Mutation and Optimal Search of Sequences in Nested Hamming Spaces

Roman V. Belavkin
School of Engineering and Information Sciences
Middlesex University, London NW4 4BT, UK
Email: R.Belavkin@mdx.ac.uk

***Abstract*—A representation of evolutionary systems is defined by sequences in nested Hamming spaces, which is analogous to variable length coding. Asexual reproduction is considered as a process of finding optimal codes, and conditions are formulated under which the optimal search operator is a simple random mutation, corresponding to binomial or Poisson process. Transition probability between spheres around an optimal sequence in each Hamming space is derived and used for optimal control of mutation rate. Several control functions are discussed, including a minimal information control that has a number of interesting properties. The theory makes a number of predictions about variability of length and mutation of DNA sequences in biological organisms.**

I. Introduction

The discovery of DNA has prompted an interest in applying information-theoretic ideas to biology. Kolmogorov suggested that probabilistic interpretation of information can be related to 'hereditary information' and adaptation based on the mechanisms of mutation [1]. A relation of information to fitness has also been explored recently [2]. Naturally, relations between optimality criteria, such as a fitness function, and coding of information have long been the subject of information and information utility theories [3], [4]. Here, we develop this approach further to understand better some properties of biological systems.

In the next section, we define a collection of nested Hamming spaces to represent organisms with variable length $l \in \mathbb{N}$ of DNA sequences. This is motivated by the fact that mutation rates in many modern biological organisms approximately equal $1/l$, but l varies significantly [5]. The theory compares evolution to a search for optimal codes in Hamming spaces of sequences with increasing lengths.

Sections III–V are devoted to optimal search operators and their controls. In particular, we show that under certain condition a simple random mutation implements an optimal search operator. Biological organisms are known to have some control over mutation rates via the mutation repair mechanisms, and in Section V we show different ways of controlling mutation rates. Optimal search in each space quickly leads to small mutation rates below $1/l$. Information theoretic approach allows us to define optimal control with a number of additional interesting properties, such as maintaining diversity in a population.

II. Representation in Nested Hamming Spaces

Let Θ be the set of all habitats, and let Ω be the set of all individual organisms, each of which is represented by a sequence $(\alpha_1, \dots, \alpha_l)$ of length l and α letters. We consider sequences with the same length as elements of a Hamming space $\mathcal{H}_\alpha^l := \{1, \dots, \alpha\}^l$ (i.e. equipped with the Hamming metric $d(a,b) := |i : a_i \neq b_i|$). Thus, the sets of sequences of length up to l and the set of all finite sequences are respectively:

$$\Omega_l = \bigcup_{k=1}^{l} \mathcal{H}_\alpha^k\,, \quad \Omega = \bigcup_{l\in\mathbb{N}} \mathcal{H}_\alpha^l$$

Clearly, $\Omega_l \subset \Omega_{l+1}$ (nested sets of Hamming spaces). The population of all living organisms can be considered as a code $\kappa : \Theta \to \Omega$ or a randomised code defined by a joint probability measure $P(\omega, \theta)$.

Many results of noiseless coding theory can be applied to this representation. For example, if habitats θ (source signals) have entropically sable distribution $P(\theta)$ (e.g. if θ are statistically independent), then for all perfectly decodable codes $\kappa(\theta) = \omega$ the expected length satisfies inequality

$$\mathbb{E}_P\{l(\kappa(\theta))\} \geq H\{\theta\}/\ln\alpha$$

where $H\{\theta\} := \mathbb{E}_P\{-\ln P(\theta)\}$ is the (expected) entropy of θ. This inequality predicts that organisms, occupying more 'exotic' or specialised habitats should generally have longer DNA. Of course, this simplification does not take into account one of the most important concepts of evolutionary theory — selection according to fitness.

Selection is a preference relation $\lesssim_\theta$ (total pre-order) on Ω, defining which organisms are better adapted to a particular habitat θ. This preference relation is represented (not uniquely) by a fitness function $f : \Omega \times \Theta \to \mathbb{R}$:

$$a \lesssim_\theta b \quad \iff \quad f(a,\theta) \leq f(b,\theta)$$

When a particular habitat θ is fixed, then we shall use simpler notation $\lesssim$ and $f(\omega)$. Fitness is usually interpreted (or measured) as a replication rate, and for this reason often defined to have non-negative values. This is, however, not essential, as the same pre-order $\lesssim$ can be induced by a function with negative values. From the point of optimisation or information theories, the values $x = f(\omega)$ can be thought of as utilities or negative costs.

