

# Clonal Selection from First Principles

Chris McEwan and Emma Hart

Edinburgh Napier University  
{c.mcewan, e.hart}@napier.ac.uk

**Abstract.** Clonal selection is the keystone of mainstream immunology and computational systems based on immunological principles. For the latter, clonal selection is often interpreted as an asexual variant of natural selection, and thus, tend to be variations on evolutionary strategies. Retro-fitting immunological sophistication and theoretical rigour onto such systems has proved to be unwieldy. In this paper we assert the primacy of *competitive exclusion* over selection and mutation; providing theoretical analysis and empirical results that support our position.

We show our algorithm to be highly competitive with well-established approximation and learning techniques; notably for large, high-dimensional problems. We further argue that it provides concise, operational abstractions of some influential theoretical immunology.

## 1 Introduction

Burnet’s Clonal Selection principle is the keystone of both mainstream theoretical immunology and immune-inspired computational systems. Briefly, antigen *select* their responding lymphocyte clones through a cyclic process of receptor-ligand binding, proliferation, mutation and competitive exclusion. Thus, randomly generated lymphocytes, with receptors proven “fit” in the pathogenic environment of the host, persist and improve.

To date, the computational perspective on clonal selection has, overwhelmingly, been based on the *genetic algorithm without crossover* model proposed almost 20 years ago by Forrest *et al.* [16]. This work was not itself motivated by solutions to computational problems, but proved to be seminal in inspiring research in that direction (e.g. [10, 4]). Despite subtle differences in implementations, these algorithms all have the same basic features: a cloning operator that reproduces solutions and a hypermutation operator that performs mutation. Although such algorithms differ from classical evolutionary algorithms in their lack of a crossover operator and in the use of operators that are a function of solution fitness, the underlying computational model remains essentially the same [30].

A second sizable branch of computational work with clonal selection has been directed at producing unsupervised learning algorithms [11, 41]. Much of the seminal research here also built upon the work of Forrest *et al.* – focusing first on the pattern-matching nature of receptor-ligand binding, then developing these ideas into clustering and classification in  $\mathbb{R}^N$ . The resulting algorithms tend to share commonalities with Self-Organising Maps, Learning Vector Quantization

or instance-based methods from statistical learning [21]. Stibor and Timmis [35] were the first to provide a negative result in this area, showing why the aiNet algorithm is incapable of achieving its stated goal of data compression. McEwan and Hart more recently elaborated several other flaws that potentially undermine the basic principles underlying this family of algorithms [27].

In contrast to the above, much of the seminal work that preceded the formation of *artificial immune systems* as a recognised field (e.g. [2, 15]) very much focused on applying and exploring computational properties of biologically plausible dynamical systems; rather than re-contextualising the immunology into existing algorithmic approaches. It is our position that this was, and still is, a better way to produce artificial immune systems. In this paper, we demonstrate how such methods (and a little hindsight) improve not only biological plausibility, but theoretical approachability and computational effectiveness.

The paper develops as follows: in the next section we motivate and introduce our dynamical system perspective of clonal selection as competitive exclusion. In Sect. 3 we theoretically analyse the approximatory capacity of our model and empirically demonstrate its effectiveness in the domain of *sparse approximation*. In Sect. 4 we extend this approximatory behaviour to decision making, cast in a statistical learning framework; and again, provide empirical support of our position. In Sect. 5 we briefly discuss the biological plausibility of our model, before concluding in Sect. 6.

## 2 Clonal Selection is Competitive Exclusion

To date, the emphasis on clonal selection as algorithm has been on the *generate-and-filter* approach typified by Darwinian selection. This emphasis has been almost exclusively placed on the *generate* aspect, with much attention given to receptor (cf. genotype) representation and mutation strategies. The *and-filter* aspect has received very little attention and any attention it has received tends to also be imported from evolutionary computing through selection and retention operators such as elitism, tournament selection and so on. Although the generation of novel and diverse components is a critical aspect of clonal (and natural) selection, here we will concentrate on filtering. In the particular case of the learning domain, poorly realised interaction mechanisms lie behind several of the negative results already discussed.

In contrast to *selection operators*, biologists have very simple and elegant models of inter-species competitive exclusion [31]. The one we will focus on is paradigmatic, originally formulated by Alfred Lotka and Vito Volterra [24, 40], and later developed by many others, notably Levins [23], Roberts [34] and Nowack [32]. This *generalised Lotka-Volterra model* is as follows: we posit each species has a *carrying capacity* that the environment can support; each species population grows towards that capacity under the classic sigmoid-shaped logistic equation; reaching this capacity is further hindered by interactions from other species competing for the same resources.

More formally, let  $\rho_i$  and  $k_i$  represent the population and carrying capacity of the  $i$ 'th species, respectively. The so-called ‘‘community matrix’’  $K$  represents the competitive effects of species  $i$  on species  $j$ , that is  $K_{ij} \geq 0$  and  $K_{ii} = 1$ . The population dynamics evolve as follows

$$\frac{d\rho_i}{dt} = \left( \frac{k_i - \sigma_i}{k_i} \right) \rho_i \quad (1)$$

where  $\sigma_i = \sum_j K_{ij}\rho_j$ . It is apparent that when  $k_i = \sigma_i$  the capacity is equal to the competitive effects and that species reaches equilibrium. If  $k_i < \sigma_i$  then the species is out-competed and declines. If  $k_i > \sigma_i$  the species grows smoothly towards its, now reduced, capacity. It is straight-forward to add additional factors such as growth-decay rates, immigration-emigration terms and predator-prey interactions, but they add little to the immediate exposition.

