Chapter 2
Information Trajectory of Optimal Learning

Roman V. Belavkin

Summary The paper outlines some basic principles of geometric and nonasymptotic theory of learning systems. An evolution of such a system is represented by points on a statistical manifold, and a topology related to information dynamics is introduced to define trajectories continuous in information. It is shown that optimization of learning with respect to a given utility function leads to an evolution described by a continuous trajectory. Path integrals along the trajectory define the optimal utility and information bounds. Closed form expressions are derived for two important types of utility functions. The presented approach is a generalization of the use of Orlicz spaces in information geometry, and it gives a new, geometric interpretation of the classical information value theory and statistical mechanics. In addition, theoretical predictions are evaluated experimentally by comparing performance of agents learning in a nonstationary stochastic environment.

2.1 Introduction

The ability to learn and adapt the behavior with respect to changes in the environment is arguably one of the most important characteristics of intelligent systems. The study of learning algorithms has become an active area of research in artificial intelligence closely related to different areas of mathematics, cognitive science, psychology and neurobiology. The optimization and information theories are of particular importance. This paper presents a geometric approach to the evolution of a learning system that is inspired by information geometry [1, 7], and it is closely related to the information value theory of Stratonovich [23].

Learning can be considered as a process of incorporating new information to improve the performance of a system. Thus, learning by this definition assumes incomplete information. On the other hand, optimization is the main motivation for learning. This duality principle of the utility and information in learning systems is

R.V. Belavkin (✉)
Middlesex University, London NW4 4BT, UK
e-mail: R.Belavkin@mdx.ac.uk

M.J. Hirsch et al. (eds.), *Dynamics of Information Systems,*
Springer Optimization and Its Applications 40, DOI 10.1007/978-1-4419-5689-7_2,

fundamental for the theory presented [4]. We now briefly outline the main principles of the classical methods and their limitations.

Without uncertainty, optimization problems are the problems of finding the extrema (minimum or maximum) of some functions, which are called utilities, costs or fitness functions depending on the particular convention. These functions represent someone's preference relation on the underlying choice set, which can be the set of lottery prizes, errors of an estimation algorithm, space-time evolutions of a dynamical system and so on. The utility function may incorporate multiple constraints and objectives in a single Lagrange function, the extreme values of which are used to find solutions to optimization problems in variational analysis and the theory of optimal control. In particular, the maximum principle [15] defines the necessary conditions of optimality in canonical form of the Euler system of differential equations. This approach to optimal control is often referred to as the *trajectory* approach. An alternative is the dynamic programming approach [6] that is solved by partial differential equations (i.e., the Hamilton–Jacobi–Bellman equation), and it is often referred to as the *wavefront* approach.

Under uncertainty, the problem is usually formulated using methods of probability theory. The elements of a choice set are drawn stochastically as the outcomes of some lottery that is represented by a probability measure over the choice set. The idea is then to 'play' a lottery that maximizes the utility (or minimizes the risk) on average. Maximization of conditional expected utility is used in Bayesian approach to stochastic optimal control and estimation [25, 26], and sequential stochastic optimization is usually solved via dynamic programming [6]. A significant development in this area was the theory of conditional Markov processes [22] that allows one to reduce the number of variables for additive utility functions and represent the space-time evolution of the system by stochastic differential equations [21].

These methods of optimal control have been also used in the design of intelligent and adaptive systems [11, 24]. One of the main challenges, however, is that these systems operate with incomplete information, and thus optimality of the described above methods (which assume a given model of the system) is no longer guaranteed. However, often one can consider asymptotic optimality under certain assumptions. Some of these assumptions are:

1. The limits of empirical distributions exist.
2. Data is obtained from independent and identically distributed samples.
3. The 'true' distributions are stationary.

The first assumption allows one to pick some priors and then update them using empirical frequencies [16]. If these frequencies converge to the 'true' distributions, then asymptotically the system becomes optimal. The first assumption, however, depends on the second (the weak law of large numbers). Its last part (identically distributed) is equivalent to the third assumption. It is now becoming increasingly apparent that these basic assumptions may be violated in learning systems.

Indeed, the last assumption may not be valid if the agents' interaction with the environment changes the underlying distributions (i.e., there is a dependency between the agents and their environment). Dropping the stationary assumption, however, is

not a problem because the Bernoulli theorem is then replaced by the Poisson theorem, where the limit of empirical frequencies is the average of the distributions. Much more important, however, is the assumption of independent trials. Note that this does not mean that the evolution of the system is a sequence of independent events. This can be a Markov or even the conditional Markov process. However, if the Markov transitional probabilities are not known, then the samples updating the empirical transitional frequencies are assumed to be independent.

To see that this assumption can be violated in a learning system, one has to consider the exchangeability concept, introduced by Bruno de Finetti [8]. Exchangeable sequences are such that their joint distributions are invariant under permutation of the sequence order. For finite sequences, there are more exchangeable distributions than independent, and they coincide only when sequences are infinite (the de Finetti's theorem). Thus, if the sequence is not exchangeable, then it is also not independent. Now, learning is a process when new information, obtained from samples, is used to adapt the system in order to improve the performance. This means that the order, in which data is sampled and used, may be important, and therefore learning sequences are generally not exchangeable and are not independent. Without this condition, the first assumption is too strong, and therefore the limit may not exist in the traditional sense (i.e., as convergence in the laws of large numbers). As an illustration of this argument, consider a cat learning the distribution of mice. What is the limit of this distribution, if the mice also learn the distribution of the cat?

The problem of incomplete prior information should not be confused with the complexity issues arising in many optimization problems, such as the 'curse of dimensionality' in sequential optimization. Given unlimited computational power, one theoretically could use the dynamic programming approach to optimize decisions even over infinite sequences, and some researchers suggested that it could resolve problems of incomplete information, such as the exploration-exploitation dilemma [24]. This idea, however, contradicts the statistical nature of information, because new information can only be obtained through measurements, and it can only be lost in transmission (such as computation). The dynamic programming method is a technique for optimization of utility over sequences, but it does not address the problem of incomplete prior information.