An optimal (top) element for habitat θ will be denoted $\top(\theta)$ or $\top$, if there is no confusion about θ, and it exists if $f(\top) = \sup f(\omega)$. We shall denote by f_l and $f_{\leq l}$ restrictions of the fitness function to subsets $\mathcal{H}_\alpha^l$ and Ω_l. Clearly, top elements exist in every finite subset $\mathcal{H}_\alpha^l$ or Ω_l, but not necessarily in Ω. The inclusion $\Omega_l \subset \Omega_{l+1}$ implies

$$\sup f_{\leq l}(\omega) \leq \sup f_{\leq l+1}(\omega)$$

Thus, if habitat θ is occupied by $\top_l \in \Omega_l$ (not necessarily unique), then other organisms cannot 'outcompete' them unless their DNA has greater length. This suggests that competition for a specific habitat should generally lead to increasing complexity. We now define the following property of a fitness function.

Definition 1 (Monotonic landscape)**.** Let (Ω, d) be a metric space, and let $f : \Omega \to \mathbb{R}$ be a function with $f(\top) = \sup f(\omega)$ for some $\top \in \Omega$. We say that f is *locally monotonic* (*locally isomorphic*) relative to metric d if for each $\top$ there exists a ball $B(\top, r) := \{\omega : d(\top, \omega) \leq r\} \neq \{\top\}$ such that for all $a, b \in B(\top, r)$:

$$-d(\top, a) \leq -d(\top, b) \quad \Longrightarrow (\iff) \quad f(a) \leq f(b)$$

We say that f is *monotonic* (*isomorphic*) relative to d if $B(\top, r) \equiv \Omega$.

We shall often assume that f_l is locally monotonic relative to (or isomorphic with) the Hamming metric on $\mathcal{H}_\alpha^l$.

Example 1 (Negative distance)**.** If f is isomorphic to d, then one can replace $f(\omega)$ by the negative distance $-d(\top, \omega)$. The number of values of such f is equal to the number of spheres $S(\top, r) := \{\omega : d(\top, \omega) = r\}$. One can easily show also that when f is isomorphic to d, then there is only one $\top$ element: $f(\top_1) = f(\top_2) \iff d(\top_2, \top_1) = d(\top_2, \top_2) = 0 \iff \top_1 = \top_2$.

Example 2 (Needle in a haystack)**.** Let $f(\omega)$ be defined as

$$f(\omega) = \begin{cases} 1 & \text{if } d(\top, \omega) = 0 \\ 0 & \text{otherwise} \end{cases}$$

This function is often used in studies of performance of genetic algorithms (GAs). In biological literature, $\top$ element is often referred to as the *wild type*, and a two-valued landscape is used to derive error threshold and critical mutation rate [5]. One can check that if for each $\top \in \Omega$ there exists $B(\top, r) \neq \{\top\}$ containing only one $\top$, then two-valued f is locally monotonic relative to any metric. Indeed, conditions of the definition above are satisfied in all such $B(\top, r) \subset \Omega$. If Ω has unique $\top$, then the conditions are satisfied for $B(\top, \infty) = \Omega$.

For monotonic f, spheres $S(\top, l)$ cannot contain elements with different values $x = f(\omega)$. We can generalise this property to *weak* or ϵ-monotonicity, which requires that the variance of $x = f(\omega)$ within elements of each sphere $S(\top, l)$ is small or does not exceed some $\epsilon \geq 0$. Monotonicity of f depends on the choice of metric, and one can define different metrics on Ω. Generally, one prefers metric d_2 to d_1 if the neighbourhoods, where f is monotonic relative to d_2, are 'larger' than for metric d_1: $B_1(\top, r) \subseteq B_2(\top, r)$ for all $B_i(\top, r)$, where f is monotonic relative to d_i. In this respect, the least preferable is the discrete metric: $d(a, b) = 0$ if $a = b$; 0 otherwise. The abundance of neutral mutations in nature supports an intuition that biological fitness landscapes are at least weakly locally monotonic relative to the Hamming metric on $\mathcal{H}_4^l$.

III. Optimality and Information

We shall now recall some basic principles of optimisation with information constraints that is used to solve the optimal coding problems in information theory, and it will allow us understand better the search problem for optimal DNA sequences.