## 2.1 Adding Immunological Detail

When modelling clonal selection, our species will be lymphocyte clones differentiated by receptor configuration. It is intuitive that the receptor configuration will determine both capacity (antigen-receptor potential) and competition (receptor-receptor redundancy) as binding-sites are a limited resource.

The common bit-string and geometric models of ligand binding have been criticised on both biological and computational grounds [3, 17]. In contrast, we will attempt a general definition, transparent to both biologists and computer scientists, by modelling *epitopes* as abstract *peptides* localised on the surface of the tertiary structure of proteins (see [28] for a deeper discussion). Let us assume  $n$  such peptides. We then conceptualise receptors  $\varphi_i$  as  $n$ -dimensional vectors with most components zero or negligible. Non-zero components indicate some affinity to that particular peptide; but binding is a function of multiple peptides being correlated on the protein surface. We model the surface of a protein as a square-symmetric matrix  $P$  where  $P_{ij}$  represents the surface correlation of peptides  $i$  and  $j$ . Our measure of *affinity*, or binding strength, is then naturally embodied in the product  $\langle \varphi_i | P | \varphi_i \rangle$  which is the magnitude of  $\varphi_i$  in the subspace of  $\mathbb{R}^n$  defined by  $P$ . We will further assume that surfaces are additive, in which case the entire antigenic environment of the immune system is concisely represented as the matrix  $Q = \sum P_k$ , the aggregation of individual protein surfaces. Thus, a clone’s capacity can now be defined as  $k_i = \langle \varphi_i | Q | \varphi_i \rangle = \sum \langle \varphi_i | P_k | \varphi_i \rangle$ . That is, magnitude in the antigenic environment.

Competitive effects between clones are quite naturally embodied in receptor redundancy, which we again represent abstractly as correlation  $\langle \varphi_i | \varphi_j \rangle \in [0, 1]$ . Thus the competitive effect on clone  $i$  is an aggregate measure of redundancy and competitor fitness  $\sigma_i = \sum_j \langle \varphi_i | \varphi_j \rangle \rho_j$ , which includes a clones ‘‘competition’’ with itself  $\langle \varphi_i | \varphi_i \rangle \rho_i$ . We collect these correlations in a matrix  $K$  that readily satisfies the conditions for Eq. (1).

### 3 Competitive Exclusion for Approximation

Our first goal is to demonstrate the empirical and theoretical efficacy of our simple model in a non-decision-making setting. Approximation is the perfect domain for this: on the one hand, it can be a precursor to good decision making; on the other, it is widely applicable in applied and theoretical science. This further illustrates clearly how the proposed method deviates from the state of the art in AIS applied to unsupervised learning.

The classical approximation problem formulation is to minimise the metric distance between a given vector or function  $x$  and its approximation  $\tilde{x}$  chosen from some set of elements. Of particular interest here will be additive expansions of basis functions  $\varphi_i \in \Phi$  such that

$$x \approx \tilde{x} = \sum_i \alpha_i \varphi_i = \Phi \alpha \quad (2)$$

The classic metric of error is the  $\ell_2$  norm, leading to a least-squares problem

$$\operatorname{argmin}_{\alpha} \|x - \Phi \alpha\|_2^2 \quad (3)$$

which, if  $\Phi$  is an orthonormal basis, has the particularly convenient solution  $\alpha = \Phi^T x$ , that is,  $\alpha_i = \langle \varphi_i | x \rangle$ . In this case the approximation is exact and easy to compute. But this convenience comes with two undesirable conditions:

1. The constraint of pairwise orthogonality severely limits the form (and amount) of components in the additive expansion [7]. This makes representing some signals extremely convoluted (e.g. representing a sharp, temporally localised wave with periodic functions). This is also a problem when the coefficients of  $\varphi_i$  are to be interpreted (e.g. representing data as a sum of latent factors). In both cases, it is desirable to expand the number and diversity of columns of  $\Phi$ , resulting in redundant, *overcomplete* representations [1].
2. Any least-squares solution  $\alpha$  will be *dense*, that is, every basis will contribute to the approximation. In many domains, assuming sparsity in the coefficients is either reasonable or highly desirable. For example, in statistics, one might appeal to parsimony of the model (i.e. *feature selection*); in signal processing, an appropriately chosen basis may induce the representation coefficients to rapidly approach zero, allowing truncation with little perceptible loss in reconstruction (i.e. *lossy compression*).

