Wright's Equation and Evolutionary Computation

H. Mühlenbein   Th. Mahnig
RWCP[1] Theoretical Foundation GMD[2] Laboratory
D-53754 Sankt Augustin
Germany
muehlenbein@gmd.de   http://ais.gmd.de/~muehlen/

Abstract: In this paper Wright's equation, formulated in 1931 is proven and applied to evolutionary computation. Wright's equation shows that evolution is doing gradient ascent in the landscape defined by the average fitness of the population. The average fitness W is defined in terms of marginal gene frequencies p_i. Wright's equation is only approximately valid in population genetics, but it is exactly describing the behavior of our Univariate Marginal Distribution Algorithm (UMDA).

Key-Words: genetic algorithms, linkage equilibrium, Wright's equation, selection, factorization of distribution

1 Introduction

The purpose of this paper is twofold. First we analyze a powerful evolutionary algorithm using univariate marginal distributions (UMDA) instead of recombination and mutation of strings as it is done by genetic algorithms. A specific instance of this algorithm can be mathematically described by a set of difference equations. We later found that these equations have been proposed in population genetics by Sewall Wright as early as 1931 [8]. Therefore the second purpose is to relate our analysis to the discussion of these equations in population genetics.

Wright himself [11] gave a short historical overview about the difference equation which later has been named Wright's equation. "Only single-gene distributions were dealt with mathematically in the 1931 paper [8] but these were merely considered to be indications of the sort of thing that is happening in the many dimensions to which the verbal discussion was devoted. In a later paper [9] a formula was given for Δp in cases of multi-factorial heredity and an intermediate optimum. This was taken up more generally in 1937 [10] with the introduction of the symbol W for selective value of the genotypes as a whole,...It was assumed that the local population in question was breeding at random and that it was sufficiently accurate that all loci were combined random." The key formula was for binary alleles as follows:

$$\Delta p_i = p_i(t+1) - p_i(t) = p_i(t)(1-p_i(t))\frac{\frac{\partial W}{\partial p_i}}{W} \qquad (1)$$

where i denotes a locus, and p_i is the frequency of gene i being 1. Wright's extension to multiple alleles is wrong and therefore omitted.

Wright's equation was criticized, especially Fisher [2]. His main argument was that the evaluation of the average fitness W needs on the order of 2^n terms. The interpretation of Wright's equation is debated till today[1]. For UMDA the difference equation is exact, whereas it is an approximation in population genetics.

The outline of the paper is as follows. In Section 2 we prove Wright's equation. It is valid in evolutionary computation for proportionate selection. Then the landscape metaphor is discussed. We give a simple formula for computing the average fitness W. In Section 4 we discuss tournament selection for a simple linear fitness function.

2 Univariate Marginal Distribution Algorithm

Let $\mathbf{x} = (x_1, \ldots, x_n)$ denote a vector, $x_i \in \Lambda_i = \{0, 1, 2, \ldots, m_i\}$. Λ_i just represents a set of $m_i + 1$ symbols, representing the different alleles. We use the following conventions. Capital letters X_i denote variables, small letters x_i assignments. Let a function $f : \boldsymbol{X} \to \mathbb{R}_{>0}$ be given. We consider the

---

[2] Real World Computing Partnership

[2] GMD - Forschungszentrum Informationstechnik

optimization problem $\mathbf{x}_{opt} = \text{argmax} f(\mathbf{x})$.

Definition: Let $p(\mathbf{x}, t)$ denote the probability of $\mathbf{x}$ in the population at generation t. Then $p_i(x_i, t) = \sum_{\mathbf{x}, X_i = x_i} p(\mathbf{x}, t)$ defines the univariate marginal distributions of variable X_i.

Note that $\sum_{x_i \in \Lambda_i} p_i(x_i, t) = 1$. This means that the parameters of the univariate marginal distributions are not independent. For notational simplicity we choose $p_i(0)$ to be the dependent parameter. It can be eliminated, if appropriate. We write $p_i(x_k)$ if just one generation is discussed. We recall two popular recombination/crossover operator used in genetic algorithms.

