Theory and Practice of Optimal Mutation Rate Control in Hamming Spaces of DNA Sequences

Roman V. Belavkin[1], Alastair Channon[2], Elizabeth Aston[2], John Aston[3] and Christopher G. Knight[4]

[1]School of Engineering and Information Sciences, Middlesex University, London NW4 4BT, UK
[2]Research Institute for the Environment, Physical Sciences and Applied Mathematics, Keele University, ST5 5BG, UK
[3]Department of Statistics, University of Warwick, CV4 7AL, UK
[4]Faculty of Life Sciences, University of Manchester, M13 9PT, UK
r.belavkin@mdx.ac.uk, {a.d.channon,e.j.aston}@epsam.keele.ac.uk, j.a.d.aston@warwick.ac.uk, chris.knight@manchester.ac.uk

Abstract

We investigate the problem of optimal control of mutation by asexual self-replicating organisms represented by points in a metric space. We introduce the notion of a relatively monotonic fitness landscape and consider a generalisation of Fisher's geometric model of adaptation for such spaces. Using a Hamming space as a prime example, we derive the probability of adaptation as a function of reproduction parameters (e.g. mutation size or rate). Optimal control rules for the parameters are derived explicitly for some relatively monotonic landscapes, and then a general information-based heuristic is introduced. We then evaluate our theoretical control functions against optimal mutation functions evolved from a random population of functions using a meta genetic algorithm. Our experimental results show a close match between theory and experiment. We demonstrate this result both in artificial fitness landscapes, defined by a Hamming distance, and a natural landscape, where fitness is defined by a DNA-protein affinity. We discuss how a control of mutation rate could occur and evolve in natural organisms. We also outline future directions of this work.

Introduction

The problem of optimal mutation rate has been studied for a long time (e.g. see Eiben et al., 1999; Ochoa, 2002; Falco et al., 2002; Cervantes and Stephens, 2006; Vafaee et al., 2010, for reviews). It relates directly to optimisation of genetic algorithms (GAs) in operations research and engineering problems (i.e. meta-heuristics). It is also related to some fundamental questions in evolutionary theory about the role of mutation in adaptation and biological mechanisms of DNA repair and mutation control.

As noted by Eiben et al. (1999), there are two trends in optimisation of parameters in GAs — optimal parameter setting and optimal parameter control. In the former, one looks for an optimal value of a parameter, which is than kept constant. Thus, Mühlenbein (1992) proposed mutation rate $\mu = 1/l$, where l is the length of sequences. The value $1/l$, as was pointed out by Ochoa et al. (1999), is related to the error threshold (Eigen et al., 1988). However, while mutation rate $1/l$ can give satisfactory performance in some problems, the advantages of using a variable rate were becoming obvious to many researchers, leading to the problem of optimal parameter control. In particular, Ackley (1987) suggested that mutation probability is analogous to temperature in simulated annealing, and should decrease with time. A gradual reduction of mutation rate was also proposed by Fogarty (1989). In a pioneering work, Yanagiya (1993) used Markov chain analysis of GAs to show that in any problem there exists a sequence of optimal mutation rates maximising the probability of obtaining global solution at each generation. A significant contribution to the field was made by Bäck (1993), who suggested that mutation rate μ should depend on fitness values rather than time. Recently, Vafaee et al. (2010) used numerical methods to optimise a mutation operator based on the Markov chain model of GA by Nix and Vose (1992). The complexity of this model, however, restricts the application of this method to small spaces and populations. Thus, the precise form of the optimal mutation rate control, as well as question about the existence of such a control in the general case, remain open problems. These problems are extremely important not only for applications of GAs, but also for biology and evolutionary theory.

In biological systems, mutation, unlike natural selection, is an evolutionary process controlled, to a degree, by the organism. This control is primarily seen in highly refined DNA repair and replication machinery (e.g. Hakem, 2008). This ensures both that physical damage to genetic material is repaired and that, in the process of cell division, the newly synthesised copies of DNA faithfully reproduce the parental sequence. The result is that biological mutation rates are very low: DNA-based organism values typically being 1/300 per genome per replication, which, for genomes frequently in the range between 10^6 and 10^{10} base-pairs, means extremely faithful repair and replication (Drake et al., 1998). Nonetheless, this observation also implies that biological mutation rates per base-pair are not minimised, since widely varying genome sizes imply very different per-base-pair rates. At a mechanistic level, some organisms do exist, such as the bacterium *Deinococcus radiodurans* with substantially more developed DNA repair or replication mechanisms than closely related species (Cox et al., 2010), implying that mutation rates elsewhere at least are not minimised.

Genetic variation also exists in mutation rates within species and in the way mutation rate changes with environment for a single genotype (Bjedov et al., 2003). Therefore, mutation rates and their variation are potentially subject to biological evolution themselves. Thus, while mutation rates are only a part of the biological evolutionary process, they merit examination independent of the vicissitudes of selection that are imposed on their products, which is what we address.