The ubiquity of these conditions leads to *sparse approximation*. Stated as an optimisation objective, it is essentially a regularised variant of Eq. (3) that can be used to finesse the over-determined nature of (1) and bias the solution of (2) towards extremal coefficient values

$$\operatorname{argmin}_{\alpha} \|\alpha\|_{\ell} \text{ s.t. } \|x - \Phi \alpha\|_2^2 < \epsilon \quad (4)$$

In principle, the sparsest solution to Eq. (4) can be quantified using the  $\ell_0$  pseudo-norm, which counts the non-zero coefficients in  $\alpha$ . Unfortunately, the

```

function mp ( $x, \Phi$ )
   $r = x$ 
   $\alpha = []$ 
  while  $\|r\|_2 > \epsilon$  do
     $i = \operatorname{argmax}_i \langle \varphi_i | r \rangle$ 
     $\alpha_i = \langle \varphi_i | r \rangle$ 
     $r = r - \alpha_i \varphi_i$ 
  end
  return  $\alpha$ 

```

**Algorithm 1:** Matching Pursuit. Repeated subtraction of the most correlated atom with the residual error. In the classification and regression setting, variations on this algorithm are Least Angle Regression and  $\ell_2$ -Boosting.

combinatorial nature of this formulation makes it NP-Hard [29]. Briefly, there have been two major thrusts at attacking this problem.

Donoho [12] was the first to show that the  $\ell_0$  and  $\ell_1$  solutions coincide when  $\|\alpha\|_0 < \frac{1+M^{-1}}{2}$ , where  $M$  is the ‘‘coherence’’ of  $\Phi$  defined as  $\max_{i \neq j} \langle \varphi_i, \varphi_j \rangle$ . Using the  $\ell_1$  norm, it is (somewhat) straight-forward to relax this combinatorial optimisation into a quadratic program with linear equality constraints (see e.g. [38, 5]). In the signal processing literature, this method is known as *Basis Pursuit* [6]; in statistical learning, it is called the *Lasso* [37]. Unfortunately, this rigorous approach is prohibitive computationally and scales very poorly.

The second approach uses heuristic, greedy algorithms to construct a sparse representation sequentially. Mallat and Zhang’s [25] *Matching Pursuit* algorithm holds a special place in the literature. It is simple, intuitive, and has a rich history within, and outside of, the field [33, 13, 20, 19]. We outline the procedure in Alg. (1): the residual error  $r$  is repeatedly stripped of structure correlated with bases until a stopping criteria is satisfied (e.g. number of chosen bases, norm of the residual etc). In regression and classification problems, this approach is known as *Forward Stepwise Regression* and  $\ell_2$ -*Boosting*, respectively. A modern variation on this idea, *Least Angle Regression* [14], avoids overly-greedy steps based on  $\langle \varphi_i | r_t \rangle$ , favouring instead to increase  $\alpha_i$  until  $\varphi_i$  is no longer the most correlated with  $r$ ; at which point a ‘‘competing’’ predictor is introduced to the representation. It is this notion of competition amongst predictors, bases or classifiers that we wish to develop here, albeit without myopic greediness.

### 3.1 Competition and Approximation

Returning to our dynamical system of Eq. (1) we now elaborate on the approximatory capacity of the repertoire under our formulation as a basis. If  $\Phi$  was orthonormal, then by definition there are no competitive effects between clones. The dynamics of  $\rho$  smoothly approaches equilibrium where  $\rho_i = k_i = \langle \varphi_i | x \rangle$ , as would be expected from any orthonormal system. For a redundant, over-complete repertoire, there is a trade-off for clones: growth requires maximising correlation with the antigenic environment (capacity) *and* minimising correla-

tion with other clones (competition). This makes sense in the approximation setting: the former reduces reconstruction error; the latter promotes “almost orthogonality”, penalising representational redundancy and rewarding diversity. In contrast to Donoho’s *coherence* (and similar measures) we need not expect our basis to satisfy almost orthogonality *a priori*; rather the competition dynamics promote satisfaction in the context of individual signal decompositions.

We can clarify the approximatory behaviour of the repertoire by simplifying capacity from  $k_i = \langle \varphi_i | X | \varphi_i \rangle$  to  $k_i = \langle \varphi_i | x \rangle^1$ . It then follows that all numerators in Eq. (1) can be written simultaneously as

$$\begin{aligned} k - K\rho &= \Phi'x - \Phi'\Phi\rho \\ &= \Phi'(x - \tilde{x}) \end{aligned} \tag{5}$$

which clearly stabilises when  $x = \tilde{x}$ . Note that this is essentially a restatement of the least-squares solution  $\rho = (\Phi'\Phi)^{-1}\Phi'x$  as  $(\Phi'\Phi)\rho = \Phi'x$ . Rather than inverting a matrix we are iterating  $\Phi'\Phi$ . The logistic equation further introducing non-linearity and “lag” into Eq. (1) and the stable configuration is *not* the least squares solution. The above derivation also shows that the competition vector  $\sigma = \Phi'\Phi\rho = \Phi'\tilde{x}$  is equivalent to receptor correlation with the implicit approximation  $\tilde{x}$ . Any clone more correlated with the approximation will decelerate and decay, again, penalising redundancy in the representation.

Extending to the matrix setting  $k_i = \langle \varphi_i | X | \varphi_i \rangle$  carries several benefits. With vectorial data, squaring the capacity promotes additional sparsity by penalising unfit and boosting fit clones (demonstrated later). Further, when  $X$  is an aggregated quantity the matrix representation more correctly represents the sum of correlated “features”, particularly in sparse data environments. Further still, it accommodates a richer matrix-based data representation, while degrading gracefully in the standard vectorial setting.