Problems of optimization with information constraints have been considered in information theory [12, 13, 18, 19] leading to optimal solutions in the form of exponential family of distributions. The dual problem of entropy maximization with linear constraints was considered in statistical mechanics [9, 10]. The information value theory was developed as an application of these results to cybernetics and statistical decisions [23]. It was shown also that the results of this theory hold for a wide class of the entropically and informationally stable systems. In particular, this class includes sequences of nonstationary, nonindependent random variables. These results, therefore, can be applied to a more general class of learning systems than those considered by the traditional methods.

This paper presents geometric approach to the analysis of learning systems, and describes also a simple experiment as an illustration. The theory defines nonasymptotic optimality conditions relative to available information, and defines the evolution of an optimal learning by a trajectory on the statistical manifold. The analysis

has similarities with the use of Orlicz spaces and exponential statistical manifolds in non-parametric information geometry to describe systems of bounded entropy [14]. Here, however, the theory is developed for more general class of convex functionals representing information. The corresponding spaces are quasi-pseudo-metric generalizing normed spaces. The approach leads to a nonasymptotic and nonparametric theory for optimization of the evolution of a learning system by empirical constraints. Some examples are closely related to information theory and statistical mechanics. The optimal trajectory also defines the utility and information bounds of a learning system, which are given by the analogue of a gradient theorem for path integrals in conservative vector fields.

2.2 Topology and Geometry of Learning Systems

In this section, we recall some elements of the theory of optimal choice under uncertainty [25] and information value theory [23] that are relevant to our representation of the learning systems. Then we define a topology on a functional space related to information dynamics in such systems.

2.2.1 Problem Statement and Basic Concepts

Fundamental concept in the theory of rational choice is the *preference* relation on a set Ω, which is a complete and transitive binary relation $\lesssim \subseteq \Omega \times \Omega$ (i.e., total pre-order). Subset of symmetric pairs $\sim \subseteq \lesssim$ is the equivalence relation, and the set of antisymmetric pairs $< \subseteq \lesssim$ is a partial order on Ω. The quotient set $\Omega/\sim$ is totally ordered. We assume that $\Omega/\sim$ can be embedded into the extended real line $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$. In this case, the preference relation can be represented by a *utility* function $u : \Omega \to \overline{\mathbb{R}}$ (i.e., $\omega_1 \lesssim \omega_2$ iff $u(\omega_1) \leq u(\omega_2)$). The rational choice (optimization) corresponds to maximization of the utility.

Under uncertainty, one considers probability measures $y : \mathfrak{R} \to \mathbb{R}$ on a σ-algebra $\mathfrak{R}(\Omega) \subseteq 2^{\Omega}$. Probability measures can be interpreted as lotteries over the choice set $(\Omega, \lesssim)$. For example, the Dirac δ-measures ($\delta_\omega(d\omega) = 1$ if $\omega \in d\omega$; 0 otherwise) correspond to the elements $\omega \in \Omega$ observed with certainty. Other probability measures are unique convex combinations of the δ-measures, and therefore the set $\mathcal{P}(\Omega)$ of *all* probability measures on $\mathfrak{R}(\Omega)$ is a simplex in some linear space L—a convex hull of the set Δ of all δ-measures on Ω ($\mathcal{P}(\Omega)$ is a Choquet simplex if Ω is infinite).

The set of all probability measures, $\mathcal{P}(\Omega)$, will be referred to as *statistical manifold*, as in information geometry, and it is the set of *all* lotteries. The choice problem under uncertainty requires an extension of the preference relation $(\Omega, \lesssim)$ onto $\mathcal{P}(\Omega)$. This extension should be compatible with $(\Omega, \lesssim)$ in the following sense $(\Delta, \lesssim) = (\Omega, \lesssim)$. One such extension is given by the *expected utility*: $E_y\{u\} = \int_\Omega u(\omega)\, dy(\omega)$. Thus, measure p is preferred to q if and only if $E_p\{u\} \geq E_q\{u\}$.

Furthermore, a fundamental result of game theory states that expected utility is the only representation that also satisfies two additional axioms: continuity and substitution independence [25].

The main difference problems of optimization under uncertainty, described above, and the learning problems is that the latter are concerned with incomplete information. In particular, this means that the probability measures on the choice set are not known exactly, or in other words that the learner does not know precisely which lotteries he plays. For an agent with limited resources, this presents a dilemma of collecting more information (exploration) or using already available information for optimal control (exploitation).

The traditional approach to solving this dilemma is to treat it as a statistical problem of estimating unknown parameters $\theta \in \mathbb{R}^m$ of some known family of distributions $y(d\omega \mid \theta)$. Points θ of the parameter space $\mathbb{R}^m$ define points on the statistical manifold, and the corresponding relations and metrics are the subject of information geometry [1, 7]. The approach taken in this work is similar to infinite-dimensional nonparametric information geometry [14], where probability measures are studied directly in the corresponding functional space. This allows for considering all families of measures and to derive nonasymptotic optimality conditions using the conjugate duality theory [17]. The topologies will be defined using quasi-norms and quasi-metrics related to information constraints, which is more appropriate for describing learning systems, and it is a generalization of the standard approach using normed spaces (i.e., Orlicz spaces in [14]).

Observe that optimization under uncertainty is concerned with at least two types of real functions on the choice set $(\Omega, \lesssim)$—utilities and probability measures. Moreover, for a fixed utility function, the expected utility is a linear functional on measures; for a fixed measure, the expected utility is a linear functional on utility functions. Thus, measures and utility functions can be represented by elements of dual linear spaces L and L^*, where the expected utility implements the pairing $\langle \cdot, \cdot \rangle : L^* \times L \to \mathbb{R}$:

$$\langle x, y \rangle = \int_{\Omega} x(\omega)\, dy(\omega) \tag{2.1}$$

Note that because we only deal with preference relations that have a utility representation, set $\Omega/\sim$ is a separable, complete, metrizable space, and therefore we only need to consider Radon measures. Such measures are finite on compact subsets $\Omega_c \subseteq \Omega$, and they are in one-to-one correspondence with linear functionals $y(f) = \langle f, y \rangle$ on the space $\mathcal{C}_c(\Omega)$ of continuous functions with compact support. Thus, we associate measures with nonnegative elements of space $L \equiv \mathcal{C}_c^*(\Omega)$, dual of $\mathcal{C}_c(\Omega)$. Utility functions are the elements of its second dual $L^* \equiv \mathcal{C}_c^{**}(\Omega)$.

