Original publication in *Optimization Letters*, 2011, Springer. DOI: 10.1007/s11590-011-0325-z

On Evolution of an Information Dynamic System and its Generating Operator*

Roman V. Belavkin

School of Engineering and Information Sciences
Middlesex University, London NW4 4BT, UK
Tel: +442084116263
R.Belavkin@mdx.ac.uk

Abstract. States of a dynamical information system can be represented by points on a statistical manifold — a subset of a vector space endowed with an information topology. An evolution of such a system is considered here with respect to changes in information rather than changes in time, because differences between states are represented by an information distance. Here we consider an optimal system maximizing a utility of an abstract information resource, and then analyze properties of information such that an optimal system is described by an evolution operator or a semigroup. The latter is generated by an operator that can be interpreted as a utility, payoff or a fitness function. We discuss the advantages and applications of the proposed approach.

1 Introduction

Recent studies in optimization of cooperative and information systems have highlighted the need for further theoretical analysis and optimization of systems with dynamic information [8,9,10]. Examples of such systems appear in cognitive science, machine learning, cybernetics, biology and artificial life. Such systems are characterized by prior uncertainty at their initial state (e.g. some statistical parameters of the underlying distributions are unknown). During their operations, however, the uncertainty is reduced due to receiving of new information, and their performance can be improved. In this paper, we shall consider general representation of such systems and their evolution within the framework of information utility (or information value) theory [20,1].

The main goal of this paper is to relate the optimal information utility dynamics with the theory of evolution operators and semigroups. The motivation for this is an observation that there does not seem to exist a time-valued function on the set of all states of an information system — given two arbitrary states, it is generally impossible to say how long does it take to transfer an information system from one state to another. This suggests that the problem of optimization of an information system evolution in time is generally ill-posed. On the other hand, there exists an abundance of information-valued functions, the so-called information distances (e.g. see [6]), and therefore the problem of optimization of the evolution in information is well-posed. Thus, understanding

* This work was supported by EPSRC grant EP/H031936/1.

which information distances lead to optimal solutions that can be described in terms of an evolution operator can be useful for optimization of an information dynamic system.

The next section introduces notation and some main formulae for solutions to the problem of maximizing expected utility of an abstract information resource. It will be shown how a specific definition of information leads to a semigroup representation of the solutions. Then we recall briefly some general facts from the theory of evolution operators and semigroups. Properties of an information functional related to evolution operators will be analyzed in the main section. It will be shown that an evolution operator and a semigroup can be constructed if subdifferential of an information functional is an injective group homomorphism. Depending on the area of application, the generating operator of the semigroup is interpreted as a utility or payoff function, negative loss, a fitness function or a replication rate. Finally, we shall discuss some potential applications and future development of this work.

2 Notation and motivation

This section introduces the notation and representation of information systems in dual linear spaces with their algebraic structures. Then we briefly recall some elements of utility theory and the problem of maximizing expected utility of information.

2.1 Representation

In what follows, Ω will denote the set of all observable states of a system under consideration. We assume that Ω is measurable with a σ-ring (or σ-algebra) of subsets $\mathscr{R} \subseteq 2^{\Omega}$. Under uncertainty, $\omega \in \Omega$ is an elementary event of probability space $(\Omega, \mathscr{R}, p)$, where $p : \mathscr{R} \to [0,1]$ is a probability measure. We denote by $\mathscr{P}(\Omega)$ the set of all probability measures on Ω. We denote by $\mathbb{E}_p\{x\}$ the expected value of a random variable $x : \Omega \to \mathbb{R}$ with respect to $p \in \mathscr{P}(\Omega)$, which in the case of countable or uncountable Ω is defined as a sum or an integral as follows:

$$\mathbb{E}_p\{x\} := \sum_{\omega \in \Omega} x(\omega) p(\omega)\,, \qquad \mathbb{E}_p\{x\} := \int_{\Omega} x(\omega)\, dp(\omega)$$

Following standard approach in functional analysis [4], we shall treat random variables and measures respectively as elements of linear spaces X and Y put in duality via bilinear form $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}$, defined by the sum or the integral above. In particular, let $C_c(\Omega)$ be the space of continuous functions with compact support in a locally compact topological space Ω. Then its dual is the space $\mathscr{M}(\Omega)$ of Radon measures, which includes σ-additive and regular Borel measures. We shall denote $X := C_c(\Omega)$ and $Y := \mathscr{M}(\Omega)$, or (with some abuse of notation) X will denote the dual of $\mathscr{M}(\Omega)$ (i.e.the second dual of $C_c(\Omega)$). Thus, $\mathbb{E}_p\{x\}$ is the value of linear functional $y(x) = \langle x, y \rangle$ (or $x(y) = \langle x, y \rangle$), where $x \in X$, $y = p \in Y = X'$ ($x \in X = Y'$, $y = p \in Y$).

Spaces X and Y have the following algebraic structures. Space $X := C_c(\Omega)$ is an ordered, commutative, linear algebra $(X, +, \cdot, \leq, \mathbb{R})$ with the pointwise binary operations and ordering.[1] The algebra X can be furnished with the Chebyshev norm

[1] More generally, $(X, +, \cdot, \leq, *, \mathbb{C})$ is a non-commutative $*$-algebra. However, the involution $*$ is trivial for $\mathbb{R} \subset \mathbb{C}$ and commutative multiplication.

$\|x\|_\infty := \sup|x(\omega)|$, and if Ω is compact, then X is complete and contains the unit element $1(\omega) = 1$. Otherwise, X is not complete, and contains only an approximate identity (i.e. a net of functions $1_i(\omega) = 1$ on compact subsets of Ω).

