In *7th IEEE International Conference on Cybernetic Intelligent Systems*, 2008.

The Duality of Utility and Information in Optimally Learning Systems

Roman V. Belavkin
School of Computing Science
Middlesex University, London NW4 4BT, UK
Email: R.Belavkin@mdx.ac.uk

***Abstract*—The paper considers learning systems as optimisation systems with dynamical information constraints, and general optimality conditions are derived using the duality between the space of utility functions and probability measures. The increasing dynamics of the constraints is used to parametrise the optimal solutions which form a trajectory in the space of probability measures. Stochastic processes following such trajectories describe systems achieving the maximum possible utility gain with respect to a given information. The theory is discussed on examples for finite and uncountable sets and in relation to existing applications and cognitive models of learning.**

I. Introduction

Learning is a process of incorporating new information by a system in order to improve its performance. Therefore, the design of such systems is closely related to theories of optimisation and information. It is not clear, however, if there exists the best way to incorporate new information, and which strategies should the system follow if information is incomplete. This paper argues that despite a very natural link between the theories of optimisation and information, many standard techniques do not take into account the information constraints, and the resulting strategies are not optimal. The optimality condition derived from the duality of utility and information can help solve problems of overfitting and the exploration-exploitation dilemma in the learning systems.

Many learning algorithms and theories are based either on statistical (asymptotic) or Bayesian approaches to optimisation under uncertainty. Without uncertainty, optimisation problems are problems of finding the extrema (minimum or maximum) of some real functions $x : \Omega \to \mathbb{R}$, which are called utilities, costs or fitness functions depending on the area of application. These functions represent someone's preference relation on the underlying choice set Ω, which can be the set of prizes, errors, states of the world and so on. Sequential optimisation can be solved in discrete or continuous time using methods of dynamic programming [5] or the maximum principle [1], [2].

Under uncertainty (i.e. when the elements of the choice set cannot be observed directly), one often applies the classical methods of statistical or Bayesian estimation. In many cases, this means maximisation of some expected utility or likelihood function, estimated from the observed data [3], [4]. Sequential problems can be solved using methods of dynamic programming [5], and the theory of conditional Markov processes allows for a significant reduction of variables [6]. These methods have inspired a number of works on statistical learning and algorithms in artificial intelligence (AI) and robotics (see [7], [8] for reviews).

It is important to note, however, that the above mentioned classical techniques were originally developed in the theory of optimal control, and they require sufficient knowledge (a model) of the systems (deterministic or stochastic). Therefore, they were not designed for the learning systems that are characterised by incomplete information. Optimality of many algorithms is asymptotic, meaning that their methods can be applied only after the probability distributions are known with sufficient precision. However, a system often has to make a decision before such information is available. Furthermore, many problems of learning and choice under uncertainty violate some of the basic assumptions of the statistical theory (e.g. the weak low of large numbers), such as the independence of trials, stationary distributions and so on. Therefore, conclusions about the convergence of the statistical techniques may no longer be valid.

It is becoming increasingly apparent that the learning theory is closely related to the theory of information [9]. The use of information theoretic concepts has significantly advanced data and signal processing algorithms [10], [11]. Recently, entropy and information dynamics has been applied in cognitive models of learning in human subjects and animals, and it helped the models to explain experimental data [9], [12]. This paper presents theoretical considerations of these observations based on a generalisation of variational problems in information theory [13], [14], [15].

In the first section of the paper, we develop some generalisations of optimisation problems under uncertainty based on methods of functional analysis and information theory. It will be shown that constraints on information define polar sets in the dual functional spaces. Then we derive optimality conditions using standard methods of convex analysis and apply them to systems with dynamical constraints. In the following section, we discuss the optimal solutions in relation to the classical (i.e. Bayesian) techniques as well as some existing parametric stochastic approximation methods. It will be shown on examples how the constraints on utility or information can be used to derive the optimal parameter values defining the optimal strategies. The final example will illustrate the totality of all optimal measures that corresponds to the trajectory of the optimal learning system in the pre-dual space.

II. Duality of Utility and Information

In this section, we shall discuss mathematical duality between the real functions $x : \Omega \to \mathbb{R}$ (e.g. the utility functions, used in optimisation problems) and their (pre)-dual functions $y : \Omega \to \mathbb{R}$ — measures representing the beliefs or information about the uncertain domain (e.g. the probability measures). This duality has been known in information theory [13], [14], [15], where it was used to define the maximum channel capacity and value of information. Here, we shall apply this duality to cybernetic systems, characterised by increasing information dynamics (i.e. learning systems), and define the optimality conditions for these systems. These conditions and their implementations will be discussed on examples in the following section.

