\omega. \quad (3.18)$$

This model allows an exchange of information between two splitting terms  $\bar{f}$  and  $\tilde{\sigma}$  which, in turn, allows the assumption (2.13) to be relaxed. This assumption is typical for deterministic models. For stochastic models neither

---

<sup>15</sup>or near the running cost if we use the terminology of optimal control theory

$$\int_{t_0}^T |\tilde{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))| d\omega = 0 \quad (3.19)$$

nor

$$\int_{t_0}^T |\tilde{\sigma}(\tau, x(\tau), u(\tau, \cdot))| d\tau = 0 \quad (3.20)$$

is assumed a-priori. In the general case no matter how small  $\Delta t > 0$  is assumed to be, the local conditions

$$\int_t^{t+\Delta t} |\tilde{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))| d\omega = 0, \quad \int_t^{t+\Delta t} |\tilde{\sigma}(\tau, x(\tau), u_{\beta_1}^{\beta_2}(\tau, \cdot))| d\tau = 0 \quad (3.21)$$

may be also violated in principle for an arbitrarily chosen  $t$  from the interval  $(t_0, T - \Delta t)$ . In such cases the connection between PMP and DPP, in the sense of the relationship (3.1), loses its meaning and should be replaced by imbedding results similar to (3.2).

The problem of their connection has its ultimate roots in the solution of the problem (1.3) by the method (1.5) which essentially uses the geometric interpretation of the problem. When applying the DPP to the derivation of the HJB type equations we are required to justify an expansion of the unknown function  $V(t + \Delta t, x + \Delta x)$  about the point  $(t, x)$  which automatically implies some a-priori smoothness assumptions on the function  $V$ . Let us initially assume that  $V(t, x) \in C^{2,2}$ . (This is an assumption that is certainly excessive for the majority of optimal control applications.) In the deterministic case the quality of the approximation of the original problem (2.5)-(2.7) by the model (3.9)-(3.12) is determined by the remainder term of the form [15]:

$$R = \frac{1}{2} \{ A(\Delta x)^2 + B\Delta x\Delta t + C(\Delta t)^2 \},$$

where  $A$  and  $C$  define second derivatives with respect to  $x$  and  $t$  respectively, and  $B$  denotes the mixed derivative of the unknown function  $V$ . If the approximation is required to be at least of the second order with respect to  $\Delta t$  we have to set

$$R = o(\Delta t) \text{ i.e. } \lim_{\Delta t \rightarrow 0} R/\Delta t = 0,$$

where  $o$  is the Landau symbol. This implies a certain connection between the partial derivatives of the function  $V$

$$\lim_{\Delta t \rightarrow 0} \{ \Delta x [A \frac{\Delta x}{\Delta t} + C \frac{\Delta t}{\Delta x}] \} = - \lim_{\Delta t \rightarrow 0} B \Delta x.$$

The validity of this depends on the topological properties of a space-time region of interest which in turn, the control function is dependent on. Hence, we have to reformulate the optimality principle (3.8) in the form of a connection between control and value functions before the actual approximation of the Hamiltonian. Alternatively, we can allow an additional degree of freedom for the Hamiltonian by introducing the Hamilton-Pontryagin function [30,39]

$$H(t, x, u, \psi, a_0) = -a_0 f_0 + \tilde{f}\psi, \quad (3.22)$$

where  $a_0$  is a parameter of normalization. In  $R^n$  the Hamilton-Pontryagin function is a direct generalization of the classic mechanics Hamiltonian (2.20) and can be written as

$$H(t, x, u, \psi, a_0) = -a_0 f_0 + \sum_{i=1}^n \psi_i f_i. \quad (3.23)$$

If it is possible to justify a-priori that  $a_0 > 0$  then the normalization condition for the Hamiltonian takes the form  $a_0 = 1$ . Otherwise a parametric normalization with respect to the quantities  $(a_0, \psi_1, \dots, \psi_n)$  in (3.23) (or with respect to  $a_0$  and  $\psi$  in (3.22)) leads to a boundary problem of the PMP for the optimal control [39]. In the general case, the a-priori assumption that the coefficient near the function  $f_0$  equals 1 is inappropriate. Similarly, in stochastic optimal control problems an approximation of the Hamiltonian by the formula (2.28) cannot be rigorously justified unless it is known that such normalization of the Hamiltonian is valid. In many cases such validation can be provided by the problem-specific information.

By assuming the possibility of a probabilistic character of evolution for a controlled dynamic system, we simultaneously require an a-priori deterministic condition for the normalization of the Hamiltonian approximation when the procedure of model construction is being undertaken. Since the continuity assumption for the function  $x(\cdot)$  is not applicable in the general stochastic case, this

implies the increasing importance of the analysis of neighborhood states of evolution. The question on how the exchange of information between such states takes place can be eventually answered by the adjustment of regularities of the control and the value function in mathematical models which we apply. In general, such an adjustment can be performed efficiently using the information which relates the mathematical model to the real-world situation.