Definition: Let two strings $\mathbf{x}$ and $\mathbf{y}$ be given. In *one-point crossover* the string $\mathbf{z}$ is created by randomly choosing a crossover point $0 < l < n$ and setting $z_i = x_i$ for $i \leq l$ and $z_i = y_i$ for $i > l$. In *uniform crossover* z_i is randomly chosen with equal probability from $\{x_i, y_i\}$.

In order to derive Wright's equation, we have to introduce a special distribution.

Definition: *Robbins' proportions* are defined by the distribution

$$\pi_p(\mathbf{x}, t) := \prod_{i=1}^{n} p_i(x_i, t) \tag{2}$$

A population in Robbins' proportions is called to be in *linkage equilibrium* in population genetics.

In [3, 6] we have shown: All complete recombination schemes lead to the same univariate marginal distributions after one step of selection and recombination. If recombination is used for a number of times without selection, then the genotype frequencies converge to linkage equilibrium. This means that *all genetic algorithms are identical if after one selection step recombination is done without selection a sufficient number of times.* This fundamental algorithm keeps the population in linkage equilibrium. Linkage equilibrium was also assumed by Wright.

Instead of performing recombination a number of times in order to converge to linkage equilibrium, one can achieve this in one step by *gene pool recombination* [7]. In gene pool recombination a new string is computed by randomly taking for each loci a gene from the distribution of the selected parents. This means that gene x_i occurs with probability $p^s(x_i)$ in the next population. $p^s(x_i)$ is the distribution of x_i in the selected parents. Thus new strings $\mathbf{x}$ are generated according to the distribution

$$p(\mathbf{x}, t+1) = \prod_{i=1}^{n} p_i^s(x_i, t) \tag{3}$$

One can simplify the algorithm still more by directly computing the univariate marginal frequencies from the data. Equation 3 is used to generate new strings. This method is used by the *Univariate Marginal Distribution Algorithm* (UMDA).

UMDA

- **STEP 0:** Set $t \Leftarrow 1$. Generate $N \gg 0$ points randomly.
- **STEP 1:** Select $M \leq N$ points according to a selection method. Compute the marginal frequencies $p_i^s(x_i, t)$ of the selected set.
- **STEP 2:** Generate N new points according to the distribution $p(\mathbf{x}, t+1) = \prod_{i=1}^{n} p^s(x_i, t)$. Set $t \Leftarrow t+1$.
- **STEP 3:** If termination criteria are not met, go to STEP 1.

Let $v = \sum_{i=1}^{n}(m_i + 1)$. UMDA formally depends on v parameters, the marginal distributions $p_i(x_i)$. We now interprete the average $\bar{f}(t) = \sum_x p(x, t) f(x)$ as a function which depends on $p_i(x_i)$. To emphasize this dependency we write (in accordance with Wright)

$$W(\mathbf{p}) = W\big(p_1(0), p_1(1), \ldots, p_n(m_n)\big) := \bar{f}(t) \tag{4}$$

We can now formulate difference equations, describing the dynamic behavior of $p_i(x_i)$.

Theorem 1. *For infinite populations and proportionate selection UMDA changes the gene* *frequencies as follows:*

$$p_i(x_i,t+1) = p_i(x_i,t)\frac{\bar{f}_i(x_i,t)}{W(\mathbf{p}(t))} \tag{5}$$

where $\bar{f}_i(x_i,t) = \sum_{\mathbf{x},X_i=x_i} f(\mathbf{x}) \prod_{j\neq i}^{n} p_j(x_j,t)$. *Using a formal derivative* $\frac{\partial W}{\partial p_i(x_k)}$, *the equations can also be written as*

$$p_i(x_k,t+1) = p_i(x_k,t) + p_i(x_k,t)\frac{\frac{\partial W}{\partial p_i(x_k)} - W}{W} \tag{6}$$

Furthermore the average fitness W never decreases.

$$W(\mathbf{p}(t+1)) \geq W(\mathbf{p}(t)) \tag{7}$$

The theorem has been proven in [3]. Note that the derivatives of W are obtained by formal differentiation of equation (4). We discuss the theorem with a simple example, a linear function.

