Appears in "Neural Computing and Applications", Springer-Verlag, 1:1, pp67-90, 1993.

Genetic Set Recombination and its Application to Neural Network Topology Optimisation

Nicholas J. Radcliffe

njr@epcc.ed.ac.uk

Edinburgh Parallel Computing Centre, University of Edinburgh, King's Buildings, EH9 3JZ, Scotland

Abstract

Forma analysis is applied to the task of optimising the connectivity of a feed-forward neural network with a single layer of hidden units. This problem is reformulated as a multiset optimisation problem and techniques are developed to allow principled genetic search over fixed- and variable-size sets and multisets. These techniques require a further generalisation of the notion of gene, which is presented. The result is a non-redundant representation of the neural network topology optimisation problem together with recombination operators which have carefully designed and well-understood properties. The techniques developed have relevance to the application of genetic algorithms to constrained optimisation problems.

1 Introduction

While genetic algorithms have been applied to a number of problems in neural networks, there are severe difficulties with this endeavour. This is significant for two reasons. First, many of the problems in neural networks are important in their own right and do not presently have any wholly satisfactory means of resolution. A good example of this is the choice of network topology. Secondly, the failure modes of the genetic algorithm seen in neural network applications are common to a broader class of problems, and their study can yield more general insights.

This paper is a study in the application of forma analysis (Radcliffe [22, 23]) to this and related problems. It begins in section 2 with a brief review and a discussion of the difficulties with previous genetic approaches to problems in neural networks. This is followed, in section 3, by a short review of schema- and forma analysis and a discussion of the "permutation problem" for neural networks. (Although the paper is intended to be self-contained, the reader may find it easier to follow after reading Radcliffe [23].)

The core of the paper, sections 4, 5 and 6 is a study of the application of forma analysis to optimisation problems for which the solution is a set or multiset. Both fixed- and variable-size sets and multisets are considered. Section 4 reviews naïve approaches to these problems, section 5 uses forma analysis to gain further insights and to develop a more satisfactory formulation of the problem, and section 6 is a study of the general phenomenon of "non-separability" of formae. The reason for studying these problems is that in section 7 neural network topology optimisation problems are reformulated as multiset optimisation problems, and the theory developed in the preceding sections becomes directly applicable. This section includes a discussion of sub-parameter-level recombination with particular reference to hidden nodes.

The paper closes with a summary and discussion of the results presented, and draws out some of the wider implications for genetic search in other domains.

2 Genetic Approaches to Neural Networks

Genetic algorithms are increasingly being applied to problems in neural networks. Rudnick [24] and Weiss [34] have produced excellent bibliographies for this field in 1990. A number of approaches can be distinguished, all of which have had limited success, and most of which have concentrated on Rumelhart-type feed-forward networks. The two primary areas of activity have been:

1. *Topology Optimisation.*
   The genetic algorithm is used to select a topology (pattern of connectivity) for the network which is then trained using some fixed training scheme, most commonly back-propagation of errors (Rumelhart [25]). This approach is inherently computationally demanding because

the complete conventional training phase (itself computationally intensive) is required simply to evaluate the fitness of a chromosome (network topology). The approach remains reasonably attractive despite this because of the paucity of principled alternative methods for selecting the network topology. Representative studies in this class include those of Miller, Todd & Hegde [15], Harp, Samad & Guha [11, 10], Whitley, Starkweather & Bogart [39], Mühlenbein [18] and Hancock [9]. This class of problems is addressed in section 7 of the paper.

2. *Genetic Training Algorithms.*
   Selecting weights for a neural network is itself an optimisation problem,[1] and a genetic algorithm can naïvely be applied to it, using an inverse error as the measure of utility (fitness). Whitley and his co-workers [35, 39, 38] have done much work in this area, and the study by Montana and Davis [16] is especially ingenious and noteworthy.

Hybrid approaches have also been discussed (Radcliffe [21]), and there have been studies in which genetic algorithms have been used to tune the parameters of other training schemes, including initial weight configurations (Belew, McInerny & Shraudolph [1]).

All of these approaches have associated problems, which have been discussed by Montana & Davis [16], Radcliffe [20, 21], Belew *et al.* [1] and Whitley *et al.* [37]. Principal among these, and appearing in many different guises, is a permutational redundancy associated with the arbitrariness of labels of topologically equivalent hidden nodes. Specifically, to take an extreme case, in a fully-connected feed-forward, layered network with a single hidden layer comprising N_h units, (figure 1) there is approximately an $N_h!$ potential redundancy associated with the indistinguishability of networks with relabelled hidden units.

If the genetic representation (whether it be topologies or weights that form the search space) distinguishes between networks which differ only by the labelling of hidden units, the search space is enormously enlarged. While optima usually become more numerous by a comparable factor, the global nature of genetic search tends to make navigation through the enlarged search space very difficult (Radcliffe [20, 21], Belew *et al.* [1], Whitley *et al.* [37] and section 3.2). Genetic algorithms are sensitive to the potential for redundant representations in a way that most other search schemes (for example, gradient techniques and stochastic processes like simulated annealing) are not. There are two (related) reasons for this:

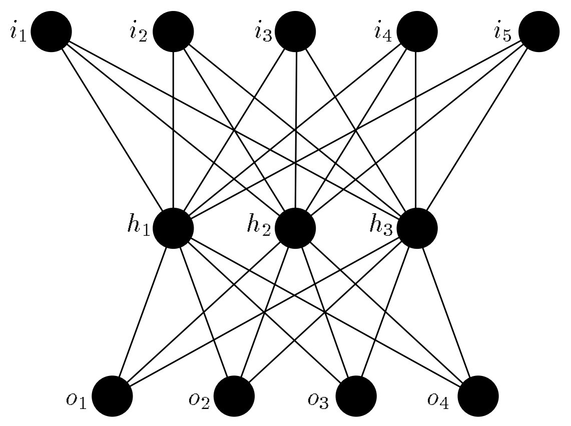

Figure 1: A fully-connected 5–3–4 layered neural network. Notice that the hidden node labels h_1, h_2, h_3 are arbitrary (providing that the hidden nodes are identical) giving, in this case, a $3!$ redundancy in labelling.

1. local techniques tend to make 'smaller' moves in the search space than those possible under genetic recombination ("crossover");
2. most techniques maintain only a single solution rather than a population of solutions.

The relationship between these two points should be clear: the danger is exemplified by the case where two equivalent networks (identical up to a re-labelling of hidden units) can be recombined to produce a child which is not equivalent to them. This is a phenomenon which is not seen in "conventional" genetic search (for example, simple parameter optimisation) and it has been strongly argued elsewhere the problem is highly detrimental to the effectiveness of the search process both in the specific context of neural networks (Radcliffe [20, 21], Belew *et al.* [1]) and more generally (Radcliffe [22, 23]).

3 Schema and Forma Analysis

3.1 Formulation and Principles

In order to understand the motivations for the ideas put forward in this paper it is necessary briefly to review some of the theory of genetic algorithms.

Holland's ground-breaking formulation and analysis of genetic algorithms introduced the theoretical

[1] albeit one which often has a rather poorly-defined objective function

framework of *schema analysis*, and the well-known (if often poorly-expressed) *Schema Theorem* (Holland [12]). This formulation applies primarily to k-ary[2] string chromosomal representations for which each locus (site) on the chromosome has a well-defined meaning.[3]

In 1985 Goldberg & Lingle [8] extended Holland's work to cover permutation-based problems (such as the well-known Travelling Sales-rep Problem) through the introduction of *o*-schemata (see also Goldberg [5]). More recently, Vose & Leipins [32, 33] and Radcliffe [22, 23] have independently further generalised Holland's results to take in very much more general objects which Vose calls *predicates* and Radcliffe terms *formae*. This paper uses and builds upon Radcliffe's formulation.

A schema may be viewed as a set of chromosomes which share some specified subset of their genes. Holland introduced a "don't care" symbol □ to aid the description of schemata, so that the schema $1\,\square 0\,\square$ is the set of all chromosomes which have a one at their first locus and a zero at their third locus. A forma, similarly, may be viewed as a set of chromosomes which are related by some (any) specific characteristic: this need not be the sharing of gene values.[4] It is convenient to regard both schemata and formae as equivalence classes of solutions under given (often implicit) equivalence relations over the representation space $\mathcal{C}$ (the set of chromosomes).

One of the central tenets of forma analysis is that formae should be chosen which group together chromosomes coding solutions which might plausibly have similar performance. Having chosen such formae, genetic operators are constructed with a view to manipulating solutions[5] in meaningful ways. Specifically, the aim is to build recombination operators which *respect* forma membership and *properly assort* formae (Radcliffe [22, 23]). These terms are explained and illustrated in figures 2 and 3 respectively, and are defined more rigorously in section 6.1.

3.2 Application to Neural Networks

In addition to increasing the size of the search space, the permutation problem described in section 2 makes

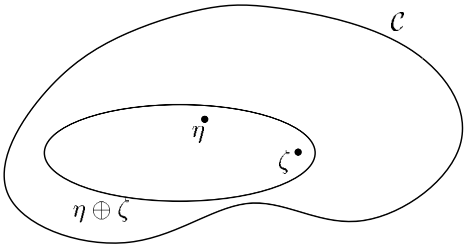

Figure 2: A recombination operator is said to *respect* a set of formae if given any pair of chromosomes η and ζ, all of their children under recombination are members of all the formae to which *both* parents belong. The *similarity set* of η and ζ, written $\eta \oplus \zeta$, is the smallest forma which contains them both. This can be constructed as the intersection of all formae containing them both parents, illustrated above. Respect amounts to the requirement that each child produced by recombination lies in the similarity set of its parents.

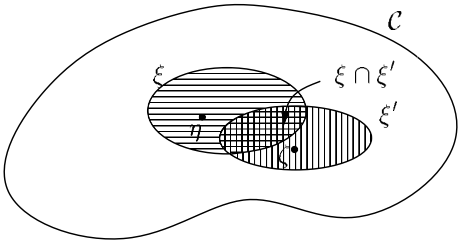

Figure 3: A key function of recombination is that it mixes features from parents so that children may exhibit traits inherited from each. Given two parents, η and η' which belong to formae ξ and ξ' respectively, it may be possible to generate children which are members of *both* ξ and ξ' (i.e. children which exhibit both the traits captured by the two formae). A recombination operator which will, with non-zero probability, produce a child in the intersection of arbitrary formae ξ and ξ' (given parents $\eta \in \xi$ and $\eta' \in \xi'$, and assuming that $\xi \cap \xi' \neq \emptyset$) is said *properly to assort* the formae under consideration. It should be noted that the requirements of respect (figure 2) and proper assortment are not always compatible, though for many sets of formae (including schemata) they are.

[2] base k, e.g. $k = 2$ gives binary, $k = 8$ octal etc.

[3] Of course, the theorem applies to *any* string-based representation given suitable coefficients quantifying the disruptive effects of the genetic operators, but the *observed* schema averages on which the theorem crucially depends will have usefully low variance only—loosely—when the loci have well-defined meanings. This is discussed in detail in Radcliffe [21].

[4] At least, not as genes are conventionally understood.

[5] strictly, their chromosomal representatives

respect and proper assortment rather hard to ensure. These differing aspects of the permutation problem will be referred to as the *numerical permutation problem* and the *navigational permutation problem* respectively, and will now be considered in turn.

Throughout the paper a distinction will be made between a "true" search space, $\mathcal{S}$, consisting of the actual structures under consideration (in the present case network topologies) and a representation space (or space of chromosomes) $\mathcal{C}$. Assume that networks with N_i input nodes, N_o output nodes and up to N_h hidden nodes are considered. Then the number of hidden node types is $2^{N_i+N_o}$, because the connection to each external node may be present or absent.

The number of network topologies (the size of $\mathcal{S}$) is given by

$$|\mathcal{S}| \approx \frac{(2^{N_i+N_o})^{N_h}}{N_h!}. \tag{1}$$

(The approximation in this expression is that all the hidden nodes have different connectivities, justifying the $N_h!$ in the denominator. This approximation is good when the number of hidden node types vastly exceeds the maximum number of hidden units, which is almost always the case.)

