International Conference on Computational Science, ICCS 2013

Immunological-based approach for accurate fitting of 3D noisy data points with Bézier surfaces

Akemi Gálvez\textsuperscript{a}, Andrés Iglesias\textsuperscript{a,b,*}, Andreina Avila\textsuperscript{a}

\textsuperscript{a}Department of Applied Mathematics and Computational Sciences, University of Cantabria, E.T.S.I. Caminos, Canales y Puertos, Avda. de los Castros, s/n, E-39005, Santander, Spain
\textsuperscript{b}Department of Information Sciences, Faculty of Sciences, Toho University, 2-2-1 Miyama, 274-8510, Funabashi, Japan

Abstract

Free-form parametric surfaces are common tools nowadays in many applied fields, such as Computer-Aided Design & Manufacturing (CAD/CAM), virtual reality, medical imaging, and many others. A typical problem in this setting is to fit surfaces to 3D noisy data points obtained through either laser scanning or other digitizing methods, so that the real data from a physical object are transformed back into a fully usable digital model. In this context, the present paper describes an immunological-based approach to perform this process accurately by using the classical free-form Bézier surfaces. Our method applies a powerful bio-inspired paradigm called Artificial Immune Systems (AIS), which is receiving increasing attention from the scientific community during the last few years because of its appealing computational features. The AIS can be understood as a computational methodology based upon metaphors of the biological immune system of humans and other mammals. As such, there is not one but several AIS algorithms. In this chapter we focus on the clonal selection algorithm (CSA), which explicitly takes into account the affinity maturation of the immune response. The paper describes how the CSA algorithm can be effectively applied to the accurate fitting of 3D noisy data points with Bézier surfaces. To this aim, the problem to be solved as well as the main steps of our solving method are described in detail. Some simple yet illustrative examples show the good performance of our approach. Our method is conceptually simple to understand, easy to implement, and very general, since no assumption is made on the set of data points or on the underlying function beyond its continuity. As a consequence, it can be successfully applied even under challenging situations, such as the absence of any kind of information regarding the underlying function of data.

Keywords: Data fitting; Bézier surfaces; immunological approach; artificial immune systems; reverse engineering

1. Introduction

In several domains such as computer aided-design and manufacturing (CAD/CAM), data usually come from real measurements of an existing geometric entity, as it typically happens in the construction of car bodies, ship hulls, airplane fuselage and other free-form objects [2, 9, 11, 13, 14, 20, 21, 27, 28]. It also appears in the shoes industry, in archeology (reconstruction of archeological assets), in medicine (computer tomography) and many others. The primary goal is to convert the real data from a physical object into a fully usable digital model, a process called reverse engineering. This allows significant savings in terms of storage capacity and processing and

*Corresponding author. Tel.: +34-942-202-062; fax: +34-942-201-703.
E-mail address: iglesias@unican.es.
manufacturing time. Furthermore, the digital models are easier and cheaper to modify than their real counterparts, and are available anytime and anywhere. These data points are usually acquired through laser scanning and other digitizing methods and are, therefore, subjected to some measurement noise and other artifacts. Consequently, a good fitting of data is generally based on approximation schemes rather than on interpolation. This is also the approach taken in this work.

The usual models for data fitting in CAD/CAM and other industrial fields are free-form parametric entities, such as Bézier, B-spline and NURBS, and best approximation methods make commonly use of least-squares techniques [2, 12, 25, 26, 28, 29, 30]. In this paper we focus particularly in the case of Bézier surfaces, where the goal is to obtain the control points of the approximating surface that fits the data points better in the least-squares sense. This problem is far from being trivial: because the surface is parametric, we are confronted with the problem of obtaining a suitable parameterization of the data points [8, 26]. As remarked in [1], the selection of an appropriate parameterization is essential for a good fitting.