The theory presented is also largely inspired by information value theory [23]. Consider two points on the statistical manifold, y_0 (prior) and y (posterior), associated with an observation of some random event. The associated change $\langle x, y - y_0 \rangle$ of the expected utility represents the value of this event, and it is different for agents with different utility functions $x \in L^*$. On the other hand, information is usually represented by some functional $F : L \to \overline{\mathbb{R}}$ as a divergence of y from fixed point

y_0 on the statistical manifold, and it does not take the utility into account. Thus, different $y \in L$ with the same divergence $F(y) = I$ may have different values for the agent. The *value of information* amount $I \in \mathbb{R}$ is defined as the maximum of the expected utility subject to information constraint $F(y) \le I$:

$$U(I) := \sup\big\{\langle x, y\rangle : F(y) \le I\big\} \tag{2.2}$$

Note that the original definition in [23] is more specific using Shannon information for $F(y)$. Clearly, an optimization of information dynamics in a learning system should be closely related to information value—the optimal system should adapt to learn only the most valuable information. We now define a topology related to information value to facilitate the analysis of such systems.

2.2.2 Asymmetric Topologies and Gauge Functions

Let L and L^* be a dual pair of linear spaces over the field $\mathbb{R}$ with bilinear form $\langle\cdot,\cdot\rangle : L^* \times L \to \mathbb{R}$ separating L and L^*: $\langle x, y\rangle = 0, \forall x \in L^*$ implies $y = 0 \in L$, and $\langle x, y\rangle = 0, \forall y \in L$ implies $x = 0 \in L^*$. We shall define topologies on L and L^* that are compatible with respect to the pairing $\langle\cdot,\cdot\rangle$, but subbases of these topologies will be formed by systems of neighborhoods of zero that are generally nonbalanced sets (i.e., $y \in M$ does not imply $-y \in M$). Thus, the spaces may fail to be topological vector spaces. The gauge functions will define quasi-norms and quasi-metrics (i.e., nonsymmetric generalizations of a norm and a metric). The main motivation for this asymmetry is to avoid nonmonotonic operations on functions, such as $x \mapsto |x|$.

First, we recall some properties that depend only on the pairing $\langle\cdot,\cdot\rangle$, and not on particular topologies chosen. Nonzero $x \in L^*$ are in one-to-one correspondence with hyperplanes $H := \{y \in L : \langle x, y\rangle = \alpha\} \subset L$, $0 \notin H$, and inequality $\langle x, y\rangle \le \alpha$ defines a closed half-space. The intersection of all closed half-spaces containing set $M \subset L$ is a *convex closure* of M denoted by $\operatorname{co} M$. Set M is a *closed convex* set if $M = \operatorname{co} M$. The *polar* of M is

$$M^* := \big\{x \in L^* : \langle x, y\rangle \le 1,\ y \in M\big\}$$

The polar set is always closed, convex and $0 \in M^*$. Also, $M^{**} = \operatorname{co}[M \cup \{0\}]$, and $M = M^{**}$ if and only if M is closed, convex and $0 \in M$. Without loss of generality, we shall assume $0 \in M$.

Set M is called *absorbing* if for each $y \ne 0 \in L$ there exists $\beta > 0$ such that $y \in \beta M$; set N is called *bounded* if $N \subset \beta M$ for all $\beta \ge \varepsilon$ and some $\varepsilon > 0$. Set M is absorbing if and only M^* is bounded (to see this, observe that $y \ne 0$ are in one-to-one correspondence with closed half-spaces in L^*).

Given a closed convex set $M \subset L$ absorbing with respect to $0 \in M$, the collection of sets $\mathfrak{M} := \{\beta M : \beta > 0\}$ is the subbasis of closed neighborhoods of zero uniquely defining a topology on L. If in addition M is bounded, then the polar $M^* \ni 0$ is also absorbing, and the collection $\mathfrak{M}^* := \{\beta^{-1} M^* : \beta^{-1} > 0\}$ is the subbasis of the polar topology on L^*.

Given set $M \subset L$, the *gauge* function $p_M : L \to \overline{\mathbb{R}}$ is defined as

$$p_M(y) := \inf\{\beta > 0 : y \in \beta M\}, \quad p_M(0) := 0$$

If M is absorbing with respect to $0 \in M$, then $p_M(y) < \infty$ for all $y \in L$, and if M is bounded, then $p_M(y) = 0$ only if $y = 0$. The gauge is positively homogeneous function of the first degree, $p_M(\lambda y) = \lambda p_M(y)$, $\lambda > 0$, and if M is convex, then it is also subadditive, $p_M(y_1 + y_2) \le p_M(y_1) + p_M(y_2)$. Thus, the gauge of an absorbing convex set satisfies all axioms of a semi-norm apart from symmetry, $p_M(y) \ne p_M(-y)$, and therefore it is a *quasi-pseudonorm*. Function $d_M(y_1, y_2) = p_M(y_1 - y_2)$ is a *quasi-pseudometric* on L. If M is also bounded, then p_M is a *quasi-norm*, and d_M is a *quasi-metric* (i.e., $d_M(y_1, y_2) \ne d_M(y_2, y_1)$).

Gauge functions are closely related to support functions. The *support* of set M is function $h_M : L^* \to \overline{\mathbb{R}}$ defined as

$$h_M(x) := \sup\big\{\langle x, y\rangle : y \in M\big\}$$

Generally, $h_M(x) = p_{M^*}(x)$, and if M is convex, then $h_{M^*}(y) = p_M(y)$ (otherwise, $h_{M^*}(y) \le p_M(y)$).

2.2.3 Trajectories Continuous in Information

Observe now that the value of information, defined by (2.2), is equal to support h_M of subset $M = \{y \in L : F(y) \le I\}$ of the statistical manifold, defined by information constraint. It is common to represent information by a closed, convex functional, and therefore M is closed and convex. For the theory of convex functions, see [17, 20]. Here, we recall some basic concepts.