For example, if there are ten input nodes, ten output nodes and a maximum of ten hidden nodes

$$|\mathcal{S}| \approx \frac{(2^{20})^{10}}{10!} \approx \frac{1.6 \times 10^{60}}{3 \times 10^{6}} \approx 4 \times 10^{53}. \tag{2}$$

The (naïve) representation space $\mathcal{C}$ consists of chromosomes which use one bit to mark the presence or absence of a connection between each external node and each *labelled* hidden node. The size of $\mathcal{C}$ is given by the numerator alone,

$$|\mathcal{C}| = (2^{20})^{10} \approx 1.6 \times 10^{60}. \tag{3}$$

While redundancy which expands the size of the search space by a factor of more than a million for even a problem of modest size may at first seem daunting, the reader may be tempted to reflect that the difference between overall sizes 10^{60} and 10^{54} seems rather less significant. This feeling may be reinforced by observing that as the size of the network increases, the rate of growth of the size of the search space (characterised by the numerator, $2^{(N_i+N_o)N_h}$) outstrips the rate of growth of the permutation problem (for likely values of node numbers N_i, N_o, and N_h) which grows only as $N_h!$. Thus, for example, increasing the size of the problem from a 10–10–10 network to an 11–11–11 network increases the size of $\mathcal{S}$ from 10^{54} to 10^{65}.

Complacency, however, would be misplaced. For suppose that some formae were constructed in the true search space $\mathcal{S}$ but that the representation used were redundant in the way described (i.e. larger by a factor of around $N_h!$). Ensuring respect and assortment of these formae would be essentially impossible.

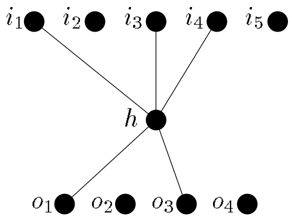

Figure 4: A hidden node h, which can be described by the set of external nodes to which it is connected, $\{\, i_1, i_3, i_4, o_1, o_3 \,\}$. This can also, of course, be conveniently described by the binary string 101101010 where a one indicates a the presence of a connection and a zero the absence, and where the nodes are laid out in order with the input nodes preceding the output nodes.

To see this, imagine that there are two beneficial sets of hidden nodes, where a hidden node is considered complete with its set of external connections (figure 4). Assume that one chromosome η represents a network which contains the first "good" set of hidden units, and that another chromosome ζ codes another network containing the second beneficial set of units. Thinking of these sets of units as "building blocks", the aim would be to bring the two together in a single chromosome by recombination. If, however, the hidden unit labels of the first beneficial set of nodes on η overlap with the labels of the second set on ζ, recombination will be unable to bring these two sets together, no matter how often it is applied[6] (figure 5).

At one level, this can be viewed simply as an example of the *numerical* permutation problem, for "all" that the genetic algorithm needs to do is to construct a similar chromosome ζ' which is like ζ but has node labels for the second "good" set of hidden units which do not overlap with those of the first set on η; in practice, the difficulty is worse.

A good way to see this is to consider an increasingly common way of implementing genetic algorithms, in which an *isolated sub-population model* is used in contrast to the traditional *panmictic* population. The isolated sub-population model consists of a number

[6] That is, the problem is not merely on of *proper* assortment, for the formae cannot be *weakly* assorted either (Radcliffe [23]).

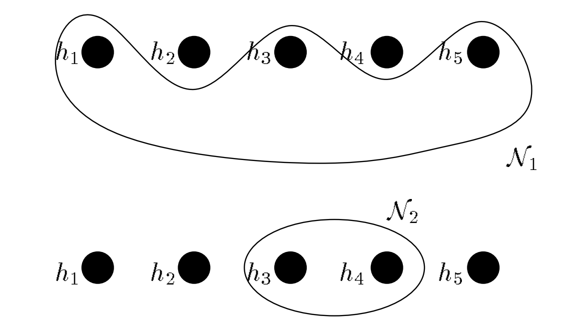

Figure 5: The hidden layers from two networks, $\mathcal{N}_1$ and $\mathcal{N}_2$, the first of which can be trained to recognise the first half of a training set reliably, and the second of which can reliably be trained to classify the second half. If the label of the hidden node h_3 on the second network were changed to h_2 recombination would be possible without disturbing either subnetwork, whereas this is impossible with the fixed but arbitrary hidden node labels shown.

of genetic algorithms each tackling the same search problem with separate populations. There is usually occasional migration of solutions between subpopulations. This approach is important both because of its convenience for parallel implementations and because the maintenance of isolated sub-populations is helpful in delaying loss of genetic diversity, encouraging "speciation" and reducing the number of evaluations required to find an acceptable solution (Norman [19], Mühlenbein [17] etc.)

Consider the relatively likely situation in which η is drawn from a population on one processor which has largely converged on network topologies which include the first beneficial set of hidden units, and ζ is drawn from a different population in which the second group of units is common. The clash of node labellings will ensure that no matter how often chromosomes from the two populations are brought together, they will be unable to be recombined to produce a solution which contains both the beneficial sets of units. This is the *navigational* permutation problem. This difficulty could, of course, arise in a similar way in the panmictic population model if the population contained two "species", each possessing one of the two beneficial sets of nodes, provided that the fitnesses of the two species were such that neither would be quickly destroyed by the population dynamics.

The foregoing discussion should convince the reader that the permutation problem is a serious impediment to genetic search. It is not isolated to neural network problems, though this is the present focus.[7] A major aim of this paper is to suggest a possible representation for neural network topology optimisation which avoids permutational redundancy and allows formae defined in a way independent of hidden unit labels (that is, formae which are well-defined in $\mathcal{S}$, as well as $\mathcal{C}$) to be respected and properly assorted. This formulation builds on the idea of regarding hidden units complete with their external connections as basic entities, and views a network topology as a collection of such hidden nodes. This is the reason for the concentration on set and multiset optimisation problems in the following sections.

4 Sets, Multisets and Formae

4.1 Preliminaries

In order to approach the problem of optimising the topology of a neural network with a genetic algorithm, it is useful first to consider set and multiset optimisation problems, which will form the solution framework. Recall that the distinction between a set and a multiset is that duplication of elements is not significant in sets, so that

$$\{\,a, a, b\,\} \equiv \{\,a, b\,\} \qquad (4)$$

whereas in multisets an element may appear more than once

$$\{\!\!\{a, a, b\}\!\!\} \neq \{\!\!\{a, b\}\!\!\}. \qquad (5)$$

(The notation $\{\!\!\{\cdots\}\!\!\}$ is used to indicate a multiset.) The difference is significant in this context.

A number of different set and multiset optimisation problems may be distinguished. In general there will be a "universal set", $\mathcal{E}$, from which elements are drawn. The aim is to construct a set or multiset consisting of elements drawn from this universal set so as to optimise some property of the resulting set or multiset. Examples could include

1. finding locations for bottle banks so as to maximise recycling in some area;
2. selecting members of a committee to make a environmental impact assessment;
3. choosing connections in a neural network to minimise its average learning time to some acceptable error;

[7] An even more pernicious form of the permutation problem is seen when graphs are being optimised for some property, for in this case there is generally a permutational redundancy of order $n!$ where n is the number of nodes in the graph, and there is not normally an analogue of the fixed "external" units.

4. choosing connections in a neural network to maximise its generalisation capability.

The first could be a set or multiset optimisation problem, according to whether multiple bottle banks were to be allowed at a single location or not. It is likely, for practical purposes that the number of bottle banks would be fixed (and that increasing this number would increase the potential for recycling) so that the size of the solution set would be known at the outset.

The second case is certainly a set rather than a multiset problem, since no human can appear more than once on a committee, and though the size may be known before-hand (perhaps because of budgetary constraints) it could be that it formed part of the optimisation. (The more people on a committee, in general, the longer decisions take to agree, and there are people whom it may be desirable not to have present or who could contribute nothing useful anyway.)

In the case of the two neural network problems, the ideal number of connections may artificially be fixed beforehand, but in general will not be known and will itself be subject to optimisation. (It should be remembered that while, in principle, the presence of a connection should never be a problem since its strength (weight) can always be set to zero, in practice a given learning scheme may well be hindered by the presence of a connection.) The trade-off is particularly acute in the case where the goal is to maximise the generalisation capability of the network. In this case, too few connections can prevent acceptable learning, while too many will tend to hinder generalisation through the phenomenon of over-learning.[8]

There may also be other complications, such as constraints on the sets, (there must be at least three bottle banks in the Prime Minister's constituency, the committee should not include arch-rivals etc.) but these will not be considered in this paper. Thus four classes of set optimisation problems will be considered—fixed-size sets, variable-size sets, fixed-size multisets and variable-size multisets.

4.2 Fixed-Size Multisets

In the case of a fixed-size multiset, the naïve approach is to use a k-ary representation, where $k = |\mathcal{E}|$, the size of the universal set, to allow each locus on a conventional linear chromosome to take any allele from $\mathcal{E}$, and to proceed as normal with a conventional genetic operators. The problems with this are obvious and profound.

1. There is a huge redundancy in the representation, i.e. the representation space $\mathcal{C}$ is much larger than the real search space $\mathcal{S}$ because of the different orders in which the members of the multiset may be written. While the number of optima is also, in general, increased dramatically (though not necessarily by the same factor[9]) navigation through this larger search space may be very difficult.
2. More specifically, *respect* of "meaningful" formae will be difficult to ensure. "Meaningful" formae must group chromosomes only according to properties of the solutions they encode: they should not distinguish between two different *representations* of a single solution. Thus, formae for sets or multisets should correspond to well-defined sets of solutions (sets or multisets) in $\mathcal{S}$, and not be defined only in the representation space $\mathcal{C}$.

   To see that conventional (one-point) crossover cannot respect any set of formae thus defined in the context of multisets, simply note that a trivial consequence of respect is that crossing a solution with itself should result in the same solution; applying conventional crossover to chromosomes ab and ba could result in aa or bb.
3. If $\mathcal{E}$ is large, this representation has a high cardinality, which is traditionally not favoured. This aspect is discussed in section 7.1.

Whitley [36] has taken essentially this approach to searching for a winning hand in a simplified form of poker, but he additionally used an operator which reversed the sequences of arbitrary portions of the chromosome.[10] This does not solve the fundamental problems (1 and 2), though it did enable him to find his chosen global optimum very easily. This was an artifact arising from the fact that the optimum happened to be five aces, a pattern which is easy to produce using this form of "inversion". This is discussed in detail in chapter 6 of Radcliffe [21].

4.3 Variable-Size Sets and Multisets

Variable-size sets and multisets can be dealt with in a more traditional fashion with fewer problems. In this case each locus on the chromosome can correspond

[8] In fact there is evidence that for learning schemes like back-propagation it may be desirable to train with a net of relatively high connectivity and then to prune nodes with highly correlated firing patterns (Seitsma & Dow [28], Burkitt [3]). This complication will not be considered in this paper.

[9] For example, optimising over a multiset of size n, if the solution be n copies of a single element, there is only one representation of the optimum, but $n!$ different representations of solutions in which every element is different.

[10] It is important to note that this is *not* inversion in the traditional sense introduced by Holland [12].

to an element of the universal set $\mathcal{E}$ from which elements are to be drawn. The gene values (alleles) can then indicate the number of copies of the element in question to be included in the multiset. In this case, binary genes correspond to set optimisation, and higher cardinality representations correspond to multisets.

This approach is both simple and traditional, and the representation scheme described contains no redundancy. It perhaps, however, requires the problem to be viewed in a slightly unconventional way. Thus in the four problems listed above, the positions on the chromosome would represent different possible locations for the bottle banks, the different possible people on the committee, and the connections in the network. In some cases this will lead to very long chromosomes, though this is not necessarily problematical.