The performance of many cybernetic systems is measured by some real functions $x : \Omega \to \mathbb{R}$ that may represent errors, costs, utilities and so on. Optimal performance is achieved by choosing elements $\omega \in \Omega$ corresponding to extreme values of these functions. In fact, the utility functions represent the *preference* relation on the domain Ω — total preorder (total, reflexive and transitive binary relation) — that is a more fundamental concept than the utility because not all preference relations can be represented by real functions. The necessary and sufficient condition for the existence of a utility representation of a preference relation on Ω is the existence of an order-dense countable subset in Ω [3]. Therefore, we shall consider utility functions as the elements of real functional space X with countable basis.

Under uncertainty, the choice between the elements of set Ω is replaced by the choice between lotteries over these elements. Given a measurable space $(\Omega, \mathcal{F})$, where $\mathcal{F} \subseteq 2^{\Omega}$ is σ-algebra of subsets, the lotteries can be represented by different measures — non-negative functions $\mu : \mathcal{F} \to \mathbb{R}$. For example, measures such that $\mu(\Omega) = 1$ are the probability measures. The choice between the lotteries is solved using the preference relation defined by the *expected utility* functional

$$E_{\mu}\{x\} = \int_{\Omega} x(\omega)\, d\mu \tag{1}$$

Thus, for any probability measures μ and ν, measure μ is preferred if $E_{\mu}\{x\} \geq E_{\nu}\{x\}$. Maximisation or minimisation of expected values is used in the Bayesian procedures of parameter estimation [3], [4] and for optimal control under uncertainty [5].

One can see from (1) that utility functions $x \in X$ are summable with respect to measures μ, ν being compared. Moreover, expected value (1) is a continuous linear functional on the space of measures, and tight (Radon) measures, such as the probability measures, correspond to non-negative linear functionals on the space $\mathcal{C}_c(\Omega)$ of continuous functions with compact support. Thus, utility functions that we need to consider are the elements of space X, dual of linear space Y containing the measures, and Y is the dual of space $\mathcal{C}_c(\Omega)$ (i.e. X is the second dual of $\mathcal{C}_c(\Omega)$, and Y is the pre-dual of X, with the inclusion $\mathcal{C}_c(\Omega) \subset Y \subset X \subset \mathbb{R}^{\Omega}$). We denote linear functionals by bilinear form $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}$ represented by inner product:

$$\langle x, y \rangle = \int_{\Omega} x(\omega) y(\omega)\, d\omega$$

where $d\omega$ is the reference (e.g. Lebesgue) measure. Other measures μ that are absolutely continuous with respect to $d\omega$ correspond to functions $y(\omega) = d\mu(\omega)/d\omega$ (the Radon-Nikodym derivatives or the density functions). Thus, optimisation problems under uncertainty can be described as maximisation of the linear functionals $\langle x, y \rangle$ over some sets $K_x \subseteq X$ or $K_y \subseteq Y$.

Convex bodies (convex sets with interior points) are defined by their support or the distance functions. The *support* function of convex body $K_x \subseteq X$ is

$$C(y) = \sup\{\langle y, x \rangle : x \in K_x\} \tag{2}$$

If $x_0 \in \mathrm{Int}(K_x)$ is some reference function, then the *distance* of $x \in X$ from x_0 is defined as

$$D(x) = \inf\{D \geq 0 : x - x_0 \in D K_{x - x_0}\} \tag{3}$$

where $x - x_0$ denotes the translation to the origin. Similarly, one can define the support and the distance functions for some convex set $K_y \subseteq Y$ in the (pre)-dual space. Convex bodies K_x and K_y are called *polar* (or *dual*) if the distance function of one is the support function of another. The constraints $C(y) \leq C < \infty$ on the support function of convex body K_x are also the constraints on the distance function of the polar set K_y, and therefore the constraints either $C(y) \leq C$ or $D(x) \leq D$ uniquely define both polar convex sets K_x and K_y. We now demonstrate how this duality can be used to describe the optimally learning systems.