Space $Y := \mathscr{M}(\Omega)$ contains commutative algebra X as a dense subset, and one can define multiplication of elements $y \in Y$ by $x \in X$ so that Y is a module over X.[2] In addition, space Y is ordered by the cone $Y_+ := \{y : \langle x, y\rangle \geq 0\,,\ \forall x \in X_+\}$, dual of $X_+ := \{x : x \geq 0\}$, and furnished with the norm of absolute convergence $\|y\|_1 := |\langle 1, y\rangle|$, dual of $\|\cdot\|_\infty$. Thus, the probability measures are positive elements $y \in Y$ with $\|y\|_1 = 1$. When the algebra X is commutative, the set $\mathscr{P}(\Omega) := \{y : y \geq 0\,,\ \|y\|_1 = 1\}$ is a simplex.

We shall return to the described above algebraic structures when we define a mapping from X into Y that is a homomorphism between the additive group $(X, +)$ and multiplicative group $(X_+, \cdot) \subset Y$. Let us discuss several results that motivated this work.

2.2 Utility

We assume that the set Ω of observable states (or events) has a preference relation $\lesssim$ (a total pre-order on Ω), which has a *utility* representation — a function $x : (\Omega, \lesssim) \to (\mathbb{R}, \leq)$ such that

$$a \lesssim b \qquad \iff \qquad x(a) \leq x(b) \qquad \forall a, b \in \Omega$$

Thus, the quotient space $\Omega/\sim$ is metrizable by $\rho([a],[b]) := |x(a) - x(b)|$, separable and complete, and therefore all probability measures on $\Omega/\sim$ are Radon (by Ulam's theorem).

The preference relation $(\Omega, \lesssim)$ can be extended to $\mathscr{P}(\Omega)$ by the expected utility:

$$q \lesssim p \qquad \iff \qquad \mathbb{E}_q\{x\} \leq \mathbb{E}_p\{x\}$$

Note that elements $\omega \in \Omega$ are identified with elementary measures $\delta_\omega \in \mathscr{P}(\Omega)$ (the Dirac δ-measures), and $\mathbb{E}_{\delta_\omega}\{x\} = x(\omega)$. The expected utility is a common measure of performance of a system under uncertainty used in numerous applications of game theory [14], optimal control [3,18] and in evolutionary systems [5,15]. In fact, $\mathbb{E}_p\{x\}$ is the only functional that extends $(\Omega, \lesssim)$ to $\mathscr{P}(\Omega)$ such that pre-order $(\mathscr{P}(\Omega), \lesssim) \subset Y$ is compatible with the vector space structure of Y (i.e. $x \lesssim v$ implies $x + w \lesssim v + w$ and $\beta x \lesssim \beta x$ for all $\beta \geq 0$ and $w \in X$) and is an Archimedian pre-order ($\beta x \lesssim v$ for all $\beta \geq 0$ implies $x \lesssim 0$) [14].

The preference relation $\lesssim$ on $\mathscr{P}(\Omega)$, induced by the expected utility, can be considered on the entire space Y with the half-space $H_x := \{y : \langle x, y\rangle \geq 0\}$ as the wedge of $\lesssim$-positive elements. The dual wedge $H_x^\circ := \{x : \langle x, y\rangle \geq 0\,,\ \forall y \in H_x\}$ is a ray $\mathbb{R}_+ x := \{\beta x : \beta \geq 0\}$, generated by utility function x. Note that utility representation $x : (\Omega, \lesssim) \to (\mathbb{R}, \leq)$ is unique only up to an order isomorphism on $(\mathbb{R}, \leq)$. In particular, given two utility representations w and x of $(\Omega, \lesssim)$, the element $\alpha w + \beta x$, $\alpha, \beta \in \mathbb{R}_+$ is also a utility representation of $(\Omega, \lesssim)$. Therefore, the set of all utility representations of $(\Omega, \lesssim)$ is a wedge $W_{\lesssim}^\circ$, and it is the union of the rays $\cup H_x^\circ = \cup \mathbb{R}_+ x$, generated by

[2] In the non-commutative case, one considers Y as a left or right module under the left or right multiplication by transposed elements of X.

all utility functions for $(\Omega,\lesssim)$. The (pre)-dual of $W^\circ_\lesssim$ is the intersection $W_\lesssim := \cap H_x$ of the corresponding half-spaces. The wedges $W^\circ_\lesssim$ and $W_\lesssim$ generate partial pre-orders on X and Y, because generally they are not reproducing (i.e. $W^\circ_\lesssim - W^\circ_\lesssim \subset X$). Two elements x and w are $\lesssim$-equivalent if and only if $x - w \in W^\circ_\lesssim$ and $w - x \in W^\circ_\lesssim$. In this case, $\langle x,y\rangle = \langle w,y\rangle$ for all $y \in W_\lesssim$. The pre-orders $(X,\lesssim)$ and $(Y,\lesssim)$, generated by $W^\circ_\lesssim$ and $W_\lesssim$, are clearly compatible with the vector space operations and are Archimedian, and if functions $x \in W_\lesssim$ are utility representations of $(\Omega,\lesssim)$, then we shall refer to $(X,\lesssim)$ and $(Y,\lesssim)$ as *utility pre-ordered* spaces [2].

2.3 Utility of information

Information associated with observation of an event is often understood as a divergence of posterior probability measure p from a prior p_0, and it is represented by some *information distance* $I : \mathscr{P} \times \mathscr{P} \to \mathbb{R}_+ \cup \{\infty\}$. Important examples of information distances include distance in variation, Fisher information metric and the Kullback-Leibler divergence defined as [12]:

$$I_{KL}(p,q) := \mathbb{E}_p\{\ln(p/q)\}$$

Unlike a measure of performance (expected utility), the choice of information measure is less certain. For this reason, one can consider a rather general concept of *information resource*, which is defined simply as some closed (lower semi-continuous) functional $F : Y \to \mathbb{R} \cup \{\infty\}$ such that its restriction to $\mathscr{P}(\Omega) \subset Y$ coincides with a chosen information distance $F(y)|_{\mathscr{P}} = I(y,y_0)$. Thus, a number of results can be obtained in general form and then specialized by associating $F(y)$ with a specific $I(y,y_0)$. In particular, an information resource associated with I_{KL} can be defined as follows:

$$F_{KL}(y) := \begin{cases} \langle \ln(y/y_0), y\rangle - \langle 1, y - y_0\rangle\,, & \text{if } y > 0 \text{ and } y_0 > 0 \\ \langle 1, y_0\rangle\,, & \text{if } y = 0 \text{ and } y_0 > 0 \\ \infty\,, & \text{otherwise} \end{cases} \tag{1}$$

Observe that the information resource can be a non-linear functional with effective domain $\operatorname{dom} F := \{y : F(y) < \infty\}$ being a proper subset of space Y.