### 3.2 Competition and the Greedy/Global Trade-Off

We have shown that, under a suitable formulation, a repertoire of competing receptors can collectively approximate the antigenic environment. We have also highlighted the similar form to the traditional least squares solution and the crucial sparsity inducing non-linearities. We will empirically validate these properties in the next section. Here we elaborate on where this method is positioned in the spectrum of greedy heuristics and global optimisations.

To simplify our analysis, we will again let  $k_i = \langle \varphi_i | x \rangle$  and retain  $K_{ij} = \langle \varphi_i | \varphi_j \rangle$ . Let  $\max(\cdot)$  return the index of the maximum component, rather than the value. Now observe that in Alg. (1) the index  $i_1$  in the first iteration will be  $\max(k)$ , simply because  $r_0 = x$ . On the second iteration

<sup>1</sup> In the traditional vectorial data-analysis setting this is reasonable. Using an outer-product to represent our “surface”  $\langle \varphi_i | x \rangle \langle x | \varphi_i \rangle$  visibly reduces to  $\langle \varphi_i | x \rangle^2$ . Thus we are simply using the square root of the matrix representation in our current analysis.

$$\begin{aligned}
i_2 &= \max(\Phi' r_1) \\
&= \max(\Phi'(x - \langle \varphi_{i_1} | x \rangle \varphi_{i_1})) \\
&= \max(\Phi' x - \langle \varphi_{i_1} | x \rangle \Phi' \varphi_{i_1}) \\
&= \max(k - \langle \varphi_{i_1} | x \rangle K_{i_1})
\end{aligned}$$

where  $K_{i_1}$  refers to the  $i_1$  column of  $K$ . In general, we have

$$\begin{aligned}
i_{t+1} &= \max(k - \sum_{j=1}^t \langle \varphi_{i_t} | x \rangle K_{i_t}) \\
&= \max(k_t)
\end{aligned} \tag{6}$$

where  $k_t = k_{t-1} - \langle \varphi_{i_t} | x \rangle K_{i_t}$ . What this derivation makes explicit is the implicit role that inter-basis correlation plays in the evolution of the Alg. (1). When a basis  $\varphi_{i_t}$  is selected, those correlated with it suffer a drop in their capacity proportional to their correlation with the signal in the subspace of  $\varphi_{i_t}$

$$k_{j(t+1)} = k_{j(t)} - \langle \varphi_j | \varphi_{i_t} \rangle \langle \varphi_{i_t} | x \rangle \tag{7}$$

Crucially, notice that we are now dealing solely the same quantities used in Eq. (1) – capacity and competition. If we expand Eq. (5) as

$$k - K\rho = k - \sum_{j \in \Phi} \rho_j K_j \tag{8}$$

then it becomes clear that while Alg. (1) greedily sums over the *current* selections, weighting by the *maximal* coefficient values (Eq. 6); in contrast, competitive exclusion sums over *all* dictionary atoms, weighting by the *current* coefficient values  $\rho_i$ . The rest of Eq. (1) simply provides an update rule to have  $\rho_i \rightarrow k_i$ , subject to competitive effects. So, in contrast to the myopic selective process of Matching Pursuit, Eq. (1) uses a more informed eliminatory process – evolving coefficients, i.e. population sizes, in parallel as a dynamical system.

Clearly, simulating the entire potential repertoire is impractical and implausible. Preliminary results (not shown) suggest that redundancy ensures results are not degraded by simulating a fixed-size repertoire with sampling from the full  $\Phi$ . Evolutionary search through  $\Phi$  is an obvious next step in development.

### 3.3 Approximation Experiments

In the following experiments we follow a standard protocol. First we randomly generate an arbitrary over-complete “basis”. From this basis we generate synthetic signals, each a sparse linear combination of basis vectors, adding a small amount of Gaussian noise. We then compare the approximations found by Matching Pursuit (greedy), Basis Pursuit (global) and competitive exclusion.

For each algorithm, we record the summary statistics (max, min, quartiles, mean and variance) averaged over 100 signals for

- **Sparsity:** number of non-zeros components  $\|\alpha\|_0$ .
- **CPU:** time to produce a representation.
- **Reconstruction Error:**  $\|x - \Phi\alpha\|_2^2$
- **Synthetic Error:**  $\|\beta - \alpha\|_2^2$ , described below.

Squared reconstruction error is the *de facto* metric in these types of experiments. However, reconstruction error is only a proxy measure *implying* that the algorithm has found a good representation. When using synthetic signals it is possible to measure the actual error in representation, that is, the error in selected coefficients and their magnitude. We refer to this as *Synthetic Error*:  $\|\beta - \alpha\|_2^2$  where  $\beta$  is the coefficients used to generate the synthetic datum. In contrast to pure approximation, this metric can be important when the bases have application-specific meaning and their coefficients are to be interpreted. This is illustrated further in Fig. (1).