Our approach is based on theories of optimal control and information. However, we believe that the key to finding solutions that are relevant not only for engineering, but also for biology, is understanding the relation between a representation space, which is a discrete space of genotypes, and its (pre)-ordering by phenotypic fitness. Biology typically understands this relation via landscape metaphors, used in a variety of ways (e.g. classically adaptive landscapes of Wright, 1932, and 'epistatic' landscapes of Waddington, 1957). However, while the underlying elements, particular alleles of genes, are acknowledged as discrete, these landscapes have almost uniformly been theorised (and visualised) in continuous space following Fisher (1930). This is problematic when one comes to the mechanistic basis of biological evolution in discrete DNA mutations. Attempts are being made to reconcile such continuous models with individual DNA mutations (Orr, 2005). However, hitherto, these attempts have maintained a continuous view of the landscape space, in contrast to the reality of its discrete domain. Discrete views have typically been restricted to abstracted biological systems, such as aptamer (Knight et al., 2009) or RNA structure evolution, where landscape analogies can be dropped in favour of networks of sequences (Cowperthwaite and Meyers, 2007) which do not lend themselves to consideration of variable mutation sizes.

This work presents elements of a theory on optimisation of asexual reproduction by a mutation rate control together with its experimental evaluation. We introduce the notion of relatively and weakly monotonic fitness landscapes, and then develop the necessary machinery for Hamming spaces of sequences with arbitrary alphabets, which are particularly relevant in biology. Then we evolve mutation rate control functions using a meta genetic algorithm, and show that they closely match our theoretical predictions.

Theory

Let Ω be a countable set of all possible individuals ω and $f : \Omega \to \mathbb{R}$ be a fitness function. Assuming that fitness value $x = f(\omega)$ is the only information available, let $P(x_{s+1} \mid x_s)$ be the conditional probability of an offspring having fitness value x_{s+1} given that its parent had value x_s at generation (time) s. This Markov probability can be represented by a left stochastic matrix T, and if $P(x_{s+1} \mid x_s)$ does not depend on s (i.e. T is stationary), then T^t defines a linear transformation of distribution $p_s := P(x_s)$ of fitness values at time s into distribution $p_{s+t} := P(x_{s+t})$ of fitness values

after $t \geq 0$ generations:

$$p_{s+1} = T p_s = \sum_{x_s} P(x_{s+1} \mid x_s) P(x_s) \ \Rightarrow \ p_{s+t} = T^t p_s$$

We denote the expected fitness at generation s as

$$\mathbb{E}\{x_s\} := \sum_{x_s} x_s P(x_s)$$

If $\mathbb{E}\{x_{s+t}\} \geq \mathbb{E}\{x_s\}$, then individuals have adapted.

Suppose that the transition probability $P_\mu(x_{s+1} \mid x_s)$ depends on a control parameter μ, so that the Markov operator $T_{\mu(x)}$ depends on the control function $\mu(x)$. Then the expected fitness $\mathbb{E}_{\mu(x)}\{x_{s+t}\}$ also depends on $\mu(x)$. We interpret $\mu(x)$ as a control function that parents use in reproduction to maximise expected fitness of their offspring based on the value of their own fitness. A particular example we shall consider here is when μ is the mutation rate parameter.

If Ω is the space $\mathcal{H}_\alpha^l := \{1, \ldots, \alpha\}^l$ of sequences of length l and α letters, then by mutation we understand here a process of independently changing each letter in a parent sequence to any of the other $\alpha - 1$ letters with probability $\mu/(\alpha - 1)$. This is point mutation, the simplest form of mutation defined by one parameter μ, called the *mutation rate*.

The main result that we present in this paper is a mutation rate control function, which is approximately optimal for maximising expected fitness $\mathbb{E}\{x_{s+t}\}$ in landscapes $f(\omega)$ that are locally monotonic relative to the Hamming metric (this property will be defined later). This mutation rate function corresponds to the cumulative distribution function (CDF) $P_e(x_r > x)$, $r \in [s, s+t]$, computed from empirical distribution $P_e(x_r)$ of observed fitness values x_r over the period $[s, s+t]$:

$$\mu_e(x) = P_e(x_r > x) = \sum_{x_r > x} P_e(x_r) \tag{1}$$

We refer to this function as *informed mutation rate*, because it uses information communicated by random variable x. We first present the theory and assumptions behind this heuristic. Then we evaluate it against nearly optimal mutation functions, evolved using a meta genetic algorithm both for artificial and natural fitness landscapes.

Problem Definition

Formally, an optimal control function (e.g. an optimal mutation rate function) is $\bar{\mu}(x)$ achieving the following optimal (supremum) value:

$$\overline{x}(\lambda) := \sup_{\mu(x)} \{\mathbb{E}_{\mu(x)}\{x_{s+t}\} : t \leq \lambda\} \tag{2}$$

Here, λ represents a time constraint. Function (2) is nondecreasing and has the following inverse

$$\overline{x}^{-1}(\upsilon) := \inf_{\mu(x)} \{t \geq 0 : \mathbb{E}_{\mu(x)}\{x_{s+t}\} \geq \upsilon\} \tag{3}$$

Here, υ is a constraint on the expected fitness at $s + t$. Thus, $\overline{x}(\lambda)$ is the maximum adaptation in no more than λ generations; $\overline{x}^{-1}(\upsilon)$ is the minimum (infimum) number of generations required to achieve adaptation υ.

Optimal solutions $\bar{\mu}(x)$, defined by function (2), depend on the constraint $t \leq \lambda$. We are interested in solutions for λ that is large enough to achieve the maximum expected fitness $\mathbb{E}\{x_{s+t}\} = \sup f(\omega)$. This can be represented dually by function (3) with constraint $\upsilon = \sup f(\omega)$. We note that $\overline{x}(\lambda) = \sup f(\omega)$, if $\lambda = \infty$. However, generally $\overline{x}^{-1}(\upsilon) \leq \infty$, even if $\upsilon = \sup f(\omega)$. Thus, our objective is to derive one optimal control function $\bar{\mu}(x)$ that can be used by each individual parent based on their fitness value throughout the entire 'evolution' $[s, s+t]$. We note also that our formulation uses only the values of fitness, and therefore it extends to the case where $f(\omega)$ is time-variable.