Convex functional $F : L \to \overline{\mathbb{R}}$ is called *proper* if its *effective domain* $\operatorname{dom} F := \{y : F(y) < \infty\}$ is nonempty and $F(y) > -\infty$. Proper convex functional is *closed* if sublevel sets $\{y : F(y) \le \lambda\}$ are closed for each $\lambda \in \mathbb{R}$. The dual functional $F^* : X \to \overline{\mathbb{R}}$ is the Legendre–Fenchel transform of F:

$$F^*(x) := \sup\big\{\langle x, y\rangle - F(y)\big\}$$

It is always closed and convex. Closed convex functionals are continuous on the (algebraic) interior of $\operatorname{dom} F$, and they have the property

$$x \in \partial F(y) \quad \Longleftrightarrow \quad \partial F^*(x) \ni y$$

where set $\partial F(y_0) := \{x \in L^* : \langle x, y - y_0\rangle \le F(y) - F(y_0),\ \forall y \in L\}$ is called *subdifferential*, and its elements are called *subgradients* (a generalization of the Gâteaux differential and gradient). In particular, $0 \in \partial F(y_0)$ implies $F(y_0) \le F(y)$ for all $y \in L$ (i.e., $\inf F = F(y_0)$). If $F(y)$ is strictly convex at y (or F^* is G-differentiable at $x \in L^*$), then $\partial F^*(x) = \{y\}$ for all $x \in \partial F(y)$. Consequently, if dual convex

functionals are both strictly convex (or G-differentiable), then $\partial F : L \to L^*$ is a bijection. Below are examples of such dual convex functionals that are used often in information theory.

Example 1 (Relative information) Given positive $y_0 \in L$, let $F : L \to \overline{\mathbb{R}}$ be:

$$F(y) = \int_\Omega \ln \frac{y(\omega)}{y_0(\omega)}\, dy(\omega) - \int_\Omega d\big[y(\omega) - y_0(\omega)\big]$$

if y is positive, $F(0) := \int_\Omega dy_0(\omega)$, and $F(y) := \infty$ for negative y. This functional is closed, strictly convex, and its G-derivative is $F'_G(y) = \ln \frac{y}{y_0}$ on the interior of $\operatorname{dom} F$. Note that $F(y) \geq 0$ for all y, because $F'_G(y_0) = 0$ and $\inf F = F(y_0) = 0$. When y and y_0 are both probability measures, then relative information is equivalent to the Kullback–Leibler divergence [13]. Relative information can be used also to represent negative entropy or Shannon mutual information.

Example 2 The dual of relative information is the following functional

$$F^*(x) = \int_\Omega e^{x(\omega)}\, dy_0(\omega)$$

Indeed, $F'_G(y) = \ln \frac{y}{y_0} = x$, and therefore $y = y_0\, e^x = F^{*\prime}_G(x)$, which is the gradient of the above functional. It is also closed, strictly convex and positive for all $x \in L^*$. Normalization of functions $y = y_0\, e^{x(\omega)}$ corresponds to transformation $F^*(x) \mapsto \ln F^*(x)$.

If $\inf F = F(0)$, then the gauge and the support functions of set $\{y : F(y) \leq I\}$ can be computed as:

$$p_F(y) = \inf\big\{\beta > 0 : F\big(\beta^{-1}\, y\big) \leq I\big\} \tag{2.3}$$

$$h_F(x) = \sup\big\{\langle x, y\rangle : F(y) \leq I\big\} \tag{2.4}$$

The support function above is the gauge of the polar set, which can also be computed as $p_{F^*}(x) = \inf\{\beta^{-1} > 0 : F^*(\beta x) \leq I^*\}$.

Thus, information functional $F : L \to \overline{\mathbb{R}}$ can be used to define a topology on the statistical manifold as the collection of all elements $y \in L$, for which set $M = \{y : F(y) \leq I\}$ is absorbing (and therefore $p_F(y) < \infty$). The topology on the dual space (the space of utility functions) is the collection of $x \in L^*$ for which the polar set is absorbing (and therefore $h_F(x) < \infty$). We shall denote these topological spaces by L_F and L^*_F.

A topology related to information $I \in \mathbb{R}$ is useful for the analysis of learning systems and their dynamics. In particular, an evolution that is continuous in information is represented by a function $y = f(I)$ that maps closed sets $(-\infty, I] \subset \mathbb{R}$ into closed sets $M = \{y : F(y) \leq I\}$ on the statistical manifold. Note that such an evolution is also order-preserving (monotonic) between $(\mathbb{R}, \leq)$ and pre-order $\lesssim$ on L_F, defined by the gauge p_F. We shall refer to such an evolution of a learning system as a continuous *information trajectory*.

2.3 Optimal Evolution and Bounds

An evolution of a learning system, even if described by a continuous trajectory, may not be optimal. As mentioned earlier, an optimal evolution is the totality of points $\bar{y} \in L_F$ maximizing information value or the expected utility $\langle x, y\rangle$ subject to information constraints. Thus, $\bar{y}$ must satisfy the extrema of (2.2) (or the support function (2.4)) for a given utility. Optimal solutions are found by the standard method of Lagrange multipliers, which we present below for completeness of exposition.

Theorem 1 (Necessary and sufficient optimality conditions) *The least upper bound* $U(I) = \sup\{\langle x, y\rangle : F(y) \le I < \infty\}$ *is achieved at* $\bar{y}$ *if and only if the following conditions are satisfied*

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

Proof The Lagrange function is $K(y, \beta^{-1}, I) = \langle x, y\rangle + \beta^{-1}[I - F(y)]$, where β^{-1} is the Lagrange multiplier corresponding to $F(y) \le I$. Zero in the subdifferential of $K(y, \beta^{-1}, I)$ gives the necessary conditions of extrema:

$$\partial_y K\big(\bar{y}, \beta^{-1}, I\big) = x - \beta^{-1}\partial F(\bar{y}) \ni 0, \quad \Rightarrow \quad \beta x \in \partial F(\bar{y})$$
$$\partial_{\beta^{-1}} K\big(\bar{y}, \beta^{-1}, I\big) = I - F(\bar{y}) \ni 0, \quad \Rightarrow \quad F(\bar{y}) = I$$

Noting that $K(\bar{y}, \beta^{-1}, I) = U(I)$, gives $\partial_I K(\bar{y}, \beta^{-1}, I) = \partial U(I) \ni \beta^{-1}$.