Let us characterise the set $K_y \subseteq Y$ that represents the domain of the optimisation problem under uncertainty. First, set K_y belongs to non-negative cone of space Y, because measures are non-negative linear functionals. Second, normalisation condition, such as $\int_{\Omega} y(\omega)\, d\omega = \mu(\Omega) = 1$ for probability measures, corresponds to a cross-section of the cone, and it is a convex hull (simplex) with the extreme points corresponding to Dirac δ-measures: $\delta_{\omega}(d\omega) = 1$ if $\omega \in d\omega$; $\delta_{\omega}(d\omega) = 0$ otherwise. Finally, the δ-measures have to be excluded from consideration under uncertainty, because they correspond to the complete certainty (i.e. $\delta_{\omega}(d\omega) \notin K_y$). Observe also that without this condition any continuous linear functional achieves the maximum on a simplex in one of its extreme points (i.e. the Dirac δ-measures).

The uncertainty excluding δ-measures can be defined by constraints on the *information divergence* (also *relative entropy*) functional [18]:

$$C_{\nu}(\mu) = \int_{\Omega} \ln \frac{d\mu}{d\nu}\, d\mu = \int_{\Omega} \ln \frac{y(\omega)}{y_0(\omega)}\, y(\omega)\, d\omega \tag{4}$$

where μ, ν are Radon measures such that μ is absolutely continuous with respect to ν. Some well-known properties of information divergence are that it is strictly convex, and its unique minimum is achieved when measure μ is proportional to ν (e.g. when $y = y_0$). If the reference measure ν is sufficiently broad (i.e. belongs to the interior of the simplex K_y), then the maximum value of information divergence (possibly infinite) is achieved when $\mu(d\omega) \propto \delta_\omega(d\omega)$. Thus, the δ-measures can be excluded from set K_y by constraints on information divergence $C_\nu(\mu) \leq C < \infty$. It will be shown in the next section how these constraints can be interpreted as the lower bound of the entropy or as the upper bound of Shannon information. We now apply the Kuhn-Tucker theorem to formulate the necessary optimality conditions for problem (2) with set K_y defined by constraints on information divergence.

Theorem 1: Let X and Y be dual Banach spaces. Given $x \in X$, the extrema $\mu^* \in K_y$ of optimisation problem $D = \sup\{\langle x, \mu\rangle : C_\nu(\mu) \leq C < \infty\}$, where $C_\nu(\mu)$ is information divergence (4), satisfy the following conditions

$$\mu^* \propto \nu\, e^{\beta x(\omega)}\,, \quad C_\nu(\mu^*) = C\,, \quad \beta = C'(D) \tag{5}$$

Proof: The Lagrangian function is

$$L(\mu, \beta) = \langle x, \mu\rangle + \beta^{-1}[C - C_\nu(\mu)]$$

where β^{-1} is the Lagrange multiplier corresponding to the constraint $C_\nu(\mu) \leq C$. The necessary conditions of extremum are

$$\begin{aligned} \partial_\mu L(\mu^*, \beta^{-1}) &= x - \beta^{-1}\partial C_\nu(\mu^*) = 0 \\ \partial_{\beta^{-1}} L(\mu^*, \beta^{-1}) &= C - C_\nu(\mu^*) = 0 \end{aligned}$$

where ∂ denotes the Gâteaux derivative, and $\partial C_\nu(\mu) = \ln\frac{\mu}{\nu}$. Thus, the first condition gives $\beta\, x = \ln\frac{\mu}{\nu}$. By considering the extreme value D as a function of constraint C, its derivative is $D'(C) = \beta^{-1}$, which gives the last condition. □

Using the fact that $C_\nu(\mu)$ is strictly convex, one can show also that the extremum of Theorem 1 is the maximum for $\beta \geq 0$ and the minimum for $\beta < 0$ (because the Lagrangian is respectively concave and convex). In addition, conditions $\beta = 0$ or $x(\omega) = \text{const}$, correspond to the minimum of information divergence $\mu^* \propto \nu$. In fact, conditions (5) can be derived by solving the following dual minimisation problem $C = \inf\{C_\nu(\mu) : \langle x, \mu\rangle \geq D > -\infty\}$.