Specific expressions for $P_\mu(x_{s+1} \mid x_s)$, defining $T_{\mu(x)}$, can be learnt or derived analytically from the domain Ω and its structure. The operator $T_{\mu(x)}$ contains all information required to compute optimal values (2) and (3). Thus, in principle, one can find an optimal control function $\bar{\mu}(x)$, if the family of operators $T_{\mu(x)}$ is known. For example, considering values $x \geq \upsilon$ as absorbing states, one can use $T_{\mu(x)}$ to compute the fundamental matrix of the corresponding absorbing Markov chain and minimise the expected convergence time to the absorbing states. Solving the complete optimisation problem, however, can be an intractable task. We shall formulate additional assumptions that will allow us to solve the problem for some important cases.

Relatively Monotonic Landscapes

First, we shall make some assumptions about fitness $f(\omega)$, which on one hand will generalise and clarify the terms 'smooth' and 'rugged' fitness landscape, and on the other hand will allow us to obtain expressions for $P_\mu(x_{s+1} \mid x_s)$. In particular, we assume that there exists optimal individual $\top \in \Omega$ (not necessarily unique) such that $\sup f(\omega) = f(\top)$. This is always true if Ω is finite. Also, we shall equip Ω with a metric $d : \Omega \times \Omega \to [0, \infty)$, so that similarity between a and $b \in \Omega$ can be measured by $d(a, b)$, and assume that there is a relation between the metric d and the fitness function f. In particular, we define f to be monotonic relative to d.

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 $a, b \in B(\top, r)$:

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

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

Example 1 (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 fitness landscape is often used in studies of GA performance. A two-valued landscape is used to derive error threshold and critical mutation rate, and elements $\top$ are referred to as the *wild type*. Such f is locally monotonic relative to any metric, if for each $\top \in \Omega$ there exists $B(\top, r) \neq \{\top\}$ containing only one $\top$. Then 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 $B(\top, \infty) = \Omega$. In a two-valued landscape, optimal function $\bar{\mu}(x)$ for any λ in (2) is defined by maximising the one-step transition probability $P_\mu(x_{s+1} = 1 \mid x_s)$.

Example 2 (Negative distance to optimum). If f is isomorphic to d, then one can replace fitness $f(\omega)$ 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: $f(\top_1) = f(\top_2) \iff d(\top_2, \top_1) = d(\top_2, \top_2) = 0 \iff \top_1 = \top_2$.

In monotonic landscapes, spheres $S(\top, r)$ cannot contain individuals with different fitness. We can generalise this property by *weak* or ε-monotonicity, which requires that the variance of fitness within individuals of each sphere $S(\top, r)$ is small or does not exceed some $\varepsilon \geq 0$. These assumptions allow us to replace fitness $f(\omega)$ by negative distance $-d(\top, \omega)$, and derive expressions for transition probability $P_\mu(x_{s+1} \mid x_s)$ using topological properties of (Ω, d).

Monotonicity of f depends on the choice of metric, and one can define different metrics on Ω. Fitness landscapes that are at least weakly locally monotonic relative to the Hamming metric seem biologically plausible given the abundance of neutral mutations in nature and redundancy in the translation of DNA to protein sequences. Thus, we focus our attention on the case when Ω is a Hamming space.

Mutation and Adaptation in a Hamming Space

First, we outline a model of asexual reproduction in metric space (Ω, d), and define the relation of parameter μ to topology on Ω. This model is a generalisation of Fisher's geometric model of adaptation in Euclidean space (Fisher, 1930). Then we shall specialise this to a Hamming space.

Let individual a be a parent of b, and let $d(a, b) = r$. We consider single-parent reproduction as a transition from parent a to a random point b on a sphere: $b \in S(a, r)$. We refer to r as a *radius of mutation*. Suppose that $d(\top, a) = n$ and $d(\top, b) = m$. We are interested in the following probability:

$$
\begin{aligned}
P(m \mid n) & := P(b \in S(\top, m) \mid a \in S(\top, n)) \\
& = \sum_{r=0}^{l} P(m \mid r, n) \, P(r \mid n) \quad (4)
\end{aligned}
$$

where the following notation was used

$$P(m \mid r, n) := P(b \in S(\top, m) \mid b \in S(a, r), a \in S(\top, n))$$
$$P(r \mid n) := P(b \in S(a, r) \mid a \in S(\top, n))$$

If mutation radius r can be controlled via parameter μ, then transition probability (4) depends on this parameter as well. Specific expressions for $P_\mu(m \mid n)$ depend on the topology of Ω. Let us consider the Hamming space.