Sufficient conditions are obtained by considering convexity. Because F is convex and $\langle x, \cdot\rangle$ is linear, the Lagrange function is concave for $\beta^{-1} > 0$ and convex for $\beta^{-1} < 0$. Therefore, $\bar{y} \in \partial F^*(\beta x)$ with $\beta^{-1} > 0$ defines the least upper bound of $U(I)$. □

Corollary 1 *The optimal trajectory* $y = \bar{y}(I)$ *is continuous in information.*

Proof The optimality condition $F(\bar{y}) = I$ implies that $\bar{y} \in \{y : F(y) \le I\}$ for any $I \in \mathbb{R}$, and therefore $y = \bar{y}(I)$ cannot map any closed set $(\infty, I] \subset \mathbb{R}$ outside closed set $\{y : F(y) \le I\}$ in L_F. □

Example 3 When F is the relative information from Example 1, the optimal solutions are in the exponential form

$$\bar{y}(\omega) = y_0(\omega) \exp\big\{\beta\, x(\omega) - \Psi(\beta)\big\}$$

where $\Psi(\beta) = \ln \int_\Omega e^{\beta x}\, dy_0(\omega)$ from the normalizing condition. If $y_0 = \text{const}$, then optimal function $\bar{y}$ is the canonical Gibbs distribution. When the utility function is $x = -|s|^2$ (i.e., negative squared deviation), then $\bar{y}$ is Gaussian with variance $\sigma^2 = (2\beta)^{-1}$ and $e^{\Psi(\beta)} = \sqrt{\pi\beta^{-1}} = \sigma\sqrt{2\pi}$.

The totality of optimal points $\bar{y}$ can be considered as one parameter family of distributions, where parameter $\beta \in \mathbb{R}$ is the gauge of $\bar{y}$ with respect to set $\{y : F(y) \leq 1\}$, and it can be determined from the information constraint $I \in \mathbb{R}$ $(F(y) \leq I)$. Note, however, that β can also be determined from the expected utility $U = \langle x, \bar{y} \rangle$. Indeed, consider function $I(U) := \inf\{F(y) : U_0 \leq U \leq \langle x, y \rangle\}$, where $U_0 = \langle x, y_0 \rangle$. Clearly, $I(U)$ is the inverse of information value $U(I)$. The Lagrange function for $I(U)$ is $K(y, \beta, U) = F(y) + \beta[U - \langle x, y \rangle]$, and the solutions are defined by

$$\bar{y} \in \partial F^*(\beta x), \quad \langle x, \bar{y} \rangle = U, \quad \beta \in \partial I(U), \quad \beta \geq 0$$

Thus, the optimal information trajectory can be parametrized by the information or by the expected utility constraints through the inverse of mappings $\beta \mapsto F(\bar{y}(\beta)) = I$ and $\beta \mapsto \langle x, \bar{y}(\beta) \rangle = U$. These mappings can be conveniently expressed by the *generalized characteristic potentials*:

$$\Phi\big(\beta^{-1}\big) := \inf\big\{\beta^{-1} I - U(I)\big\}, \qquad \Psi(\beta) := \sup\big\{\beta U - I(U)\big\}.$$

The potentials are real functions, and the extrema in their definitions are given by conditions $\beta^{-1} \in \partial U(I)$ and $\beta \in \partial I(U)$. One can show also that $\Phi(\beta^{-1}) = -\beta^{-1}\Psi(\beta)$. The parametrization is based on the following theorem.

Theorem 2 (Parametrization) *Parameter $\beta \in \mathbb{R}$ defining solutions $\bar{y}$ to problems $U = \sup\{\langle x, y \rangle : F(y) \leq I\}$ and $I = \inf\{F(y) : U \leq \langle x, y \rangle\}$ is related to the constraints $I \in \mathbb{R}$ or $U \in \mathbb{R}$ by the following relations*

$$I \in \partial\, \Phi\big(\beta^{-1}\big), \qquad U \in \partial\, \Psi(\beta)$$

$$I \in \beta\, \partial\, \Psi(\beta) - \Psi(\beta), \qquad U \in \beta^{-1}\, \partial\, \Phi\big(\beta^{-1}\big) - \Phi\big(\beta^{-1}\big)$$

Proof Consider the Legendre–Fenchel transforms of Φ and Ψ:

$$U(I) = \inf\big\{\beta^{-1} I - \Phi\big(\beta^{-1}\big)\big\}, \qquad I(U) = \sup\big\{\beta U - \Psi(\beta)\big\}$$

The extrema are satisfied when $I \in \partial\Phi(\beta^{-1})$ and $U \in \partial\Psi(\beta)$, which is the first pair of relations. Substituting them into the Legendre–Fenchel transforms gives the second pair. □

Subdifferentials in Theorem 2 are replaced by derivatives $\Phi'(\beta^{-1})$ and $\Psi'(\beta)$ if Ψ and Φ are differentiable. This is the case when $F(y)$ is strictly convex.

Example 4 When solutions $\bar{y}$ are in the exponential from (Example 3), one obtains $U = \langle x, \bar{y} \rangle = \int x\, e^{\beta x - \Psi(\beta)}\, dy_0$, and condition $U = \Psi'(\beta)$ gives

$$\Psi(\beta) = \ln \int_{\Omega} e^{\beta\, x(\omega)}\, dy_0(\omega)$$

The above is the *cumulant generating function* of measure y_0. Potential $\Phi(\beta^{-1}) = -\beta^{-1}\Psi(\beta)$ in this case is the *free energy*.