Roughly this approach to neural network topology optimisation was taken by Miller, Todd & Hegde [15]. In this case, however, there is a further complication already mentioned, which is the equivalence of different hidden nodes under re-labelling. Thus while Miller *et al.* directly manipulated the binary connection matrices for the neural network, a connection $c_{ij} \in \{0, 1\}$ between nodes i and j contributes to redundancy in the representation if either i or j is a hidden node. This problem will be reconsidered in section 7.

4.4 Fixed-Size Sets

Fixed-size sets present more of a problem for traditional schemes. The (wholly inadequate) approach described for dealing with fixed-size multisets could be used if the recombination operator were altered to ensure that multiple copies of elements were never generated. This could be fairly easily achieved if the chromosomes were sorted, though the resulting recombination operator would have to be carefully designed to ensure that it was unbiased.

Similarly, the approach described for variable-sized sets and multisets could be adopted for fixed-size sets, but with the additional constraint that the sum of the genes should be the number of elements in the set. This could be ensured in a number of slightly unprincipled ways, including random 'helpful' mutations after crossover.

The approach described in the next section obviates the need for such manipulation, and can be extended to deal with the other classes of set and multiset optimisation problems discussed.

5 Set and Multiset Recombination

5.1 Random Respectful Recombination

Radcliffe [23] introduced a class of *random, respectful recombination* (R^3) operators, which are defined with respect to specific sets of formae and which are guaranteed both to respect and properly to assort those formae whenever these two conditions are compatible.[11] The R^3 operator simply makes a uniform, random choice over the members of the similarity set (figure 2) of the two parents undergoing recombination.

It is easy to see that R^3 fulfills these claims. It plainly *respects* the formae since respect amounts precisely to the condition that all children be members of the similarity set of their parents (figure 2). The requirement that respect and proper assortment be compatible is therefore precisely the requirement that the solutions which proper assortment requires be capable of generation by the recombination operator lie in the similarity set of the two parents. Thus *any* operator which chooses every element from the similarity set with some non-zero probability, and never generates any other, must respect and properly assort the formae, provided that this is possible.[12]

In the light of this, the essential requirement is to construct suitable formae for set and multiset optimisation problems. The four classes of set and multiset problems identified in section 4.1 will now be discussed in turn, suitable formae for them will be suggested and the R^3 operator and others will be constructed.

5.2 Fixed-Size Sets (Attempt I)

The most obvious definition of formae for set problems (whether fixed-size or otherwise) is that they specify elements which the solution must contain. This is very convenient, but requires rather careful notation to avoid subtleties.

Let the universal set (from which all elements are drawn) be $\mathcal{E}$. Then, assuming that the set size is fixed to be N, the search space $\mathcal{S}$ is a subset of the power set[13] $\mathbb{P}(\mathcal{E})$. Specifically,

$$\mathcal{S} = \left\{ \eta \subset \mathcal{E} \;\middle|\; |\eta| = N \right\}. \tag{6}$$

[11] A set of formae which can be simultaneously respected and properly assorted is said to be *separable,* and a recombination operator which achieves this is said to *separate* the formae.

[12] This is not the same as saying that every operator which separates a set of formae *must* be capable of generating every solution in the similarity set of the two parents; the condition is sufficient but not necessary.

[13] The power set of a set A is the set of all subsets of A.

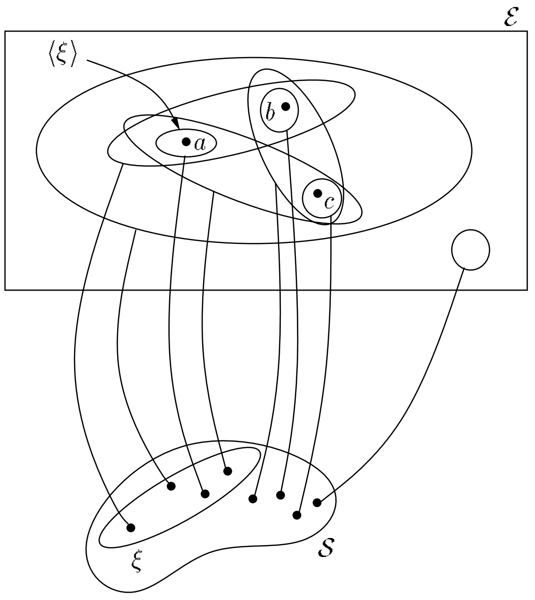

Figure 6: The universal set $\mathcal{E}$ (top) contains, in this example, three elements. The power set $\mathbb{P}(\mathcal{E})$ of $\mathcal{E}$ is the set of all subsets of $\mathcal{E}$: each subset is also drawn in the top figure. The search space $\mathcal{S}$ is identical to the power set, and is illustrated in the lower figure. The forma ξ has the description set $\langle\xi\rangle = \{a\}$, as shown. This forma consists of all the sets in $\mathcal{S}$ which contain $\langle\xi\rangle$, as is shown in the bottom figure.

A forma ξ is then a subset of $\mathcal{S}$, and the set Ξ of all formae is subject to

$$\Xi \subset \mathbb{P}(\mathcal{S}). \tag{7}$$

More precisely, Ξ will be defined by

$$\Xi = \left\{ \xi \subset \mathcal{S} \;\middle|\; \exists \mathcal{E}_\xi \subset \mathcal{E} : (\eta \in \xi \iff \mathcal{E}_\xi \subset \eta) \right\}. \tag{8}$$

This says that every forma ξ has an associated *description set* $\mathcal{E}_\xi$. This description set consists of the members of $\mathcal{E}$ which a solution must contain in order to be an instance of the forma ξ. It will be convenient to use the notation $\langle\xi\rangle$ for the description set $\mathcal{E}_\xi$. Figure 6 illustrates the general idea.

An example of a low precision[14] forma ξ with the description set $\langle\xi\rangle = \{a\}$ is

$$\xi = \left\{ \eta \in \mathcal{S} \;\middle|\; \eta \supset \{a\} \right\}. \tag{9}$$

Notice that the description set of the intersection of two formae is the union of their description sets,

$$\langle \xi \cap \xi' \rangle \equiv \langle\xi\rangle \cup \langle\xi'\rangle. \tag{10}$$

Recall that the similarity set of two solution sets is the intersection of all formae which contain them. It should be clear that this, the smallest forma containing solution sets η and ζ, is the forma whose description set is their intersection:

$$\langle \eta \oplus \zeta \rangle \equiv \eta \cap \zeta. \tag{11}$$

Recall also that, quite generally, the R^3 operator produces a child by randomly selecting a member from the similarity set of the two parents. Thus, in this case, given two parents η and ζ, R^3 chooses a random set which contains their intersection. For example, if

$$\mathcal{E} = \{\, a, b, c, d, e, f \,\} \tag{12}$$

and $N = 3$, the similarity set of $\{a, b, c\}$ and $\{a, d, e\}$ is described by

$$\langle \{a, b, c\} \oplus \{a, d, e\} \rangle = \{a, b, c\} \cap \{a, d, e\} = \{a\} \tag{13}$$

so that

$$\begin{aligned} \{a, b, c\} \oplus \{a, d, e\} = \; & \big\{ \{a, b, c\}, \{a, b, d\}, \{a, b, e\}, \\ & \{a, b, f\}, \{a, c, d\}, \{a, c, e\}, \\ & \{a, c, f\}, \{a, d, e\}, \{a, d, f\}, \\ & \{a, e, f\} \big\}. \end{aligned} \tag{14}$$

Thus R^3 applied to $\{a, b, c\}$ and $\{a, d, e\}$ picks one of these ten sets, each with probability one tenth.

This operator may seem a little odd, in that it can produce a solution set containing an element which neither of the parent sets contains: this is addressed in section 5.4. Of more immediate concern is the observation that R^3 fails properly to assort the formae as specified. To see this, simply observe that $\{a, b, c\}$ is a member of the forma described by $\langle\xi\rangle = \{b, c\}$, and $\{a, d, e\}$ is a member of the forma described by $\langle\xi'\rangle = \{d\}$, but that R^3 cannot produce a member of the intersection of these formae $\xi \cap \xi'$, because $\langle \xi \cap \xi' \rangle = \{b, c, d\}$, and R^3 will always pick a member of the similarity set given in equation 14. This is not a failing of the *operator*, but rather reflects the fact that the formae Ξ are not separable, i.e. they *cannot* simultaneously be respected and properly assorted. The general problem of non-separable formae is discussed in section 6.

5.3 Variable-Size Sets

The non-separability of the formae encountered in the consideration of fixed-size sets in the previous section

[14] The precision of a forma is similar to the order of a schema, and is defined in Radcliffe [22, 23]. Essentially, high precision formae are small (contain few members) and low precision formae are large (contain many members). The precision of a schema is 2^o, where o is the order of the schema.

can be seen to be the direct result of the restriction to fixed size. All of the definitions of the previous section carry over to the case of variable-size sets with the exception of the definition of the search space (equation 6), which is replaced by

$$\mathcal{S} = \mathbb{P}(\mathcal{E}). \tag{15}$$

This changes the similarity sets and consequently the random, respectful recombination operator R^3. Considering the "same" example as before, the description set for $\{a, b, c\} \oplus \{a, d, e\}$ is unchanged (equation 13),

$$\langle \{a, b, c\} \oplus \{a, d, e\} \rangle = \{a, b, c\} \cap \{a, d, e\} = \{a\}.$$

The similarity set itself is, however, quite different,

$$\{a, b, c\} \oplus \{a, d, e\} = \{ \eta \subset \mathcal{E} \mid a \in \eta \}, \tag{16}$$

where there is now no restriction on the size of η. To verify that proper assortment is satisfied if a random member of this similarity set is selected, simply note that the union of the two parents is always a member of the similarity set and that the intersection of any pair of formae containing the children contains this union.

Thus variable-size sets are simpler than their fixed-size counterparts and have separable formae, ensuring that the R^3 operator both respects and properly assorts them. In common with the fixed-size case, however, children may be produced which contain elements which belong to neither of the parents. This is addressed next.

5.4 Gene Transmission and Basic Formae

In response to situations like the ones above, in which R^3 succeeds in respecting formae, and (if they are separable) in properly assorting them, but generates solutions which bear rather less relation to their parents than might be deemed desirable, the concept of a *complete orthogonal basis* and a formal concept of *gene* were introduced in Radcliffe [23]. These ideas will now be re-examined in the context of the examples above.

In their original conception, formae were introduced as equivalence classes induced by arbitrary *equivalence relations* over the search space[15] $\mathcal{S}$ (Radcliffe [22, 23]). Thus, the idea was to try to choose equivalence relations which would group solutions into equivalence classes which might reasonably be expected to contain solutions with correlated (similar) performance. In this way, the formae arose as secondary objects, induced by equivalence relations. It is important to note that more than one equivalence relation is used at a time in this analysis, which is slightly unusual and constitutes a potential source of confusion.

It should be clear that regarding formae as equivalence classes of equivalence relations is not a restriction on their generality, since a forma representing an arbitrary subset ξ of the search space $\mathcal{S}$ can be induced by constructing an equivalence relation $\sim$ according to the rule

$$\eta \sim \zeta \iff (\eta, \zeta \in \xi \text{ or } \eta, \zeta \notin \xi). \tag{17}$$

Traditional schemata over strings of length n can be viewed as equivalence classes of equivalence relations described by members of

$$\Psi = \{ \square, \blacksquare \}^n. \tag{18}$$

Here $\square$ is the traditional "don't care" character introduced by Holland [12], and $\blacksquare$ is a "care" character.[16] An equivalence relation from this set then relates those chromosomes which agree (have a common allele value) at every position in which the description of the equivalence relation has the "care" character $\blacksquare$; positions in which the equivalence relation has $\square$ are not considered. Thus the equivalence relation $\square\blacksquare\square\blacksquare$, defined for a binary representation, has four equivalence classes (formae), $\square 0 \square 0$, $\square 0 \square 1$, $\square 1 \square 0$ and $\square 1 \square 1$, which are, of course, ordinary schemata.