Let Ω be a space $\mathcal{H}_\alpha^l := \{1, \ldots, \alpha\}^l$ — a space of sequences of length l and α letters and equipped with the Hamming metric $d(a, b) := |\{i : a_i \neq b_i\}|$. Then, given probability of mutation $\mu(n) \in [0, 1]$ of each letter in the parent sequence $a \in S(\top, n)$, the probability that $b \in S(a, r)$ is

$$P_\mu(r \mid n) = \binom{l}{r} \mu(n)^r (1 - \mu(n))^{l-r} \qquad (5)$$

Probability $P(m \mid r, n)$ is defined by the number of elements in the intersection of spheres $S(\top, m)$ and $S(a, r)$:

$$P(m \mid r, n) = \frac{|S(\top, m) \cap S(a, r)|_{d(\top, a) = n}}{|S(a, r)|} \qquad (6)$$

where cardinality of the intersection $S(\top, m) \cap S(a, r)$ with condition $d(\top, a) = n$ is computed as follows

$$|S(\top, m) \cap S(a, r)|_{d(\top, a) = n} = \qquad (7)$$
$$\sum (\alpha - 2)^{r_0} \binom{n - r_-}{r_0} (\alpha - 1)^{r_+} \binom{l - n}{r_+} \binom{n}{r_-}$$

where the triple 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 $r_0 + r_+ + r_- = \min\{r, m\}$. These conditions are based on metric inequalities for r, m and n (e.g. $|n - m| \leq r \leq n + m$). The number of sequences in $S(a, r) \subset \mathcal{H}_\alpha^l$ is

$$|S(a, r)| = (\alpha - 1)^r \binom{l}{r} \qquad (8)$$

Substituting equations (5)–(8) into (4) we obtain the expression for $P_\mu(m \mid n)$ in Hamming space $\mathcal{H}_\alpha^l$.

Analytical Solutions for Special Cases

If fitness f is isomorphic to the Hamming metric, then transition probabilities $P_\mu(x_{s+1} \mid x_s)$ are completely defined by $P_\mu(m \mid n)$ with $x_{s+1} = -m$ and $x_s = -n$. The corresponding Markov operator $T_{\mu(n)}$ is then an $(l + 1) \times (l + 1)$ matrix completely defining the evolution on $[s, s+t]$, $t \leq \lambda$, for a given mutation rate function $\mu(n)$, if all individuals are allowed to reproduce (with selection, one has to compose $T_{\mu(n)}$ with a selection operator). For example, one can show that for $\lambda = 1$, the optimal mutation rate is a step function:

$$\mu_1(n) := \begin{cases} 0 & \text{if } n < l(1 - 1/\alpha) \\ \frac{1}{2} & \text{if } n = l(1 - 1/\alpha) \\ 1 & \text{otherwise} \end{cases}$$

Unfortunately, analytical or numerical solutions to optimisation problems (2) or (3) are not available or tractable for $\lambda > 1$ and large l. However, analysis allows us to derive some main features of an optimal control function $\bar{\mu}(n)$.

Minimisation of the convergence time to state $m = 0$ is related to maximisation of probability $P_\mu(m = 0 \mid n)$. Because $r = n$ and $|S(\top, 0) \cap S(a, n)|_{d(\top, a) = n} = 1$, it has the following expression:

$$P_\mu(m = 0 \mid n) = (\alpha - 1)^{-n} \mu^n (1 - \mu)^{l-n} \qquad (9)$$

Mutation rate maximising this probability is obtained by taking its derivative P_μ' over μ to zero, and together with condition $P_\mu'' \leq 0$, this gives $n - l\mu = 0$ or

$$\mu_2(n) = \frac{n}{l} \qquad (10)$$

This linear mutation control function has very intuitive interpretation — if sequence a has n letters different from the optimal sequence $\top$, then substitute n letters in the offspring. One can show that the linear function (10) is optimal for two-valued fitness landscapes with one optimal sequence, such as the Needle in a Haystack discussed in Example 1. This is because expected fitness $\mathbb{E}_{\mu(x)}\{x_{s+t}\}$ in this case is completely defined by probability (9). For other fitness landscapes that are monotonic relative to the Hamming metric, function (10) is an approximation of the optimal control, because it does not take into account transition probabilities $P_\mu(m \neq 0 \mid n \neq 0)$ between other (transient) states, which may influence the expected time of convergence to $m = 0$. As a result, the convergence can be very poor in the initial stages of evolution on $[s, s + t]$.

Bäck (1993) derived probability $P_\mu(m < n \mid n)$ of 'success' in the space $\mathcal{H}_2^l$ of binary sequences, and then considered mutation rates $\hat{\mu}$ maximising its value for each $n = d(\top, \omega)$. Our equations (4)–(8) allow us to perform such optimisation for arbitrary α. This method makes significant improvement over the linear control for the speed of convergence in the initial stages of evolution on $[s, s+t]$. We note, however, that the resulting mutation controls do not achieve optimal values (2) or (3). One can show that maximisation of $P_\mu(m < n \mid n)$ is equivalent to maximisation of conditional expectation $\mathbb{E}\{u(m, n) \mid n\} = \sum_m u(m, n) P_\mu(m \mid n)$ of a two-valued utility function: $u(m, n) = 1$ if $m < n$; 0 otherwise. This function has only two values, and such optimisation of $\mu(n)$ is not precise for fitness functions with more than two values. In fact, analysis using absorbing Markov chains shows that linear control (10) achieves shorter expected times of convergence into absorbing state $m = 0$.