In the next section, we propose the mathematical formalization of such an adjustment. We assume that both the control and value functions are from the Banach space  $L^1(Q_T)$ , where  $Q_T = \{(x, t) : t_0 \leq t \leq T, x_0 \leq x \leq x(T)\}$  is a space-time region of interest which can be thought to be a topological approximation of the set  $\Omega_0^0$ . Although the theoretical possibility that  $T \rightarrow \infty$  is not excluded by Theorem 4.1, any specific mathematical model of optimal control with the infinite horizon requires a justification of the Hamiltonian normalization. In turn, this requires a certain trade-off between averaging and discounting procedures that is achievable at least in principle by relating the model to the real-world situation.

## 4 Regularization algorithm for mathematical models of optimal control.

Let  $f_0 \in L^1(Q_T)$ . We assume that  $t_0 \leq t \leq t^0 \leq T$  and denote the topological space that approximates  $\Omega_0^0$  by

$$Q_0^0 = \{(t, x) : t_0 \leq t \leq t^0, x_0 \leq x \leq x^0(t^0)\}. \quad (4.1)$$

Let us assume that a mathematical model of optimal control has been constructed and  $(x(\cdot), u_{\beta_1}^{\beta_2}(\cdot, \cdot))$  is the process described by this model, where dots represent the dependency of the process on time, neighborhood states, and the goal functional defined, for example, by (2.7) or by (2.26). Then the governing equation for the system dynamics can be considered in the form of (2.5)(as a partial case by (2.11)) or (2.25) respectively.

To reflect the dependency of the control function on the neighborhood states at a given moment of time  $\tau$  we will write that

$$u_{\beta_1}^{\beta_2}(\tau, \cdot) = u_{\beta_1}^{\beta_2}(\tau, x(h(\tau))),$$

where  $h(\tau)$  is a recursive function of time referred to as the function of the neighborhood states when the model is specified. This leads to the interpretation of control as a family of functions that form a control population, and consequently to the coupling between the topology of  $Q_0^0$  and the definition of the function  $h$  which is not known a-priori from the mathematical model itself.

We assume that for any two control populations  $\hat{u}_{\beta_1}^{\beta_2}, \tilde{u}_{\beta_1}^{\beta_2} \in \mathcal{U}$  that are realizable with the same goal function  $f_0$  the Lipschitz-type condition holds in  $L^1(Q_0^0)$ , thus

$$\|f_0(t, x, \hat{u}_{\beta_1}^{\beta_2}) - f_0(t, x, \tilde{u}_{\beta_1}^{\beta_2})\|_{L^1(Q_0^0)} \leq L \|\hat{u}_{\beta_1}^{\beta_2} - \tilde{u}_{\beta_1}^{\beta_2}\|_{L^1(Q_0^0)} \quad (4.2)$$

(here  $L$  may be dependent of the topologies  $\beta_1$  and  $\beta_2$ ). Then we define the performance measure for the mathematical model by the Lebesgue integral as

$$V^h(x, t, u_{\beta_1}^{\beta_2}) = \int_t^{t^0} f_0(\tau, x(h(\tau)), u_{\beta_1}^{\beta_2}(\tau, x(h(\tau))) d\tau. \quad (4.3)$$

Let us assume that for the model under consideration the function  $h$ , as a function of time, can be specified, at least in principle, in the topological space  $Q_0^0$  as a whole. We introduce the following definition.

**Definition 4.1** A pair of positive functions  $\chi_1$  and  $\chi_2$  that are dependent on the function  $h$  is called the conjugate pair of probabilistic weights in  $Q_0^0$  if it satisfies the inequality

$$\chi_1 + \chi_2 \leq 2.$$

Let us further assume that there exist two points  $(t, x)$  and  $(t', x')$  (which may coincide) in the topological space  $Q_0^0$  such that  $t_0 < t \leq t' < t^0$ . We define the topological spaces  $Q_L$  and  $Q_R$  by the following inequalities respectively

$$Q_L = \{(t, x) : t_0 \leq t \leq t', x_0 \leq x \leq x'(t')\}, \quad (4.4)$$

and

$$Q_R = \{(t', x') : t \leq t' \leq t^0, x \leq x' \leq x^0(t^0)\}, \quad (4.5)$$

where all values  $x$  and  $x'$  may be dependent on the function  $h$ .

Such an interpretation of optimal control problems allows us to relate the mathematical model to the problem specific-information which is formalized through the function  $h$ . This provides the basis on which the choice of the algorithm should be made [1]. Since  $u_{\beta_1}^{\beta_2}$  is a population of all realizable controls with the domain of definition that can be imbedded in  $Q_0^0$ , the values of  $u_{\beta_1}^{\beta_2}$  depend on the topology  $\mathcal{T}$  (provided  $\mathcal{X}$  is specified we approximate the topological product  $\Omega_0^0$  by  $Q_0^0$ ) as well as on the choice of measures  $\omega$  in the resulting topological approximation.