Observe that Ω by our definition is the union of finite sets $\mathcal{H}_\alpha^l$, and therefore it is countable. Thus, we can consider sequences as elements of a probability space $(\Omega, \mathcal{R}, P)$. Measures on a σ-algebra $\mathcal{R}(\Omega)$ can be conveniently represented as linear functionals on an algebra of functions. Let $X := C_c(\Omega, \mathbb{R}, \|\cdot\|_\infty)$ be the normed algebra of continuous real functions with compact support in a locally compact topological space Ω. Its dual is the Banach space $Y := \mathcal{M}(\Omega, \mathbb{R}, \|\cdot\|_1)$ of real Radon measures [6]. Here, the duality is with respect to bilinear form $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y \rangle := \sum_{\omega \in \Omega} x(\omega)\, y(\omega)$$

Functions $v(\omega) = f_{\leq l}(\omega, \theta_1)$ and $w(\omega) = f_{\leq l}(\cdot, \theta_2)$, corresponding to fitness in different habitats and restricted to finite subsets Ω_l, are elements of algebra X. Probability measures on $\mathcal{R}(\Omega)$ are positive elements of Y with norm one: $\|y\|_1 = \sum_\Omega y(\omega) = 1$. The set of all probability measures is

$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0\,,\ \|y\|_1 = 1\}$$

This set is weakly compact and convex, and for commutative algebra X it is a simplex. We equip $\mathcal{P}(\Omega)$ with information topology induced by an *information distance* $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R} \cup \{\infty\}$. Observe also that expected fitness $\mathbb{E}_P\{x\}$ is a linear functional $x(p) = \langle x, p \rangle$. We define the following *optimal value* functions:

$$\phi^*(x; \lambda) := \sup\{\mathbb{E}_P\{x\} : I(p, q) \leq \lambda\} \tag{1}$$

$$\psi^*(x; v) := \inf\{I(p, q) : \mathbb{E}_P\{x\} \geq v\} \tag{2}$$

These functions are inverse of each other, when considered as functions of the constraint: $v = \phi^*(x; \lambda) = (\psi^*)^{-1}(x; \lambda)$. Optimality conditions have been obtained in the general case, when $I(\cdot, q)$ is a closed (lower semicontinuous), but not necessarily differentiable or convex function [7]. The most important and unique in a certain sense (see [8] for analysis from the point of evolution operators) is the classical Kullback-Leibler information divergence [9]:

$$I_{KL}(p, q) := \langle \ln[p/q], p \rangle \tag{3}$$

As is well-known, it is strictly convex and Gâteaux differentiable, and the extreme values of functions (1) (or (2))

with respect to I_{KL} are achieved if and only if the following conditions are satisfied:

$$P(\omega;\beta) = e^{\beta\, x(\omega)-\psi(x;\beta)}\, Q(\omega) \quad (4)$$

$$I_{KL}(p,q) = \lambda\,, \qquad \beta^{-1} = \frac{d}{d\lambda}\phi^*(x;\lambda) \quad (5)$$

$$\Big(\mathbb{E}_p\{x\} = \upsilon\,, \qquad \beta = \frac{d}{d\upsilon}\psi^*(x;\upsilon)\Big) \quad (6)$$

where $\psi(x;\beta) = \ln\langle e^{\beta\, x}, q\rangle$ is the cumulant generating function. Note that $\psi(x;\beta)$ and free energy $\phi(x;\beta^{-1}) = -\beta^{-1}\psi(x;\beta)$ are Legendre-Fenchel dual functions of $\phi^*(x;\lambda)$ and $\psi^*(x;\upsilon)$ [7]:

$$\phi(x;\beta^{-1}) = \inf\{\beta^{-1}\lambda - \phi^*(x;\lambda)\} \quad (7)$$

$$\psi(x;\beta) = \sup\{\beta\upsilon - \psi^*(x;\upsilon)\} \quad (8)$$

The following relations are useful for natural parametrisation of family of solutions (4):

$$\upsilon = \frac{d}{d\beta}\psi(x;\beta)\,, \qquad \lambda = \beta\frac{d}{d\beta}\psi(x;\beta) - \psi(x;\beta) \quad (9)$$

IV. Optimal Mutation Operators

Let $a \in \mathcal{H}_\alpha^l$ be a parent sequence of b. We refer to $r = d(a,b)$ as *mutation radius*. Let $q \in \mathcal{P}(\mathcal{H}_\alpha^l)$ be the probability distribution $Q(\omega)$ of parent sequences. We now consider the following optimisation problem:

$$\psi^*(r;\upsilon) := \inf\{I_{KL}(p,q) : \mathbb{E}_P\{r\} \geq \upsilon\} \quad (10)$$

Its solution defines optimal distribution $p \in \mathcal{P}(\mathcal{H}_\alpha^l)$, minimising information divergence from q and with expected mutation radius no less than υ. The following proposition defines conditions for the optimal operator as simple point mutation with binomial distribution.