Empirically Informed Mutation Rate

Another approach to optimal control of parameters in evolutionary systems is based on theories of information and optimal coding. In brief, one can reformulate problems (2)
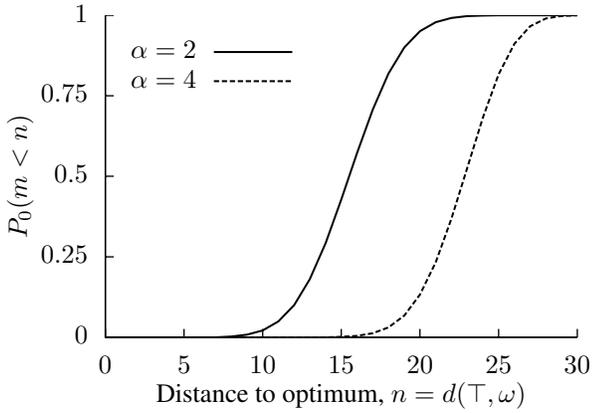
Figure 1: Cumulative distribution functions $P_0(m < n)$ of distances to optimum under random distribution of sequences in $\mathcal{H}_2^{30}$ and $\mathcal{H}_4^{30}$.

and (3) by replacing the time constraint $t \leq \lambda$ with a constraint on information 'distance' $\mathbb{E}_{s+t}\{\ln(p_{s+t}/p_s)\} \leq \lambda$ of distribution $p_{s+t} = T^t p_s$ from p_s. Although minimisation of information distance is not equivalent to minimisation of convergence time, this formulation has the advantage that the corresponding optimal values can be computed exactly and used to evaluate various control functions.

Our evaluation shows that adaptation $\mathbb{E}\{x_{s+t}\} \geq v$ with the least information distance of p_{s+t} from p_s is achieved if mutation rate is identified with the CDF of the 'least informed' distribution $P_0(x)$ of fitness values. In particular, assuming a uniform distribution $P_0(\omega) = \alpha^{-l}$ of sequences in $\mathcal{H}_\alpha^l$, the distribution $P_0(n) := P_0(\omega \in S(\top, n))$ of their distances from $\top$ can be obtained by counting sequences in the spheres $S(\top, n) \subset \mathcal{H}_\alpha^l$. One can show also that this corresponds to binomial distribution with $\mu = 1 - 1/\alpha$:

$$ P_0(n) = \binom{l}{n}\left(1 - \frac{1}{\alpha}\right)^n \left(\frac{1}{\alpha}\right)^{l-n} = \binom{l}{n}\frac{(\alpha-1)^n}{\alpha^l} $$

In this case, $\mathbb{E}\{n\} = l\mu = l(1 - 1/\alpha)$. Under the minimal information distance assumption, the offspring will have very similar distribution, and the probability $P_0(m < n)$ that an offspring is closer to $\top$ is given by the CDF of $P_0(n)$, which can be used to control the mutation rate:

$$ \mu_0(n) = P_0(m < n) = \sum_{m=0}^{n-1} P_0(m) \qquad (11) $$

This mutation control function has the following interpretation — if sequence a has n letters different from the optimal sequence $\top$, then substitute each letter in the offspring with the 'least informed' probability of improvement relative to the current value $n = d(\top, a)$. Figure 1 shows $P_0(m < n)$ for $\mathcal{H}_2^{30}$ and $\mathcal{H}_4^{30}$. We note that minimisation of information distance of p_{s+t} from $p_s := P_0$ corresponds to maximisation of entropy of p_{s+t}, but adaptation $\mathbb{E}\{x_{s+t}\} \geq \mathbb{E}\{x_s\}$

leads to increasing the distance and decreasing the entropy (i.e. slow 'cooling' as in simulated annealing).

In the next section, we present nearly optimal mutation rate functions, obtained experimentally, and find that they correspond to CDFs of distributions that are skewed towards the optimum compared to the CDFs of the least informed distributions P_0 (i.e. skewed to the left compared to those used in Figure 1). This can be explained by the fact that the offspring sequences do not have a uniform distribution in $\mathcal{H}_\alpha^l$ during long intervals $[s, s + t]$ due to adaptation $\mathbb{E}_{\mu(n)}\{m\} \leq \mathbb{E}\{n\} = l(1 - 1/\alpha)$. Therefore, the probabilities of improvement relative to the current fitness are higher than $P_0(m < n)$, and they can be approximated by empirical functions $P_e(m < n)$, observed during $[s, s + t]$. Thus, we refer to such a control as 'informed'.

Finally, we note that if fitness is monotonic relative to the Hamming metric, then function $P_e(m < n)$ can be replaced by function $P_e(x_r > x)$ for fitness values. We conjecture that the corresponding control (1) of mutation rate should achieve good performance also in landscapes that are only weakly or ε-monotonic. Our experiments with an aptamer landscape (Rowe et al., 2010) support this hypothesis.

Evolving Optimal Mutation Rates

To evaluate our theoretically derived mutation control functions, we have evolved such functions independently using a meta-genetic algorithm (Meta-GA). Populations of the Meta-GA comprised individual functions $\mu(x)$, which were then used to control mutation rates of another GA, referred to as Inner-GA. We first give some details about the Inner- and Meta-GAs, and then describe results of the experiments.

Inner-GA

The Inner-GA is a simple generational genetic algorithm that uses no selection and no recombination. Each genotype in the Inner-GA is a sequence $\omega \in \mathcal{H}_\alpha^l$, and we used populations of 100 individuals. The initial population had equal numbers of individuals at each fitness value, and all runs within the same Meta-GA generation were seeded with the same initial population. Individuals were evolved by the Inner-GA for $t = 500$ generations using simple mutation. The objective was to maximise a fixed fitness function $f(\omega)$. Here, we report results of the following three experiments:

1. $\mathcal{H}_2^{30}$ (i.e. $\alpha = 2$, $l = 30$) and fitness $f(\omega) = -d(\top, \omega)$, where d is Hamming metric.