Notice that the similarity set of two chromosomes defined with respect to schemata is the schema which has the "don't care" symbol $\square$ at every position at which the two chromosomes disagree, and their common value at each remaining locus, so that, for example,

$$1010 \oplus 1001 = 10\square\square. \tag{19}$$

It can be seen, therefore, that the R^3 operator defined with respect to schemata, in the case of binary chromosomes, reduces to precisely the familiar *uniform crossover* operator[17] (e.g. Syswerda [30]). In the case of k-ary genes with $k > 2$, however, R^3 makes a random choice over the *whole* allele set for genes at which the two parents disagree: this may not be desirable. The introduction of the notion of a *complete orthogonal basis* for a set of equivalence relations allows this possibility to be disbarred.

The concept of a basis is rather simple, and is motivated by familiar definitions from the algebra of linear spaces. The idea is that the equivalence relations with

[15] In practice formae tend to be defined over the representation space $\mathcal{C}$ of chromosomes: the distinction, though important, is not of great relevance to this discussion.

[16] Similar notation can be used to describe Walsh partitions in the analysis of deception. See Goldberg [6].

[17] Strictly, uniform crossover is parameterised by the probability of drawing each gene from the "first" parent: in this paper this probability is always assumed to be 0.5 unless an explicit statement is made.

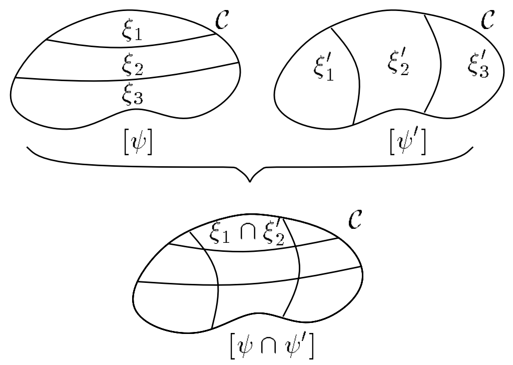

Figure 7: The set of formae induced by the equivalence relations ψ, ψ' and $\psi \cap \psi'$. The formae induced by $\psi \cap \psi'$ are intersections of those induced by ψ and ψ'.

only a single defining position (i.e. only one ■ symbol) can be used to generate all the higher-order equivalence relations. The definitions that follow will allow the set

$$E = \{ \blacksquare\square\square\square,\ \square\blacksquare\square\square,\ \square\square\blacksquare\square,\ \square\square\square\blacksquare \} \qquad (20)$$

to be interpreted as a basis for the equivalence relations which induce schemata (equation 18) with $n = 4$.

Intersection of equivalence relations will be defined in a natural way (see figure 7) so that, for example:

$$\blacksquare\square\square\square \cap \square\blacksquare\square\square = \blacksquare\blacksquare\square\square. \qquad (21)$$

The basic equivalence relations in E can then be identified as *genes*, and the basic formae (equivalence classes) as *alleles*. Having made these identifications, it is possible to insist that in general, as in the familiar case, each of a child's genes be inherited from one or other parent. That is, the child should be in the same basic forma as one of its parents for each of the basic equivalence relations in E. This principle is called *strict transmission of genes* (Radcliffe [23]). If R^3 is modified to obey this principle, yielding the *inheritance crossover* operator, then uniform crossover is recovered for k-ary string representations with $k > 2$. It is important to appreciate that the purpose of this rather laborious construction of a simple operator is that the construction is valid for *any* set of formae induced by a set of equivalence relations for which a complete orthogonal basis can be found.

A more rigorous formulation of complete orthogonal basis than the foregoing is now presented, based closely on Radcliffe [23].

First, intersection is defined for equivalence relations. For these purposes an equivalence relation $\sim$ is conveniently described by a binary function

$$\psi : \mathcal{C} \times \mathcal{C} \longrightarrow \{ 0, 1 \} \qquad (22)$$

which returns 1 if its arguments are equivalent and 0 if they are not:

$$\psi(\eta, \zeta) = \begin{cases} 1, & \text{if } \eta \sim \zeta, \\ 0, & \text{otherwise.} \end{cases} \qquad (23)$$

The intersection of two equivalence relations $\psi, \psi' \in \Psi$ can then be defined by

$$(\psi \cap \psi')(\eta, \zeta) = \begin{cases} 1, & \text{if } \psi(\eta,\zeta) = \psi'(\eta,\zeta) = 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (24)$$

Given this, a subset $E \subset \Psi$ will be said to constitute a *complete orthogonal basis* for Ψ provided that

- (Completeness)
All relations $\psi \in \Psi$ can be constructed as the intersection of some subset of the basic relations:
$$\forall \psi \in \Psi\ \exists E_\psi \subset E : \bigcap E_\psi = \psi. \qquad (25)$$
- (Orthogonality)
Every forma ξ induced by every basic relation $\psi \in E$ is compatible with every forma ξ' induced by every other basic relation $\psi' \in E$:
$$\begin{aligned} &\forall \psi, \psi' \in E\ (\psi \neq \psi') \\ &\qquad \forall \xi \in [\psi]\ \forall \xi' \in [\psi'] : \xi \cap \xi' \neq \emptyset\,, \end{aligned} \qquad (26)$$
where $[\psi]$ is the set of equivalence classes (formae) induced by ψ.

The relationship between these definitions and their counterparts in linear algebra should be clear. The notion of completeness is essentially identical, and expresses the fact that the basic equivalence relations *span* the space of equivalence relations under consideration, while orthogonality ensures that alleles (membership of basic formae) can be freely mixed. It will become apparent in later sections that the definition of orthogonality can be relaxed to some degree; this will be necessary in order for a suitable basis to be found for some classes of multiset formae.

5.5 Fixed-Size Sets (Attempt II)

Having defined genes in terms of a complete orthogonal basis for some equivalence relations the task is now to find equivalence relations which induce the set formae described in section 5.2 and to find a complete orthogonal basis for them.

Recall that these formae were characterised by a set of elements which a solution must contain in order to be an instance of the forma in question. Thus, a simple forma is described by

$$\langle \xi \rangle = \{a\}. \tag{27}$$

Clearly various equivalence relations could be constructed which have ξ as one of their equivalence classes. One such can be generated simply by using the trivial rule expressed by equation 17 as follows:

$$\psi_{\{a\}}(\eta, \zeta) = \begin{cases} 1, & \text{if } (a \in \eta \cap \zeta \text{ or } a \notin \eta \cup \zeta), \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

This equivalence relation induces two equivalence classes, one comprising the solutions containing the element a and another comprising those which do not. Thus, a second equivalence class, which had not originally been specified, has also been induced by ξ.

There is clearly an equivalence relation $\psi_{\{x\}}$ of the form described by equation 28 for *each* $x \in \mathcal{E}$. Moreover, these are intuitively natural candidates for a basis for a set Ψ of equivalence relations which might generate all the formae of the type described. As will now be demonstrated, if the rule for intersection of equivalence relations described by equation 24 is followed, the set

$$E = \left\{ \psi_{\{x\}} \;\middle|\; x \in \mathcal{E} \right\} \tag{29}$$

does indeed form a complete orthogonal basis for a set of equivalence relations Ψ which induce all the formae in Ξ as defined in equation 8, together with others.

To see this, consider the intersection of $\psi_{\{a\}}$ and $\psi_{\{b\}}$, which will be denoted $\psi_{\{a,b\}}$. According to the definition of intersection for equivalence relations (equation 24)

$$(\psi_{\{a\}} \cap \psi_{\{b\}})(\eta, \zeta) = \begin{cases} 1, & \text{if } \psi_{\{a\}}(\eta, \zeta) \\ & \quad = \psi_{\{b\}}(\eta, \zeta) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

This equivalence relation induces four equivalence classes, which might conveniently be written

$$\begin{aligned} \xi_{ab} &= \{\eta \in \mathcal{C} \mid a \in \eta, b \in \eta\}, \\ \xi_{a\bar{b}} &= \{\eta \in \mathcal{C} \mid a \in \eta, b \notin \eta\}, \\ \xi_{\bar{a}b} &= \{\eta \in \mathcal{C} \mid a \notin \eta, b \in \eta\}, \\ \xi_{\bar{a}\bar{b}} &= \{\eta \in \mathcal{C} \mid a \notin \eta, b \notin \eta\}. \end{aligned} \tag{31}$$

The generalisation of this is rather obvious. A general equivalence relation, $\psi \in \Psi$, has a description set, conveniently written $\langle \psi \rangle$, which is a subset of the universal set $\mathcal{E}$. Members of the search space (themselves subsets of $\mathcal{E}$) are then equivalent under ψ precisely if they contain the same subset of the members of the description set $\langle \psi \rangle$. Formally,

$$\psi(\eta, \zeta) = \begin{cases} 1, & \text{if } \langle \psi \rangle \cap \eta = \langle \psi \rangle \cap \zeta, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

It is clear that E (defined in equation 29) does indeed form a basis for the equivalence relations.[18] A forma ξ induced by an equivalence relation $\psi \in \Psi$ is then characterised by a partition of the description set $\langle \psi \rangle$. It then becomes convenient to describe a forma by a 2-tuple

$$\langle \xi \rangle = (\xi^+, \xi^-) \tag{33}$$

where

$$\xi^+ \cap \xi^- = \emptyset \tag{34}$$

and

$$\xi^+ \cup \xi^- = \langle \psi \rangle \tag{35}$$

with the interpretation

$$\eta \in \xi \iff \left(\eta \cap \xi^+ = \xi^+ \text{ and } \eta \cap \xi^- = \emptyset \right). \tag{36}$$

Having made these identifications, it is possible to define the similarity set of two chromosomes with respect to the formae Ξ induced by Ψ. This will allow the random respectful recombination operator R^3 to be constructed. Using the notation for the description sets of formae just introduced, this gives

$$\langle \eta \oplus \zeta \rangle = \big(\eta \cap \zeta, \mathcal{E} - (\eta \cup \zeta) \big), \tag{37}$$

where the minus sign denotes set subtraction. The R^3 operator makes a random (uniform) selection from this similarity set. Returning to the example used earlier, (equation 12, with $N = 3$),

$$\begin{aligned} &\langle \{a, b, c\} \oplus \{a, d, e\} \rangle \\ &\quad = \Big(\{a, b, c\} \cap \{a, d, e\}, \mathcal{E} - \big(\{a, b, c\} \cup \{a, d, e\} \big) \Big) \\ &\quad = \Big(\{a\}, \{f\} \Big). \end{aligned} \tag{38}$$

This describes the forma containing those sets which contain a and exclude f. With $N = 3$, this gives

$$\begin{aligned} &\{a, b, c\} \oplus \{a, d, e\} \\ &\quad = \big\{ \{a, b, c\}, \{a, b, d\}, \{a, b, e, \}, \\ &\qquad \{a, c, d\}, \{a, c, e\}, \{a, d, e, \} \big\}. \end{aligned} \tag{39}$$

Thus, R^3 for these formae can be understood as an operator which

1. copies all the elements which are common to the two parents into the child;
2. fills the remaining places in the child with a random selection of the unused elements from the two parents.

[18] Technically, there is a problem given the definition of orthogonality, when equivalence relations with description sets of size greater than or equal to the fixed size of solution sets are considered, but this is a very minor consideration.

So a child θ of η and ζ has the natural properties

$$\eta \cap \zeta \subset \theta \subset \eta \cup \zeta. \tag{40}$$

It is clear, therefore, that in this case R^3 strictly transmits genes where a gene corresponds to an element of $\mathcal{E}$ and an allele to the presence or absence of that element (equation 28). Notice, however, that the counterexample used at the end of section 5.2 remains valid, so that that the formae are not separable, with the consequence that R^3 cannot properly assort them. This again arises directly from the restriction to fixed-size sets.

An alternative way of viewing this operator is to imagine a conventional linear chromosome in which every position represents an element from the universal set, and to imagine an operator like uniform crossover, but constrained so that the total number of 1's on the child is constant and equal to N, the fixed size of the set.