Hence, globally in the region  $Q_0^0$ , the control is a function of time, realizable states, topology and measure

$$u_{\beta_1}^{\beta_2}(t, x) = u(t, x; \mathcal{T}, \omega).$$

At the same time, locally in the neighborhood of any specified point, the control is a function of time and states. Since this function is dependent on the definition of the neighborhood states, which in turn are dependent on  $h$ , in general we will write  $u_{\beta_1}^{\beta_2} = u^h(t, x)$ . Of course, at any given point  $(t, x) \in Q_0^0 \approx \Omega_0^0 = \mathcal{T} \otimes \mathcal{X}$ , we may have many control functions that may belong to different control populations. They may be different in the sense of the definition of their neighborhood state functions  $h$ , however their values at a specified point of  $Q_0^0$  are expected to be the same, given the model has been constructed.

To measure the quality of control with respect to the goal function we introduce the operator of controlled measures by the formula

$$\Phi^h(t, x) = u^h(t, x) - \int_t^{t^0} f_0(\tau, x(h(\tau)), u^h(\tau, x)) d\tau, \quad (4.6)$$

In general, this operator may be unbounded subject to the definition of the goal function  $f_0$  and the function of neighborhood states  $h$ .

The main result of this section provides a constructive approximation of measures to guarantee boundedness properties of the operator  $\Phi^h(t, x)$ . It is shown below that, by the appropriate choice of measures in  $Q_0^0$ , the  $L^1$ -norm of the operator (4.6) can be made arbitrarily close to zero under quite mild a-priori assumptions on the functions  $u^h$  and  $V^h$ . As a result of our construction, the operator  $\Phi^h(t, x)$  becomes arbitrarily close to an absolutely continuous function.

**Theorem 4.1** *Let  $u^h(t, x) \in L^1(Q_0^0)$  and  $V^h(t, x) \in L^1(Q_0^0)$ , where  $Q_0^0$  is defined by (4.1). Let us assume that for any  $\epsilon > 0$  there exist such functions  $\chi_1(M) > 0$ ,  $\chi_2(N) > 0$  of integer numbers  $M$  and  $N$  respectively that*

$$\chi_1(M) + \chi_2(N) \leq 2, \quad (4.7)$$

*and that the following inequalities for the Lebesgue measure  $\omega$  defined in  $Q_0^0$*

$$\omega(Q_L) < \delta_L, \quad \omega(Q_R) < \delta_R, \quad (4.8)$$

*hold where subregions  $Q_L$  and  $Q_R$  in  $Q_0^0$  are defined by (4.4) and (4.5) respectively and*

$$0 < \delta_L < \frac{\epsilon \chi_1(M)}{N+1}, \quad 0 < \delta_R < \frac{\epsilon \chi_2(N)}{M+1}, \quad (4.9)$$

*Then*

$$\|\Phi^h(t, x)\|_{L^1(Q_0^0)} < \epsilon. \quad (4.10)$$

**Proof.**

The proof of the theorem consists of two parts which correspond to the forward and backward procedures of the model construction respectively.

**I. Forward procedure.** First we assume that a moment  $t_0$  in the initial condition of the model<sup>16</sup> is given with the probability exactly 1 (if the model is based on the governing equation (2.5), then with the deterministic certainty of (2.6)). No other assumptions with the same probability are made a-priori. We introduce the sequence of sets  $Q_{t_0^0}, Q_{t_0^1}, \dots$  by the formula

$$Q_{t_0^i} = \{(t, x) \in Q_0^0 : i \leq |\Phi^h(t, x)| < i + 1\}, \quad i = 0, 1, \dots \quad (4.11)$$

Then for the Lebesgue measure  $\omega$  in  $Q_0^0$  we have

$$\int_{Q_0^0} |\Phi^h(t, x)| d\omega = \sum_{i=0}^{\infty} \int_{Q_{t_0^i}} |\Phi^h(t, x)| d\omega \quad (4.12)$$

due to the  $\sigma$ -additivity property of the Lebesgue integral.

Let us consider two cases.