For normalised measures, the first equation of conditions (5) becomes

$$\mu^* = \nu\, e^{\beta x - \Gamma(\beta)}$$

where function $\Gamma(\beta)$ is found from the following integral equation $\int d\mu = e^{-\Gamma(\beta)}\int d\nu e^{\beta x} = 1$, and therefore $\Gamma(\beta) = \ln\int e^{\beta\, x}\, d\nu$. Thus, $\Gamma(\beta)$ is the *semi-invariant generating* function of the reference measure ν. One can see that the optimal measures belong to exponential family. Parameter $\beta \in \mathbb{R}$ is determined by the conditions $C_\nu(\mu^*) = C$, $\beta = C'(D)$ or $\langle x, \mu^*\rangle = D$, $\beta^{-1} = D'(C)$.

Theorem 2: Maximisation of linear functional $\langle x, \mu\rangle$ under the constraints on information divergence $C_\nu(\mu) \leq C$ corresponds to the canonical equations

$$D = \Gamma'(\beta)\,, \quad \beta = C'(D) \tag{6}$$

where $C = C(D)$ is the Legendre-Fenchel transform of potential $\Gamma(\beta) = \ln\int e^{\beta\, x}\, d\nu$.

Proof: Let $\langle x, \mu^*\rangle = D$ be the value of the linear functional for the optimal measure μ^*, and let $C = C(D)$ be the Legendre-Fenchel transform of function $\Gamma(\beta)$:

$$C(D) = \sup[D\beta - \Gamma(\beta)]\,, \quad \Gamma(\beta) = \sup[\beta D - C(D)]$$

The first expression above is, in fact, the integral of the necessary condition of extremum: $\ln\frac{\mu^*}{\nu} = \beta\, x - \Gamma(\beta)$ over normalised measure $\mu^*(\Omega) = 1$. Canonical equations (6) follow from the common properties of the Legendre-Fenchel transform. □

The analysis shows that potential $\Gamma(\beta)$ is strictly convex, and therefore its derivative is a strictly increasing function. Thus, the optimal value of parameter β can be determined from the constraint $\langle x, \mu\rangle \geq D$ using the inverse function $\beta = (\Gamma')^{-1}(D)$. Similarly, β^{-1} can be determined from constraint $C_\nu(\mu) \leq C$ using potential $F(\beta^{-1}) = -\beta^{-1}\Gamma(\beta)$, also know as the *free energy*. The corresponding Legendre-Fenchel transforms are

$$D(C) = \inf[C\beta^{-1} - F(\beta^{-1})],\ F(\beta^{-1}) = \inf[\beta^{-1}C - D(C)]$$

which give the following conditions: $C = F'(\beta^{-1})$, $\beta^{-1} = D'(C)$. The free energy $F(\beta^{-1})$ is both convex and concave (depending on the sign of β^{-1}). Thus, its derivative has both increasing and decreasing branches, and its inverse $\beta^{-1} = (F')^{-1}(C)$ is a relation.

The necessary conditions of optimality (5) and potentials $\Gamma(\beta)$ or $F(\beta^{-1})$ define the optimal measures that minimise information amount with respect to a given expected utility constraint, or equivalently maximise the expected utility for the given amount of information. Let us now extend this duality to learning systems, which are characterised by an increasing dynamics of information and expected utility.

Theorem 3: Let $C = C(t)$, $D = D(t)$ be monotone functions describing the constraints $C_\nu(\mu) \leq C < \infty$ on information divergence or $\langle x, \mu\rangle \geq D > -\infty$ on the linear functional in a dynamical system $\mu = \mu(t)$, $t \in [t_1, t_2]$. Then

$$\begin{aligned} \int_{\mu(t_1)}^{\mu(t_2)} \langle x, \mu(t)\rangle\, d\mu(t) &\leq \Gamma(\beta_2) - \Gamma(\beta_1) \\ \int_{\mu(t_1)}^{\mu(t_2)} C_\nu(\mu(t))\, d\mu(t) &\geq F(\beta_1^{-1}) - F(\beta_2^{-1}) \end{aligned}$$

where β_1, β_2 are determined from $C(t_1)$, $C(t_2)$ or $D(t_1)$, $D(t_2)$ using functions $\beta^{-1} = (F')^{-1}(C)$ and $\beta = (\Gamma')^{-1}(D)$.