2. $\mathcal{H}_4^{10}$ (i.e. $\alpha = 4$, $l = 10$) and fitness $f(\omega) = -d(\top, \omega)$, where d is Hamming metric.

3. $\mathcal{H}_4^{10}$ (i.e. $\alpha = 4$, $l = 10$) and fitness $f(\omega)$ defined by a complete DNA-protein affinity landscape for 10-base-pair sequences (Rowe et al., 2010), which we refer to as the aptamer landscape.
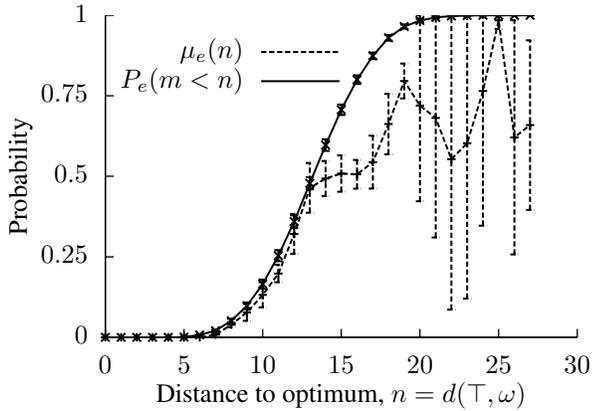
Figure 2: Average of evolved mutation functions $\mu_e(n)$ and CDF $P_e(m < n)$ for fitness $f(\omega) = -d(\top, \omega)$ in $\mathcal{H}_2^{30}$.

Meta-GA

The Meta-GA is a simple generational genetic algorithm that uses tournament selection (a good choice when little is known or assumed about the structure of the landscape). Each genotype in the Meta-GA is a mutation rate function $\mu(x)$, which is a sequence of $l + 1$ real values $\mu \in [0, 1]$ representing per-locus probabilities of mutation. We used populations of 100 individual functions, which were initialised to $\mu(x) = 0$.

The Meta-GA evolved functions $\mu_e(x)$ for $t = 5 \cdot 10^5$ generations to maximise the average fitness in the final generation of the Inner-GA. The Meta-GA used the following selection, recombination and mutation:

- Randomly select three individuals from the population and replace the least fit of these with a mutated crossover of the other two; repeat until all individuals from the population have been selected.

- Crossover (recombination) uses a single cut point chosen randomly (excluding the possibility of being at either end, so that there are no clones).

- Mutation adds a uniform-random number $\Delta\mu \in [-.1, .1]$ to one randomly selected value μ (mutation rate) on the individual (mutation rate function), but then bounds that value to be within $[0, 1]$.

The Meta-GA returned the fittest mutation rate function $\mu_e(x)$. In addition, we recorded empirical frequencies $P_e(x)$ of fitness values $x = f(\omega)$, observed during running the Inner-GA for t generations on the relevant landscape and using that mutation rate function. We note that empirical frequencies $P_e(x)$ counted only the number of phenotypic mutations (i.e. genetic mutations that result in a change in fitness). Empirical frequencies $P_e(x)$ were then used to compute the cumulative distribution functions $P_e(x_r > x)$, which we then compared to the evolved $\mu_e(x)$.
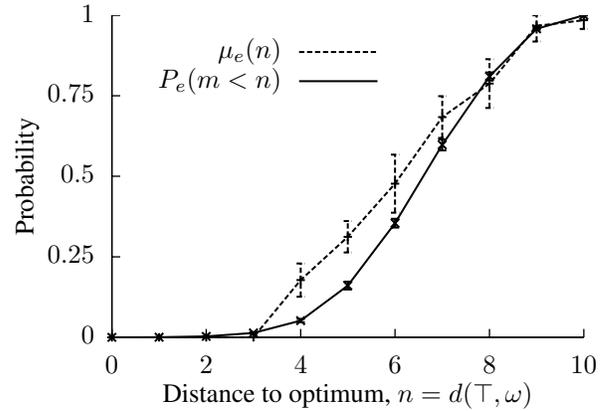

Figure 3: Average of evolved mutation functions $\mu_e(n)$ and CDF $P_e(m < n)$ for fitness $f(\omega) = -d(\top, \omega)$ in $\mathcal{H}_4^{10}$.
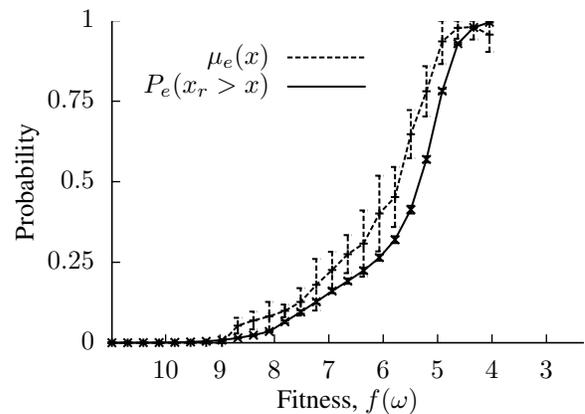

Figure 4: Average of evolved mutation functions $\mu_e(x)$ and CDF $P_e(x_r > x)$ for fitness $f(\omega) = x$ from the aptamer landscape (Rowe et al., 2010) in $\mathcal{H}_4^{10}$.