Proposition 1. *Assuming a uniform distribution $Q(\omega) = \alpha^{-l}$ of parent sequences in Hamming space $\mathcal{H}_\alpha^l$, the optimal distribution of child sequences satisfying condition (10) is achieved by independently mutating each letter with probability*

$$\mu = \frac{\upsilon}{l}$$

The asymptotic of this mutation operator as $l \to \infty$ corresponds to Poisson process with mutation rate υ.

Proof: Assuming $x(b) = d(a,b)$ and $Q(a) = \alpha^{-l}$, the extremal distribution (4) is

$$P(b;\beta) = e^{\beta\, d(a,b)-\psi(r;\beta)}$$

where $\psi(r;\beta)$ is the cumulant generating function

$$\begin{aligned}\psi(d(a,b);\beta) &= \\ &= \ln\sum_{b\in\mathcal{H}_\alpha^l} e^{\beta\, d(a,b)} = \ln\sum_{r=0}^{l}(\alpha-1)^r\binom{l}{r}e^{\beta\, r} \\ &= l\ln[1+(\alpha-1)e^{\beta}]\end{aligned}$$

We used the fact that the number of sequences in sphere $S(a,r) := \{\omega : d(a,\omega) = r\}$ is

$$|S(a,r)| = (\alpha-1)^r\binom{l}{r} \quad (11)$$

The extremal distributions of equivalence classes is $P(r;\beta) = |S(a,r)|P(b;\beta)$. The relation between β and $\upsilon = \mathbb{E}\{r\}$ is

$$\upsilon = \frac{d}{d\beta}\psi(x;\beta) = \frac{l}{1+e^{-\beta}/(\alpha-1)}$$

the inverse of which gives parametrisation

$$\beta = \ln\frac{\upsilon}{l-\upsilon} - \ln(\alpha-1)$$

Substituting the above expression for β into $P(r;\beta)$ we obtain

$$P(r;\upsilon/l) = \binom{l}{r}\left(\frac{\upsilon}{l}\right)^r\left(1-\frac{\upsilon}{l}\right)^{l-r}$$

Thus, $P(r;\beta)$ corresponds to binomial distribution with parameter $\mu = \upsilon/l$.

As is well-known, the asymptotic of $P(r;\upsilon/l)$ as $l \to \infty$ is Poisson process:

$$\lim_{l\to\infty} P(r;\upsilon/l) = \frac{\upsilon^r}{r!}\,e^{-\upsilon}$$

which follows from the following facts

$$\lim_{l\to\infty}\frac{l!}{(l-\upsilon)^r(l-r)!} = 1\,, \qquad \lim_{l\to\infty}\left(1-\frac{\upsilon}{l}\right)^l = e^{-\upsilon}$$

Parameter $\upsilon = \mathbb{E}_p\{r\}$ is interpreted as the expected *rate* — the average number of mutations in a unit interval. □

We emphasise that binomial and Poisson distributions were obtained here as solutions to optimisation problem (10) with constraint $I_{KL}(p,q) \leq \lambda$ (or its inverse with $\mathbb{E}_p\{x\} \geq \upsilon$) and a constant reference measure. The latter corresponds to $\beta = 0$, which gives $\mu_0 = 1 - 1/\alpha$. For $l \to \infty$, this corresponds to a Poisson process with infinite rate.

V. Optimal Mutation Rates

Let us consider the problem of finding optimal sequence $\top$ in each $\mathcal{H}_\alpha^l \subset \Omega$. We assume here that fitness f_l is isomorphic to Hamming distance $d(\top,\omega)$. The search process is preformed by a simple mutation of the parent sequences. The main tool for optimising mutation rates is Markov probability of transitions between different levels of fitness, or for $f_l(\omega) = -d(\top,\omega)$ between spheres around $\top$.