Proof: is shown trivially by substituting the optimality conditions $\langle x, \mu^*\rangle = D = \Gamma'(\beta)$ and $C_\nu(\mu^*) = C = F'(\beta^{-1})$ into the integrals, and then applying the Newton-Leibniz formula. □

III. Applications and Examples

As discussed earlier, the Bayesian approach to optimisation under uncertainty is to find the extrema of conditional expectations [3], [4] (i.e. minimisation of risk or maximisation of the expected utility). In particular, many applications can be described by set $\Omega = S \times A$, where set S is interpreted as some input variable (e.g. observations of some random phenomenon), and set A as some output variable (e.g. parameter estimates, actions or control functions). In this case, the optimal $a^* \in A$ is often defined by the maximum of conditional expectation

$$a^* = \arg\max_{a \in A} \left\{ E\{x \mid a\} = \int_S x(a,s) P(ds \mid a) \right\} \qquad (7)$$

where $x : A \times S \to \mathbb{R}$ is some utility function (e.g. minus error or minus cost). If the maximum exists, then in this case one has to always choose the optimal value a^*. Thus, the marginal probability measure on set A is a δ-measure $P(da) = \delta_{a^*}(da)$ (i.e. $P(da) = 1$ if $a^* \in da$; $P(da) = 0$ otherwise). This is clearly a deterministic strategy. We now show that this strategy does not satisfy the optimality conditions taking into account the constraints on information (as in previous section).

Indeed, if $P(da) = \delta_{a^*}(da)$, then $P(ds, da) = P(ds \mid a)P(da) = 0$ for all da such that $a^* \notin da$. According to equations (5), the optimal measure is zero only if either $\nu(ds, da) = 0$, $x(s, a) = -\infty$ or $\beta = \pm\infty$. The former two conditions are not feasible (the reference measure is usually chosen such that $\nu(ds, da) > 0$); the latter condition implies $\beta^{-1} = D'(C) = 0$, which means that the maximum of the expected utility does not change with respect to a change of information. In fact, condition $\beta \to \infty$ corresponds to the maximum value of information divergence (which can be infinite). This means that deterministic strategies, such as the decisions simply maximising the expected utility, are optimal only in systems that have collected all (possibly infinite) information. In learning systems that do not have full information $\beta < \infty$, and therefore such deterministic strategies are not optimal. This observation may explain the problems of 'overfitting' in many estimation methods based on the classical Bayesian inference procedures, because they do not take into account information constraints.

The constraints on information divergence have very important statistical interpretations. When μ is a standard probability measure, and ν is a constant (e.g. $\nu(d\omega) = d\omega$), then information divergence corresponds to minus entropy. The maximum of entropy is achieved when probability measure is also constant (i.e. uniform distribution). Therefore, condition $-C_\nu(\mu) \geq -C$ is the lower bound of the uncertainty.

Another interpretation of information constraints is as the constraints on the Shannon information amount between random variables $a \in A$ and $s \in S$ [13]:

$$I_{a,s} = \int_{A \times S} P(da, ds) \ln\left[\frac{P(ds \mid a)}{P(ds)}\right] \qquad (8)$$

One can see that the expression above corresponds to information divergence $C_\nu(\mu)$, where μ is the conditional (or joint) probability measure, and ν is the marginal (product of marginals) probability. The minimum of Shannon information is achieved when variables a and s are independent ($P(ds \mid a) = P(ds)$). Note that optimisation of function (7) requires at least some dependency between random variables s and a. In this case, $P(ds \mid a) \neq P(ds)$ and $I_{a,s} > 0$. In particular, functional (deterministic) dependency $s = f(a)$ corresponds to δ-functions $P(ds \mid a)$ and the maximum information amount (possibly infinite). Thus, the optimisation problem with constraints on Shannon information are of a particular interest. Its solution can be written using the general measure form (5) as follows

$$P_\beta(da \mid s) = P(da)\, e^{\beta x(a,s) - \Gamma(\beta,s)} \qquad (9)$$

where $\Gamma(\beta, s) = \ln \int_A e^{\beta x(a,s)} P(da)$, which follows from $\int_A P(da, ds) = P(ds)$. In particular, for $P(da) = \text{const}$, the optimal solution is the canonical Gibbs distribution

$$P_\beta(da \mid s) = \frac{e^{\beta x(a,s)}}{\int_A e^{\beta x(a,s)} da} \qquad (10)$$

where da is the Lebesgue measure. Note that for $\beta \geq 0$ and $x(a, s) > -\infty$, the probability above $P_\beta(da \mid s) > 0$ for all $a \in A$. Thus, in any optimisation problem with incomplete information the deterministic strategies are indeed not optimal.