Experimental Results

We performed multiple runs of each experiment collecting multiple versions of evolved mutation control functions $\mu_e(x)$ and cumulative distribution functions $P_e(x_r > x)$ of observed fitness values. Figures 2, 3 and 4 show the average of these functions from 20 runs together with standard deviations. Figures 2 and 3 are for the experiments in $\mathcal{H}_2^{30}$ and $\mathcal{H}_4^{10}$ respectively, and with fitness $f(\omega)$ defined by the negative Hamming distance $-d(\top, \omega)$ to a fixed optimum $\top$. Figure 4 is for the experiment in $\mathcal{H}_4^{10}$, but with fitness $f(\omega)$ defined by the complete aptamer landscape from (Rowe et al., 2010). The evolved functions $\mu_e(x)$ are approximated fairly by the cumulative distribution functions $P_e(x_r > x)$, supporting heuristic (1). The mismatch in the areas of low fitness can be explained by slower convergence of functions $\mu_e(x)$ in this part of the space $\mathcal{H}_\alpha^l$ due to limited exploration of it by populations of 100 individuals in the Inner-GA, which are small relative to $|\mathcal{H}_\alpha^l| = \alpha^l$.

Discussion

In this work, we have made some progress towards understanding optimal control of mutation rate, and some general principles can be formulated. It appears that choice of a representation space and its topology is crucial, as it defines the monotonic property of a fitness landscape. Our analysis was performed for a Hamming space, but the ideas can be extended to other spaces, such as a space of variable or infinite sequences with p-adic metric. If the right representation has been found, then specific formulae can be derived using geometric analysis in the representation space. These principles can also be extended to sexual reproduction and control of recombination. Our analysis and experiments suggest that an optimal control of mutation rate is based on statistical information about the distribution of fitness values.

The existence of optimal mutation rates that vary depending upon an individual's fitness raises a number of questions about the existence and control of variable mutation rates in biological organisms. For mutation rate control to have evolved in nature, a first prerequisite is that biological mutation rates can vary and are not simply minimised. There is ample evidence that mutation rates do vary in nature, between distantly (Drake et al., 1998) and closely-related (Matic et al., 1997) organisms, between regions of genomes (Lang and Murray, 2008) and even within an organism in stressful versus benign environments (Bjedov et al., 2003). However, the question of whether there may be an adaptive trait, allowing an individual organism to affect the number of mutations between itself and its offspring, dependent upon environmental cues, remains an open question. This would be an example of 'higher-order' selection, that is selection not on the immediate fitness of an individual, but on its ability to produce fitter descendants, potentially many generations later. Such higher order effects have always been questioned in biology, since they might be expected, in real populations, to be swamped by direct selective effects (Pigliucci, 2008). However, discussion has intensified recently over the concepts of 'robustness' and 'evolvability' (Masel and Trotter, 2010). These are higher order effects of somewhat unclear definition; the latter potentially relates directly to the control of mutation rate considered here. Very recent results from experimental evolution of microbial populations show that higher order evolvability effects can indeed play an important part in the evolution of real biological populations (Woods et al., 2011). However, in mechanistic terms, only the gross evolution of mutation rate itself (rather than mutation rate control) in 'mutator' strains has been identified in such experiments (Arjan et al., 1999).

If one moves from complete organisms to viruses and *in silico* quasi-biological evolution, there is more work on optimal mutation rates and their evolution. Optimal mutation rates can be identified (Kamp et al., 2002), relating to the concept of an 'error threshold' (Ochoa et al., 1999) the mutation rate at which selection can no longer be sufficient to balance the deleterious effects of mutation (Biebricher and Eigen, 2005). However, Clune et al. (2008) used digital organisms to show that natural selection does not always effectively evolve optimal mutation rates for adaptation in the long-term, and this fact is particularly apparent when evolution occurs on a rugged fitness landscape. There is evidence that, in nature, epistasis is widespread (e.g. Costanzo et al., 2010), leading to rugged fitness landscapes. This potentially reduces the biological relevance of work, such as ours, with simple fitness functions. Nonetheless, even in rugged landscapes, biological evolution is, empirically, able to occur via locally monotonic accessible paths (Poelwijk et al., 2007), and we find a good agreement between the evolved and theoretical functions, even for a fitness landscape known to be rugged (e.g. Fig. 4 in Rowe et al., 2010). Similarly, temporal variation in fitness landscapes has been highlighted as biologically important (Costanzo et al., 2010), which, while it calls into question the biological relevance of optimal mutation rates in static landscapes, leads back to the potential biological importance of mutation rate variation in response to environmental cues (Stich et al., 2010).

Finally, we observe that understanding of evolution and dynamical systems, such as populations of organisms, may be facilitated by theories of information and information dynamics. In particular, optimisation problems, defined by functions (2) and (3), can be reformulated by replacing time with an information distance between probability distributions. Analytical solutions for such problems can be obtained (e.g. Belavkin, 2010), providing an alternative way to evaluate control functions. Although we do not report such evaluation here, we have observed that these information-theoretic optimal values are achieved when the mutation rate corresponds to a CDF of the 'least informed' distribution of fitness values. Understanding this relation between mutation rate control and information, along with its biological relevance, are some of the directions of our future work.

References

Ackley, D. H. (1987). An empirical study of bit vector function optimization. In Davis, L., editor, *Genetic Algorithms and Simulated Annealing*, chapter 13, pages 170–204. Pitman.

Arjan, J. A., Visser, M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L., and Lenski, R. E. (1999). Diminishing returns from mutation supply rate in asexual populations. *Science*, 283(5400):404–6.