Let a be the parent of b, and denote $n = d(\top,a)$, $m = d(\top,b)$ and $r = d(a,b)$. We define the following probabilities:

$$\begin{aligned}P(r \mid n) &:= P(b \in S(a,r) \mid a \in S(\top,n)) \\ P(m \mid n,r) &:= P(b \in S(\top,m) \mid a \in S(\top,n), b \in S(a,r)) \\ P(m \cap r \mid n) &:= P(S(\top,m) \cap b \in S(a,r) \mid a \in S(\top,n)) \\ P(m \mid n) &:= P(b \in S(\top,m) \mid a \in S(\top,n))\end{aligned}$$

Proposition 2. *The probability of transition between spheres of radii n and m around $\top$ in a Hamming space $\mathcal{H}_\alpha^l$ under simple mutation with rate $\mu \in [0,1]$ is*

$$P_\mu(m \mid n) = \sum_{r=0}^{l} \frac{|S(\top,m) \cap S(a,r)|_n}{(\alpha-1)^r} \mu^r (1-\mu)^{l-r} \quad (12)$$

where

$$|S(\top,m) \cap S(a,r)|_n = \quad (13)$$
$$\sum (\alpha-2)^{r_0} \binom{n-r_-}{r_0} (\alpha-1)^{r_+} \binom{l-n}{r_+} \binom{n}{r_-}$$

Here, $|\cdot|_n$ denotes condition $d(\top,a) = n$, and summation runs over r_0, r_+ and r_- satisfying $r_+ \in [0,(r+m-n)/2]$, $r_- \in [0,(n-|r-m|)/2]$, $r_- - r_+ = n - \max\{r,m\}$ and $r_0 + r_+ + r_- = \min\{r,m\}$.

Proof: Expansion of probability $P_\mu(m \mid n)$ gives

$$P_\mu(m \mid n) = \sum_{r=0}^{l} P_\mu(m \cap r \mid n) = \sum_{r=0}^{l} P(m \mid n,r) P_\mu(r \mid n)$$

For simple mutation, probability $P_\mu(r \mid n)$ is binomial distribution with parameter μ (Proposition 1). Probability $P(m \mid n,r)$ is defined by the number of elements in the spheres $S(a,r)$, $S(\top,m)$ and their intersection as follows:

$$P(m \mid n,r) = \frac{|S(\top,m) \cap S(a,r)|_n}{|S(a,r)|} \quad (14)$$

Number $|S(a,r)|$ is given by equation (11). Equation (13) is obtained by counting the numbers r_- of deleterious, r_0 neural and r_+ advantageous mutations in r, guarded by conditions based on metric inequalities for r, m and n (e.g. $|n-m| \le r \le n+m$). □

The collection of mutation rates $\mu(n)$ for all $n = d(\top,a)$ is a mutation rate control function. The optimal control can be defined in different ways. Traditional approach to sequential optimisation seeks a control maximising expected fitness over a sequence $\omega_0,\ldots,\omega_t,\ldots$, where t represents generation number or time. One can show that for $t=1$ (one generation), the optimal mutation rate function, minimising conditional expected value $\mathbb{E}_\mu\{m \mid n\}$, is the following step function:

$$\mu(n) := \begin{cases} 0 & \text{if } n < l(1-1/\alpha) \\ \frac{1}{2} & \text{if } n = l(1-1/\alpha) \\ 1 & \text{otherwise} \end{cases} \quad (15)$$

This control is not optimal for $t>1$, as it does not change the distribution if $d(\top,a) < l(1-1/\alpha)$ for all parents sequences.

Another approach is to minimise the expected time of convergence to $\top$. This problem can be solved using fundamental matrix of an absorbing Markov chain for $P_\mu(m \mid n)$ and numerical methods. However, a useful and simple approximation is given by maximising probability of mutating to $\top$:

$$P_\mu(m=0 \mid n) = (\alpha-1)^{-n} \mu^n (1-\mu)^{l-n}$$

Taking its derivative over μ to zero gives $n - l\mu = 0$ or

$$\mu(n) = \frac{n}{l} \quad (16)$$

This linear control is optimal for two-valued fitness functions, such as discussed in Example 2, because conditional expected fitness depends only on the above probabilities. For other relatively monotonic or isomorphic landscapes this control gives good expected times of convergence to $\top$. However, its performance in the initial stages is quite poor.