It has been known for a long time that stochastic techniques are generally more successful in optimisation problems, where the standard asymptotic techniques preform poorly. However, the traditional stochastic techniques [16], [17] are parametric, and the optimal values of parameters controlling randomisation are not known. In machine learning literature, this is known as the exploration-exploitation dilemma. The optimality conditions (5) not only define the optimal family of probability measures, but also the optimal values of parameter β controlling randomisation (β^{-1} is sometimes called the *temperature* in analogy with thermodynamics, where the Gibbs or Boltzmann distributions were first applied). The potentials $\Gamma(\beta)$ and $F(\beta^{-1})$ can be used to derive functions to compute β from the constraints. We now illustrate the method on two examples, where $C_\nu(\mu)$ is understood as minus entropy.

Example 1: Let Ω be a finite set, $\mu(\omega) = P(\omega)$ be the probability distribution, and let $\nu(\omega) = 1$. Thus, $C_\nu(\mu)$ is minus entropy. Let $\Omega = \{\omega_1, \omega_2\}$ and $x(\omega) = \{c - d, c + d\}$. In this case, the corresponding functions become

$$\begin{aligned} Z(\beta) &= e^{\beta(c-d)} + e^{\beta(c+d)} = 2e^{\beta c}\cosh(\beta d) \\ \Gamma(\beta) &= \ln Z(\beta) = \ln 2 + \beta c + \ln\cosh(\beta d) \\ D(\beta) &= \Gamma'(\beta) = c + d\tanh(\beta d) \end{aligned}$$

Figure 1 shows the graph of $D(\beta)$ for $c = 0$ and $d = 1$.

Example 2: Let the range of function $x : \Omega \to \mathbb{R}$ be the interval $[c - d, c + d]$ of the real line, and let us consider the divergence of probability measure $P(dx)$ with respect to Lebesgue measure dx. Thus, $C_{dx}(P(dx))$ is minus differential entropy of probability density function $p(x)$, $x \in [c-d, c+d]$. In this case, the optimal density function is the Gibbs distribution $p(x) = e^{\beta x - \Gamma(\beta)}$, where $\Gamma(\beta) = \ln Z(\beta)$, and

$$\begin{aligned} Z(\beta) &= \int_{c-d}^{c+d} e^{\beta x} dx = \frac{2e^{\beta c}}{\beta}\sinh(\beta d) \\ \Gamma(\beta) &= \beta c - \ln|\beta/2| + \ln|\sinh(\beta d)| \\ D(\beta) &= c - \beta^{-1} + d\coth(\beta d) \end{aligned}$$

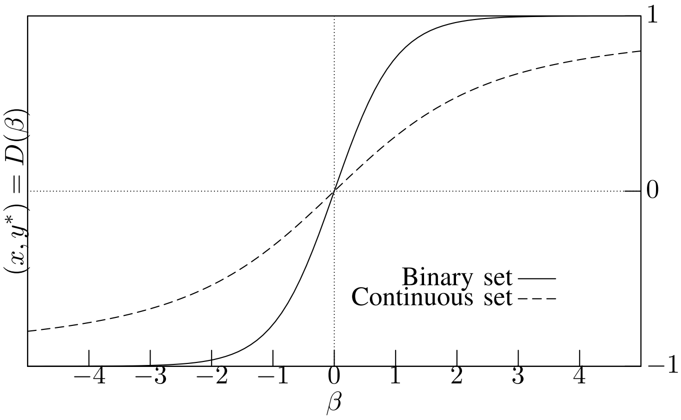

Fig. 1. Functions $D(\beta)$ in Examples 1 (binary set) and 2 (continuous set).

Figure 1 shows the graph of $D(\beta)$ for $c = 0$ and $d = 1$. When information divergence changes from I_1 to I_2 corresponding to β_1, β_2, the maximum change of utility is

$$\Gamma(\beta_2) - \Gamma(\beta_1) = \ln \frac{\beta_1}{\beta_2} + c(\beta_2 - \beta_1) + \ln \left[\frac{\sinh(\beta_2\, d)}{\sinh(\beta_1 d)} \right]$$

Clearly, zero change of information gives zero change of the expected utility.