Some recent papers have shown that the application of Artificial Intelligence (AI) techniques can achieve remarkable results regarding this parameterization problem [9, 17, 18, 19, 20, 21, 24]. Most of these methods rely on some kind of neural networks, such as standard neural networks [17] and Kohonen's SOM (Self-Organizing Maps) nets [18, 19]. In the case of surfaces, the network is used exclusively to order the data and create a grid of control vertices with quadrilateral topology [19]. After this preprocessing step, any standard surface reconstruction method has to be applied. The generalization of these techniques to the (more general) functional networks is analyzed in [3, 9, 20, 21, 22, 23]. The work in [10] describes the application of genetic algorithms and functional networks to this problem. Other papers reporting pretty good results for data fitting with free-form curves are given in [29, 30] for genetic algorithms (GA), in [12] for particle swarm optimization (PSO), in [31] for estimation of distribution algorithms, and in [16] for artificial immune systems. These works have been extended to the case of free-form polynomial surfaces in [13] by using genetic algorithms iteratively, then in [14] to free-form rational surfaces through particle swarm optimization, and finally in [15] to surfaces in manufacturing by using a hybrid GA-PSO approach.

Our strategy for tackling the problem also belongs to this group of AI techniques. In this paper we analyze the application of a powerful bio-inspired paradigm called Artificial Immune Systems (AIS), which is receiving increasing attention from the scientific community during the last few years because of its appealing computational features [4]. The AIS can be understood as a computational methodology based upon metaphors of the biological immune system of humans and other mammals. As such, there is not one but several AIS algorithms. In this paper we focus on the clonal selection algorithm (CSA), which explicitly takes into account the affinity maturation of the immune response. This algorithm is applied to to the accurate fitting of 3D noisy data points with Bézier surfaces.

The structure of this chapter is as follows: in Section 2 we provide a gentle overview about the artificial immune systems and their main features and advantages. The section also describes the fundamentals of the clonal selection algorithm in detail. Section 3 introduces the problem to be addressed along with the proposed method to solve it. Then, some simple yet illustrative examples of its application along with some implementation details are reported in Section 4. The paper closes with the main conclusions of this contribution and our plans for future work in the field.

2. Artificial Immune Systems

In recent years there has been an increasing interest in the development of new computational algorithms for optimization inspired in biological systems. Amongst them, the Artificial Immune Systems (AIS) are receiving increasing attention from the scientific community because of their ability to solve complex optimization problems in several fields. Roughly speaking, AIS are a group of computationally intelligent systems inspired by the principles and processes that typically happen at the level of the immune system of humans and other vertebrate.

The immune system is a complex network composed of specialized cells, tissues and organs and is responsible for protecting the organism against diseases caused by pathogenic agents. It is comprised of two distinct parts: the innate immune system and the adaptive immune system. The former is responsible for powerful immediate but non-specific defenses that prevent or limit infections by most pathogenic microorganisms, called antigens and represented in this paper as Ag. If pathogens successfully evade the innate response, there is a second layer of
protection, the adaptive immune system, which is activated by the innate response. The adaptive immune system uses somatically generated antigen receptors. The general design of the adaptive immune response is based upon the clonal selection of lymphocytes expressing receptors with particular specificities. Through this second line of defense, the immune system adapts its response during an infection to improve its recognition of the pathogen. This improved response is preserved even after the pathogen has been eliminated so that any future attack of this pathogen finds a much faster and stronger answer over the time. In other words, the adaptive immune system has some kind of immunological memory, a valuable feature that can be extended to AIS.

There are also some other properties of the immune system of potential interest for computer scientists and engineers [5]:

- **uniqueness**: each individual possesses its own immune system, with its particular features and vulnerabilities;
- **recognition of foreigners**: the molecules that are not native to the body are recognized and eliminated by the immune system;
- **anomaly detection**: the immune system can detect and react to pathogens that the body has never encountered before;
- **distributed detection**: the cells of the system are distributed all over the body and, most importantly, are not subject to any centralized control;
- **imperfect detection**: an absolute recognition of the pathogens is not required, hence the system is flexible;
- **reinforcement learning and memory**: the system can “learn” the structures of pathogens, so that future responses to the same pathogens are faster and stronger.

2.1. The clonal selection algorithm

In 2000, De Castro and Von Zuben [5, 6, 7] proposed the Clonal Selection Algorithm (CSA), an AIS method based on the clonal selection principle, a widely accepted theory used to explain the basic features of an adaptive immune response to an antigenic stimulus. When we are exposed to an Ag, our immune system responds by producing antibodies (represented by Ab onwards). Ab’s are molecules attached primarily to the surface of B cells whose aim is to recognize and bind to Ag’s. Each B cell secretes a single type of Ab, which is relatively specific for the Ag. Under the clonal selection theory, only those cells that recognize the antigens are selected to proliferate. The selected cells are subject to an affinity maturation process, which improves their affinity to the selective Ag’s over the time.