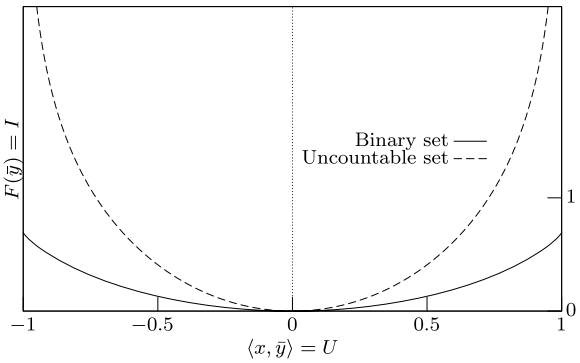

Fig. 2.1 Parametric dependencies of $I = F(\bar{y})$ on $U = \langle x, \bar{y} \rangle$ in Examples 5 and 6

Information amount is often represented by negative entropy, which corresponds to relative information F minimized at some uniform measure $y_0 = 1/|\Omega|$ (if Ω is finite) or a Lebesgue measure $dy_0 = d\omega / \int d\omega$ (if Ω is compact). Potential $\Psi(\beta)$ in these cases is

$$\Psi(\beta) = \ln \sum_{\Omega} e^{\beta x(\omega)} - \ln |\Omega| \quad \text{or} \quad \Psi(\beta) = \ln \int_{\Omega} e^{\beta x(\omega)}\, d\omega - \ln \int_{\Omega} d\omega$$

The following examples give expressions for $U(\beta)$ in two important cases.

Example 5 (Binary utility) Let $\Omega = \{\omega_1, \omega_2\}$, and $x : \Omega \to \{c - d, c + d\}$. Then using $e^{\beta(c-d)} + e^{\beta(c+d)} = 2e^{\beta c} \cosh(\beta d)$, we obtain

$$\Psi(\beta) = \beta c + \ln \cosh(\beta d), \qquad U(\beta) = c + d \tanh(\beta d)$$

Example 6 (Uncountable utility) Let Ω be compact, $x : \Omega \to [c - d, c + d] \subset \mathbb{R}$ such that $dx/d\omega = 1$. Then $\int_{\Omega} e^{\beta x(\omega)}\, d\omega = \int_{c-d}^{c+d} e^{\beta x}\, dx = 2\beta^{-1} e^{\beta c} \sinh(\beta d)$, $\int_{\Omega} d\omega = \int_{c-d}^{c+d} dx = 2d$, and we obtain

$$\Psi(\beta) = \beta c + \ln\big|\sinh(\beta d)\big| - \ln|\beta d|, \qquad U(\beta) = c + d \coth(\beta d) - \beta^{-1}$$

Functions $U = \Psi'(\beta)$ and $I = \beta \Psi'(\beta) - \Psi(\beta)$ define parametric dependency between U and I in a system evolving along the optimal information trajectory $y = \bar{y}(t)$, and it defines the following bounds on learning systems: $U(I)$ is the maximum expected utility for a given information amount; $I(U)$ is the least information amount required to achieve a given expected utility. Figure 2.1 shows $I(U)$ for functions in Examples 5 and 6 with $c = 0$ and $d = 1$.

Continuity in information, introduced earlier, allows us to consider path integrals of expected utility and information along a continuous trajectory. The upper and lower bounds on these quantities can be expressed in the following convenient form [5]. Here, we assume that Ψ and Φ are differentiable.

Theorem 3 (Optimal bounds) *Let $y = y(t)$, $t \in [t_1, t_2]$ be a continuous information trajectory of a learning system such that information $F(y) = I(t)$ and expected utility $\langle x, y\rangle = U(t)$ are increasing functions. Then*

$$\int_{y_1}^{y_2} \langle x, y\rangle \, dy \le \Psi(\beta_2) - \Psi(\beta_1)$$

$$\int_{y_1}^{y_2} F(y) \, dy \ge \Phi\left(\beta_1^{-1}\right) - \Phi\left(\beta_2^{-1}\right)$$

where $y_1 = y(t_1)$, $y_2 = y(t_2)$, and β_1, β_2 are determined from $I(t_1)$, $I(t_2)$ or $U(t_1)$, $U(t_2)$ using functions $\beta^{-1} = (\Phi')^{-1}(I)$ or $\beta = (\Psi')^{-1}(U)$, respectively.

Proof The first path integral is bounded above by a path integral along the optimal information trajectory $y = \bar{y}(t)$. Similarly, the second integral is bounded below. These path integrals exist, because the optimal trajectory is continuous in topology L_F (Corollary 1). The expected utility, $\langle x, \bar{y}\rangle = U$, in the optimal system is given by $U = \Psi'(\beta)$, where $\beta^{-1} = (\Phi')^{-1}(I)$ (Theorem 2). Similarly, the information amount, $F(\bar{y}) = I$, in the optimal system is given is $I = \Phi'(\beta^{-1})$, where $\beta = (\Psi')^{-1}(U)$. Because $I = I(t)$ and $U = U(t)$ are monotonic, the integrals do not change if the trajectory is parametrized by $\beta \in [\beta_1, \beta_2]$. Thus, path integrals along the optimal trajectory are equal to Riemann integrals $\int_{\beta_1}^{\beta_2} d\Psi(\beta)$ and $\int_{\beta_2^{-1}}^{\beta_1^{-1}} d\Phi(\beta^{-1})$. The final expressions are obtained by applying the Newton–Leibniz formula. □

2.4 Empirical Evaluation on Learning Agents

The optimal learning trajectory is not an algorithm for optimal learning. It, however, describes the equivalence class of evolutions of learning systems that is optimal with respect to a utility function x and some measure of information F. Subdifferential $\partial F^*(x)$ of its dual defines the family of optimal distributions, which depends also on the prior corresponding to the minimum of information. The points on the optimal trajectory are then computed using the amount of empirical information $I \in \mathbb{R}$ or empirical expected utility $U \in \mathbb{R}$. Moreover, because I or U are local constraints, the optimality is not asymptotic. Thus, an algorithm for nonasymptotic optimal learning in the described above sense should be such that the evolution of the system were as close as possible to the optimal information trajectory.

Here, we evaluate this idea in an experiment using an architecture for comparing different action-selection strategies in agents, described in [3]. The architecture consists of an agent placed in a virtual environment, and the main goal of the agent is to find and collect as many rewards as possible. The rewards appear in the environment stochastically according to some probability law that is unknown to the agent. The probabilities of rewards depend on some predefined initial pattern of the

environment and also on the previous actions of the agent (recall the cat and mice problem). Thus, the probability law defining the rewards is nonstationary.