Note that set Ω in Example 1 corresponds to an elementary algebra of two events, and the corresponding utility functions and measures are the elements of two-dimensional function spaces X and Y. However, in Example 2, set Ω is a continuum, and the corresponding function spaces are infinite-dimensional. Figure 1 illustrates the similarity between functions $D(\beta)$ for these cases. These functions can be used to compute parameter β from the constraints $\langle x, \mu \rangle \geq D$ on the expected utility or constraints $C_\nu(\mu) \leq C$ on information. The constraints can be derived from data at each time moment. The totality of optimal functions with respect to the dynamic constraints defines the trajectory $\mu^* = \mu^*(t)$ of probability measures in the (pre)-dual space Y corresponding to the optimal learning process, because it maximises the expected utility (or minimises information) at every point.

The properties of the expected utility and information divergence functionals are illustrated on Figure 2 using probability triangle, which can be explained as follows. Consider probability measures defined on a set of three elements $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Thus, any probability measure on Ω is a set of three elements $P(\omega) = \{P_1, P_2, P_3\}$. The charts on Figure 2 represent P_1 on abscissa and P_3 on the ordinate ($P_2 = 1 - P_1 - P_3$). The totality of all probability distributions $P(\omega)$ is the triangle (projection of the simplex onto the $P_1 P_3$-plane). Let the corresponding set of utility values be $x \in \{0, 1, 2\}$ representing preference relation $\omega_1 \leq \omega_2 \leq \omega_3$. The conditions $E\{x\} = \text{const}$ define the level sets of the expected utility, which are shown by straight lines on the left chart of Figure 2. Let $C_\nu(\mu) = -H\{p\}$ be the entropy. The conditions $H\{p\} = \text{const}$ define the level sets of the entropy shown on the right chart of Figure 2.

The left chart Figure 3 shows the extrema. The totality of all extreme measures is shown on the right chart of Figure 3, and they are exponential distributions parametrised by β. Clearly, if information constraints change from some initial value C_1 to C_2, then the optimal trajectory for $\beta \geq 0$ belongs to this set of exponential distributions. One can also see that the trajectory belongs to the interior of the probability triangle. Thus, $P(\omega) > 0$ for all $\omega \in \Omega$, which confirms that optimal learning is a stochastic process.

IV. Discussion

This paper considered the problem of optimal learning as a variational problem of expected utility maximisation with dynamical information constraints. This work builds on information theory, applying it to the problems of optimal learning. The optimal solutions form a trajectory in the (pre)-dual space, and it has some relation to stochastic approximation and optimisation techniques.

This work is related to the previous research of the author on cognitive models of learning in human subjects and animals [9], [19]. In these works, the entropy feedback from the posterior probability was used to control β parameter in action selection algorithm of the architecture. This modification alone has significantly improved the performance of the cognitive models [9], [19]. A similar stochastic control has been implemented in the agent architecture studying optimal action selection strategies and adaptation of agents in stochastic environments [20]. The duality methods and information-theoretic analysis, applied in this paper, allow for a solid theoretical justification of these results.

It is important to mention the relation of the presented theory to the methods based on the classical procedures of Bayesian risk minimisation or expected utility maximisation. These methods include sequential optimisation under uncertainty based on dynamic programming [5]. It was shown here that the resulting strategies are optimal only in the case of complete information; taking into account incomplete information leads to different solutions. This observation may explain problems encountered in many systems, generally referred to as overfitting and the exploration-exploitation dilemma. Note that sequential optimisation does not address the problem of information constraints because new information about probability distributions can only be obtained through measurements, not through computations.

This work shows how information constraints provide additional criteria for optimisation. The resulting theory can be applied to all optimisation problems under uncertainty including sequential optimisation and learning. This work is therefore complimentary to the existing techniques, and their integration should be the subject of future research.

Acknowledgement

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

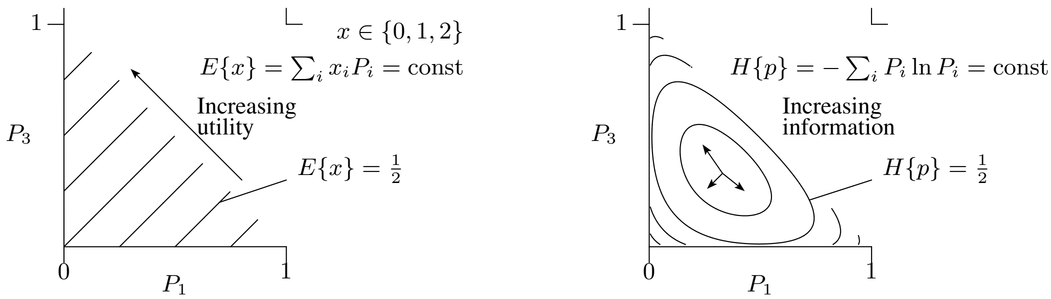

Fig. 2. Probability triangle representation of the expected utility ($E\{x\}$, left chart) and entropy ($H\{p\}$, right chart). The lines and curves correspond to the level sets of the functionals (i.e. $E\{x\} = \mathrm{const}$ on the left and $H\{p\} = \mathrm{const}$ on the right chart respectively).