**Case ( $\alpha$ ): the deterministic step followed by the probabilistic approach.**

Let the function  $f_0$  and the measure  $\omega$  be given in  $Q_0^0$ . Then for arbitrarily small  $\epsilon > 0$  and arbitrarily large integer number  $N'$  let  $\delta_L$  be such that

$$0 < \delta_L < \frac{\epsilon}{2(N' + 1)}. \quad (4.13)$$

Since (4.7) implies  $\chi_1(M) < 2$ , we have from (4.9) that

$$\omega(Q_L) < \delta_L, \quad (4.14)$$

where

$$Q_L = \bigcup_{i=0}^{N'} Q_{t_0^i}. \quad (4.15)$$

Then from (4.8), (4.9) and (4.11)-(4.15) we conclude that

$$\int_{Q_L} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2}. \quad (4.16)$$

If  $N'$  is a-priori fixed then this allows us to approximate the process described by the model, at least up to the time  $t_0^{N'}$ , which defines a lower time-bound for the limit of the model predictability. The value  $t_0^{N'}$  depends on the topology  $\mathcal{T}$  and the measure  $\omega$  which are subject to our a-priori choice. Such a choice implies probabilistic features of our knowledge on the evolution of the process from  $t_0^{N'}$  to  $t^0$ . The possible existence of the process related to the evolution from  $t_0^{N'}$  to  $t^0$  [4] can be formalized mathematically by introducing in the model a probabilistic weight function  $\bar{\chi}_1(M)$  and by assuming a possibility of such evolution by a-priori choice of the topology  $\mathcal{T}$  and the measure  $\omega$  in the topological space  $Q_0^0$ . Let us define a remainder set by  $Q_{t_0}^{N'+1} = Q_0^0 \setminus Q_L$ . Then we choose  $N'$  in such a way that

$$\int_{Q_{t_0}^{N'+1}} |\Phi^h(t, x)| d\omega = \sum_{i=N'+1}^{\infty} \int_{Q_{t_0}^i} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2} \bar{\chi}_1(M). \quad (4.17)$$

The dependency of the probabilistic weight function  $\bar{\chi}_1$  on  $M$  reflects the fact of possibility of model improvement when additional information on the problem becomes available. From the probabilistic point of view we expect that

$$\text{if } M \rightarrow \infty, \text{ then } \bar{\chi}_1(M) \rightarrow 1. \quad (4.18)$$

**Case ( $\beta$ ): the probabilistic step followed by the deterministic approach.**

Alternative reasoning is based on the probabilistic assumption (4.18). If (4.18) holds then given  $\mathcal{T}$  and  $\omega$  for any  $\epsilon > 0$  we can choose  $N'$  such that

$$\int_{Q_{t_0}^{N'+1}} |\Phi^h(t, x)| d\omega = \sum_{i=N'+1}^{\infty} \int_{Q_{t_0}^i} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2}. \quad (4.19)$$

---

<sup>16</sup>for example, with the governing equation (2.25)

This step challenges the problem on the limit of predictability for backward evolution. Specifically, since  $t_0^{N'}$  (implicitly defined by the set  $Q_{t_0}^{N'+1}$ ) depends on the topology  $\mathcal{T}$  (that a-priori introduced in  $Q_0^0$ ) and the measure  $\omega$ , we have to formalize mathematically a possible reversibility of evolution from  $t_0$  to  $t_0^{N'}$ . Such a formalization is performed by introducing in the model a probabilistic weight  $\hat{\chi}_1(M)$  that satisfies the assumptions (4.8), (4.9) of the theorem. As a result we have

$$\omega(Q_L) < \delta_L \text{ for } 0 < \delta_L < \frac{\epsilon \hat{\chi}_1(M)}{2(N' + 1)}. \quad (4.20)$$

Then from (4.15) we conclude

$$\int_{Q_L} |\Phi^h(t, x)| d\omega = \sum_{i=0}^{N'} \int_{Q_{t_0}^i} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2} \hat{\chi}_1(M). \quad (4.21)$$

**II. Backward procedure.** Let us assume now that the moment  $t^0$  is given, from the set  $Q_0^0$ , or may at least be given in principle with probability exactly 1 (or with the deterministic certainty of (2.24) or (3.10)). This type of assumptions is typical when backward evolution equations are formulated with the terminal (Cauchy type) conditions in a similar way to the model (2.27).

Analogous to the forward procedure this situation gives two cases to be considered.

**Case (α): the deterministic step followed by the probabilistic approach.**

Let

$$0 < \delta_R < \frac{\epsilon}{2(M' + 1)}, \quad (4.22)$$

where  $\epsilon$  and  $M'$  can be an arbitrarily small positive real number and an arbitrarily large integer number respectively. Similar to (4.11), (4.12) we use the  $\sigma$ -additivity property of the Lebesgue integral to conclude that

$$\int_{Q_0^0} |\Phi^h(t, x)| d\omega = \sum_{j=0}^{\infty} \int_{Q_{t_j^0}} |\Phi^h(t, x)| d\omega \quad (4.23)$$

where

$$Q_{t_j^0} = \{(t, x) \in Q_0^0 : j \leq |\Phi^h(t, x)| < j + 1\}. \quad (4.24)$$