In the immune system, the learning process involves raising the relative population size and affinity of those lymphocytes that have been valuable in recognizing a given Ag. For practical problems, it is convenient to keep a small set of best individuals rather than using a large number of candidate solutions. A clone will be created temporarily and those progenies with low affinity will be discarded. Typically, an organism would be expected to encounter a given Ag repeatedly during its lifetime. The initial exposure to an Ag that stimulates an adaptive immune response is handled by a small number of low-affinity B cells, each producing an Ab type of different affinity. The effectiveness of the immune response to secondary encounters is enhanced considerably by the presence of memory cells associated with the first infection, capable of producing high-affinity Ab’s just after subsequent encounters.

Ab’s in a memory response have, on average, a higher affinity than those of the early primary response. This phenomenon, which is restricted to T-cell-dependent responses, is referred to as the maturation of the immune response. This maturation requires the Ag-binding sites of the Ab molecules to be structurally different from those present in the primary response. Then, random changes (mutation) are introduced into the genes responsible for the Ag-Ab interactions, leading occasionally to an increase in the affinity of the Ab. A rapid accumulation of mutations is necessary for a fast maturation of the immune response, even although most of the changes lead to poorer Ab’s. Indeed, when a B cell recognizes an antigen, it is stimulated to divide (or proliferate). During proliferation, the B cell receptor locus undergoes an extremely high rate of somatic mutation that is at least five or six orders of magnitude greater than the normal rate of mutation across the genome. This process, referred to as somatic hypermutation (SHM), is regulated by the fact that cells with low-affinity receptors may be further mutated and die if they do not improve their clone size or antigenic affinity, while in cells with high-affinity Ab receptors, hypermutation may become gradually inactive.
Fig. 1. Flow chart of the clonal selection algorithm as described in [5, 7] for pattern recognition purposes.

Fig. 1 shows the flowchart of the original clonal selection algorithm, as described in [5, 7]. The algorithm considers two repertoires (populations): a set of antigens $A_{\text{org}}$ and a set of antibodies $A_{\text{Ant}}$. For the sake of clarity, cardinality is indicated by the subindexes within brackets. The latter is further divided into two subsets: memory Ab repertoire $A_{\text{Mem}}$ and remaining Ab repertoire $A_{\text{Rem}}$, such that $m + r = N$. We also keep track of two other sets: the set $A_{\text{High}}$ of the $n$ Ab’s with the highest affinities to a given Ag, and the set $A_{\text{New}}$ of the $d$ new Ab’s that will replace the low-affinity Ab’s from $A_{\text{Rem}}$. The algorithm can be summarized as follows:

1. Random choice of an antigen $A_{\text{org}}$. It is presented to all antibodies of $A_{\text{Ant}}$.
2. Compute the vector affinity $F = (F_1, F_2, \ldots, F_N)$ where $F_i = Af(A_{\text{Mem}}, A_{\text{org}})$.
3. Select the $n$ highest affinity components of $F$ to generate $A_{\text{High}}$.
4. Elements of $A_{\text{High}}$ will be cloned adaptively. The number of clones is proportional to the affinity: the higher the affinity, the higher the number of clones. Such amount is given by: $N_c = \sum_{h=1}^{n} \text{round}\left(\frac{\beta \cdot N}{h}\right)$ where $N_c$ represents the number of clones, $\beta$ is a positive number that plays the role of a multiplying factor, $N$ is the total number of Ab’s and $\text{round}(\cdot)$ is the operator that rounds its argument toward the closest integer.
5. The clones in the set resulting from the previous step are subjected to somatic hypermutation. The affinity maturation rate is inversely proportional to the antigenic affinity: the higher the affinity, the smaller the maturation rate.
6. Compute the vector affinity of $A_{\text{org}}$ with respect to the new matured clones.
7. From this set of matured clones, select the one with the highest affinity to be candidate to enter into the set $A_{\text{High}}$. If $Af(A_{\text{Mem}}, A_{\text{org}}) > Af(A_{\text{Rem}}, A_{\text{org}})$ for a given $A_{\text{Rem}} \in A_{\text{Rem}}$, then $A_{\text{Mem}}$ will replace $A_{\text{Rem}}$.
8. Replace the $d$ Ab’s with lowest affinity in $A_{\text{Rem}}$ by new individuals, inserted into $A_{\text{Rem}}$.