We consider systems with a performance measure and information used as a bounded resource. The former is given by a linear functional $x(\cdot) = \langle x,\cdot\rangle$, and the latter by a closed functional $F(\cdot)$. An optimal learning system maximizes the performance for a given information amount. Thus, our interest will be in the systems with performance defined by the following optimal value function $\bar{x} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$:

$$\bar{x}(\lambda) := \sup\{\langle x,y\rangle : F(y) \le \lambda\} \tag{2}$$

with $\bar{x}(\lambda) := -\infty$ if $\lambda < \inf F$, and $\bar{x}(\infty) = \infty$ for some x. Function $\bar{x}(\lambda)$ is a generalization of the Stratonovich utility of information [20].

Optimal values $\bar{x}(\lambda)$ are conditional extrema $\langle x, y(\beta)\rangle$ that are achieved by solutions $y(\beta)$ to problem (2). The following are necessary and sufficient optimality conditions, which can be found using the standard method of Lagrange multipliers (see [1] for derivation):

$$y(\beta) \in \partial F^*(\beta x)\,, \quad F(y(\beta)) = \lambda\,, \quad \beta^{-1} \in \partial \bar{x}(\lambda)\,, \quad \beta^{-1} > 0 \tag{3}$$

where $\partial F^* : X \to 2^Y$ is subdifferential $\partial F^*(x) := \{y : F^*(w) \geq F^*(x) + \langle w - x, y\rangle\,,\ \forall w \in X\}$ of convex functional $F^*(x) := \sup\{\langle x,y\rangle - F(y)\}$ (i.e. the Legendre-Fenchel dual of F). In particular, if F is defined by equation (1), then $\partial F_{KL}(y) = \ln(y/y_0)$. Thus, $\partial F^*_{KL}(x) = e^x y_0$, and the dual of F_{KL} can be defined as

$$F^*_{KL}(x) := \langle 1, e^x y_0\rangle \tag{4}$$

The first of conditions (3) give optimal solutions as a one-parameter exponential family

$$y(\beta) = e^{\beta x} y_0 \tag{5}$$

Normalization $p := y/\|y\|_1$ of optimal measures (5) with respect to $\|\cdot\|_1 = |\langle 1,\cdot\rangle|$ gives a one-parameter exponential family of probability distributions with density

$$p(\beta) = e^{\beta x - \Psi_x(\beta)} y(0) \tag{6}$$

where $\Psi_x(\beta) := \ln F^*_{KL}(\beta x)$ is the cumulant generating function. Its first derivative $\Psi'_x(\beta)$, in particular, is the expected value $\mathbb{E}_{p(\beta)}\{x\} = \langle x, p(\beta)\rangle$.

Equations (5) and (6) correspond to the following differential equations

$$y'(\beta) = x y(\beta)\,, \qquad p'(\beta) = [x - \langle x, p(\beta)\rangle]\, p(\beta) \tag{7}$$

The first equation is linear, while the second defines the so-called *log-linear* dynamics, and it is known as the *replicator equation* in population dynamics [15]. Note that $x(\omega)$, interpreted as a replication rate or a fitness function, may depend on $p(\omega)$ in general case. Thus, we have obtained these dynamics as particular cases of the optimal information utility evolution. An important difference, however, is that the evolution is considered with respect to parameter β, which is the inverse of Lagrange multiplier related to the information constraint $F(y) \leq \lambda$.

At this point we note that exponential formula (5) is the Crandall-Liggett representation of a semigroup of operators, generated by operator x. Thus, optimal dynamics of the Kullback-Leibler information can be considered from the point of the theory of evolution operators and semigroups for the corresponding Cauchy problem. In the next section, we shall recall some basic definitions of this theory. Then we shall identify properties of the information resource functional that lead to an evolution operator or a semigroup. We then consider several potential applications.

3 Evolution operators and semigroups

Theory of evolution operators and semigroups is an important tool in the study of solutions to differential equations and dynamical systems they describe. Here, we recall some basic definitions and facts [17,16].

Definition 1 (Evolution operator). *Let $D(t) \subset Y$ denote a subset of linear space Y at $t \in [0,T]$. An evolution operator $U(t,s) : D(s) \to D(t)$ is a family $\{U(t,s),\ 0 \leq s \leq t \leq T\}$ of (possibly non-linear) operators, which satisfy the following properties:*

1. $U(s,s) = \mathrm{id}_Y$ *for* $s \in [0,T]$.

2. $U(t,s)\circ U(s,r)=U(t,r)$ *for* $0\leq r\leq s\leq t\leq T$.
3. *The mapping* $t\mapsto U(t,s)\,y$ *is continuous on* $[s,T]$ *for each* $y\in D(s)$.

The theory of evolution operators studies conditions when $y(t)=U(t,s)\,z$ is an integral solution of the following Cauchy problem

$$y'(t)\in A(t)\,y(t)\,,\qquad y(s)=z\in\overline{D(A(s))}\,,\qquad s\leq t\leq T \tag{8}$$

where $A(t):D(A(t))\subset Y\to 2^Y$ is a (possibly non-linear and multivalued) operator with t-dependent domain. The case when A is t-independent is called *autonomous*, and in this case $S(t):=U(t,0)$ is a (possibly nonlinear) semigroup of operators generated by operator A.

Definition 2 (Semigroup). *Let* $D\subset Y$ *be non-empty. A semigroup is a function* $S(t):D\to D$, $t\geq 0$, *which satisfies the following properties:*

1. $S(0)=\mathrm{id}_Y$.
2. $S(t+s)=S(t)\circ S(s)$ *for* t, $s\geq 0$.
3. $\lim_{t\downarrow 0}S(t)y=y$ *for all* $y\in\overline{D}$.