Example: $f(x) = \sum_i a_i x_i, \quad x_i \in \{0,1\}$

After some tedious manipulations one obtains:

$$\begin{aligned} W(\mathbf{p}) &= \sum_i a_i p_i(1) \\ \frac{\partial W}{\partial p_i(1)} &= a_i + \sum_{j\neq i} a_j p_j(1) \end{aligned}$$

We obtain the difference equation

$$\Delta p_i(1) = p_i(1,t)\big(1-p_i(1,t)\big)\frac{a_i}{\sum_i a_i p_i(1,t)} \tag{8}$$

This equation has been approximately solved in [4]. □

This example shows that the expressions for W and its derivatives can be surprisingly simple. $W(\mathbf{p})$ can formally be obtained from $f(x)$ by exchanging x_i with $p_i(1)$. But the formal derivative of $W(p)$ cannot be computed from the simple $W(p)$ expression! This problem will be investigated in the next section.

Wright's equation is obtained from equation 6 by using real derivatives in Euclidian spaces. Thus we have to eliminate one parameter for each locus. In order to minimize possible confusion later we denote the average fitness $W(\mathbf{p}) = \bar{f}(t)$ with parameters $p_i(0)$ eliminated as $\tilde{W}$.

Definition: Let $\bar{\Lambda}_i = \Lambda_i \setminus \{0\}$. If we eliminate $p_i(X_i=0,t)$ in $W(\mathbf{p})$ by inserting $1 - \sum_{x_i\in\bar{\Lambda}_i} p_i(x_i,t)$ then we obtain $\tilde{W}(\mathbf{p})$.

We can now formulate the main theorem.

Theorem 2 (Wright's Equation). *For infinite populations and proportionate selection UMDA changes the gene frequencies as follows:*

$$\begin{aligned} p_i(x_i,t+1) = p_i(x_i,t) + p_i(x_i,t)\cdot \\ \cdot\frac{\frac{\partial \tilde{W}}{\partial p_i(x_i)} - \sum_{y_i\in\bar{\Lambda}_i} p_i(y_i,t)\frac{\partial \tilde{W}}{\partial p_i(y_i)}}{\tilde{W}(\mathbf{p})} \end{aligned} \tag{9}$$

Proof: Note that

$$\frac{\partial \tilde{W}}{\partial p_i(x_i)} = \bar{f}_i(x_i,t) - \bar{f}_i(0,t)$$

Furthermore we have

$$\sum_{x_i\in\Lambda_i} p_i(x_i,t)\bar{f}_i(x_i,t) = \tilde{W}(t)$$

We next compute for $x_i \in \bar{\Lambda}_i$

$$\begin{aligned} \bar{f}_i(x_i,t) - \tilde{W}(\mathbf{p}(t)) =& \frac{\partial \tilde{W}}{\partial p_i(x_i)} - \bar{f}_i(0,t) \\ & - \sum_{y_i\in\Lambda_i} p_i(y_i,t)\bar{f}_i(y_i,t) \end{aligned}$$

$$\begin{aligned} \sum_{y_i\in\Lambda_i} p_i(y_i,t)\bar{f}_i(y_i,t) =& \sum_{y_i\in\bar{\Lambda}_i} p_i(y_i,t)\frac{\partial \tilde{W}}{\partial p_i(y_i)} \\ & + \bar{f}_i(0,t) \end{aligned}$$

Inserting this equation into the difference equation 5 gives the conjecture. □

This is the exact formulation of Wright's equation for multiple alleles at n loci. For binary alleles we obtain equation 1. The above equations completely describe the dynamics of UMDA with proportionate selection.