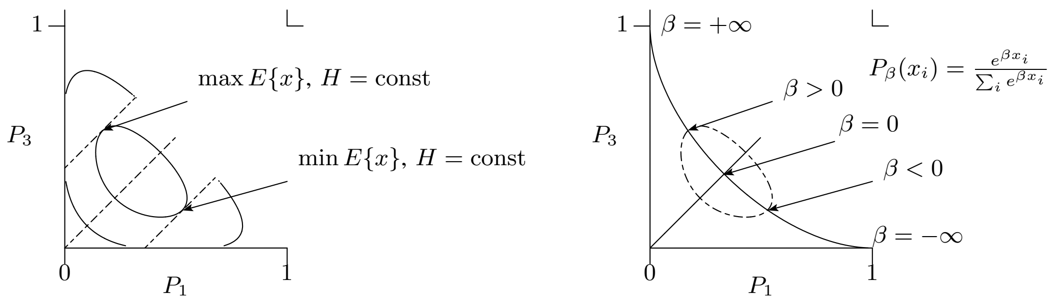

Fig. 3. Left: Solutions of the variational problem corresponding to the extrema of $E\{x\}$ for $H\{p\} = \mathrm{const}$; Right: The totality of all optimal solutions parametrised by β for all values of $H\{p\}$.

REFERENCES

[1] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The mathematical theory of optimal processes*. John Wiley and Sons, 1962, translated from Russian.

[2] I. M. Gelfand and S. V. Fomin, *Calculus of variations*. Prentice–Hall, 1963, translated from Russian.

[3] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, 1st ed. Princeton, NJ: Princeton University Press, 1944.

[4] A. Wald, *Statistical Decision Functions*. New York: John Wiley & Sons, 1950.

[5] R. E. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton University Press, 1957.

[6] R. L. Stratonovich, "Conditional Markov processes," *Theory of Probability and its Applications (USSR)*, vol. 5, no. 2, pp. 156–178, 1960.

[7] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[8] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 1998.

[9] R. V. Belavkin, "On emotion, learning and uncertainty: A cognitive modelling approach," PhD thesis, The University of Nottingham, Nottingham, UK, August 2003.

[10] A. Hyvärinen and E. Oja, "A fast fixed point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[11] E. Learned-Miller and J. Fisher, "ICA using spacings estimates of entropy," *Journal of Machine Learning Research*, vol. 4, pp. 1271–1295, 2003.

[12] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[14] A. N. Kolmogorov, "The theory of information transmission," in *Meeting of the USSR Academy of Sciences on Scientific Problems of Production Automatisation, 1956*. USSR Academy of Sciences, 1957, pp. 66–99, in Russian.

[15] R. L. Stratonovich, "O tsennosti informatsii," *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, vol. 5, pp. 3–12, 1965, in Russian.

[16] H. Robbins and S. Monro, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, September 1951.

[17] S. Kirkpatrick, C. D. Gelatt, and J. M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[18] S. Kullback, *Information Theory and Statistics*. John Wiley and Sons, 1959.

[19] R. V. Belavkin and F. E. Ritter, "The use of entropy for analysis and control of cognitive models," in *Proceedings of the Fifth International Conference on Cognitive Modeling*, F. Detje, D. Dörner, and H. Schaub, Eds. Bamberg, Germany: Universitäts–Verlag Bamberg, April 2003, pp. 21–26, iSBN 3-933463-15-7.

[20] R. V. Belavkin, "Acting irrationally to improve performance in stochastic worlds," in *Proceedings of AI–2005, the 25th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, ser. Research and Development in Intelligent Systems, M. Bramer, F. Coenen, and T. Allen, Eds., vol. XXII, BCS. Cambridge: Springer, December 2005, pp. 305–316, iSBN 1-84628-225-X.