A generating operator of the semigroup $S(t)$ is $A:D\subset Y\to 2^Y$ such that:

$$A\,y=\lim_{t\downarrow 0}\frac{S(t)\,y-y}{t}$$

for those $y\in D$ for which the limit exists. An important role in the theory is played by accretive and dissipative operators. Thus, A is accretive and $-A$ is dissipative if

$$\|y_1-y_2\|\leq\|y_1-y_2+\lambda(Ay_1-Ay_2)\|$$

for all $y\in D$, $\lambda\geq 0$.

Accretivity is closely related to monotonicity. If X and Y are linear spaces in duality, then an operator (subset) $A\subset X\times Y$ is said to be *monotone* if

$$\langle x_1-x_2,y_1-y_2\rangle\geq 0\,,\qquad\forall\,(x_i,y_i)\in A$$

and it is *maximal monotone* if $B\subset A$ for any other monotone $B\subset X\times Y$. An example of a maximal monotone operator is subdifferential of a convex function [11,19]. In the case when $X=Y$ (i.e. if X is a Hilbert space), monotone and accretive operators coincide.

Finally, we remind the fundamental result of Crandall-Liggett on the generation of semigroup $S(t)$: If $A-\lambda I$ is dissipative (here I is the identity operator), and satisfies the range condition $R(I-\lambda A)\supset\overline{D(A)}$ for small $\lambda>0$, then there exists a semigroup $S(t)$ generated by A via the following exponential formula:

$$S(t)\,y=\lim_{n\to\infty}\left(I-\frac{t}{n}A\right)^{-n}y=e^{tA}\,y\,,\qquad y\in\overline{D(A)}\,,\quad t\geq 0 \tag{9}$$

The function $y(t)=S(t)\,z$ is an integral solution to the Cauchy problem $y'(t)=Ay(t)$, $y(0)=z\in\overline{D(A)}$, $t\geq 0$. We can now consider the optimal dynamics of an information system from the point of evolution operators and semigroups.

4 Properties of information

As mentioned earlier, we consider information as a resource defined by some closed functional $F : Y \to \mathbb{R} \cup \{\infty\}$ such that its restriction to $\mathscr{P}(\Omega)$ coincides with an information distance $I(p,q) = F(p)$, $\inf F|_{\mathscr{P}} = F(q)$. We shall now analyze properties of information such that optimal solutions are described by an evolution operator or a semigroup.

4.1 Monotonicity

The dual functional $F^*(x) := \sup\{\langle x,y\rangle - F(y)\}$ to F is always closed and convex [13,19], and therefore $\partial F^* \subset X \times Y$ is a maximal monotone operator. This fact allows us to prove the first important property of the optimal value function (2), because optimal solutions (3) are defined by ∂F^* evaluated on the ray $\{\beta x : \beta \geq 0\}$. We shall denote by $\sup_x F$ the supremum of F in the direction of x, which is the value $F(\delta_x)$ at $\delta_x \in \operatorname{dom} F$ such that $\bar{x}(\sup F) = \langle x, \delta_x\rangle$. Note that generally $\sup_x F \leq \sup F$.

Theorem 1. *The optimal value function $\bar{x}(\lambda)$, defined by equation (2) for closed $F : Y \to \mathbb{R} \cup \{\infty\}$ and $x \neq 0$, is concave and strictly increasing on the open interval $\lambda \in (\inf F, \sup_x F)$.*

Proof. Let $y(\beta_1)$, $y(\beta_2)$ be solutions to problem (2) with constraints $\lambda_1 \leq \lambda_2$ respectively. We use the facts that F^* is convex, $y(\beta) \in \partial F^*(\beta x)$, and function $\bar{x}(\lambda) = \langle x, y(\beta)\rangle$ is isotone (non-decreasing) by the inclusion $\{y : F(y) \leq \lambda_1\} \subseteq \{y : F(y) \leq \lambda_2\}$ for any $\lambda_1 \leq \lambda_2$.

$$
\begin{aligned}
& \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, && \forall y_i \in \partial F^*(x_i) \\
\Longrightarrow\ & (\beta_2 - \beta_1)\langle x, y(\beta_2) - y(\beta_1)\rangle \geq 0\,, && \text{by } y(\beta) \in \partial F^*(\beta x) \\
\Longrightarrow\ & \beta_1 \leq \beta_2\,, && \text{by } \langle x, y(\beta_2) - y(\beta_1)\rangle \geq 0 \\
\Longrightarrow\ & \lambda \mapsto \beta^{-1} \in \partial \bar{x}(\lambda) \text{ is antitone} && \\
\Longrightarrow\ & \bar{x}(\lambda) \text{ is concave} && \\
\Longrightarrow\ & \bar{x}(\lambda) \text{ is strictly increasing if } \lambda \in (\inf F, \sup_x F) &&
\end{aligned}
$$

□

Strict monotonicity of function $\bar{x}(\lambda)$ means that it is order-isomorphism between $\lambda \in (\inf F, \sup_x F)$ and the optimal values $\langle x, y(\beta)\rangle$. We shall now show that this property is related to the fact that subdifferentials ∂F^* or ∂F^{**} possess a much stronger property than just monotonicity — they implement pre-order isomorphisms between $\operatorname{dom} F^* \subset (X, \lesssim)$ and $\operatorname{dom} F \subset (Y, \lesssim)$, where pre-orders $(X, \lesssim)$ and $(Y, \lesssim)$ are induced from preference relation $(\Omega, \lesssim)$. Recall that dual spaces X and Y can be utility pre-ordered using the wedge $W^\circ_\lesssim \subset X$ of utility functions for $(\Omega, \lesssim)$ and $W_\lesssim \subset Y$. The proof uses Galois connection between pre-ordered sets, and we remind its definition.