In Figure (2) we graph the performance of the algorithms in approximating 100-dimensional signals, each generated from 10 bases selected at random from a 1000 elements basis. Each population starts with the same initial value (see supplementary code for further details). We also plot two variants of our algorithm to illustrate the trade-off discussed in Sect. 3.1: *dp* uses the regular dot product as capacity; *dp2* squares the dot product. In general, the former is more accurate but slower and denser, the latter is faster and sparser. It is apparent that competitive exclusion is, true to its name, very competitive across metrics. Although simulating dynamics will obviously be computationally more expensive than greedy approximation, the difference is not as large as one might expect. Further it is significantly faster than performing optimisation by linear programming – Basis Pursuit’s CPU time was over 200 seconds and is well outside the bounds of the graph. It is interesting to note that in all other respects *dp* performs similarly to Basis Pursuit, at a fraction of the computational effort. At the other extreme, we see that although Matching Pursuit achieves very low reconstruction error, it does so at significant cost to synthetic error and sparsity. Recall, each signal is a linear combination of 10 bases. Matching Pursuit, Basis Pursuit and our own algorithm without quadratic capacity all significantly under-estimate the true sparsity, employing between 60 and 90 bases. In contrast, the quadratic capacity version of our algorithm, as suggested in Sect. 3.1, is able to drive the sparsity down to around 20, with negligible effect of synthetic error and a large improvement in time. The cost here is a notable increase in reconstruction error.

## 4 Competitive Exclusion for Learning

Statistical learning is simply another application of approximation, so we do not have to digress too much. The essential difference is that we are now approximating an unknown function of our signal  $y = f(x)$ , rather than approximating

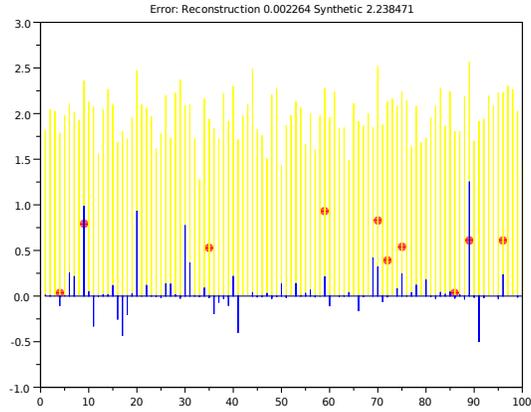


Fig. 1: Illustration motivating the synthetic error metric. Light (yellow) background bars represent each bases correlation with a signal. Heavy (blue) foreground bars represent coefficients recovered by Matching Pursuit. Dots (red) represent generating coefficients. Low reconstruction error does not imply correct recovery of representation.

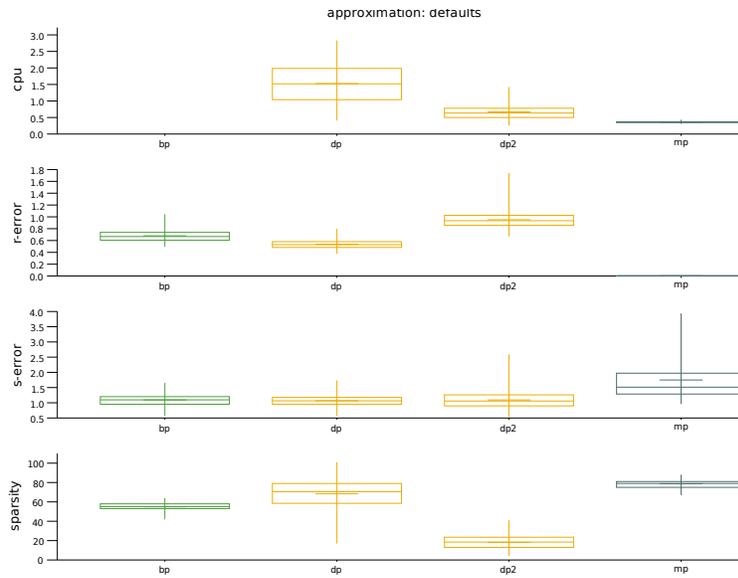


Fig. 2: Approximation results for Matching Pursuit (greedy), Basis Pursuit (global) and variants of our competitive exclusion algorithm. See text for discussion.

$x$  itself. However, we may still make use of the coefficients of  $\varphi_i$  in approximating  $x$  to weigh its contribution to approximating the decision surface  $y$ .

In [28], the authors introduce a novel derivation of the linear classifier, with immunological interpretation, based on the spectral decomposition of the data’s feature correlation matrix. The reader is directed to the primary references for the full mathematical and biological justification, here we simply recreate the resultant decision function

$$\hat{y} = f(\hat{x}) = \langle \hat{x} | X X'^{-1} | X y \rangle = \sum_i \langle \varphi_i | \hat{x} \rangle \frac{\langle \varphi_i | \tilde{w} \rangle}{\langle \varphi_i | G | \varphi_i \rangle} \quad (9)$$

If  $\varphi_i$  were the eigenvectors of  $G = X X' = \sum |x\rangle\langle x|$  then this would be the ubiquitous linear classifier. For the same reasons given in Sect. 3, we are interested in non-orthogonal, overcomplete bases. Technically, only the eigenvectors are blessed with the property that  $G = \Phi \Lambda \Phi'$  and  $G^{-1} = \Phi \Lambda^{-1} \Phi'$  because  $\Lambda$  is the diagonal matrix of eigenvalues  $\Lambda_{ii} = \langle \varphi_i | G | \varphi_i \rangle$ . Using an argument similar to that for justifying nearest-neighbour decisions, we assert that this technical omission can be ignored if the benefits of expanding  $\Phi$  overcome any costs in inaccurately approximating  $G^{-1}$ . For classification this is entirely plausible, as approximation errors do not necessarily imply classification errors.