5.6 Fixed-Size Multisets

The extension of the previous case from sets to multisets is in essence simple, but involves one complication. The basic idea will be that rather than specify whether or not an element is a member of the multiset under consideration, a forma will specify the *multiplicities* of some elements. Formally, let $\mathbb{P}_m(\mathcal{E})$ be the *multipower set* of $\mathcal{E}$, that is, the set of all multisets whose elements are drawn from $\mathcal{E}$. Then the *multiplicity* function

$$m : \mathcal{E} \times \mathbb{P}_m(\mathcal{E}) \longrightarrow \mathbb{Z}^+ \cup \{0\} \tag{41}$$

is defined so that $m(x, \eta)$ is the number of copies of x in the multiset η.

A forma for multisets could either specify exact multiplicities for elements or could give bounds on their multiplicities. Since the former is a special case of the latter, where the bounds are maximally tight, the more general case will be examined.

A forma is now conveniently described by a set of 3-tuples of the form $(x, N_x^\downarrow, N_x^\uparrow)$ each of which is understood to specify that the multiplicity $m(x, \eta)$ of the element x in the set η lies in the inclusive range $N_x^\downarrow$ to $N_x^\uparrow$. For example, a forma ξ with the description set

$$\langle \xi \rangle = \{(a, 0, 0), (b, 1, 3)\} \tag{42}$$

contains all those multisets over $\mathcal{E}$ of size N which contain no copies of a and contain between 1 and 3 copies of b (figure 8).

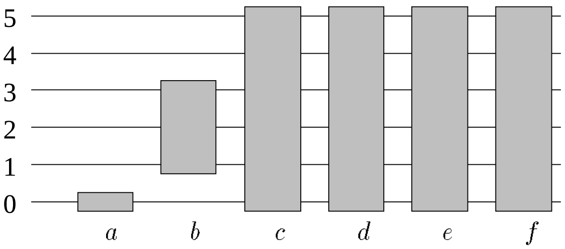

Figure 8: A visualisation of the forma ξ with description set $\langle \xi \rangle = \{(a, 0, 0), (b, 1, 3)\}$. The full ranges for elements c to f correspond to the "don't care" character familiar from conventional schemata.

As usual, there are a number of sets of equivalence relations which could be constructed to generate these formae, and again an obvious starting point is equivalence relations based on the lowest-order formae. Thus the equivalence relation ψ which induces the forma described by $\langle \xi \rangle = \{(x, N_x^\downarrow, N_x^\uparrow)\}$ would have the same description set

$$\langle \psi \rangle = \big\{(x, N_x^\downarrow, N_x^\uparrow)\big\} \tag{43}$$

and would be defined by

$$\psi(\eta, \zeta) = \begin{cases} 1, & \text{if } m(x,\eta), m(x,\zeta) < N_x^\downarrow \\ & \text{or } m(x,\eta), m(x,\zeta) \in [N_x^\downarrow, N_x^\uparrow] \\ & \text{or } m(x,\eta), m(x,\zeta) > N_x^\uparrow, \\ 0, & \text{otherwise,} \end{cases} \tag{44}$$

where

$$[N_x^\downarrow, N_x^\uparrow] = \big\{ n \in \mathbb{Z} \bigm| N_x^\downarrow \le n \le N_x^\uparrow \big\}. \tag{45}$$

As was the intention, formae can now specify a range of multiplicities for any element; a single equivalence relation can in fact be seen to suffice to define up to three ranges simultaneously. The problem of finding a basis for these equivalence relations will now be considered.

The natural candidates to form a basis are the equivalence relations which divide the range of multiplicities for a single element into a lower portion and an upper portion, as shown in figure 9,

$$E = \big\{ \psi \in \Psi \bigm| \langle \psi \rangle = \{(x, 0, N_x^\uparrow)\}, x \in \mathcal{E}, N_x^\uparrow \in [0, N^\star] \big\} \tag{46}$$

where $N^\star$ is the maximum allowed multiplicity for an element. These equivalence relations can easily be seen to be complete, for any "first order" equivalence relation[19] with a description set $\{(x, N_x^\downarrow, N_x^\uparrow)\}$

[19] Once genes have been defined, order can be defined for formae in a way similar to the definition for schemata.

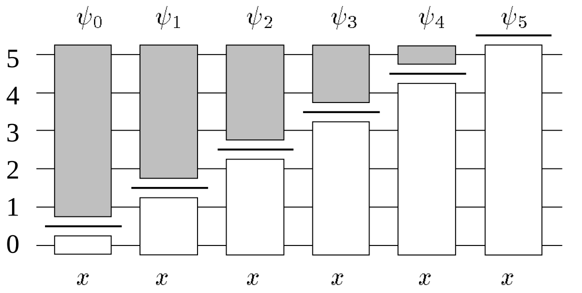

Figure 9: The set of equivalence relations with description sets of the form $\{(x, 0, N)\}$ divide the range $[0, N^\star]$ into a lower and and upper portion as shown: the equivalence relations may be thought of as simple dividing lines at integer-plus-half values.

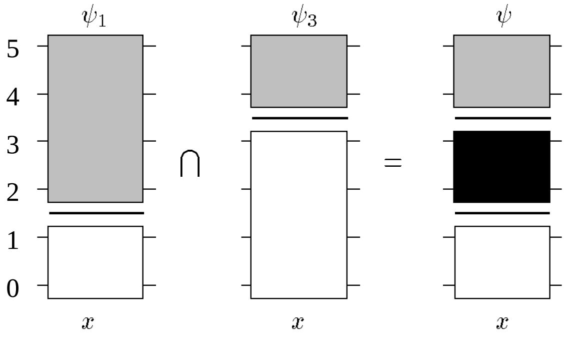

Figure 10: Any equivalence relation ψ defined on a single element can be constructed as an intersection of the elements in E as defined in equation 46. Here ψ has the description set $\langle\psi\rangle = (x, 2, 3)$ and is constructed as the intersection of ψ_1 with description set $\langle\psi_1\rangle = (x, 0, 1)$ and $\langle\psi_3\rangle = (x, 0, 3)$.

can be constructed as an intersection of the relations with description sets $\{x, 0, N_x^\uparrow\}$ and $\{(x, N_x^\downarrow, N^\star)\}$ (figure 10). Higher order equivalence relations can then be constructed trivially by intersection. It is equally easy, however, to see that the relations in E do not satisfy the condition of orthogonality specified in equation 26. To verify this, simply note that if a set is a member of the forma with description set $\{(x, 0, 2)\}$ it cannot also be a member of the forma with the description set $\{(x, 4, N^\star)\}$, as would be required if E were orthogonal (equation 26).

Rather than abandon this potential basis, it is instructive to return to the analogy with linear algebra which led to the original formulation of the conditions on a basis, namely completeness and orthogonality. In linear algebra there is a weaker notion than orthogonality known as *linear independence*: a set of vectors is said to be linearly independent if no one of them can be expressed as a linear combination of the others. Following the analogy, a set of equivalence relations will be said to be *independent* if no one of them can be constructed as an intersection of some of the others. The set E defined in equation 46 satisfies this weaker condition.[20]

The purpose of introducing the notion of a complete orthogonal basis for a set of equivalence relations was to generalise the notion of a gene and allow a principle of strict gene transmission to be extended to more general formae. It will be demonstrated below that the weaker notion of a *complete non-orthogonal basis* suffices for the definition of genes, and thus is adequate for the original purpose. The notion of independence is formalised as follows:

- (Independence)
  A set E of equivalence relations will be said to be *independent* if no one of the relations $\psi \in E$ can be expressed as the intersection of some subset of the others, i.e.

$$\nexists\, \psi \in E \;\exists\, E_\psi \subset E : \bigcap E_\psi = \psi. \qquad (47)$$

Using the same definition of genes and alleles for non-orthogonal bases as for orthogonal bases, (i.e. genes are the basic equivalence relations and alleles are the basic equivalence classes) it is now possible to construct the *inheritance crossover* operator induced by the basis E for Ψ, described by equation 46.

The inheritance crossover operation can be defined in a way similar to random respectful recombination, the difference being that instead of selecting from the similarity set of the two parent chromosomes it selects from the subset of chromosomes in the similarity set which have every gene in common with one or other parent. This subset, which for parents η and ζ is written $\eta \odot \zeta$, is called their *inheritance set,* and defined by

$$\eta \odot \zeta = \left\{ \theta \in \eta \oplus \zeta \;\middle|\; \forall \psi \in E : \theta \in [\eta]_\psi \cup [\zeta]_\psi \right\} \qquad (48)$$

where $[\eta]_\psi$ is the equivalence class induced by ψ to which η belongs. The inheritance crossover operator picks each element in the inheritance set of the parents with equal probability, and both strictly transmits genes and properly assorts formae provided that these conditions are compatible. (The proof of this is identical in form to the proof that R^3 respects and properly assorts a set of formae, section 5.1.)

[20] A rather minor point which should nevertheless be made in passing is that the formae now being considered violate the closure discussed in Radcliffe [22, 23]: this turns out to be unimportant.

The formalism developed above can now be applied to the problem of recombining fixed-size multisets. The similarity set of two chromosomes (now multisets) is the forma with the description set

$$\begin{aligned}\langle \eta \oplus \zeta \rangle = \ & \big\{ (x, N_x^{\downarrow}, N_x^{\uparrow}) \;\big|\; N_x^{\downarrow} = \min\big(m(x,\eta), m(x,\zeta)\big), \\ & \qquad N_x^{\uparrow} = \max\big(m(x,\eta), m(x,\zeta)\big) \big\}.\end{aligned} \tag{49}$$

This similarity set contains all those multisets of the given fixed size N which have at least as many copies of each element as the parent with fewer copies, and no more than the number held by the parent with more. For example, if the chosen fixed size for the multisets is five, and the universal set $\mathcal{E}$ is given by equation 12, then given

$$\eta = \{\!|a, a, a, b, c|\!\} \tag{50}$$

and

$$\zeta = \{\!|a, b, b, c, d|\!\} \tag{51}$$

the similarity set is described by

$$\begin{aligned}\langle \eta \oplus \zeta \rangle = \ & \big\{ (a,1,3), (b,1,2), (c,1,1), \\ & (d,0,1), (e,0,0), (f,0,0) \big\}.\end{aligned} \tag{52}$$

The similarity set itself contains those multisets containing $\{\!|a, b, c|\!\}$ together with exactly two elements from $\{\!|a, a, b, d|\!\}$. The inheritance set of any η and ζ is, for these equivalence relations, identical to the similarity set. To see this, consider any basic equivalence relation ψ with the description set

$$\langle \psi \rangle = \{ (x, 0, N) \}. \tag{53}$$

This has two equivalence classes, described by

$$\begin{aligned}\langle \xi_1 \rangle &= \{ (x, 0, N) \}, \\ \langle \xi_2 \rangle &= \{ (x, N+1, N^{\star}) \}.\end{aligned} \tag{54}$$

If both parents belong to the same basic forma, then their similarity set is clearly a subset of this forma. If, however, they belong to different basic formae, then since there are only two of these, the requirement that their similarity set lie in their union is no restriction at all. Thus inheritance sets for these equivalence relations are indeed identical to similarity sets[21] and so it can be seen that strict gene transmission is in this case no stronger a requirement than respect.

5.7 Variable-Size Multisets

Variable-size multisets can be dealt with simply by relaxing the constraint of fixed size as discussed in the previous section. The formae then arrived at are separable and the inheritance crossover operator (which is in this case identical to the R^3 operator) not only properly assorts and respects the formae, but also strictly transmits genes as a direct consequence of the identity of the similarity and inheritance sets.

In summary, R^3 for variable-size multisets simply inserts a number of copies of each element from the universal set which is bounded by the number of copies in the two parents, and in doing so strictly transmits genes and properly assorts the formae induced by the equivalence relations generated by the complete spanning basis of equation 46.

6 Non-Separability of Formae

Before going on to apply the results of section 5 to the problem of optimising neural network topologies, it is appropriate to consider the general problem of non-separability of formae, which arises from some of the formae, bases and recombination operators constructed thus far.