Definition 3 (Galois connection). *The pair (f,g) of mappings $f : X \to Y$ and $g : Y \to X$ is said to set up an isotone Galois connection between pre-ordered sets $(X, \lesssim)$ and $(Y, \lesssim)$ if*

$$f(x) \lesssim y \quad \Longleftrightarrow \quad x \lesssim g(y)$$

where mappings f, g of (f,g) are called lower and upper adjoints respectively.

We point out that mappings f and g of isotone Galois connection (f,g) are (expectedly) isotone: $w \lesssim x$ implies $f(w) \lesssim f(x)$ and $z \lesssim y$ implies $g(z) \lesssim g(y)$. In addition, the lower adjoint preserves all existing suprema, while the upper adjoint preserves all existing infima in their domains (e.g. see Proposition 7.31 in Chapter 7 of [7]). A pre-order isomorphism, however, should preserve both suprema and infima. The proof is based on the fact that f and g are pre-order isomorphisms if and only if both (f,g) and (g,f) are isotone Galois connections (i.e. each of f, g are both lower and upper adjoints).

Theorem 2. *Let $(X,\lesssim)$ and $(Y,\lesssim)$ be a dual pair of linear spaces with dual Archimedian pre-orders.*[3] *Let $F: Y \to \mathbb{R}\cup\{\infty\}$ be a closed functional, F^* and F^{**} be its first and second dual respectively. Then the subdifferential mappings $\partial F^* : X \to 2^Y$ and $\partial F^{**} : Y \to 2^X$ implement pre-order isomorphisms between the sets $(\operatorname{dom} F^*,\lesssim)$ and $(\operatorname{dom} F,\lesssim)$.*

Proof. Let us first show that ∂F^* is a lower and ∂F^{**} is an upper adjoint of isotone Galois connection $(\partial F^*, \partial F^{**})$ in the sense that for any $z \in \partial F^*(x)$ and $w \in \partial F^{**}(y)$:

$$\partial F^*(x) \ni z \lesssim y \quad \iff \quad x \lesssim w \in \partial F^{**}(y)$$

Observe that z satisfies conditions (3) (i.e. $z \in \partial F^*(\beta x)$, $\beta = 1$, $F(z) = \lambda$), and therefore it is a solution to problem $\bar{x}(\lambda) = \sup\{\langle x,y\rangle : F(y) \le F(z)\}$ with utility $x \in W^\circ_\lesssim$.

$(\Rightarrow)$ Condition $\partial F^*(x) \ni z \lesssim y$ implies $y - z \in W_\lesssim$ and $\langle x, y-z\rangle \ge 0$, because $x \in W^\circ_\lesssim$. Also, $\langle w, y-z\rangle \ge \langle x, y-z\rangle \ge 0$ for any $w \in \partial F^{**}(y)$ by monotonicity $\langle w-x, y-z\rangle \ge 0$, and therefore $w \in W^\circ_\lesssim := \{x : \langle x, y-z\rangle \ge 0\,,\ \forall y-z \in W_\lesssim\}$. Thus, x and w are comparable. Because $(X,\lesssim)$ is Archimedian, we can consider $\beta x \sim w \in \partial F^{**}(y)$ (i.e. $\beta x - w \in W^\circ_\lesssim$ and $w - \beta x \in W^\circ_\lesssim$). Thus, $\langle w, y-z\rangle = \langle \beta x, y-z\rangle$, and using monotonicity $\langle w-x, y-z\rangle \ge 0$ we have $(\beta-1)\langle x, y-z\rangle \ge 0$. Thus, $\beta \ge 1$ and $x \lesssim \beta x \sim w \in \partial F^{**}(y)$.

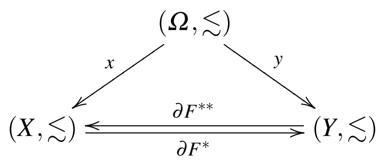
$(\Leftarrow)$ Condition $x \lesssim \beta x \sim w \in \partial F^{**}(y)$ implies $\beta \ge 1$. Therefore, $(\beta-1)\langle x, y-z\rangle \ge 0$ and $\langle x, y-z\rangle \ge 0$, and in fact $\langle w, y-z\rangle \ge 0$ for any $w \in W^\circ_\lesssim$.[4] Thus, $y - z \in W_\lesssim$ and $\partial F^*(x) \ni z \lesssim y$.

It is shown in the same manner that $\partial F^{**}(y) \lesssim x \iff y \lesssim \partial F^*(x)$, and therefore ∂F^* is an upper and ∂F^{**} is a lower adjoint of isotone Galois connection $(\partial F^{**}, \partial F^*)$. Therefore, because each of the subdifferentials ∂F^* and ∂F^{**} is both lower and upper adjoint of isotone Galois connections, they preserve all existing suprema and infima, and therefore they are pre-order isomorphisms of $\operatorname{dom} F^* \subset (Y,\lesssim)$ and $\operatorname{dom} F^{**} \subset (X,\lesssim)$. Observe also that $\operatorname{dom} F \subseteq \operatorname{dom} F^{**}$, which proves the theorem. □

[3] Should be 'with dual utility pre-orders'. That is, each element $x \in X$ (a utility function on $(\Omega,\lesssim)$) is considered as pre-order unit ($x \in \operatorname{Int}(W^\circ_\lesssim)$), and then $(X,\lesssim)$, $(Y,\lesssim)$ are dual pre-orders, generated by the corresponding wedges.

[4] It is assumed here that $0 \lesssim x$. If $x \lesssim 0$, then $0 \lesssim -x$ and $-\langle x, y-z\rangle \ge 0$.

The diagram below shows pre-order isomorphisms between the set Ω of elementary events, set X of utility functions and set Y of measures.



Note that subdifferential mappings $\partial F^* : X \to 2^Y$ and $\partial F^{**} : Y \to 2^X$ generally correspond to multi-valued mappings between $\operatorname{dom} F^* \subset X$ and $\operatorname{dom} F^{**} \subset Y$, and therefore they may not establish a bijection between these sets.