The experiments, reported here, compare the performance of three agents in an environment with five states $\mathcal{Y} = \{y_1, \ldots, y_5\}$ and rewards with a binary utility $x(y) \in \{0, 1\}$. The results are reported for rewards distributed according to two initial patterns $\{p, 0, p, 0, p\}$ and $\{p, 0, 0, 0, p\}$, where $p \in [0, 1]$ is the probability $P(x = 1 \mid y_i, x = 0)$ of a reward appearing at state $y_i \in \mathcal{Y}$ with no current reward. Thus, p defines the average *reward frequency* in a state. The agent has three actions $\mathcal{Z} = \{z_1, z_2, z_3\}$—moving left, right or do nothing.

The agent selected actions based on estimates $\tilde{x}(y, z)$ of receiving a reward by taking action $z \in \mathcal{Z}$ in state $y \in \mathcal{Y}$ (i.e., $z(y) = \arg\max_z \tilde{x}(y, z)$). These estimates were computed using empirical probability $P_e(x \mid y, z)$ based on joint empirical distribution $P_e(x, y, z)$ stored in the agent's memory. Using different methods to compute $\tilde{x}(y, z)$ may result in the agent selecting different actions in the same states leading to differences in performance and empirical distributions $P_e(x, y, z)$. The empirical distribution $\bar{P}_e(x, y, z)$ of an optimal system should evolve along the optimal learning trajectory.

Three agents were compared using the following estimation methods:

$$\tilde{x}(y, z) = E\{x \mid y, z\} \tag{2.5}$$

$$\tilde{x}(y, z) = E\{x \mid y, z\} + \xi, \quad \xi \in \mathcal{N}\left(0, \sigma^2\right), \quad \sigma^2 = \mathrm{Var}\{x \mid y, z\} \tag{2.6}$$

$$\tilde{x}(y, z) = \bar{F}^{-1}(\xi), \quad \xi \in \mathrm{Rand}(0, 1), \quad \bar{F}(x) = \int_{-\infty}^{x} d\bar{P}(t \mid y, z) \tag{2.7}$$

The first agent, referred to as 'max $E\{u\}$' (max expected utility), estimates the utilities by their empirical expectations. This strategy is known to be suboptimal in some problems, and is often referred to as a *greedy* strategy. Note that max $E\{u\}$ corresponds to optimization without information constraints. Indeed, the maximum of information gives $\beta^{-1} = 0$ in Theorem 1, and the Lagrange function reduces to the expected utility. Thus, the greedy strategy 'overestimates' the amount of empirical information.

The second agent, referred to as 'Noisy $E\{u\}$', uses stochastic strategy, where the conditional expectation is randomized by ξ, sampled from zero-mean normal distribution with empirical variance. Thus, this method does not use statistics of order higher than two. Generally, this corresponds to using less information than the empirical distribution contains.

The third agent, referred to as 'Rand $ML(u)$' (for 'random maximum likelihood'), uses stochastic estimates sampled from probability measure $\bar{P}(x \mid y, z)$ that is optimal with respect to empirical information constraints. Sampling is performed using the inverse distribution function method. Note that $\bar{P}$ can be also parametrized by the empirical expected utility $U \in \mathbb{R}$, and for binary utility function $x \in \{0, 1\}$ there is only one distribution such that $E\{x\} = U$. Thus, for binary utility $\bar{P} = P_e$, and $\tilde{x}(y, z)$ are sampled directly from $P_e(x \mid y, z)$.

The results are reported on Figs. 2.2, 2.3 and 2.4. Charts on the left are for pattern $\{p, 0, p, 0, p\}$ and on the right for $\{p, 0, 0, 0, p\}$. All the points on the charts are

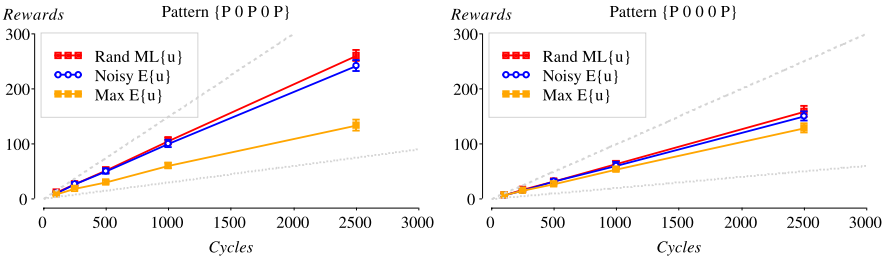

Fig. 2.2 Average numbers of rewards collected (ordinates) as a functions of cycles (abscissae) for three strategies

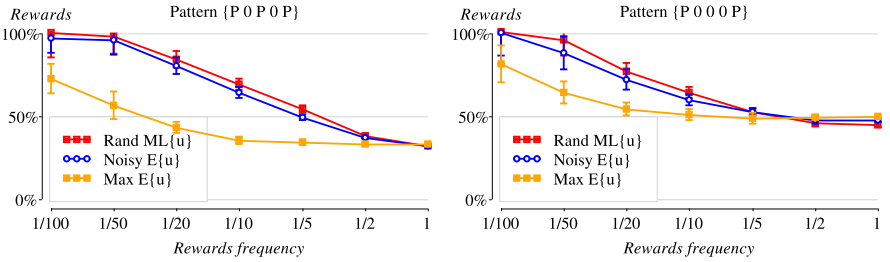

Fig. 2.3 Percentage of rewards collected as a function of rewards' frequency

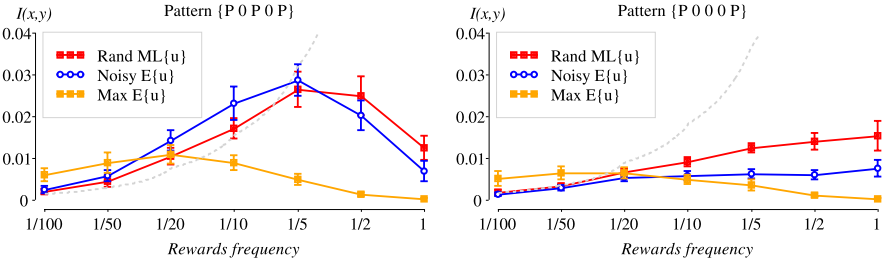

Fig. 2.4 Posterior information amount as a function of rewards' frequency

the average values from 30 experiments. The error bars on all charts are standard deviations.