Each execution of the previous steps for all given Ag’s is called a generation. The algorithm is repeated for a certain number of generations, $N_{\text{gen}}$, a parameter that depends on the specific problem under analysis. This algorithm has proved to be very well suited for optimization problems, having been successfully applied to problems such as character recognition and multimodal optimization with very good performance.
3. Our Method

We assume that the reader is familiar with the main concepts of free-form parametric surfaces [8]. A free-form parametric Bézier surface of degree \((m,n)\) is defined as:

\[
S(u, v) = \sum_{i=0}^{m} \sum_{j=0}^{n} P_{i,j} B^m_i(u) B^n_j(v)
\]

where \(P_{i,j}\) are vector coefficients in \(\mathbb{R}^3\) (usually referred to as the control points), \(B^d_k(t)\) are the Bernstein polynomials of index \(k\) and degree \(d\), given by:

\[
B^d_k(t) = \binom{d}{k} t^k (1-t)^{d-k}
\]

where

\[
\binom{d}{k} = \frac{d!}{k! (d-k)!}
\]

and \(u, v\) are the surface parameters, defined on the finite interval \([0, 1]\). Note that in this paper vectors are denoted in bold. By convention, \(0! = 1\).

Let us suppose that we are given a set of data points \(\{Q_{i,j}\}_{i=1,...,M; j=1,...,N}\). The goal is to obtain the Bézier parametric surface \(S(u, v)\) that fits the data points better in the discrete least-squares sense. To do so, we have to compute the control points of the approximating surface by minimizing the least-squares error, \(E\), defined as the sum of squares of the residuals:

\[
E = \sum_{i=1}^{M} \sum_{j=1}^{N} \left( Q_{i,j} - \sum_{i=0}^{m} \sum_{j=0}^{n} P_{i,j} B^m_i(u_i) B^n_j(v_j) \right)^2
\]

where we need to obtain the parametric values \((u_i, v_j)\) to be associated with each data point \(Q_{i,j}\). Due to the fact that the blending functions \(B^m_i(u)\) and \(B^n_j(v)\) are nonlinear in \(u\) and \(v\) respectively, the least-squares minimization of the errors is a strongly nonlinear problem, with a high number of unknowns for large sets of data points, a case that happens very often in practice. Our strategy for solving the problem in the general case consists of applying the CSA methodology to determine suitable parameter values for the data points, and then calculating the best least-squares fitting coefficients. To this purpose, the affinity measure function corresponds to the evaluation of the least-squares function, given by Eq. (2). Similarly, in our formulation each antibody represents a potential solution of the problem.

However, in order to apply the clonal selection algorithm to our problem, some modifications on the original CSA are needed. First of all, the original CSA algorithm was intended for supervised problems, where the target solution is already known \textit{a priori}, as it happens, for instance, in pattern-recognition problems. Clearly, this is not our case, since no information about the approximating surface is known in advance. On the other hand, there is no need to maintain the subset of memory Ab’s, since no specific Ag has to be recognized (note that our problem is an example of unsupervised learning, as we do not know the optimal solution for this problem \textit{a priori}). Finally, several Ab’s with high affinity are selected in step 7 of our algorithm, rather than just a single one, with the effect of increasing the convergence rate.

With these modifications, the CSA algorithm can now be applied to our problem. However, before searching a solution to the problem, some control parameters should be set up. These control parameters are:

- the number of control points for each parametric direction \(u\) and \(v\), given by \(m\) and \(n\), respectively
- the total number of antibodies, \(N_{ab}\)
- the number of antibodies, \(d\), to be replaced in \(Ab_{[d]}\)
- the number of antibodies to be cloned, \(n_{best}\)
- the number of clones for the selected antibodies, \(N_c\)
- the somatic hypermutation rate, and
• the number of iterations (generations), $N_{\text{iter}}$.