4.2 Continuity

Let $\{y(\lambda)\}_x \subset Y$ be the family of optimal solutions to problem (2) for all values $\lambda = F(y)$. The optimality conditions (3) allow us to identify each $y(\lambda)$ with some elements of the family $\{y(\beta)\}_x := \{y(\beta) \in \partial F^*(\beta x) : \beta \geq 0\}$. Indeed, $y(\lambda) \in F^{-1}(\lambda) \cap \{y(\beta)\}_x$. Note that for each λ (β), there are generally several elements of the family $\{y(\lambda)\}_x$ ($\{y(\beta)\}_x$). Consider functions $\gamma : \lambda \mapsto y(\lambda) \in \{y(\lambda)\}_x$ and $\sigma : \beta \mapsto y(\beta) \in \{y(\beta)\}_x$. We shall refer to $\gamma(\lambda)$ and $\sigma(\beta)$ respectively as *natural* and *canonical* parametrization of solutions to problem (2). The conditions, under which these parametrizations are continuous functions, are as follows.

Theorem 3. *Let X and Y be locally convex spaces in duality, and let $F : Y \to \mathbb{R} \cup \{\infty\}$ be a strictly convex function that is finally bounded on the neighborhood of some point. Then natural and canonical parametrizations $\gamma(\lambda)$ and $\sigma(\beta)$ of solutions to problem $\bar{x}(\lambda) = \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ are unique and continuous for each $x \in X$. If in addition $F^*(x) = \sup\{\langle x, y\rangle - F(y)\}$ is strictly convex, then $\gamma(\lambda)$ and $\sigma(\beta)$ are isomorphic.*

Proof. We remind that if $F : Y \to \mathbb{R} \cup \{\infty\}$ is a proper convex function that is finitely bounded on the neighborhood of some point, then F is continuous on the interior of its effective domain [13]. Thus, because $F(y) = \lambda$ for any solution y, continuity of F implies that any natural parametrization $\gamma(\lambda) = y$ must be an open and closed mapping. Furthermore, $\gamma(\lambda)$ is also injective, because $F(y) = \lambda$ is single-valued. Thus, $\gamma(\lambda)$ is continuous if its inverse is also injective (i.e. if $\gamma(\lambda)$ is an open and closed bijection).[5] Assume that γ^{-1} is not injective: $\gamma^{-1}(y) = \gamma^{-1}(z)$ and $y \neq z$. Because y and z are solutions for some λ, they are elements of subdifferential $\partial F^*(\beta x)$ by conditions (3). In this case, F is not strictly convex on the interval $[y, z] \subset \partial F^*(\beta x)$. Dually, strict convexity of F implies that γ^{-1} is injective, and natural parametrization $\gamma(\lambda)$ is continuous. Moreover, it is unique, because for strictly convex F subdifferentials $\partial F^*(\beta x)$ are singletons for each $\beta x \in \operatorname{dom} F^*$.

[5] The inverse function $\gamma^{-1}(y) = \lambda$ is continuous, its image is Hausdorff, and its domain is a collection of weakly compact sets $\partial F^*(\beta x) \subset Y$. In this case, the fact that it is a bijection implies continuity of $\gamma(\lambda) = y$.

Canonical parametrization $\sigma(\beta) = y \in \partial F^*(\beta x)$ may be discontinuous only if subdifferential ∂F^* at some βx contains more than one element. Indeed, $\partial F^*(\beta x) \subset Y$ is a non-empty, closed convex set if $\beta x \in \operatorname{dom} F^*$ (e.g. see [19], page 34), and if $\partial F^*(\beta x)$ contains more than one element, then $\operatorname{Int}(\partial F^*(\beta x))$ is non-empty and open, but its pre-image $\{\beta x\}$ is closed. Conversely, if F is strictly convex, then $\partial F^*(\beta x)$ is a singleton set for each $\beta x \in \operatorname{dom} F^*$, and in fact F^* is Gâteaux differentiable for each $\beta x \in \operatorname{Int}(\operatorname{dom} F^*)$ (e.g. see [21], Chapter 2, Section 4.1). Therefore, strict convexity of F implies that canonical parametrization $\sigma(\beta)$ is continuous and unique.

Uniqueness of parametrizations $\gamma(\lambda)$ and $\sigma(\beta)$ of optimal solutions implies that their images coincide. Their domains are related by the mapping $\lambda \mapsto \beta^{-1} \in \partial \bar{x}(\lambda)$, where $\partial \bar{x}(\lambda)$ is a supdifferential of the concave function $\bar{x}(\lambda)$. For strictly convex F this mapping is injective, because $\partial F^*(\beta x)$ is a singleton $\{y(\beta)\}$, and $F(y(\beta)) = \lambda$ is single-valued. The inverse mapping $\beta^{-1} \mapsto \lambda = F(y(\beta))$ is injective only if $\partial \bar{x}(\lambda)$ is a singleton $\{\beta^{-1}\}$ for each λ (or if $\partial F(y(\beta)) = \{\beta x\}$). In particular, this is the case when F^* is strictly convex. Therefore, strict convexity of F and F^* implies that there is a bijection between λ and β^{-1}, so that $\gamma(\lambda)$ and $\sigma(\beta)$ are isomorphic. □

4.3 Additivity

In this section, we show that monotone operator ∂F^* with an additional algebraic property defines an evolution operator acting in Y. Let us equip spaces X and Y with associative binary operations $\oplus : X \times X \to X$ and $\odot : Y \times Y \to Y$, and let $G \subseteq X$ and $H \subseteq Y$ be subsets, where $\oplus$ and $\odot$ admit inverse elements:

$$x \oplus x^{-1} = x^{-1} \oplus x = e_{\oplus}\,, \qquad y \odot y^{-1} = y^{-1} \odot y = e_{\odot}$$

Thus, $(G, \oplus)$ and $(H, \odot)$ are groups, in which $e_{\oplus}$ and $e_{\odot}$ are the units (neutral elements). The additional property we shall require is that $\partial F^* : X \to 2^Y$ is a group homomorphism:

$$\partial F^*(x \oplus w) \ni \partial F^*(x) \odot \partial F^*(w)\,, \qquad \partial F^*(e_{\oplus}) \ni e_{\odot}\,, \qquad \partial F^*(x^{-1}) \ni (\partial F^*(x))^{-1}$$

Note that the first property implies the other two. Using the property $\partial F^*(x) \ni y$ if and only if $x \in \partial F^{**}(y)$, one can show that $\partial F^{**} : Y \to 2^X$ is also group homomorphism. An evolution operator can be constructed as follows.