The relation to Eq. (1) lies in the denominator  $\langle \varphi_i | G | \varphi_i \rangle$ , which is just the same “surface representation of the antigenic environment” used in Sect. 2.1. Thus we see that the decision function involves two other quantities not present in the purely approximatory setting; but both are still measures of approximatory capacity, receptor correlation or affinity:  $\langle \varphi_i | \hat{x} \rangle$  is an intuitive measure of correlation with the test datum and is thus not so different to what has come before;  $\langle \varphi_i | \tilde{w} \rangle$  is a measure of correlation with  $\tilde{w} = \sum y_i x_i$ , the vector measuring feature-label correlations; that is, the current bias individual features have towards one decision or the other. The immunological interpretation of this vector is discussed in Sect. 5, here we concentrate on the numerical details.

The leap from linear classifier to non-linear dynamical model of an immune response now rests on one simple idea. With a slight abuse of notation, let  $\langle \cdot \rangle$  be upper-bounded by the dot product. That is, as before, this upper-bound is the *capacity*; the actual value  $\langle \cdot \rangle$  takes in Eq. (9) is the equilibrium population emerging from inter-clonal competition dynamics. Thus, according to Eq. (9), the immune response is the integration across the repertoire of three key pieces of information about the fitness of each receptor in competing for (i.e. approximating) different environmental resources: the target of the response  $\hat{x}$ ; the antigenic environment of the host  $G$ ; and what we will call the *context* of the response  $\tilde{w}$ . Notice that  $G$  is in the denominator, thus large values are penalised.

#### 4.1 Learning Experiments

It is important to note that we do not think it reasonable to expect Eq. (9) to perform well in arbitrary classification tasks; it is designed to exploit the properties of sparse, high-dimensional problems. To validate our algorithm we

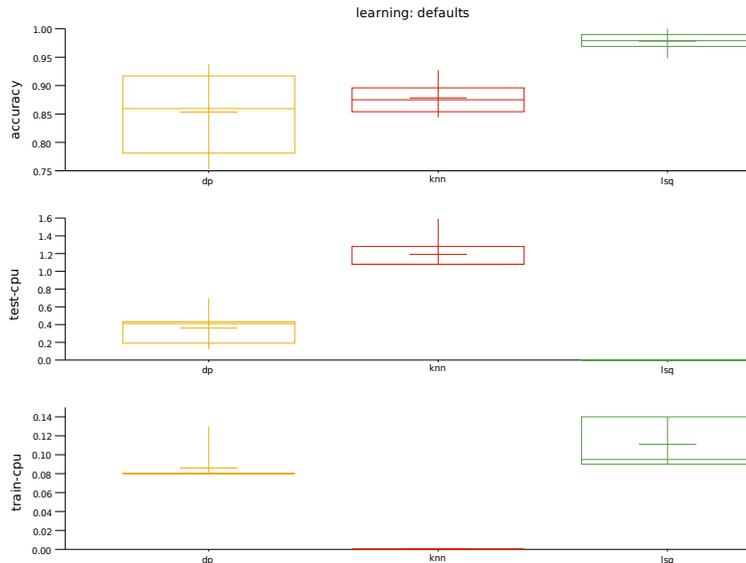


Fig. 3: Learning using competitive exclusion with randomly generated receptors.

use a subset of the UCI *newsgroups* dataset, producing a task of discriminating `comp.graphics` from `alt.atheism` postings. The high dimensions, size, sparsity and noise of text data can antagonise classical parametric and non-parametric techniques. In [27], the immune-inspired algorithm AIRS was shown to perform no better than random guessing on a similar dataset.

As before we record summary statistics for each algorithm, this time classification accuracy, testing and training CPU time. These are recorded over a 10-fold, unstratified partitioning of 1000 documents. We compare performance of our algorithm against the 3-nearest neighbour classifier and the linear classifier because, much like the comparison in Sect. 3.3, these algorithms represent extremes between which our approach lies. Unlike before, we have no *a priori* determined basis to represent the repertoire. In place of more sophisticated methods, we simply generate 10,000 basis vectors, each from a 3-step uniformly random walk on  $G$ . Thus each basis has three non-zero components. Note that  $10,000 \approx 5 \times 10^{-7}$  of the receptor space and we perform no additional searching. We also do not *a priori* assess receptors with respect to producing good results. We stress that our goal here is demonstrating the efficacy of competitive exclusion on randomly generated receptors, not optimising performance metrics.

We plot our results in Fig. (3). At one extreme, the linear classifier's low complexity is robust in high dimensions and its parametric model fast to compute decisions. Numerically it is also most accurate, but at the significant cost of inverting either  $XX'$  or  $X'X$ , neither of which is small. Including the linear classifier is the only reason why the dataset has been limited to 1000 documents.