These parameters are selected empirically. To this aim, we perform a number of executions for different values of the parameters and then, select the best ones to be used in our final experiments, as reported in next section. After the selection of those parameters, the CSA is executed for the prescribed number of iterations. The process is performed iteratively while the evolution of the parameters does not stabilize the minimization of the error. The antibody with the best (i.e., minimum) affinity value is selected as the best solution to the problem.

4. Experimental results

This section discusses the performance of our CSA-based method described above through three simple yet illustrative examples. Many other examples have also been tested with excellent results in all cases. They are not reported here, however, because of limitations of space.

First example corresponds to a set of 2500 data points arranged in a matrix of $50 \times 50$ elements. They are fitted by using a Bézier surface of degree $(m, n) = (3, 3)$. The other parameters are as follows: $N_{\text{ab}} = 100$, $d = 10$, $n_{\text{best}} = 10$, and $N_{\text{iter}} = 3$. The number of clones was taken proportionally to the affinity measure, such that the antibodies with better affinity would yield more clones. In particular, we select 60 clones for the antibody with the best affinity, 40 clones for the second best, 20 clones for the third best and 10 for the other ones. The total number of clones thus becomes $N_c = 190$. These clones then undergo real-valued somatic hypermutation at a rate that is inversely proportional to the affinity.

Our results for these parameter values are depicted in Figure 2, where the original data points are displayed on the left as small red spheres whereas the reconstructed surface is displayed on the right. Note from Fig. 2(left) that our data points are noisy. This implies that, even with an initial choice of $50 \times 50$ data points, they are not perfectly aligned on iso-parametric lines; therefore, the strategy of computing approximating Bézier curves for the iso-parametric lines is not applicable here. Instead, we focus on the set of data points as a whole and seek for the approximating surface in the sense of least-squares overall. Note also how the method yields an approximating surface that reproduces the shape of the cloud of data points with high visual quality. In fact, we got a RMSE fitting error of $7.83 \times 10^{-4}$ for this example, indicating that the reconstructed Bézier surface fits the data points with high accuracy.

![Fig. 2. Applying the clonal selection algorithm to Bézier surface fitting of data in Example 1: (left) Original data points; (right) approximating Bézier surface of degree (3, 3).](image)

Second example consists of a set of 3600 data points obtained from an explicit algebraic surface of equation $z = x^3 - 3xy^2$, usually known as the monkey saddle surface. This surface has an isolated umbilical point with zero Gaussian curvature at the origin, and several changes of curvature accounting for three depressions. This makes it a very good candidate to evaluate the behavior of our approach. To check the robustness of our method against the noise, the initial data points have been perturbed by an additive uniform random noise of mean 0 and variances $\sigma_x = 0.15$, $\sigma_y = 0.15$, and $\sigma_z = 0.6$ for the three coordinates, respectively. The perturbed data points are then
fitted by using a Bézier surface of degree \((m, n) = (4, 4)\). The other parameters are similar to those used in the previous example and, hence, are not reported here to avoid unnecessary duplication of material.

Our results for this example are depicted in Figure 3. The figure on the left shows the original data points displayed as red emptied diamonds whereas the figure on the right shows the reconstructed Bézier surface. Once again, we obtained an excellent matching between the original data points and the reconstructed points from the approximating surface. In this case we obtained a RMSE fitting error of \(3.39 \times 10^{-3}\). From it, we conclude that the reconstructed Bézier surface fits the data points extremely well. We also computed the mathematical properties of the fitting surface and found that it exhibits the same properties as the original surface at the origin. In particular, we also obtained the umbilical point of zero Gaussian curvature at that point. Besides, that point is not a saddle point, because it does not behave as a minimum and/or maximum for different section planes through the origin (i.e. sections of equation \(y = k \times x, k \in \mathbb{R}\)), but as an inflection point instead. This behavior is also observed in the reconstructed surface, meaning that our approach is amazingly reliable in capturing these subtle details of the original surface. Despite of the fact that the original explicit algebraic surface is approximated by a free-form parametric surface, we obtained similar analytical results for the approximating surface (the corresponding calculations are out of the scope of this paper and are, therefore, not included here for brevity). From this, we also conclude that the approximating surface is a very good model of the original one.