Theorem 4. *Let $\partial F^* : X \to 2^Y$ be a group homomorphism between $(G, \oplus) \subseteq \operatorname{dom} F^*$ and $\operatorname{dom} F^{**} \subseteq (H, \odot)$ such that $\partial F^*(e_{\oplus}) = \{e_{\odot}\}$, and let $x(t) \subset \operatorname{dom} F^*$ be a continuous function on $t \in [0, T]$. Then the mapping $U(t, s) : \operatorname{dom} F^{**} \to \operatorname{dom} F^{**}$ defined as*

$$U(t,s)\,z := \partial F^*(x(t) \oplus (x(s))^{-1}) \odot z\,, \qquad 0 \le s \le t \le T$$

is an evolution operator. If function $x(t)$ is group homomorphism $x : (\mathbb{R}, +) \to (G, \oplus)$, then the function $S(t) := U(t, 0)$ is a semigroup.

Proof. Condition $\partial F^*(e_{\oplus}) = \{e_{\odot}\}$ implies that the inverse group homomorphism $\partial F^{**} : Y \to 2^X$ is injective. Indeed, its kernel is the singleton set $\{e_{\odot}\}$. Therefore, $\partial F^*(x) =$

$\{y\}$ for each $x \in \operatorname{dom} F^*$, and canonical parametrization $\sigma(\beta) = y \in \partial F^*(\beta x)$ is continuous and unique for each x (Theorem 3). Because $x(t)$ is continuous by our assumption, we conclude that the mapping $t \mapsto \partial F^*(x(t))$ is continuous.

Using $x \oplus x^{-1} = e_\oplus$ and properties of group homomorphism, we have:

$$\begin{aligned} U(s,s)\,z &= \partial F^*(x(s) \oplus (x(s))^{-1}) \odot z = \partial F^*(e_\oplus) \odot z = e_\odot \odot z = z \\ U(t,s) \circ U(s,r)\,z &= \partial F^*(x(t) \oplus (x(s))^{-1}) \odot \partial F^*(x(s) \oplus (x(r))^{-1}) \odot z \\ &= \partial F^*(x(t) \oplus (x(s))^{-1} \oplus x(s) \oplus (x(r))^{-1}) \odot z \\ &= \partial F^*(x(t) \oplus (x(r))^{-1}) \odot z = U(t,r)\,z \end{aligned}$$

Thus, $U(t,s)$ is an evolution operator.

If $x(t+s) = x(t) \oplus x(s)$, then

$$\begin{aligned} U(t,s)\,z &= \partial F^*(x(t) \oplus (x(s))^{-1}) \odot z \\ &= \partial F^*(x(t) \oplus (x(s))^{-1} \oplus (x(s))^{-1} \oplus x(s)) \odot z \\ &= \partial F^*(x(t-s) \oplus (x(s-s))^{-1}) \odot z = U(t-s,0)\,z \end{aligned}$$

In this case, $S(\beta) := U(\beta,0)$, $\beta = t - s \geq 0$, is a semigroup. □

One can define adjoint evolution operator $U^*(t,s) : X \to X$ using group homomorphism $\partial F^{**} : (Y,\odot) \to (X,\oplus)$ as follows

$$U^*(t,s)\,w := \partial F^{**}(y(t) \odot (y(s))^{-1}) \oplus w\,, \qquad 0 \leq s \leq t \leq T$$

and with additional conditions $\partial F^{**}(e_\odot) = \{e_\oplus\}$ and $y(t) \subset Y$ being continuous.

Let us consider the commutative algebra $X := C_c(\Omega)$, where $\oplus$ and $\odot$ are the usual pointwise operations. As discussed earlier, its dual space $Y := \mathscr{M}(\Omega)$ is a module over $X \subset Y$. In this case, $G = (X,\oplus)$ with $e_\oplus = 0 \in X$, and $H = (X \setminus \{0\},\odot) \subset Y$ with $e_\odot = 1 \in Y$ (called the reference measure). We remind that an exponential function $\exp : X \to X$ is the unique (up to the base constant) homomorphism between the additive and multiplicative groups of X. This fact suggests that functionals F on $Y := \mathscr{M}(\Omega)$ such that their subdifferential operators are $\partial F = \ln$, the inverse of $\partial F^* = \exp$, are the only representations of information resource such that the optimal information dynamics is described by an evolution operator. One such representation is F_{KL}, defined by equation (1).

We note the condition of continuity of $x(t)$ in Theorem 4. For example, if $x(t) = tx$, then $(sx)^{-1} = -sx$ and $tx \oplus (sx)^{-1} = (t-s)x$. In this case, $S(\beta) := U(\beta,0)$, $\beta = t - s \geq 0$, is a semigroup of operators acting in Y. Similarly, if $y(t) = y^t$, then $(y^s)^{-1} = y^{-s}$ and $y^t \odot y^{-s} = y^{t-s}$. In this case, $S^*(\beta) = U^*(\beta,0)$ is a semigroup of operators acting in X. Thus, in the autonomous case, the semigroups have the following forms

$$S(\beta)\,z = e^{\beta x}\,z\,, \qquad S^*(\beta)\,w = \beta \ln y + w$$

Semigroup $S(\beta) = e^{\beta x}$ defines canonical parametrization $\sigma(\beta) = S(\beta)\,z$ of optimal solutions (5) to problem (2) maximizing the utility of information F_{KL}. One can see from the Crandall-Liggett exponential formula (9) that element $x \in X$ is the generator of $S(\beta)$, and it is the gradient of affine functional $\varphi(y) = \langle x, y \rangle + \alpha$. The generator x of the semigroup can have various interpretations. In economics and game theory, x is a utility or a payoff function; in mathematical statistics and estimation, $-x$ is called a loss or cost function; in evolutionary theory, x has the meaning of a fitness function or reproduction success.