The equation has been proposed by Wright [8] for theoretical analysis of population genetics. There has been a fierce battle between Wright and Fisher concerning the importance of Wright's equation. Fisher especially criticized Wright's average fitness function W. He wrote [2]: "Prof. Wright confuses the number of genotypes, e.g. 3^{1000}, (for 1000 loci) which may be distinguished among individuals, with the continuous field of variation of gene frequencies. Even if a potential function, such as W is supposed to be, really existed, the large number of genotypes supplies no reason for thinking that even one peak, maximal for variations of all gene ratios should occur in this field of variation."

Wright was not able to refute this argument. In 1963 he wrote [11]: "The summation in the formula for $\tilde{W}$ has, however, as many terms as there are kinds of genotypes, 3^{1000} for 1000 pairs of alleles. This, of course, points to a practical difficulty in calculating Δp for more than two or three pairs of interacting factors, unless a regular model is postulated. There was no confusion."

We will show in the next section that both, Fisher and Wright, overlooked that if the fitness function is simple, $\tilde{W}$ can be easily computed. This is discussed next.

3 Average Fitness and the Landscape Metaphor

Wright is also the father of the landscape metaphor in population genetics. But Wright actually used two quite different definitions for the landscape, apparently without realizing the fundamental distinction between them. The first describes the relation between the genotypes and their fitness, while the second describes the relation between the allele frequencies in a population and its mean fitness. The first definition is just the fitness function $f(x)$ used in evolutionary computation, the second one is the average fitness $W(p)$. The second definition is much more useful, because it lends to a quantitative description of the evolutionary process, i.e. Wright's equation. Theorem 1 further shows that for binary alleles the population evolves uphill on the $W(p)$ landscape at a rate proportional to its gradient.

For notational simplicity we only show how to compute $\tilde{W}$ for binary alleles. Let $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_i \in \{0,1\}$ be a multi-index. We define with $0^0 := 1$:

$$\mathbf{x}^\alpha := \prod_i x_i^{\alpha_i}$$

Definition: The representation of a binary discrete function using the ordering according to function values is given by

$$f(\mathbf{x}) = f(0,\dots,0)(1-x_1)\cdots(1-x_n) + \dots \\ + f(1,\dots,1)x_1\cdots x_n \quad (10)$$

The representation using the ordering according to variables is

$$f(\mathbf{x}) = \sum_\alpha a_\alpha x^\alpha \quad (11)$$

$\max\{|\alpha|_1 = \sum_i \alpha_i : \ a_\alpha \neq 0\}$ is called the order of the function.

In both representations the function is linear in each variable x_i. The following lemma is obvious.

Lemma: *The two representations are unique. There exist a unique matrix A of dimension $2^n * 2^n$ such that*

$$a_\alpha = (Af)_\alpha$$

We now use this result for $\tilde{W}$. Let $S_i = \{q_i | \sum_{k\in\bar{\Lambda}_i} q_i(x_i) \leq 1;\ 0 \leq q_i(x_i) \leq 1\}$ and $S = \prod_i S_i$ the Cartesian product. For binary alleles we have $S = [0,1]^n$.

Lemma: $\tilde{W}(\mathbf{p}) := \bar{f}(t)$ *is an extension of $f(x)$ to S. There exist two representations for $\tilde{W}(p)$. These are given by*

$$\tilde{W}(\mathbf{p}) = f(0,\dots,0)(1-p_1)\cdots(1-p_n) + \dots \\ + f(1,\dots,1)p_1\cdots p_n \quad (12)$$

$$\tilde{W}(\mathbf{p}) = \sum_\alpha a_\alpha p^\alpha \quad (13)$$

Equation 12 is the definition of average fitness. The computation is exponential in n. But if the function is given in analytical form (Equation 11) and the order of the function is bounded by a constant independent of n, then $\tilde{W}(\mathbf{p})$ can be computed in polynomial time. Equation 13 can also be used to compute the derivative of $\tilde{W}$. It is given by

$$\frac{\partial \tilde{W}(p)}{\partial p_i(1)} = \sum_{\alpha|\alpha_i=1} a_\alpha p^{\alpha'} \quad (14)$$

with $\alpha'_i = 0, \alpha'_j = \alpha_j$.