On the basis of (4.22) and (4.8), (4.9) we have

$$\omega(Q_R) < \delta_R, \quad (4.25)$$

where

$$Q_R = \bigcup_{j=0}^{M'} Q_{t_j^0}. \quad (4.26)$$

This allows us to conclude that

$$\int_{Q_R} |\Phi^h(t, x)| d\omega = \sum_{j=0}^{M'} \int_{Q_{t_j^0}} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2}. \quad (4.27)$$

In view of the fact that  $M'$  can be arbitrarily large, we fix it to satisfy the following inequality with the probabilistic weight  $\bar{\chi}_2(N)$

$$\int_{Q_{t_{M'+1}^0}} |\Phi^h(t, x)| d\omega = \sum_{j=M'+1}^{\infty} \int_{Q_{t_j^0}} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2} \bar{\chi}_2(N), \quad (4.28)$$

where  $0 < \bar{\chi}_2(N) < 2$ . By (4.28) we formalize mathematically the possibility of recovering a backward evolution from  $t_{M'}^0$  to  $t_0$  for the specified mathematical model. To reflect the possibility of the model improvement it is expected that

$$\text{if } N \rightarrow \infty, \text{ then } \bar{\chi}_2(N) \rightarrow 1. \quad (4.29)$$

**Case (β): probabilistic step followed by the deterministic approach.**

If we a-priori assume that the limiting situation (4.29) holds for the mathematical model, then we can always choose  $M'$  in such a way that the following estimate

$$\int_{Q_{t_{M'+1}^0}} |\Phi^h(t, x)| d\omega = \sum_{j=M'}^{\infty} \int_{Q_{t_j}^0} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2} \quad (4.30)$$

will be satisfied for any arbitrarily small  $\epsilon$ . However, to justify the mathematical model we need an assumption of reversibility of evolution from  $t_{M'}^0$  to  $t^0$ . Since the definition of  $t_{M'}^0$  is given implicitly through the definition of the set  $Q_{t_{M'+1}^0}$  and is not independent of the measure  $\omega$  and the topology in  $Q_0^0$ , we choose a probabilistic weight function  $\hat{\chi}_2(N)$  to satisfy the assumptions of the theorem for any arbitrarily large  $M'$  chosen to satisfy (4.30). Hence, the choice of  $\delta_R$  such that inequality

$$0 < \delta_R < \frac{\epsilon \hat{\chi}_2(N)}{2(M' + 1)} \quad (4.31)$$

holds, implies that

$$\omega(Q_R) < \delta_R. \quad (4.32)$$

Therefore

$$\int_{Q_R} |\Phi^h(t, x)| d\omega = \sum_{j=0}^{M'} \int_{Q_{t_j}^0} |\Phi^h(t, x)| d\omega < \frac{\epsilon}{2} \hat{\chi}_2(N). \quad (4.33)$$

To combine both above procedures we introduce the following functions

- for the forward procedure

$$\chi_1(M) = \begin{cases} \bar{\chi}_1(M) & \text{for the case } (\alpha), \\ \hat{\chi}_1(M) & \text{for the case } (\beta), \end{cases}$$

- and for backward procedure

$$\chi_2(N) = \begin{cases} \bar{\chi}_2(N) & \text{for the case } (\alpha), \\ \hat{\chi}_2(N) & \text{for the case } (\beta). \end{cases}$$

These functions are coupled by the condition (4.7) that allows us to define the conjugate pair of probabilistic weights  $(\chi_1, \chi_2)$  which characterizes the mathematical model under consideration. Then from estimates (4.16), (4.17), (4.19), (4.21) and (4.27), (4.28), (4.30), (4.33) we derive (4.10) and this completes the proof. ■

**Remark 4.1** *The case  $(\alpha)$  of the forward procedure is essentially an approximation of the boundary of the set  $Q_L$  whereas the case  $(\beta)$  is an approximation of the boundary of the supplement of  $Q_L$  to the whole set  $Q_0^0$ , that is an approximation to the boundary of the set  $Q_0^0 \setminus Q_L$ . Therefore, the approaches of both cases are equivalent in the sense that they attempt to approximate the boundary between  $Q_L$  and  $Q_0^0 \setminus Q_L$  separated by a surface that characterizes system dynamics at  $t = t_0^{N'}$ . The a-priori unknown topology of this surface in the general case leads to the limiting considerations  $t \rightarrow \infty, N' \rightarrow \infty$ , which cannot be justified within the forward procedure itself.*

**Remark 4.2** *Specifying  $t^0$  either explicitly or implicitly (by the set  $Q_0^0$ ) we use the backward procedure to approximate the boundary of  $Q_0^0 \setminus Q_R$  or the boundary of  $Q_R$  in the cases  $(\alpha)$  and  $(\beta)$  respectively. In general, the surface separating these two sets is not reducible to a point, though such a reduction can be always effectively performed under the assumption that the initial conditions in the mathematical model are given either precisely or with the probability exactly 1. If such an assumption is made a-priori, then the resulting mathematical models manifest a competition between the Markovian character of evolution and classical conservation laws [21]. In the framework of theorem 4.1 this corresponds to the situations when (4.18) and (4.29) both hold.*