Some improvement in the initial stage (but increased time of convergence) is achieved by $\mu(n)$ maximising probability of 'success':

$$P_\mu(m < n \mid n) = \sum_{m=0}^{n-1} P_\mu(m \mid n)$$

Mutation rate control for this criterion was obtained by Bäck for the case of binary sequences [10]. Transition probability (12) can be used to compute such a control for arbitrary alphabets.

Finally, we point out that sequential optimisation can be performed with respect to cumulative expected fitness (i.e. the sum of expected fitness values for all generations). Because the objective function is additive and transition probabilities are Markov, this problem can be formulated as dynamic programming. The complexity of the problem, however, increases rapidly with the length of sequences making the precise solution intractable. A good approximation can be performed using the following information-theoretic approach.

The main idea is to replace the problem of sequential optimisation in time by optimisation in information distance, as defined by functions (1) or (2). One can show that optimal evolution in information I_{KL} is achieved if Markov transition probability has the form:

$$P_\beta(m \mid n) = e^{\beta(n-m) - \psi(n-m;\beta)} P(m) \quad (17)$$

Expression $\beta(n-m) - \psi(n-m;\beta)$ represents minimal information subject to $\mathbb{E}\{n-m \mid n\} \ge \upsilon$. This can be seen from equation (8), which shows that function ψ is Legendre-Fenchel dual of minimal information ψ^*, defined by equation (2). In fact, any conditional probability is expressed as $P(m \mid n) = e^{\iota(m,n)} P(m)$, where $\iota(m,n) = \ln P(m \mid n) - \ln P(m)$ is random information transmitted by n about m. The fact that $\iota(m,n) = \psi^*(n-m;\upsilon)$ is related to $\mu(n)$ can be understood by representing transition probability (12) in the exponential form:

$$P_\mu(m \mid n) = \sum_{r=0}^{l} \frac{|S(\top,m) \cap S(a,r)|_n}{(\alpha-1)^r} e^{r \ln \frac{\mu}{1-\mu} + l \ln[1-\mu]}$$

Our evaluation shows that optimal values $\lambda = \psi^*(x;\upsilon)$, or their inverse values $\upsilon = \phi^*(x;\lambda)$ (see equation (1)), are achieved if $\mu(n)$ is defined by the cumulative distribution function of $P_0(m)$ assuming a uniform distribution of sequences in $\mathcal{H}_\alpha^l$:

$$\mu(n) = P_0(m < n \mid n) = \sum_{m=0}^{n-1} P_0(m) \quad (18)$$

Distribution $P_0(m)$ can be obtained by counting sequences in spheres $S(\top,m)$, and it is equivalent to $\top$ mutating with

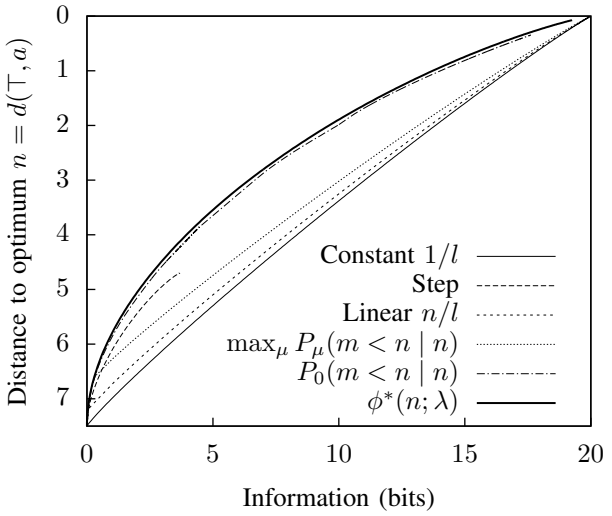

Fig. 1. Expected distance to optimum $\top \in \mathcal{H}_4^{10}$ as a function of information divergence λ from initial distribution. Different curves correspond to different controls $\mu(n)$ of mutation rate; $\phi^*(n; \lambda)$ represents theoretical optimum.