6.1 Background on Formae

There were a number of motivations for the forma analysis developed by Radcliffe [22, 21, 23], the most important of which can be summarised as follows:

1. *Nature of Representation.*
   Holland's schemata are defined for fixed-length strings for which each locus has a well-defined allele set, with the implicit assumption that all distributions of alleles over loci represent valid solutions. For many problems, including those from graph theory, set theory, constrained optimisation and neural networks, no useful coding of this form is known. In any case, it is frequently very much more convenient to store and manipulate structures in a non-string form, as is the case, for example, with Koza's evolution of Lisp programs (Koza [13, 14]). A generalisation was thus required.

2. *Genotype-Phenotype Mapping.*
   Schemata are defined in the representation space and thus group together *genotypes* (chromosomes). Where the genotype and phenotype spaces are isomorphic (that is, every chromosome corresponds to exactly one solution in the real search space $\mathcal{S}$ and conversely every solution in $\mathcal{S}$ is represented by exactly one chromosome in $\mathcal{C}$) this is probably acceptable, but where the mapping is more complex than this it may be desirable to define formae in the true search space $\mathcal{S}$.

[21] This is not, of course, true in general.

3. *Redundancy.*
   A particular example of the inadequacy of schemata arises when the coding introduces redundancy, as tends to be the case in graph and set optimisation problems (see section 2).

4. *Generality.*
   The groupings of chromosomes which can be expressed by schemata are variegated and have been shown to be sufficiently general to be useful in many problems. Nevertheless, there are many other cases where it is desirable to be able to use other partitionings of the search space. Forma analysis allows this.

5. *Intrinsic Parallelism.*
   The counting argument which is sometimes used to claim that binary representations are more powerful than those of higher cardinality applies only if attention is restricted to traditional schemata. If more general formae are considered, the argument no longer holds. The degree of intrinsic parallelism which can be inferred is defined entirely by the selection of formae (Radcliffe [22, 21]).

Having shown that the "schema theorem" applies to general formae in exactly the same way as to schemata given suitable expressions for the disruption coefficients (Radcliffe [22],Vose [32]) the question became how to manipulate formae sensibly. Guidance was taken from studies of the traditional crossover operators for conventional linear chromosomes.

The three characteristics of recombination operators which were suggested to be desirable are as follows:

- *Respect.*
  The formulation of the principle of respect was motivated by a desire to ensure that in cases where the parents share some attribute, children are guaranteed to inherit that attribute also. (This is qualified only by mutation, which is traditionally understood to serve the important but secondary rôle of ensuring that the entire search space remains accessible (Holland [12]), though see also Schaffer and Eschelman [27] and references therein.) Thus respect requires that whenever two parents are both a member of some forma, all their offspring be members of that forma also. This principle has been independently formulated by Vose [32, 33].

- *Gene Transmission*
  Respect alone is not enough to ensure that every gene possessed by a child is taken from one or other parent. The introduction of the notion of a *complete orthogonal basis* for a set of equivalence relations which induce the chosen formae allowed a general notion of gene to be formalised (Radcliffe [23]) and thus allowed a principle of (strict) gene transmission to be formulated. The introduction of non-orthogonal bases in this paper allows further application of the principle. It should be noted that gene transmission implies respect.

- *Assortment*
  The notion of assortment can be viewed as an extension and formalisation of the "building-block hypothesis" (Goldberg [5]) which expresses some of the most fundamental beliefs about the way in which genetic search proceeds. The key idea is that by recombining two solutions it is sometimes possible to piece together a solution which combines properties of the two parents. A recombination operator is said *properly to assort* a set of formae if it is the case that whenever one parent η is a member of one forma ξ, and another parent η' is a member of another forma ξ', then provided that the intersection of the two formae is non-empty (that is, provided that some chromosome exists which is a member of both formae) it is *possible* that the recombination will produce a child θ which is a member of both formae, i.e. $\theta \in \xi \cap \xi'$.

  It is interesting to note that while traditional one-point crossover (or indeed, n-point) does not properly assort *schemata* in the absence of an inversion operator, when inversion is present, it does.[22] Uniform crossover, on the other hand, does properly assort schemata. Moreover, while one-point crossover does not *properly* assort schemata in the absence of inversion, it does *weakly* assort them in the sense that given a finite number of generations and applications, it does have that ability to assemble a chromosome in the intersection of any two compatible schemata, given suitable parents.

The problem faced in the case of fixed-size sets and multisets with the formae discussed in section 5 is that the requirements of respect and assortment are incompatible, so that the formae are said to be *non-separable.* This is illustrated in figure 11.

6.2 Examples of Non-Separability

Examples of non-separable formae have already been seen in sections 5.2, 5.5 and 5.6, and the previous difficulties in even respecting reasonable formae for prob-

[22] This assumes that linkage is taken to be unspecified.

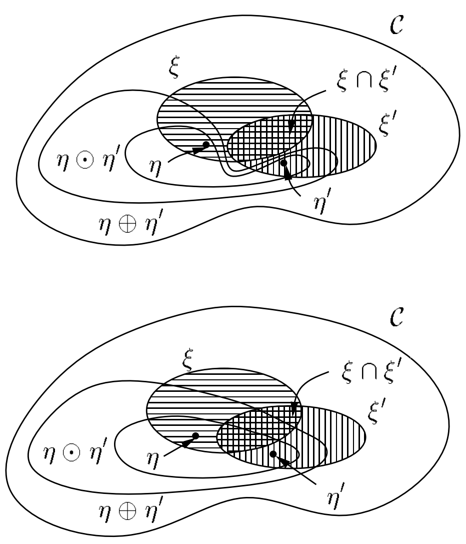

Figure 11: The top figure shows non-separable formae. Notice that the intersection of ξ and ξ', while non-empty, does not lie within the similarity set of chromosomes $\eta \in \xi$ and $\eta' \in \xi'$. The bottom figure shows separable formae.

lems in neural networks has also been mentioned in section 2. The travelling sales-rep problem (TSP) provides further interesting examples of non-separability. Whitley [40] and Radcliffe [23] have both argued that in tackling the TSP it is essential to focus attention on *edges* rather than nodes.

There are two obvious sets of formae which might be constructed for the TSP which are based on edges, the difference being whether the edges are considered to be directed or undirected. In either case, a forma is characterised by a set of edges which a tour must contain in order to be an instance of the forma. Radcliffe [23] has show that non-directed edge formae are not separable using the example in figure 12. The example in figure 13, which is due to Vose [31], suffices to show that directed edge formae are also non-separable. It should be noted, however, that this second example relies on the introduction of a cycle, which is only permitted if the cycle forms the entire tour. In this sense, the problem with directed edges is perhaps less severe than with non-directed edges.

Whitley has constructed a *genetic edge recombination operator* with the specific aim of ensuring high transmission of edges from parents to children.[23] The first version of this operator (Whitley *et al.* [40]) did not en-

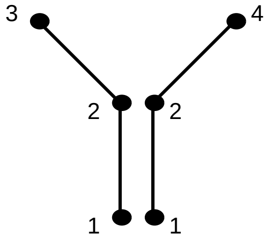

Figure 12: The tour fragment on the left is a member of the forma described by $\{\overleftrightarrow{23}\}$, which contains all tours which include the 2–3 or the 3–2 edge. Similarly, the tour fragment on the right is a member of the forma described by $\{\overleftrightarrow{24}\}$. Since both tours are also a member of the forma described by $\{\overleftrightarrow{12}\}$, however, if these formae are to be respected all of their children must contain the 1–2 edge, thus preventing a child being produced which is a member of the forma described by $\{\overleftrightarrow{23}, \overleftrightarrow{24}\}$, the intersection of those described by $\{\overleftrightarrow{23}\}$ and $\{\overleftrightarrow{24}\}$. Thus these formae are non-separable.

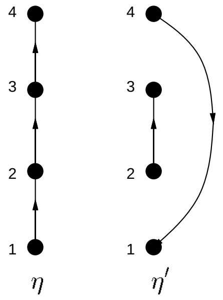

Figure 13: η is a member of the *directed edge forma* described by $\langle\xi\rangle = \{\overrightarrow{12}, \overrightarrow{34}\}$, containing those tours which include the directed edges $\overrightarrow{12}$ and $\overrightarrow{34}$. Similarly, η' is a member of the forma described by $\langle\xi'\rangle = \{\overrightarrow{41}\}$. Since, however, both η and η' are members of the directed edge forma described by $\{\overrightarrow{23}\}$, respect requires that all their children contain the $\overrightarrow{23}$ edge, thus preventing the construction of a member of $\xi \cap \xi'$, which has the description set $\{\overrightarrow{12}, \overrightarrow{34}, \overrightarrow{41}\}$.

[23] It should be noted that transmission of edges is subtly different from transmitting edge-formae (see section 6.4).

sure respect since it allowed children to be constructed which did not possess an edge common to both parents. It also failed to ensure proper assortment, by concentrating exclusively on high transmission rates for edges. A second version of the operator was constructed specifically to ensure respect[24] (Whitley *et al.* [41]) and in so doing guaranteed that proper assortment was violated.

The important point to note here is that if it is accepted that edges (whether directed or otherwise) hold the key to the TSP, the problem of non-separability arises immediately.

6.3 Exploitation and Exploration

In cases such as fixed-size set recombination and the travelling sales-rep problem it is clear that some accommodation between respect and assortment will be required if the formae suggested are to be manipulated effectively. The conflict can be viewed as an unusually sharp form of the familiar trade-off between *exploitation* of information already gathered (encapsulated by respect and gene transmission) and adequate *exploration* of the search space (encapsulated by assortment), as discussed by Holland [12].

The counter-examples used to show that the fixed-size set- and multiset formae are not separable and to show the same for both directed and non-directed edge formae for the TSP share rather similar characteristics, so that focusing on a single example will have relevance to them all. For simplicity, the example for fixed-size set formae used at the end of section 5.2 will be revisited. Recall that the problem is that the sets

$$\eta = \{\, a, b, c \,\} \tag{55}$$

and

$$\eta' = \{\, a, d, e \,\} \tag{56}$$

of size three are respective members of the formae ξ and ξ' described by

$$\langle \xi \rangle = \{\, b, c \,\} \tag{57}$$

and

$$\langle \xi' \rangle = \{\, d \,\}, \tag{58}$$

but that the sole member of the intersection $\xi \cap \xi'$ is

$$\theta = \{\, b, c, d \,\}, \tag{59}$$

which does not lie in the similarity set $\eta \oplus \zeta$ described by

$$\langle \eta \oplus \zeta \rangle = \{ a \}. \tag{60}$$

The effect of giving primacy to respect (as do all R^3 operators) is to make impossible the construction of the solution θ from these parents. This extremely worrying, not least because experience shows that premature convergence is a common problem with genetic algorithms. Thus if, in the present example, the presence of a in a solution is generally beneficial in the early stages of genetic search, but θ, which does *not* contain a, is the optimum, it is quite possible that a will become represented in every chromosome in the population early on so that the presence of solutions such as η and ζ in a later population would not allow the optimum to be constructed even though all of the necessary “building blocks” would appear to be present. Genetic search would be entirely dependent in this circumstance on a mutation which eradicated the element a from a solution, a situation which though not irretrievable seems better avoided. This prospect, which could equally easily manifest itself in more realistic, larger-scale examples, is sufficiently worrying to suggest that assortment should be given precedence over respect, the lack of which would seem likely to do no more than delay progress towards an optimum, rather than imposing mutation-dependent barriers. Similar comments apply equally to the edge formae discussed above.

6.4 Assortment

While the preceding discussion has suggested that assortment should take priority over respect when there is a conflict, this does not mean that respect should be altogether discarded in such situations. A parameterised operator will now be introduced which allows the priority given to the two considerations to be varied, so that with the parameter set at one end of the scale respect is complete (and proper assortment is violated) and as the parameter is adjusted ever-less regard is paid to the requirements of respect.

Consider the R^3 operator for fixed-size sets. This has been described as having two stages: the first constructs a partial child which contains only the intersection of the two parents; the remaining spaces in the child are chosen at random from the remaining elements which the parents contain. An alternative approach, which would ensure proper assortment but drastically violate respect, would involve discarding the first stage and simply picking elements from the union of the two parents at random. This approach would attach no weight at all to the fact that some elements were present in both parents and in this sense would have entirely disregarded respect.