## 5 Logical issues related to the probabilistic weight functions.

Although the possibility of simultaneous realization of (4.18) and (4.29) is not excluded by theorem 4.1, we have not required the assumptions

$$\chi_1(M) \leq 1, \quad \chi_2(N) \leq 1 \quad (5.1)$$

in the proof. Nevertheless, if (5.1) holds, it is possible to interpret the probabilistic weight functions as a pair of conjugate (or coupled) random functions that are associated with the model itself. Due to the requirement of positivity of  $\chi_1$  and  $\chi_2$  neither part of such a conjugate pair vanishes although each of them can be arbitrarily small.

On the other hand, the possibility of equalities in (5.1) is induced by the assumption that a point on the phase space trajectory of a dynamic system may be specified at least in principle with the probability exactly 1. Physically, this assumption requires invariance of the density along the trajectory and is often referred to as conservation of extension in phase. Theorem 4.1 includes this possibility, yet the realization of this possibility excludes one logically possible case in the comparison of two arbitrary sets in (2.3). In order to include all four possibilities, we have introduced a recursive dependency of control on the structure of the a-priori defined topological space and measures that are used. Such a consideration allows us to take into account incompleteness of information available a-priori, and as a result better reflects the reality of modelling in mathematical control theory.

When constructing mathematical models of optimal control it is important to take into account that the Hamiltonian (Lagrangian) of dynamic systems is always an approximation which is dependent on

- approximation of initial conditions for the system;
- approximation of system-environment boundary interface.

Such dependency can be relaxed by appropriate assumptions on the topology of the state space and initial conditions of the system. Such relaxation is based on the two logical steps that form the core of mathematical models construction:

- the notion of the empty set;
- the definition of a point or a surface with probability exactly 1.

Typically, in the theory of optimal control we require both of these steps to be undertaken simultaneously. Theorem 4.1 provides a constructive way for the sequential interchange of information between these two arguments. Using conjugate pairs of probabilistic weights, it is possible to analyse mathematical models of optimal control on the basis of the associated partial differential equations without a-priori assumptions related to the Hamiltonian normalization on the boundary interface (similar to (2.23) or (2.28)). In many applications of the theory of optimal control the assumption of the existence of a point  $(t, x)$ , for which the following probabilistic equality holds,

$$P\{|x(t + \tau) - x(t)| < \epsilon_1 \wedge |u(t + \tau, \cdot) - u(t, \cdot)| < \epsilon_2\} = 1, \quad (5.2)$$

can be justified for sufficiently small  $\tau$  and arbitrarily small  $\epsilon_1 > 0$ , but  $\epsilon_2$  not necessarily small. Although this equality is natural in classical problems of mechanical control when dynamic motion is idealized by a continuous phase space trajectory, the possibility of its realization becomes questionable when we approach the solution of singular control problems.

Similarly, it is often possible to reduce the dynamic programming equation to a partial differential equation with the given approximation to the Hamiltonian  $H$  and the unknown value function  $V$ . In such cases it is often straightforward to assume that there exists sufficiently small  $\tau$  and  $\hat{\epsilon}_1$  such that the equality

$$P\{|V(t + \tau, \cdot) - V(t, \cdot)| < \hat{\epsilon}_1 \wedge |H(t + \tau, \cdot) - H(t, \cdot)| < \hat{\epsilon}_2\} = 1 \quad (5.3)$$

holds for some value  $\hat{\epsilon}_2$  (not necessarily small). In the general case, neither control changes nor the Hamiltonian increment in that small period of time  $\tau$  are required to be small, and the values of  $\epsilon_2$  and  $\hat{\epsilon}_2$  can, in principle, be arbitrarily large.

The a-priori continuity assumption imposed on at least one of the functions appearing in (5.2) and (5.3) leads to the possibility of theoretical justification of the mathematical model. In the context of Theorem 4.1 this assumption is equivalent to the “freeze” of one of the functions  $\bar{\chi}_1, \hat{\chi}_1, \bar{\chi}_2$  or  $\hat{\chi}_2$  at a constant level, and the investigation of the dynamics induced by the resulting approximation. The quality of such approximations is eventually defined by a comparison between

$$|\epsilon_2 - \hat{\epsilon}_1| \text{ and } |\hat{\epsilon}_2 - \epsilon_1|. \quad (5.4)$$

With increasing values of  $\epsilon_2$  and  $\hat{\epsilon}_2$  such approximations become inappropriate when the absolute value of the difference between  $\epsilon_1$  and  $\hat{\epsilon}_1$  also increases. In order to improve such approximations

the regularities of the control function  $u$  and the value function  $V$  should be adjusted by a sequential procedure based on estimates of the quantities (5.4). The freezing of one of the functions in (5.2), (5.3) determines the boundary interface between two sets that often includes *internal boundary conditions in a recursive manner* which requires the solution of a multiphase regularity problem. As a result, the construction of the model in general and the definition of the topology of the boundary interface in particular ultimately defines the choice of the method for an efficient solution of the problem.