The corresponding Cauchy problems are defined by differential equations (7). It is important to note, however, that these equations describe evolution in parameter β, which is the inverse of Lagrange multiplier $\beta^{-1} = d\bar{x}(\lambda)/d\lambda$, where $\lambda = F_{KL}(y)$ is information.

5 Discussion

In this paper, we studied the relation between evolution operators and optimal solutions to a problem of maximizing expected utility of an abstract information resource. We established that solutions are defined by an evolution operator if subdifferential operator of a functional representing information resource is an injective group homomorphism. For the algebra of continuous functions with compact support, which plays an important role in applications of mathematical statistics and probability theory, this property is satisfied only if the subdifferential is a logarithm function, such as subdifferential of the Kullback-Leibler information (or relative entropy). Note that this property also implies additivity of information (i.e. $I_{KL}(yz, y_0 z_0) = I_{KL}(y, y_0) + I_{KL}(z, z_0)$).

In the autonomous case, optimal dynamics in information is defined by a semigroup. Its generating operator can be interpreted as a utility, payoff, fitness function or reproduction rate depending on the area of application. States (positive measures) of an information system may represent subjective probabilities, belief functions, strategies or frequencies of individuals in populations. The corresponding differential equations suggest that states of an optimal information systems change with information proportionally to the generator.

One interesting application of this relation is elicitation or estimation of an unknown utility function of a player in a game. In a zero-sum game between two players A and B, utility u_A of one player completely defines utility u_B of another, because $u_A + u_B = 0$. However, this is not the case in a game with more than two players or in a non zero-sum game. The semigroup property of information dynamics suggests that utility functions can be estimated by observing changes in strategies of different players. Similarly, one can estimate a fitness function in a biological system by observing changes in frequencies of certain genes in a population.

Problems of optimal control of an information system can be understood better if their evolution is considered with respect to information rather than time. Indeed, the traditional approach of maximization of their performance in time is an extremely difficult or even intractable task. This difficulty can be related to the fact that sates of an information system are represented by positive measures (e.g. normalized measures in $\mathscr{P}(\Omega)$), and generally it is impossible to say how long does it take to transfer an information system from one state to another. On the other hand, one can easily define an information distance between the states, and the problem of optimal evolution in information can be solved at least in principle. One future direction of this work is optimal control of reproduction operators in evolutionary systems (i.e. biological or artificial life systems), such as optimal control of mutation rate and recombination strategies. The methods developed can help in optimization of other systems with dynamic information.

References

1. Belavkin, R.V.: Information trajectory of optimal learning. In: M.J. Hirsch, P.M. Pardalos, R. Murphey (eds.) Dynamics of Information Systems: Theory and Applications, *Springer Optimization and Its Applications Series*, vol. 40. Springer (2010)
2. Belavkin, R.V.: Utility and value of information in cognitive science, biology and quantum theory. In: L. Accardi, W. Freudenberg, M. Ohya (eds.) Quantum Bio-Informatics III, *QP-PQ: Quantum Probability and White Noise Analysis*, vol. 26. World Scientific (2010)
3. Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton, NJ (1957)
4. Bourbaki, N.: Eléments de mathématiques. Intégration. Hermann (1963)
5. Broom, M., Cannings, C., Vickers, G.T.: Sequential methods in patterns of ESS's. Journal of Mathematical Biology **22**, 597–615 (1994)
6. Chentsov, N.N.: Statistical Decision Rules and Optimal Inference. Nauka, Moscow, U.S.S.R. (1972). In Russian, English translation: Providence, RI: AMS, 1982
7. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order, second edn. Cambridge University Press (2002)
8. Grundel, D., Murphey, R., Pardalos, P., Prokopyev, O. (eds.): Cooperative Systems: Control and Optimization, *Lecture Notes in Economics and Mathematical Systems*, vol. 588. Springer (2007)
9. Hirsch, M.J., Commander, C., Pardalos, P.M., Murphey, R. (eds.): Optimization and Cooperative Control Strategies: Proceedings of the 8th International Conference on Cooperative Control and Optimization, *Lecture Notes in Control and Information Sciences*, vol. 381. Springer (2009)
10. Hirsch, M.J., Pardalos, P.M., Murphey, R. (eds.): Dynamics of Information Systems, *Springer Optimization and Its Applications*, vol. 40. Springer (2010)
11. Kachurovskii, R.I.: Nonlinear monotone operators in Banach spaces. Russian Mathematical Surveys **23**(2), 117–165 (1968)
12. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics **22**(1), 79–86 (1951)
13. Moreau, J.J.: Functionelles Convexes. Lectrue Notes, Séminaire sur les équations aux derivées partielles. Collége de France, Paris (1967)
14. von Neumann, J., Morgenstern, O.: Theory of games and economic behavior, first edn. Princeton University Press, Princeton, NJ (1944)
15. Nowak, M.A.: Evolutionary Dynamics: Exploring the Equations of Life. Harvard University Press, Cambridge, London (2006)
16. Pavel, N.H.: Nonlinear Evolution Operators and Semigroups: Applications to Partial Differential Equations, *Lecture Notes in Mathematics*, vol. 1260. Springer (1987)
17. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations, *Applied Mathematical Sciences*, vol. 44. Springer (1983)
18. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The mathematical theory of optimal processes. John Wiley and Sons (1962). Translated from Russian
19. Rockafellar, R.T.: Conjugate Duality and Optimization, *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 16. Society for Industrial and Applied Mathematics, PA (1974)

20. Stratonovich, R.L.: On value of information. Izvestiya of USSR Academy of Sciences, Technical Cybernetics **5**, 3–12 (1965). In Russian
21. Tikhomirov, V.M.: Analysis II, *Encyclopedia of Mathematical Sciences*, vol. 14, chap. Convex Analysis, pp. 1–92. Springer-Verlag (1990)