These two extremes can be interpolated between by attaching a *weight* to elements of the union, with elements of the intersection being accorded a higher

[24] though it was not discussed in these terms

element	a	b	c	d	e	f
weight	2	1	1	1	1	0
probability	1/3	1/6	1/6	1/6	1/6	0

Table 1: Weights for a parameterised assorting recombination operator of weight 2, given parent sets $\eta = \{a, b, c\}$ and $\zeta = \{a, d, e\}$. The last line shows the probability that each element will be included in the child.

weight than those present in only one parent. The probability of picking an element could then be made proportional to its weight. It would seem reasonable to set the weight of elements in the intersection to at least twice that of other elements since these elements were present twice, once in each parent. In the case of the example used above (equations 55–60) this would lead to the weights shown in table 1. Clearly higher weights than two could be used to ensure a greater degree of respect, but the higher the weight is made, less assortment will be performed. The generalisation to multisets is given in section 9.1.3.

Constructing similarly parameterised operators for the TSP which could be computed efficiently is more difficult, though a paper specification for them is possible.

7 Neural Network Topologies

The professed purpose of developing the machinery of set and multiset formae was to aid the application of genetic search to problems in neural networks, though the set and multiset optimisation problems are of independent interest. The formulation of network topology optimisation given in section 2 was chosen to bring out the multiset-like nature of the problem.

Recall that if attention is restricted to feed-forward networks with a single layer of hidden units then a network topology can be described as a multiset of hidden units each of which is specified by its set of external connections (figure 4). This is an entirely non-redundant representation to which the results of the forma analysis of earlier sections can immediately be applied. In principle, this should allow genetic search to proceed efficiently.

There is, however, an obvious complication. Suppose that a modest network with ten input nodes and ten output nodes is to be considered, and that up to ten hidden units will be employed (the example used in section 2). In this case the size of the search space is about 4×10^{53} (equation 2). The problem is not the size of this search space (which is fairly modest by the standards of genetic algorithms) but the fact that each chromosome will contain at most ten hidden units, while the number of hidden unit types is $2^{20} \approx 10^6$. Thus, even if a population of 10,000 were employed, less than one percent of the available node types could be included in the population. If the node types are considered as atomic and not available for recombination it will be extremely difficult for the genetic algorithm to make progress.

Of course, this situation is not unfamiliar in genetic search, for in the classic case of parameter optimisation exactly the same predicament arises. If, for example, ten parameters are to be optimised, each coded using twenty bits, the similarity should immediately be clear (though in this case the search space would be larger, because each of the parameters would normally have a meaning, so that this is not a set optimisation problem). The solution usually employed is to allow recombination to take place at the sub-parameter level either through employing binary or other low-cardinality encodings or by using recombination operators which make use of knowledge of the high-level meaning of parameters. (For examples of the latter see Davis [4] and the discussions of range formae in Radcliffe [22, 23]; for a sceptical view see Goldberg [7]).

By exploiting this analogy it will be possible both to construct a sensible genetic approach to network topology optimisation and to shed a little more light on traditional parameter optimisation.

7.1 High Cardinality Representations and Gene Recombination

Consider the classic parameter optimisation problem where the search space is comprises vectors

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) \tag{61}$$

so that

$$\mathcal{S} = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n \tag{62}$$

with

$$I_i = [a_i, b_i] \tag{63}$$

and the intervals I_i are understood to be discrete approximations to their continuous counterparts. Using k bits per parameter, the typical chromosomal representation would employ

$$\mathcal{C} = \mathbb{B}^{kn} \tag{64}$$

where

$$\mathbb{B} = \{0, 1\}. \tag{65}$$

A chromosome $\eta \in \mathcal{C}$ would be given by

$$\begin{aligned} \eta = \quad (&\eta_{11}, & \eta_{12}, & \quad \ldots, & \eta_{1k}, \\ &\eta_{21}, & \eta_{22}, & \quad \ldots, & \eta_{2k}, \\ &\vdots & \vdots & \quad \ddots & \vdots \\ &\eta_{n1}, & \eta_{n2}, & \quad \ldots, & \eta_{nk}), \end{aligned} \tag{66}$$

where v_i is represented by $(\eta_{i1}, \eta_{i2}, \ldots, \eta_{ik})$. If schemata are defined in the space $\mathcal{C}$ of chromosomes then uniform crossover respects and properly assorts the schemata

$$\Xi_{\mathcal{C}} = \{\, 0, 1, \Box \,\}^{nk} \tag{67}$$

and strictly transmits genes. On the other hand, if schemata are defined in parameter space $\mathcal{S}$, as defined by equation 62, so that

$$\Xi_{\mathcal{S}} = \left(\prod_{i=1}^{n} \Big(I_i \cup \{\Box\} \Big) \right) \tag{68}$$

then uniform crossover in chromosome space $\mathcal{C}$ neither respects the schemata in $\Xi_{\mathcal{S}}$ nor transmits parameters, though it does assort them.

Thus transmission of genes and respect can in these cases be seen to be sub-parameter level concepts. A number of observations follow:

1. Using one-point crossover in the absence of inversion, most parameters will be transmitted whole: a maximum of one will be crossed.
2. Using one-point crossover in the presence of inversion, any number of parameters may be crossed.
3. Similar comments apply when using two-point crossover (except that up to two parameters may then be crossed), and this is true even using Booker's reduced surrogate form (Booker [2]).
4. With uniform crossover, any number of parameters may be recombined.
5. Although the parameter-space schemata $\Xi_{\mathcal{S}}$ *are* properly assorted by uniform crossover applied in the space of binary chromosomes $\mathcal{C}$, the probability of generating a child in the intersection $\xi \cap \xi'$ given

   $$\eta \in \xi \in \Xi_{\mathcal{S}} \tag{69}$$

   and

   $$\eta' \in \xi' \in \Xi_{\mathcal{S}} \tag{70}$$

   is much lower than if uniform crossover is applied in the parameter space.[25] For example, in $\mathcal{S}$

   $$(7, 8) \oplus (5, 3) = \{(7, 3), (7, 8), (5, 8), (5, 3)\}, \tag{71}$$

   having four elements, whereas in $\mathcal{C}$

   $$(0111, 1000) \oplus (0101, 0011) = (01\Box 1, \Box 0 \Box \Box), \tag{72}$$

   where the right hand side is understood to be a set with sixteen elements.

It is useful, especially in the current context, to think of uniform crossover as applied at the bit-level as having two components, the first of which randomly selects each parameter from one or other parent, while the second crosses these parameters. (Of course, the fact that each component is subsequently crossed means that the initial choice is irrelevant.) Similarly, one-point crossover in $\mathcal{C}$ (without inversion) can be thought of as first randomly crossing the whole *parameter* string in $\mathcal{S}$, and then crossing the parameter at the interface at the bit level. The two approaches could be combined.

This way of thinking about conventional operators suggests a solution to the problem of the extremely high cardinality of the neural network topology representation suggested above.

7.2 Crossing Hidden Nodes

The proposed way of tackling neural network topology optimisation makes use of the previously discussed breakdown of recombination into two stages. Most of the nodes in the child (say all but one) will be generated using the multiset recombination operators introduced in section 5. The final node or nodes will be produced by directly crossing nodes from the two parents. This turns out to be an extremely simple matter because the hidden nodes have already been conveniently described as binary strings (figure 4). Moreover, this binary representation of a hidden unit is highly meaningful and completely non-redundant: a one indicates the present of a connection to an external node and a zero indicates its absence. Thus uniform crossover is the natural operator to use to perform the sub-node level cross, since it transmits genes and properly assorts the schemata which are the natural formae for describing hidden units.

7.3 Direct Input-Output Connections

The formulation thus far has concentrated on strictly layered networks so that direct connections from the input nodes to the output nodes have not been considered. Each such connection can, however, be described perfectly and without redundancy by a binary gene, and thus if such connections are to form part of the optimisation they may simply be added to the

[25] This is true if 1-point crossover with inversion is applied in $\mathcal{C}$ also, though in this case the linkages should in principle make useful crosses more likely.

chromosome. Such portions of the gene can be recombined using uniform or some other standard crossover operator without further complication.

8 Linkage and Forma Disruption

The discussion of recombination operators in the previous sections has not considered two much-discussed and intricately related aspects of recombination—linkage between the components of a solution and disruption rates for formae. These will now briefly be reviewed.

When Holland introduced genetic algorithms with his seminal 1975 book (Holland [12]), he listed three generic operators—crossover, mutation and inversion. While there have been experiments with inversion, which are summarised in Goldberg [5], and while there is still a belief among some workers that as more complex problems are tackled the inclusion of inversion will be found to be more helpful, inversion operators are rarely now used in practice. The issue of linkage is, however, relevant not only to traditional linear chromosomes. In the case of the set and multiset operations which form one focus of this paper it would be easy to add linkage information in the usual way, to manipulate this information through inversion or other operators, and to use it to modify the probability of groups of genes being transferred to children *en masse.*

Similarly, the uniform crossover operator with parameter half[26] is widely criticised for being unduly disruptive both of short schemata and of longer ones, the latter effect arising because uniform crossover with parameter half is biased towards taking half the genetic material for a child from each parent (Schaffer *et al.* [26], Syswerda [30], Radcliffe [21]). It has been pointed out, however, by Spears and De Jong [29] that by using parameters other than half with uniform crossover its degree of disruptiveness can be completely controlled. Whether or not uniform crossover appears more attractive is therefore primarily a function of whether there is any *a priori* reason to believe that the arrangement of the genes on the chromosome groups together bits which should be tightly coupled. (Where parameters are laid out on continuous segments of the chromosome, there is a strong argument that this is so.)

Just as uniform crossover can be parameterised to control the amount of genetic material it tends to take from each parent, the weights of the elements used to determine the child in the manner described in section 6.4 can be used to control the degree of forma disruption exhibited by the recombination operators discussed in this paper.

9 Discussion and Summary

This paper contains a number of disparate threads, principal among which are the application of genetic algorithms to problems in neural networks and the application of forma analysis to the construction of suitable operators and representations for (multi)set recombination. These considerations have required a number of sub-discussions, including the introduction of non-orthogonal bases for equivalence relations and a consideration of sub-parameter level recombination with particular reference to crossing hidden nodes in neural network topologies. Each of these subjects is now set in context and—where appropriate—summarised.

9.1 Set and Multiset Recombination

Set recombination can be regarded as a special case of multiset recombination in which the maximum number of copies of any element is one. A distinction must, however, be made between fixed- and variable-size multisets.

9.1.1 Formae and Equivalence Relations

Generic formae for multisets specify ranges of values for the multiplicities of elements drawn from a universal set $\mathcal{E}$. These formae are induced by equivalence relations Ψ over the search space of which a low precision example is given by equation 44:

$$\psi(\eta,\zeta) = \begin{cases} 1, & \text{if } m(x,\eta), m(x,\zeta) < N_x^{\downarrow} \\ & \text{or } m(x,\eta), m(x,\zeta) \in [N_x^{\downarrow}, N_x^{\uparrow}] \\ & \text{or } m(x,\eta), m(x,\zeta) > N_x^{\uparrow}, \\ 0, & \text{otherwise.} \end{cases}$$

This equivalence relation is described by the *description set* $\{(x, N_x^{\downarrow}, N_x^{\uparrow})\}$. A basis for these equivalence relations is given by equation 46:

$$E = \bigl\{ \psi \in \Psi \bigm| \langle\psi\rangle = \{(x, 0, N_x^{\uparrow})\}, x \in \mathcal{E}, N_x^{\uparrow} \in [0, N^{\star}] \bigr\}.$$

9.1.2 Variable-Size (Multi)set Recombination

The inheritance/R^3 operator for these formae generates a child θ, given parents $\eta, \zeta \in \mathcal{C}$ $[\equiv \mathbb{P}_m(\mathcal{E})]$ as follows.