The *a-priori* assumptions (4.18) and (4.29) may play a decisive role in mathematical justification of the model in those cases when an efficient algorithm for solving a problem described by the mathematical model is not known. Finally, these assumptions allow us to claim that two distinct states of a dynamic system that correspond to different times of its evolution can be determined with the same probability (exactly 1) using a mathematical model. The theoretical possibility of the existence of such a model is covered by Theorem 4.1 as a limiting case.

## 6 Concluding remarks and future directions.

Each algorithm is model-specific in the sense that its quality essentially depends on the consistency of the mathematical model to the real-world situation [37]. Since constructing a model is an art whereas deriving an algorithm is a science, it would be reasonable to combine these two processes.

This idea suggests future directions of this work. Since the Hamiltonian of any dynamic system can be given only approximately in general, continuity arguments may not be appropriate for construction of mathematical models with the unknown value function in Hausdorff topological spaces. The partial differential equations resulting from such arguments do acquire hyperbolic features as is expected from the general principles of extended thermodynamics [25,13]. The results on the derivation of algorithms from the computational models that have been obtained on the basis of partial differential equations that contain a hyperbolic mode (generalized energy equations [21]) in stochastic, nonsmooth and purely deterministic cases will be the subject of a separate publication. Some preliminary results in this field can be found in [21]. The underlying numerical procedures are essentially based on the Markov Chain approximation method [19], that allows us to define an approximation to the mapping  $H$  between the two topological spaces  $\Omega_0^0 = \mathcal{T} \otimes \mathcal{X}$  and  $\mathcal{E}$ . For any mapping  $H$ , given *a-priori*, the possible existence of a homeomorphism between these topological spaces can not be established without additional *a-posteriori* information. Therefore, the approach based on an approximation of such a mapping provides a natural, and the most general, framework for the mathematical modelling of dynamic systems and evolutionary processes. This framework leads one to view the problem of control as much broader than the current field of optimal control theory [16,32].

## Acknowledgements

I wish to acknowledge the support of the School of Mathematics at the University of South Australia. I would like also to thank Dr John Boland for his helpful comments and suggestions of improvement and Dr Jemery Day for helpful assistance at the final stage of preparation of this paper.

## References

- [1] Anderssen, R.S. *Global optimization*, in Optimization, University of Queensland Press, 1972, 26-48.
- [2] Benes, V.E., Shepp, and Witsenhausen, H.S. *Some solvable stochastic control problems*, Stochastics, 4, 1980, 181-207.
- [3] P. Diamond, P. Kloeden and A. Pokrovskii, *Degree theory on finite lattices: discretizations of dynamical systems*, Numer. Funct. Anal. and Optimiz. vol.16, no. 1&2, 1995, 43-52.
- [4] D. Dubois and H. Prade, *Possibility Theory*, Plenum Press, New York, 1988.
- [5] Falcone, M., Ferretti, R., *Discrete time high-order schemes for viscosity solutions of Hamilton-Jacobi-Bellman equations*, Numerische Mathematik, 67, 1994, 315-344.