We will now characterize the attractors of UMDA.

Theorem 3. *The stable attractors of UMDA with proportionate selection are at the corners of S, i.e.* $p_i \in \{0,1\}$, $i = 1,\ldots,n$. *The stable attractors are local maxima of* $f(x)$. *A local maximum is a maximum with respect to one bit changes. In the interior there are only saddle points where grad* $W(p)) = 0$. *UMDA solves the continuous optimization problem* $\operatorname{argmax}\{\tilde{W}(\mathbf{p})\}$ *in S by gradient ascent.*

Proof: $\tilde{W}$ is linear in p_i, therefore it cannot have any local maxima in the interior. Points with *grad* $W(p) = 0$ are unstable fix points of UMDA.

We next show that boundary points which are not local maxima of $f(x)$ cannot be attractors. We prove the conjecture indirectly. Without loss of generality we choose the boundary point $\hat{p} = (1,\ldots,1)$. We now take an arbitrary neighbor, i.e $p^* = (0,1,\ldots,1)$. The two points are connected at the boundary by

$$p(z) = (1-z, 1, \ldots, 1) \qquad z \in [0,1]$$

We know that $\tilde{W}$ is *linear* in the parameters p_i. Because $\tilde{W}(p^*) = f(0,1,\ldots,1)$ and $\tilde{W}(\hat{p}) = f(1,\ldots,1)$ we have

$$\tilde{W}(p(z)) = f(1,\ldots,1) + z \cdot [f(0,1,\ldots,1) - f(1,\ldots,1)]. \quad (15)$$

If $f(0,1,\ldots,1) > f(1,\ldots,1)$ then $\hat{p}$ cannot be an attractor of UMDA. The mean fitness increases with z. □

The extension of the above lemmata to multiple alleles and multivariate distributions is straightforward, but the notation becomes difficult. Multivariate distributions are used by an extension of UMDA, the *Factorized Distribution Algorithm* FDA [5, 6]

4 The Selection Problem

Fitness proportionate selection is the undisputed selection method in population genetics. It is considered to be a model for *natural selection.* But this selection method strongly depends on the fitness values. When the population approaches an optimum, selection gets weaker and weaker, because the fitness values become similar. This behavior can be derived from Equation 6. The step-size decreases if p_i approaches 0 or 1. This slows down the convergence near the corners.

Therefore breeders of livestock use other selection methods. For large populations they mainly apply *truncation selection.* It works as follows. A truncation threshold $\tau < 1$ is fixed. Then the τN best individuals are selected as parents for the next generation. These parents are then randomly mated.

We mainly use truncation selection in our algorithms. Another popular scheme is *tournament selection of size k.* Here k individuals are randomly chosen. The best individual is taken as parent. Unfortunately the mathematical analysis for both selection methods is more difficult than for proportionate selection. Analytical results for tournament selection have been first obtained by Mühlenbein [3].