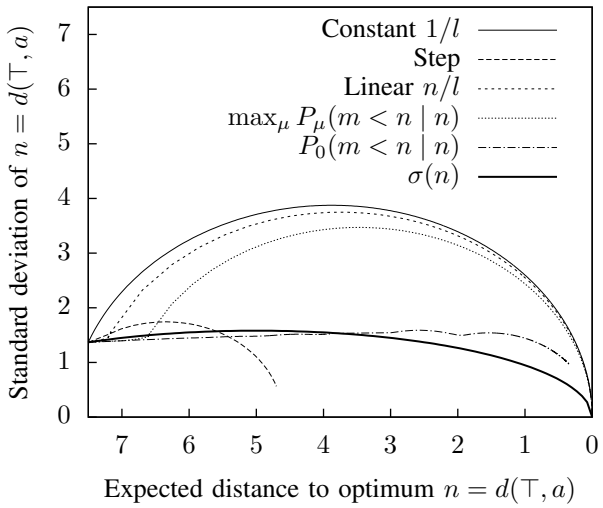

Fig. 2. Standard deviation of distance to optimum $\top \in \mathcal{H}_4^{10}$ as a function of its expected value. Different curves correspond to different controls $\mu(n)$ of mutation rate; $\sigma(n)$ corresponds to mutation of $\top$ with $\mu = n/l$.

probability $\mu_0 = 1 - 1/\alpha$. This method gives exceptional performance in the initial stages as well as good overall performance in terms of cumulative expected fitness.

Figure 1 shows expected distance $\mathbb{E}_p\{d(\top, \omega)\}$ as function of information divergence $I_{KL}(p, q)$ achieved by different controls $\mu(n)$ of mutation rate in Hamming space $\mathcal{H}_4^{10}$. Results for other Hamming spaces are very similar. One can see that the *minimum information* control (18) achieves almost perfectly theoretical values $\phi^*(n; \lambda)$, shown by a bold solid curve. Figure 2 shows standard deviations σ from $\mathbb{E}_p\{d(\top, \omega)\} = \upsilon$ in populations, as υ decreases (i.e. getting closer to $\top$). The minimum information control (18), unlike other functions, maintains the smallest deviation. Moreover, it matches closely theoretical curve $\sigma(\upsilon) = \sqrt{\upsilon(1 - \upsilon/l)}$, shown in bold, which is derived assuming simple mutation of $\top$ sequence with mutation rate $\mu = \upsilon/l$. Thus, the minimum information control (18) maintains distributions of sequences in $\mathcal{H}_\alpha^l$, that are 'spherical' around $\top$.

VI. Discussion

We have shown how biological organisms can be represented by codes of variable length in the set of nested Hamming spaces, and evolution by asexual reproduction has been compared to a processes of finding optimal codes. The model presented has a number of simplifications. For example, optimality of simple mutation is based on additivity of the fitness function with respect to letters in a sequence. This would imply that biological organisms can be characterised by a set of weakly interacting random variables, which is not really the case, and it contradicts the epistasis phenomenon in nature. Furthermore, probabilities of mutation between different letters are not equal in biological DNA, which shows that a one-parameter mutation operator is not truly biological. A more realistic model can be obtained by considering a metric on sequences that depends on different letters in the alphabet. We also have not considered other search operators, such as recombination of sequences (i.e. sexual reproduction).

The approach presented, however, allowed us to make some predictions about mutation mechanisms, variability of mutation rates and lengths of sequences. It also demonstrated how information theoretic ideas can facilitate our understanding of nature. Mathematical tools developed can also be useful in engineering problems.

Acknowledgement

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

References

[1] A. N. Kolmogorov, "Three approaches to the definition of mutual information," *Problems of Information Transmission*, vol. 1, no. 1, pp. 3–11, 1965, in Russian.

[2] C. T. Bergstrom and M. Lachmann, "Shannon information and biological fitness," in *Proceedings of the IEEE Workshop on Information Theoery*, 2004.

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

[4] R. L. Stratonovich, "On value of information," *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, vol. 5, pp. 3–12, 1965, in Russian.

[5] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, London: Harvard University Press, 2006.

[6] N. Bourbaki, *Eléments de mathématiques. Intégration*. Hermann, 1963.

[7] R. V. Belavkin, "Information trajectory of optimal learning," in *Dynamics of Information Systems: Theory and Applications*, ser. Springer Optimization and Its Applications Series, M. J. Hirsch, P. M. Pardalos, and R. Murphey, Eds. Springer, 2010, vol. 40.

[8] ——, "On evolution of an information dynamic system and its generating operator," *Optimization Letters*, pp. 1–14, 2011, 10.1007/s11590-011-0325-z.

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

[10] T. Bäck, "Optimal mutation rates in genetic search," in *Proceedings of the 5th International Conference on Genetic Algorithms*, S. Forrest, Ed. Morgan Kaufmann, 1993, pp. 2–8.