- [6] Falcone, M., Giorgi, T, and P. Loreti, *Level sets of viscosity solutions: some applications to fronts and rendez-vous problems*, SIAM J. App. Math., Vol.54, No.5, 1994, 1335-1354.
- [7] Fleming, W. and H. Soner, *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, 1993.
- [8] Frankowska, H. *Lower semicontinuous solutions of Hamilton-Jacobi-Bellman equations*, SIAM J. Control and Optimiz., vol.31, no.1, pp.257-272, 1993.
- [9] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*, Dover, New York, 1960.
- [10] Goldstein, S. *Sufficient Conditions to Single out the Gibbs Measure from other Time-Invariant Measures*, in Long-Time Prediction in Dynamics. Ed. by C.W.Horton, Jr., L.E.Reichl, and V.G.Szebehely, John Wiley & Sons, 1983, 71-78.
- [11] Hawking, S., *The edge of spacetime*, in The New Physics, Ed. Paul Davis, Cambridge; New York: Cambridge University Press, 1989, 61-69.
- [12] Johnson, C. *A New Paradigm for Adaptive Finite Element Methods*, in The Mathematics of Finite Elements and Applications. Ed. by J. R. Whiteman, John Wiley & Sons, 1994, 105-120.
- [13] Jou, D., Casas-Vazquez, J. and G. Lebon, *Extended Irreversible Thermodynamics*, Springer-Verlag, 1993.
- [14] Kawahara, M., Anju, A. *Control, identification and estimation in finite elements in fluids*, The Third World Congress on Computational Mechanics, V.1, 1994, 300-305.
- [15] Kirk, D.E. *Optimal Control Theory*, Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- [16] Kirilik, A., Miller, R.A., and Jagacinski, J., *Supervisory Control in a Dynamic and Uncertain Environment: A Process Model of Skilled Human-Environment Interaction*, IEEE Trans. on Systems, Man, and Cybernetics, vol.23, no.4, 1993, 929-951.
- [17] A. Kolmogorov and S. Fomin, *Fundamentals of Theory of Functions and Functional Analysis*, Moscow: Nauka, 1989, (or by Dover Publications, New York, 1975).
- [18] Krylov, N.V., *Controlled Diffusion Processes*, Springer-Verlag, 1980.
- [19] Kushner, H., and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer-Verlag, 1992.
- [20] Mackey, M.C. *Time's Arrow: The Origins of Thermodynamic Behaviour*. Springer-Verlag, 1992.
- [21] Melnik, V.N., *Nonconservation Law Equation In Mathematical Modelling: Aspects of Approximation*, Proceedings of the International Conference AEMC'96, Sydney, pp.423-430, 1996.
- [22] Melnik, V.N., *Convergence of the operator-difference scheme to generalized solutions of a coupled field theory problem*, to appear in Journal of Difference Equations and Applications, 1997.
- [23] Milner, F.A., Park, E.J. *Mixed finite element methods for Hamilton-Jacobi-Bellman-type equations*, IMA Journal of Numerical Analysis, 16, 1996, 399-412.
- [24] Misra, B., and I. Prigogine, *Time, Probability, and Dynamics*, in Long-Time Prediction in Dynamics, Ed. by C.W.Horton, Jr., L.E.Reichl, and V.G.Szebehely, John Wiley & Sons, 1983, 21-43.
- [25] Muller, I. and T. Ruggeri, *Extended Thermodynamics*, Springer-Verlag, 1993.
- [26] Newton, I., *A tretise of the method of fluxions and infinite series, with its application to the geometry of curve lines*, in V.1, Mathematical Works, New York: Johnson Reprint Corp., 1964-1967.
- [27] M. Nisio, *Optimal control for stochastic partial differential equations and viscosity solutions of Bellman equations*, Nagoya Math. J., vol.123, pp.13-37, 1991.
- [28] Okuda, H., Yagawa, G. and Yashiki, T. *Three-dimensional incompressible flow analysis using adaptive finite element method*, The Third World Congress on Computational Mechanics, V.1, 1994, 167-168.
- [29] Osher, S., Shu, C.-W., *High-Order essentially nonoscillatory schemes for Hamilton-Jacobi equations*, SIAM J. Numer. Anal., Vol.28, No.4, 1991, 907-922.

- [30] Pontryagin, L.S. et al, *The Mathematical Theory of Optimal Processes*, New York: Gordon and Breach Science Publishers, 1986.
- [31] Reyna, L.G., Ward, M.J., *On exponential ill-conditioning and internal layer behavior*, Numer. Funct. Anal. and Optimiz., 16 (3 &4), 1995, 475-500.
- [32] Rouse, W. B., *Systems Engineering Models of Human-Machine Interaction*, New York: North Holland, 1980.
- [33] Russell, B., *Human Knowledge. Its Scope and Limits*. London: Allen & Unwin, 1966.
- [34] H. M. Soner and S. E. Shreve, *Regularity of the value function for a two-dimensional singular stochastic control problem*, SIAM J. Control and Optimiz., vol.27, no.4, 1989, 876-907.
- [35] Sontag, E.D. *Mathematical Control Theory*, Springer-Verlag, 1990.
- [36] Struwe, M., *Variational Methods*, Springer-Verlag, 1990.
- [37] A. Sucharev, *Minimax Models in the Theory of Numerical Methods*, Kluwer Academic Publishers, 1992.
- [38] Vainberg, M.M. and Aizengendler, P.G. *The Theory and Methods of Investigation of Branch Points of Solutions*. in Progress in Mathematics, Vol.2, Ed. R.V. Gamkrelidze, New York: Plenum Press, 1968, pp.1-72.
- [39] Vasiliev, F.P. *Numerical Methods for the solutions of extremal problems*, Moscow, Nauka, 1988.
- [40] Zeidler, E. *Applied Functional Analysis*, Springer-Verlag, 1995.
- [41] Zhou, X.Y., *Maximum principle, dynamic programming, and their connection in deterministic control*, Journal of Optimization Theory and Applications, Vol.65, No.2, 1990, 363-373.
- [42] Zhou, X.Y., *The connection between the maximum principle and dynamic programming in stochastic control*, Stochastics and Stochastics Reports, Vol.31, 1990, 1-13.