We have been able to compute Wright's equation for tournament selection for one of the simplest cases, the linear function $OneMax(n) = \sum_i x_i$. If we consider the special case of $p_1(0) = p_2(0) = \ldots = p_n(0) := p(0)$ we obtain a difference equation with just one parameter [3]. We only give the difference equation for 3 variables.

$$p(t+1) = p(t) + p(t)(1-p(t) \cdot \cdot \left(1 - 2p(t) + 4p(t)^2 - 4p(t)^3 + 2p(t)^4\right) \quad (16)$$

The highest exponent of p is $2*3$, for general n the exponent is $2n$. This shows that even for the linear $OneMax$ function we now encounter the problem Fisher mentioned. The computation of $W(\mathbf{p})$ becomes very difficult. We have not been able to solve the above equation analytically. Using the theory of breeding we have computed an approximate difference equation. It is given by [3]

$$p(t+1) = p(t) + \frac{I}{n}\sqrt{np(t)(1-p(t))} \quad (17)$$

where I is called *selection intensity.* For tournament selection of size 2 we have $I = 1/\sqrt{\pi}$. The reader interested in this approximation is referred to [3].

For proportionate selection we have the simple expression $\tilde{W} = \sum_i p_i = np$. This gives

$$p(t+1) = p(t)\big(1-p(t)\big)\frac{1}{np(t)} = \frac{1-p(t)}{n} \quad (18)$$

The three difference equations 16,17, 18 are compared in Figure 1. The initial value is $p(0) = 0.1$. Proportionate selection moves at first stronger than tournament selection to the attractor, but then it slows down. At generation 7 tournament selection overtakes proportionate selection. The approximation given by Equation 17 is fairly good. It even gets better for larger n.

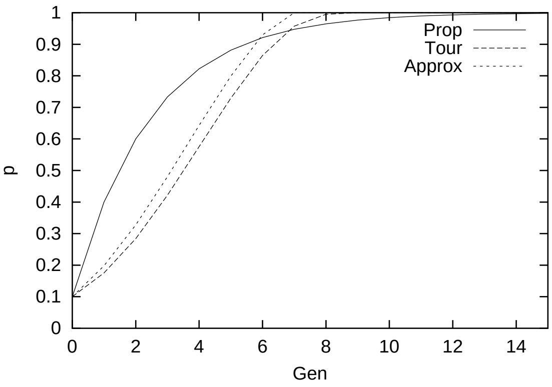

Figure 1: Comparison of selection methods for OneMax(3), Approx. equation 17

5 Summary and Outlook

The paper has shown that the interaction between classical population genetics and evolutionary computation can be fruitful. But it is not a one way. Evolutionary computation can contribute to population genetics and vice versa.

Because of space limitations, we discuss the application of Wright's equation in a separate paper.

References

[1] A. W. F. Edwards. Sewall wright's equation. *Theor. Population Biology*, 57:67–70, 2000.

[2] R.A. Fisher. Average excess and average effect of a gene substitution. *Annals of Eugenics*, 11:53–63, 1941.

[3] H. Mühlenbein. The equation for the response to selection and its use for prediction. *Evolutionary Computation*, 5(3):303–346, 1997.

[4] H. Mühlenbein and Th. Mahnig. Convergence theory and applications of the factorized distribution algorithm. *Journal of Computing and Information Technology*, 7:19–32, 1999.

[5] H. Mühlenbein and Th. Mahnig. FDA – a scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary Computation*, 7(4):353–376, 1999.

[6] H. Mühlenbein, Th. Mahnig, and A. Rodriguez Ochoa. Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, 5:215–247, 1999.

[7] H. Mühlenbein and H.-M. Voigt. Gene pool recombination in genetic algorithms. In J.P. Kelly and I.H Osman, editors, *Metaheuristics: Theory and Applications*, pages 53–62, Norwell, 1996. Kluwer Academic Publisher.

[8] S. Wright. Evolution in mendelian populations. *Genetics*, 16:97–159, 1931.

[9] S. Wright. The analysis of variance and the correlations between relatives with respect to deviation from an optimum. *J. Genetics*, 30:243–256, 1935.

[10] S. Wright. The distribution of gene frequencies in populations. *Proc. Nat. Acad. Sci*, 24:253–259, 1937.

[11] S. Wright. Stochastic processes in evolution. In J. Gurland, editor, *Stochastic Models in Medicine and Biology*, 1964.