[26] i.e. equal probabilities of taking each gene from either parent

- For each element $x \in \mathcal{E}$ set $m(x, \theta)$ so that

$$\begin{aligned} \min\bigl(m(x,\eta), m(x,\zeta)\bigr) &\le m(x,\theta) \\ &\le \max\bigl(m(x,\eta), m(x,\zeta)\bigr), \end{aligned} \tag{73}$$

  where $m(x, \eta)$ is the number of copies of x in η, and in the case of sets (as opposed to multisets) $m(x, \eta) \in \{0, 1\}$. In each case a *uniform random* choice is made between the maximum and minimum allowed values for the multiplicity.

This operator respects and properly assorts the formae induced by the equivalence relations in Ψ, and transmits the genes identified with the basic equivalence relations in E. The operator can be biased towards the multiplicity of one or other parent without violating these properties. Notice that

$$\eta \cap \zeta \subset \theta \subset \eta \cup \zeta. \tag{74}$$

9.1.3 Fixed-Size (Multi)set Recombination

The corresponding inheritance crossover operator for multisets of fixed size N generates a child θ from parents η and ζ as follows:

1. Let
$$n_x^{\downarrow}(\eta,\zeta) = \min\left(m(x,\eta), m(x,\zeta)\right) \tag{75}$$
   and
$$n_x^{\uparrow}(\eta,\zeta) = \max\left(m(x,\eta), m(x,\zeta)\right). \tag{76}$$
   Then for each $x \in \mathcal{E}$ set the *initial* values for the multiplicity to be the minimum of that of the two parents:
$$\forall x \in \mathcal{E}: \; m(x, \theta_i) = n_x^{\downarrow}(\eta,\zeta), \tag{77}$$
   where the subscript i on θ_i indicates that this is the initial child only.
2. Fill the remaining
$$N - \sum_{x \in \mathcal{E}} n_x^{\downarrow}(\eta,\zeta)$$
   places in the child θ by increasing the multiplicity of elements at random subject to the following conditions on the *final* child θ,
$$\forall x \in \mathcal{E}: \; m(x,\theta) \le n_x^{\uparrow}(\eta,\zeta), \tag{78}$$
   and
$$|\theta| = N. \tag{79}$$

This operator respects formae and strictly transmits genes but does not assort them because the formae are non-separable. For this reason, the following alternative *assorting recombination operator* is to be preferred:

1. Assign to each element x in $\mathcal{E}$ a weight
$$w_x = m(x,\eta) + m(x,\zeta). \tag{80}$$
2. Fill *each* available place in the child in turn picking each element x with probability
$$p(x) = w_x \Big/ \sum_{x' \in \mathcal{E}} w_{x'} \tag{81}$$
   subject to the constraints on the final child θ given by equations 78 and 79.

9.1.4 Mutation

In the case of multiset (as opposed to set) recombination, there is a bias in the operator towards the middle of the range of allowed multiplicities for elements. This should be redressed by the use of *end-point mutation operators* as discussed in Radcliffe [22, 23]. These, with low probability, set the multiplicity of some element in a child to 0 or $N^{\star}$ (the maximum allowed value), making any other necessary adjustments if multisets are of fixed size.

9.2 Neural Networks

By framing three-layer neural network topology optimisation as a multiset problem over hidden nodes, each complete with external connections, the following have been achieved:

- a non-redundant representation;
- the multiset recombination operator discussed above may be applied;
- suitable formae exist which can be separated.

Because, however, of the very large size of the universal set $\mathcal{E}$, consisting of all possible hidden unit types (2^n for networks with n external (input/output) nodes), recombination of hidden units is also advisable. If the hidden nodes are described by binary strings, with each bit representing an external connection, conventional crossover operators (uniform, n-point etc.) can be used to perform this cross in the controlled ways advocated in section 7.2.

Experimental verification of the efficacy of this approach remains to be demonstrated. An integrated scheme for optimising connectivity and weights also requires further work and might be regarded as the “Holy Grail” in this area.

9.3 Non-Orthogonal Bases

The introduction of non-orthogonal bases is a significant innovation for forma analysis and allows a deeper study of problems for which schemata are not appropriate and for which an orthogonal basis for equivalence relations inducing appropriate formae is unavailable.

Given a complete *orthogonal* basis E for a set Ψ of equivalence relations which induce the chosen formae, a chromosomal representation can be constructed by allocating one locus to each basic equivalence relation in E:

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	$\cdots$	e_n

where $|E| = n$. The basic equivalence classes (basic formae) then serve as the alleles. Uniform crossover strictly transmits these genes and properly assorts the formae. The critical point here is that with an *orthogonal* basis, *any* chromosome of this form represents some legal solution in $\mathcal{S}$ uniquely and every solution in $\mathcal{S}$ has one such chromosomal representative.

In the case of a non-orthogonal basis, not all chromosomes of this type are legal: constraints exist on the permitted combinations of gene values (alleles). The work in section 5.6 shows, however, that while traditional crossover operators (uniform, n-point etc.) cannot manipulate these formae and solutions effectively, the inheritance crossover operator will respect the formae and transmit genes while properly assorting whenever these conditions are compatible. In cases where gene transmission prevents proper assortment, operators such as the *assorting crossover* of section 6.4 can provide reasonable alternatives.

It is hoped that the examples in this paper convincingly demonstrate the flexibility of forma analysis, and the insights which it can provide.

Acknowledgements

I would like to thank Peter Hancock for a conversation which revived my interest in and consideration of the application of genetic algorithms to problems in neural networks, and thus led indirectly to the to work in this paper. I would also like to thank Mike Norman for continuing occasional discussions of genetic algorithms, which form the backdrop to the thinking in this paper.

References

[1] Richard K. Belew, John McInerny, and Nicol N. Schraudolph. Evolving Networks: Using the Genetic Algorithm with Connectionist Learning. Technical Report CS90–174, UCSD (La Jolla), 1990.

[2] Lashon Booker. Improving Search in Genetic Algorithms. In Lawrence Davis, editor, *Genetic Algorithms and Simulated Annealing*. Pitman (London), 1987.

[3] Anthony N. Burkitt. Optimisation of the Architecture of Feed-forward Neural Networks with Hidden Layers by Unit Elimination. *Complex Systems*, 5(4):371–380, 1991.

[4] Lawrence Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold (New York), 1991.

[5] David E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley (Reading, Mass), 1989.

[6] David E. Goldberg. Genetic Algorithms and Walsh Functions: Part I, A Gentle Introduction. *Complex Systems*, 3:129–152, 1990.

[7] David E. Goldberg. Real-coded Genetic Algorithms, Virtual Alphabets, and Blocking. Technical Report IlliGAL Report No. 90001, Department of General Engineering, University of Illinois at Urbana-Champaign, 1990.

[8] David E. Goldberg and Robert Lingle Jr. Alleles, Loci and the Traveling Salesman Problem. In *Proceedings of an International Conference on Genetic Algorithms*. Lawrence Erlbaum Associates (Hillsdale), 1985.

[9] Peter J. B. Hancock. GANNET: Design of a Neral Net for Face Recognition by a Genetic Algorithm. Technical report, University of Stirling, 1990.

[10] Steven Alex Harp, Tariq Samad, and Aloke Guha. The Genetic Synthesis of Neural Networks. Technical Report CSDD–89–I4852–2, Honeywell, 1989.

[11] Steven Alex Harp, Tariq Samad, and Aloke Guha. Towards the Genetic Synthesis of Neural Networks. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[12] John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press (Ann Arbor), 1975.

[13] John R. Koza. Genetic Programming: A Paradigm for Genetically Breeding Populations

of Computer Programs to Solve Problems. Technical Report STAN-CS-90-1314, Stanford University, 1990.

[14] John R. Koza. Evolving a Computer to Generate Random Numbers Using the Genetic Programming Paradigm. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 37–44. Morgan Kaufmann (San Mateo), 1991.

[15] Geoffrey F. Miller, Peter M. Todd, and Shailesh U. Hegde. Designing Neural Networks using Genetic Algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[16] David J. Montana and Lawrence Davis. Training Feedforward Neural Networks Using Genetic Algorithms. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 762–767, 1989.

[17] H. Mühlenbein. Parallel Genetic Algorithms, Population Genetics and Combinatorial Optimization. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[18] H. Mühlenbein and J. Kindermann. The Dynamics of Evolution and Learning — Towards Genetic Neural Networks. In P. Pfiefer, Z. Schreter, F. Fogelman-Soulié, and L. Steels, editors, *Connectionism in Perspective*. North-Holland, 1989.

[19] Michael Norman. A Genetic Approach to Topology Optimisation for Multiprocessor Architectures. Technical report, University of Edinburgh, 1988.

[20] Nicholas J. Radcliffe. The Permutation Problem. (unpublished manuscript), 1988.

[21] Nicholas J. Radcliffe. *Genetic Neural Networks on MIMD Computers*. PhD thesis, University of Edinburgh, 1990.

[22] Nicholas J. Radcliffe. Equivalence Class Analysis of Genetic Algorithms. *Complex Systems*, 5(2):183–205, 1991.

[23] Nicholas J. Radcliffe. Forma Analysis and Random Respectful Recombination. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 222–229. Morgan Kaufmann (San Mateo), 1991.

[24] Mike Rudnick. A Bibliography of the Intersection of Genetic Search and Neural Networks. Technical Report CS/E 90-001, Oregon Graduate Centre, 1990.

[25] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Representations by Back-Propagating Errors. *Nature*, 323, 1986.

[26] J. David Schaffer, Richard A. Caruna, Larry J. Eshelman, and Rajarshi Das. A Study of the Control Parameters Affecting Online Performance of Genetic Algorithms for Function Optimisation. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[27] J. David Schaffer and Larry J. Eshelman. On Crossover as an Evolutionarily Viable Strategy. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 61–68. Morgan Kaufmann (San Mateo), 1991.

[28] J. Seitsma and R. J. F. Dow. Neural Net Pruning — Why and How. In *Proceedings of the IEEE Conference on Neural Networks*, volume II, 1988.

[29] William M. Spears and Kenneth A. De Jong. On the Virtues of Parameterised Uniform Crossover. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 230–236. Morgan Kaufmann (San Mateo), 1991.

[30] Gilbert Syswerda. Uniform Crossover in Genetic Algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[31] Michael D. Vose. Personal Communication, 1991.

[32] Michael D. Vose. Generalizing the Notion of Schema in Gentic Algorithms. *Artificial Intelligence*, (in press).

[33] Michael D. Vose and Gunar E. Liepins. Schema Disruption. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 237–243. Morgan Kaufmann (San Mateo), 1991.

[34] Gerhard Weiss. Combining Neural and Evolutionary Learning: Aspects and Approaches. Technical Report FKI–132–90, Forschungsberichte Künstliche Intelligenz, 1990.

[35] Darrell Whitley. Applying Genetic Algorithms to Neural Network Problems: A Preliminary Report. (Unpublished manuscript).

[36] Darrell Whitley. Using Reproductive Evaluation to Improve Genetic Search and Heuristic Discovery. In *Proceedings of the Second*

International Conference on Genetic Algorithms. Lawrence Erlbaum Associates (Hillsdale), 1987.

[37] Darrell Whitley, Stephen Dominic, and Rajarshi Das. Genetic Reinforcement Learning With Multilayer Neural Networks. In *Proceedings of the Fourth International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1991.

[38] Darrell Whitley and Thomas Hanson. The Genitor Algorithm: Using Genetic Algorithms to Optimize Neural Networks. Technical Report CS-89-107, Colorado State University, 1989.

[39] Darrell Whitley, Timothy Starkweather, and Christopher Bogart. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. Technical Report CS-89-117, Colorado State University, 1989.

[40] Darrell Whitley, Timothy Starkweather, and D'Ann Fuquay. Scheduling Problems and Traveling Salesmen: The Genetic Edge Recombination Operator. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[41] Darrell Whitley, Timothy Starkweather, and Danial Shaner. The Traveling Salesmen and Sequence Sheduling: Quality Solutions Using Genetic Edge Recombination. In Lawrence Davis, editor, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold (New York), 1991.