Figure 2.2 shows the numbers of rewards against the number of cycles in the experiments with $p = .1$. One can see that the best performance was achieved by the Rand $ML(u)$ agent, the second is the Noisy $E\{u\}$ agent, and the least number of reward was collected by the max $E\{u\}$ agent, as expected.

Figure 2.3 shows the percentage of rewards collected by the agents after 1000 cycles in different experiments with the control probability of rewards $p \in [.01, 1]$, shown on the horizontal axis. Figure 2.4 shows, for the same experiments, the amount of Shannon information $I_{x,y}$ between rewards and states computed from the empirical distribution $P_e(x, y) = \sum_z P_e(x, y, z)$. One can see that the agent col-

lecting the greatest number of rewards also often requires the least amounts of information (particularly for $p \in [.01, .05]$). These empirical results agree with the theory, presented in previous sections.

2.5 Conclusion

This paper presented geometric representation of evolution of learning systems. The representation is related to the use of Orlicz spaces in infinite-dimensional nonparametric information geometry, but the topology considered here is based on more general convex functions on linear spaces. The duality plays a very important role. In particular, subdifferentials of dual convex functionals are (generally multi-valued) monotone operators between the dual spaces, and they set up Galois connection preserving pre-orders on the topological spaces. Monotone transformations are very desirable in our theory, because when applied to utility functions, they also preserve the preference relation (complete pre-order) on the space of outcomes. Note that pre-order (order) is not symmetric (antisymmetric) binary relation, and preserving this property was our main motivation for considering asymmetric topologies on the statistical manifold.

The topology related to information allows for the definition of continuous trajectories representing the evolution of a learning system. Optimality conditions have been formulated using the information value theory, and generalized characteristic potentials have been defined to parametrize the optimal information trajectory by empirical constraints. Path integrals along the optimal trajectory define theoretical bounds for a learning system that can be computed as a difference of the potentials at the end points of the trajectory. This result has some similarity to the gradient theorem about path independence of the integral in a conservative vector field.

The theory was illustrated not only on several theoretical examples, but also evaluated in an experiment. The results suggest that the theory can be very useful in many applications of machine learning, such as nonasymptotic optimization of systems with dynamic information, optimization of communication networks based on information value and optimization of the 'exploration-exploitation' balance in statistical decisions. The latter problem has been often approached using stochastic methods based on Gibbs distributions with unknown parameter β^{-1} (temperature). Optimality conditions $\beta^{-1} \in \partial U(I)$ or $\beta \in \partial I(U)$ define the parameter from empirical constraints, and with it the optimal level of exploration. Previously, the author applied the relation between parameter β^{-1} and information to cognitive models of human and animals' learning behavior [2], and it improved significantly the correspondence between the models and experimental data. Further development of the theory and its applications to machine learning problems is the subject of ongoing research.

Acknowledgement This work was supported in part by EPSRC grant EP/DO59720.

References

1. Amari, S.I.: Differential-geometrical methods of statistics. Lecture Notes in Statistics, vol. 25. Springer, Berlin (1985)
2. Belavkin, R.V.: On emotion, learning and uncertainty: A cognitive modelling approach. PhD thesis, The University of Nottingham, Nottingham, UK (2003)
3. Belavkin, R.V.: Acting irrationally to improve performance in stochastic worlds. In: Bramer, M., Coenen, F., Allen, T. (eds.) Proceedings of AI–2005, the 25th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence. Research and Development in Intelligent Systems vol. XXII, pp. 305–316. Springer, Cambridge (2005). BCS
4. Belavkin, R.V.: The duality of utility and information in optimally learning systems. In: 7th IEEE International Conference on 'Cybernetic Intelligent Systems'. IEEE Press, London (2008)
5. Belavkin, R.V.: Bounds of optimal learning. In: 2009 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning, pp. 199–204. IEEE Press, Nashville (2009)
6. Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton (1957)
7. Chentsov, N.N.: Statistical Decision Rules and Optimal Inference. Nauka, Moscow (1972). In Russian, English translation: Am. Math. Soc., Providence (1982)
8. de Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. Henri Poincaré **7**, 1–68 (1937). In French
9. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957)
10. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **108**, 171–190 (1957)
11. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. J. Artif. Intell. Res. **4**, 237–285 (1996)
12. Kolmogorov, A.N.: The theory of information transmission. In: Meeting of the USSR Academy of Sciences on Scientific Problems of Production Automatisation, 1956, pp. 66–99. Akad. Nauk USSR, Moscow (1957). In Russian
13. Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)
14. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Stat. **23**(5), 1543–1561 (1995)
15. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Wiley, New York (1962). Translated from Russian
16. Robbins, H.: An empirical Bayes approach to statistics. In: Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 157–163 (1956)
17. Rockafellar, R.T.: Conjugate Duality and Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 16. SIAM, Philadelphia (1974)
18. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Techn. J. **27**, 379–423 (1948)
19. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 623–656 (1948)
20. Showalter, R.E.: Monotone Operators in Banach Space and Nonlinear Partial Differential Equations. Mathematical Surveys and Monographs, vol. 49. Am. Math. Soc., Providence (1997)
21. Stratonovich, R.L.: Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise. Radiofizika **2**(6), 892–901 (1959)
22. Stratonovich, R.L.: Conditional Markov processes. Theory Probab. Appl. **5**(2), 156–178 (1960)
23. Stratonovich, R.L.: On value of information. Izv. USSR Acad. Sci. Techn. Cybern. **5**, 3–12 (1965). In Russian
24. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. Adaptive Computation and Machine Learning. MIT Press, Cambridge (1998)
25. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior, 1st edn. Princeton University Press, Princeton (1944)
26. Wald, A.: Statistical Decision Functions. Wiley, New York (1950)

http://www.springer.com/978-1-4419-5688-0

Dynamics of Information Systems
Theory and Applications
Hirsch, M.; Pardalos, P.; Murphey, R. (Eds.)
2010, XIV, 372 p. 123 illus., 78 illus. in color., Hardcover
ISBN: 978-1-4419-5688-0