Bäck, T. (1993). Optimal mutation rates in genetic search. In Forrest, S., editor, *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 2–8. Morgan Kaufmann.

Belavkin, R. V. (2010). Information trajectory of optimal learning. In Hirsch, M. J., Pardalos, P. M., and Murphey, R., editors, *Dynamics of Information Systems: Theory and Applications*,

volume 40 of *Springer Optimization and Its Applications Series*. Springer.

Biebricher, C. K. and Eigen, M. (2005). The error threshold. *Virus Res*, 107(2):117–27.

Bjedov, I., Tenaillon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., and Matic, I. (2003). Stress-induced mutagenesis in bacteria. *Science*, 300(5624):1404–9.

Cervantes, J. and Stephens, C. R. (2006). "optimal" mutation rates for genetic search. In Cattolico, M., editor, *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 1313–1320, Seattle, Washington, USA. ACM.

Clune, J., Misevic, D., Ofria, C., Lenski, R. E., Elena, S. F., and Sanjuan, R. (2008). Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. *PLoS Comput Biol*, 4(9):e1000187.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., et al. (2010). The genetic landscape of a cell. *Science*, 327(5964):425–31.

Cowperthwaite, M. C. and Meyers, L. A. (2007). How mutational networks shape evolution: Lessons from rna models. *Annual review of Ecology, Evolution and Systematics*, 38:203–230.

Cox, M. M., Keck, J. L., and Battista, J. R. (2010). Rising from the ashes: DNA repair in Deinococcus radiodurans. *PLoS Genet*, 6(1):e1000815.

Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, 148(4):1667–86.

Eiben, A. E., Hinterding, R., and Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 3(2):124–141.

Eigen, M., McCaskill, J., and Schuster, P. (1988). Molecular quasispecies. *Journal of Physical Chemistry*, 92:6881–6891.

Falco, I. D., Cioppa, A. D., and Tarantino, E. (2002). Mutation-based genetic algorithm: performance evaluation. *Applied Soft Computing*, 1(4):285–299.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Fogarty, T. C. (1989). Varying the probability of mutation in the genetic algorithm. In Schaffer, J. D., editor, *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 104–109. Morgan Kaufmann.

Hakem, R. (2008). DNA-damage repair; the good, the bad, and the ugly. *Embo J*, 27(4):589–605.

Kamp, C., Wilke, C. O., Adami, C., and Bornholdt (2002). Viral evolution under the pressure of an adaptive immune system: Optimal mutation rates for viral escape. *Complexity*, 8(2):28–33.

Knight, C. G., Platt, M., Rowe, W., Wedge, D. C., Khan, F., Day, P. J., McShea, A., Knowles, J., and Kell, D. B. (2009). Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. *Nucleic Acids Res*, 37(1):e6.

Lang, G. I. and Murray, A. W. (2008). Estimating the per-base-pair mutation rate in the yeast saccharomyces cerevisiae. *Genetics*, 178(1):67–82.

Masel, J. and Trotter, M. V. (2010). Robustness and evolvability. *Trends Genet*, 26(9):406–14.

Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., Denamur, E., and Elion, J. (1997). Highly variable mutation rates in commensal and pathogenic Escherichia coli. *Science*, 277(5333):1833–4.

Mühlenbein, H. (1992). How genetic algorithms really work: Mutation and hillclimbing. In einhard Männer and Manderick, B., editors, *Parallel Problem Solving from Nature 2*, pages 15–26, Brussels, Belgium. Elsevier.

Nix, A. E. and Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5(1):77–88.

Ochoa, G. (2002). Setting the mutation rate: Scope and limitations of the $1/l$ heuristics. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2002)*, pages 315–322, San Francisco, CA. Morgan Kaufmann.

Ochoa, G., Harvey, I., and Buxton, H. (1999). Error thresholds and their relation to optimal mutation rates. In *Proceedings of the Fifth European Conference on Artificial Life (ECAL'99)*, volume 1674 of *Lecture Notes in Artificial Intelligence*, pages 54–63, Berlin. Springer-Verlag.

Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nat Rev Genet*, 6(2):119–27.

Pigliucci, M. (2008). Is evolvability evolvable? *Nat Rev Genet*, 9(1):75–82.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., and Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445(7126):383–6.

Rowe, W., Platt, M., Wedge, D. C., Day, P. J., and Kell, D. B. (2010). Analysis of a complete DNA-protein affinity landscape. *Journal of Royal Society Interface*, 7(44):397–408.

Stich, M., Manrubia, S. C., and Lazaro, E. (2010). Variable mutation rates as an adaptive strategy in replicator populations. *PLoS One*, 5(6):e11186.

Vafaee, F., Turán, G., and Nelson, P. C. (2010). Optimizing genetic operator rates using a Markov chain model of genetic algorithms. pages 721–728. ACM.

Waddington, C. H. (1957). *The strategy of the genes; a discussion of some aspects of theoretical biology*. Allen & Unwin, London.

Woods, R. J., Barrick, J. E., Cooper, T. F., Shrestha, U., Kauth, M. R., and Lenski, R. E. (2011). Second-order selection for evolvability in a large escherichia coli population. *Science*, 331(6023):1433–6.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, 1:356–366.

Yanagiya, M. (1993). A simple mutation-dependent genetic algorithm. In Forrest, S., editor, *Proceedings of the 5th International Conference on Genetic Algorithms*, page 659. Morgan Kaufmann.