At the other extreme,  $k$ -NN is obviously superior in training time (because there is none) but test time suffers due to nearest neighbour calculations. The negative effects of many, noisy dimensions also outweigh any benefit in producing non-linear decision boundaries. Our own algorithm performs comparably to the  $k$ -nearest neighbour, albeit with notable variance caused by the simple-minded receptor generation process. However, this variance is not enough to call into question that learning is indeed taking place, even under such severe conditions as uniformly random receptors. Although we generate 10,000 receptors, only about 50-1500 survive the competitive exclusion process. It is perhaps remarkable that so few, 3-dimensional random projections are capable of retaining sufficient representation to compete at all. Notice that, in contrast to the other algorithms, our scaling properties are insulated from data size and, to some extent, dimension; being dominated by the size of the repertoire (in time) and the density of  $XX'$  (in space). Further, our approach directly tackles the curse of dimensionality inasmuch as the repertoire is not attempting to fill  $\mathbb{R}^n$  and the decision function is the aggregation of many low-dimensional decisions.

## 5 Discussion on Immunological Interpretation

With some exceptions [9, 18, 22, 36] the ecology of lymphocytes has been given little attention in the immunology literature. Our own path to the presented model was driven by a desire to make operational some of the influential rhetoric that has inspired artificial immune systems to date. We discuss this now.

Francisco Varela was the original proponent of the cognitive view of the immune system [39]. He posited that the immune system constructs and asserts an internal representation of the immunological self. Quite simply, constructing representations requires building blocks, *not prototypical instances*, and this is precisely what basis functions are. The approximatory capacity of *receptors-as-bases* is our attempt to quantify these largely philosophical arguments. Similarly, Irun Cohen has promoted cognitive analogies. The central idea behind his *co-responsence* [8] – that coherent systemic decisions can emerge from the interactions of randomly generated, contradictory and individually weak components – is supported by our results in Sect. 4. Cohen is also a proponent of the blessings of *degeneracy* and *redundancy* [42]. Our abstraction of receptors as bases makes these these notions operational too. The benefits of redundant bases is what motivated sparse approximation in Sect. 3. Degeneracy naturally follows from our notion of affinity as magnitude in a projected subspace: all  $x$  will intersect with several, but not all, subspaces of the  $\varphi_i$  (i.e. *poly-clonality*); and each  $\varphi_i$  will intersect with several, but not all,  $x$  (i.e. *poly-recognition*). Degeneracy is impossible when abstracting affinity as pointwise distance in metric shape-space.

Lastly, Polly Matzinger’s Danger theory [26] has been contentious in both the immunological and computational literature. Based on our distinction between peptide-fragments and compound epitopes, the so-called *context* vector  $\tilde{w}$  in Eq. (9) may provide some insight here. In conjunction with epitope binding  $\langle \varphi_i | P | \varphi_i \rangle$  inducing B-Cells to proliferate and mutate, the so-called *second*

*signal* that activates B-Cells is triggered, under Matzinger’s hypothesis, by the pro- and anti-response feedback driven by T-Helper and Dendritic cells that, respectively, recognise and sample peptide fragments and chemical indicators of pathology from the tissues. Although we do not model the dynamics of this process, it is easy to see that the feature-class correlations in vector  $\tilde{w}$ , which the  $\varphi_i$  also compete over, fulfil much the same role as the second signal. Certainly,  $\tilde{w} = \sum x_i y_i$  is a gross simplification, but our model is indifferent to how the  $\tilde{w}$  vector is produced. Seen in the light of the decision function Eq. (9), the Danger Theory seems, at least, a pragmatic alternative to more convoluted or teleological explanations of the self-nonsel distinction.

Of course, we are not immunologists. How well our abstractions may yield to biological realism remains to be seen. But we are not aware of another model that captures so many aspects of the immune response, so concisely.

## 6 Conclusion

We have reconsidered the foundational principle behind clonal selection, highlighting the primacy of competitive exclusion over selection and mutation. Theoretical and empirical analysis support that our model is computationally effective and formally approachable. In addition, we have argued that it represents a step towards a level of plausibility and sophistication that is lacking in the ad-hoc hybridisation of evolutionary and instance-based algorithms.

To be clear, we do not wish to abandon the evolutionary aspect of clonal selection. On the contrary, stochastic search still has a lot to offer in terms of adapting the repertoire which, in both biology and computation, cannot typically be realised whole and evolved in unison. But we think it better that this be an additional feature of artificial immune systems, rather than the foundational principle from which all else follows. In the future we hope to take this work in two directions: integrating evolutionary search back into clonal selection; and developing the dynamical system beyond the rudimentary model presented here.

## Reproducible Research

All source code used to produce the results and figures in this paper is available from <http://www.dcs.napier.ac.uk/~cs268/mcewan2010principles.tgz>.

## References

1. M Aharon, M Elad, and a Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear Algebra and its Applications*, 416(1):48–67, 2006.
2. H. Bersini. Immune Network and Adaptive Control. In Paul Bourguine and Francisco Varela, editors, *Toward a practice of autonomous systems, Proceedings of the First European Conference on Artificial Life*, page 217–226, 1991.