The two previous examples have a similar complexity and sampling rate at both parametric directions. Our last example is aimed at checking the behavior of our method when this condition no longer applies. Now, we consider a set of 2400 data points arranged in a matrix of \(40 \times 60\) elements. These data points are obtained from a surface exhibiting a qualitatively different behavior for different parametric directions, therefore, requiring a different degree for each direction for an optimal fitting of data. In particular, we consider a Bézier surface fitting of degree \((m, n) = (5, 7)\). The other parameters of the method are similar to those of previous examples.

Figure 4 shows our results: on the left, we display the original data points as small red spheres; on the right, the reconstructed Bézier fitting surface is displayed. A simple visual inspection of the surface shows that it has a complicated shape, with several valleys and hills and very oscillating surface boundaries. Still, similarly to the previous examples, we also obtain an excellent matching between the original and the reconstructed data points, with a RMSE fitting error of \(5.26 \times 10^{-4}\). This shows that our method is able to yield a very good fitting surface even in the cases of surfaces with asymmetric behavior along different parametric directions.

4.1. Implementation details

All computations in this paper have been performed on a 2.4 GHz. Intel Core 2 Duo processor with 4 GB. of RAM. The source code has been implemented by the authors in the native programming language of the popular
Fig. 4. Applying the clonal selection algorithm to Bézier surface fitting of data in Example 3: (left) Original data points; (right) approximating Bézier surface of degree (5, 7).

scientific program Matlab, version 2010b. In our opinion, Matlab is a very suitable tool for this task: it is fast and provides reliable, well-tested routines for efficient matrix manipulations. It also contains a bulk of resources regarding the solving of systems of equations. In this paper, Matlab has been primarily used since it provides the best numerical answer in the sense of least-squares for those cases in which the exact solution of the systems of equations is not possible. To this purpose, Matlab provides us with the command mldivide to solve the least-squares equation for our resulting over-determined systems.

Regarding the computation times, they obviously depend on many factors, such as the degree of the approximating surface, the size of the sets of data points, the noise intensity and the selection of parameters for the method, making it hard to determine in advance how long does it take for given example to be obtained. However, in our trials we found our method to be very affordable in terms of CPU times: all simulations in this paper took less than 30 seconds to be obtained.

5. Conclusions and Future Work

This paper addresses the problem of obtaining a good fitting surface to a given set of 3D noisy data points. This problem arises very often in many industrial environments such as in reverse engineering for CAD/CAM. In this paper, input data points are fitted by using a Bézier surface of a given degree, a problem far from being trivial as soon as no parameterization of data points is assumed a priori. In our approach the parameterization is computed by using a popular bio-inspired methodology called artificial immune systems. Rather than a single method itself, AIS is a computational paradigm supporting different techniques inspired by the behavior of the biological immune system of humans and other vertebrate. In this work, we focus on the clonal selection algorithm, arguably the most popular AIS approach. This algorithm is applied to determine the Bézier surface that fits the given data points better in the least-squares sense. The method is discussed through its application to three illustrative examples. Our experimental results show that the method is able to yield an extremely accurate fitting of data points in all cases. Furthermore, this high accuracy does not come at the price of large computation times. In fact, from our experiments we found that our method is very competitive regarding the CPU time, although not well suited for real-time computations.

Our future work includes the extension of this method to other families of parametric surfaces, such as the B-splines and NURBS, where the existence of additional parameters (such as knots and weights) can modify our procedure significantly. We also would like to carry out a detailed theoretical analysis about some important issues related to our approach, such as the convergence of the method and the parameter tuning. The application of this method to some interesting real-world problems in industrial settings (such as blending, offsetting, and surface intersection) is also part of our plans for future work. Finally, we are also interested in the hybridization of this method with other optimization strategies in order to further improve the general performance of the method and the computation times.
Acknowledgements

This research has been kindly supported by the Computer Science National Program of the Spanish Ministry of Economy and Competitiveness, Project Ref. #TIN2012-30768, and the University of Cantabria (Santander, Spain). This work has been done during the sabbatical stay of second author at Toho University (Funabashi, Japan), whose academic support has been very valuable to produce this paper. We would also like to thank the three anonymous reviewers for their encouraging comments which helped us to improve the first version of the paper.

References