3. Jorge Carneiro and John Stewart. Rethinking Shape Space: Evidence from simulated docking suggests that steric shape complementarity is not limiting for antibody-antigen recognition and idiotypic interactions. *J.Theor.Biol*, 169:391–402, 1994.
4. F. Castiglione, S. Motta, and G. Nicosia. Pattern Recognition by primary and secondary response of an Artificial Immune System. *Theory in Biosciences*, 2(120):93–106, 2001.
5. K.M. Cheman. *Optimization Techniques for Solving Basis Pursuit Problems*. Masters, 2006.
6. Scott Shaobing Chen, David L. Donoho, and Michael a. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Review*, 43(1):129, 2001.
7. Ole Christensen. *Frames and Bases: An Introductory Course (Applied and Numerical Harmonic Analysis)*. Birkhauser, illustrate edition, July 2008.
8. Irun R Cohen. Real and artificial immune systems: computing the state of the body. *Group*, 7(July):569–574, 2007.
9. Rob J. De Boer and Alan S. Perelson. T Cell Repertoires and Competitive Exclusion. *Journal of Theoretical Biology*, 169:375–390, 1994.
10. L. N. de Castro and F. J. Von Zuben. The Clonal Selection Algorithm with Engineering Applications. 2000.
11. L. N. de Castro and F. J. Von Zuben. *Data Mining: A Heuristic Approach*, book chapter/section aiNet: An, pages 231–259. Idea Group Publishing, 2001.
12. David L Donoho. For Most Large Underdetermined Systems of Linear Equations the Minimal 1 -norm Solution is also the Sparsest Solution. *Statistics*, 40698:1–28, 2004.
13. D.L. Donoho, I. Drori, Y. Tsaig, and J.L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, 2006.
14. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–451, 2004.
15. J. D. Farmer, N. H. Packard, and A. S. Perelson. The Immune System, Adaptation and Machine Learning. *Physica*, 22:187–204, 1986.
16. S. Forrest, B. Javornik, R.E. Smith, and A.S. Perelson. Using genetic algorithms to explore pattern recognition in the immune system. *Evolutionary computation*, 1(3):191–211, 1993.
17. A.A. Freitas and J. Timmis. Revisiting the Foundations of Artificial Immune Systems for Data Mining. *Evolutionary Computation, IEEE Transactions on*, 11(4):521–540, 2007.
18. Antonio A. Freitas and Benedita Rocha. Population Biology of Lymphocytes: The Flight for Survival. *Annual Review of Immunology*, 18:83–111, 2000.
19. Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2001.
20. J.H. Friedman and J.W. Tukey. A Projection Pursuit Algorithm for Exploratory Data Analysis. *Computers, IEEE Transactions on*, C-23(9):881–890, August 2006.
21. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2001.
22. Kalet Leon, Rolando Perez, Agustin Lage, and Jorge Carneiro. Modelling T-Cell-Mediated Suppression Dependent on Interactions in Multicellular Conjugates. *Journal of Theoretical Biology*, 207:231–254, 2000.
23. R Levins. Evolution in changing environments, 120 pp. *Princeton Univ. Press, Princeton, NJ*, 1968.

24. A.J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences of the United States of America*, 6(7):410, 1920.
25. S.G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 41(12):3397–3415, 1993.
26. P. Matzinger. Tolerance, danger, and the extended family. *Annual review of immunology*, 12(1):991–1045, 1994.
27. Chris McEwan and Emma Hart. On AIRS and Clonal Selection for Machine Learning. In *Proceedings of 8th Annual Conference in Artificial Immune Systems (ICARIS)*. Springer, 2009.
28. Chris McEwan and Emma Hart. Representation in the (Artificial) Immune System. *Journal of Mathematical Modelling and Algorithms*, 8:125–149, 2009.
29. B. K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, 24(2):227, 1995.
30. J. Newborough and S. Stepney. A generic framework for population-based algorithms, implemented on multiple FPGAs. In *Artificial Immune Systems*, page 43–55. Springer, 2005.
31. Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press of Harvard University Press, September 2006.
32. K Page and M Nowak. Unifying Evolutionary Dynamics. *Journal of Theoretical Biology*, 219(1):93–98, November 2002.
33. Y.C. Pati, R. Rezaifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44 vol.1, 1993.
34. A. Roberts. The stability of a feasible random ecosystem. *Nature*, 251:607–608, 1974.
35. T. Stibor and J. Timmis. An Investigation on the Compression Quality of aiNet. In *Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium on*, pages 495–502, 2007.
36. E.R. Stirk, C. Molina-Paris, and H.A. Van Den Berg. Stochastic niche structure and diversity maintenance in the T cell repertoire. *Journal of Theoretical Biology*, 255:237–249, 2008.
37. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, page 267–288, 1996.
38. Joel A Tropp. Just relax: convex programming methods for subset selection and sparse approximation, 2004.
39. Francisco J. Varela and Antonio Coutinho. Second generation immune networks. *Immunology Today*, 12(5):159–166, 1991.
40. V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science*, 3(1):3, 1928.
41. Andrew Watkins, Jon Timmis, and Lois Boggess. Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm. *Genetic Programming and Evolvable Machines*, 5(3):291–317, September 2004.
42. K.W. Wucherpfennig, P.M. Allen, F. Celada, I.R. Cohen, R. De Boer, K.C. Garcia, B. Goldstein, R. Greenspan, D. Hafler, P. Hodgkin, and Others. Polyspecificity of T cell and B cell receptor recognition. *Seminars in immunology*, 19(4):216–224, 2007.