

## DYNAMIC STATISTICAL CLASSIFICATION

## CLASIFICACIÓN ESTADÍSTICA DINÁMICA

JAVIER PULIDO–CEJUDO\*    CARLOS CUEVAS–COVARRUBIAS†

*Received: 3/Mar/2016; Revised: 25/Aug/2016;  
Accepted: 7/Oct/2016*

---

*Revista de Matemática: Teoría y Aplicaciones* is licensed under a Creative Commons  
Reconocimiento-NoComercial-Compartirigual 4.0 International License.  
Creado a partir de la obra en <http://www.revistas.ucr.ac.cr/index.php/matematica>



\*CIEMA, Universidad Anáhuac, México D.F., México.

E-Mail: [javierpulidocejudo@gmail.com](mailto:javierpulidocejudo@gmail.com)

†Misma dirección que/Same address as: J. Pulido. E-Mail: [ccuevas@anahuac.mx](mailto:ccuevas@anahuac.mx)

### Abstract

We consider the statistical supervised classification problem from a dynamical systems approach. We assume that two classes exist and that, for each one, a multivariate normal distribution determines the probability to be in a certain region in the  $n$  dimensional real vector space. These density functions are the potentials of corresponding gradient vector fields for each class; we construct a “classifying vector field” as a suitable weighted mean of them. From data known in the literature, we estimate the population parameters, and the classes are successfully distinguished; we compute and present confusion matrices. A one and two-dimensional analysis is given.

**Keywords:** supervised statistical classification; multivariate normal distribution; vector fields; attractors; bifurcation; dynamical systems.

### Resumen

Consideramos el problema de la clasificación estadística supervisada desde la perspectiva de los sistemas dinámicos. Suponemos la existencia de dos clases cuyas poblaciones se distribuyen como normales en un espacio real de dimensión  $n$ . Estas funciones de densidad son los potenciales de los correspondientes campos gradientes de cada clase; construimos el “campo vectorial clasificante” como un campo promedio con pesos apropiados. A partir de datos bien conocidos en la literatura, estimamos los parámetros poblacionales y las clases son clasificadas exitosamente. Un análisis en dimensiones uno y dos es presentado.

**Palabras clave:** Clasificación estadística supervisada; distribución normal multivariada; campos vectoriales; atractores; bifurcación; sistemas dinámicos

**Mathematics Subject Classification:** 37B25, 62G32.

## 1 Introduction

The present work introduces a new method to tackle the statistical classification problem, based on the idea by René Thom [7], that any concept is an attractor, we define a suitable chosen vector field which, in fact, has attractors close to the means of each population that is intended to be classified. Each attractor has a basin of attraction, when data has values in each of those basins, the classifying vector field will yield to the class it supposedly belongs. Though we can apply this technique to problems with any number of variables, and for any number of classes, we only consider in this paper two classes and dimensions one or two. Basins have boundaries that separate data, and in our analysis those

boundaries become apparent, they correspond to usual phase portraits in one and two dimensions.

Since it is not common to use the theory of dynamical systems in this context of statistical classification, a brief introduction of the concepts used is given and a good reference for this discipline is the book [4] (Hirsch, Smale, & Devaney, 2013). For a thorough treatment of attractors the paper [6] (Milnor, 1985) is an excellent reference for the matter, in that paper anyone pursuing more references will find them. For all matters classical in statistical classification, we refer the reader to [2] (Duda, Hart, & Stork, 2001) and [8] (Trejos, Castillo, & González, 2014).

We treat classification in dimension one for two classes, we analyze this situation for two normal populations in a general setting, and a picture of the power of this technique is apparent. A different insight of what happens when the relative variance of one class is much bigger than the other, suggests that no classification is possible since all data is attracted to just one class, this is a new approach that may lead to distinguishing among classifiable and none situations.

In the two dimensional case, we do not give a general result of the kind we do in the case of one dimension. Nevertheless, we apply the method to popular data in this field [1] (Campbell & Mahon, 1974), and we analyze the phase portrait we get by using this method. We obtain a different insight and the boundary of the two basins, called a separatrix, appears in a natural way as the stable manifold of a hyperbolic point. In addition, we use data from the pioneer paper [3] (Fisher, 1936) and present the confusion matrix for our classification method. We can extend the approach in numerous ways in which not only in the parametric case, but also in the non-parametric one, we will get a better insight of the difficulties and boundaries of the classification attempt, though that is not included in this paper.

We want to express our gratitude to the Research Office of the Anahuac University for sponsoring this project. The authors are also indebted to Luis Calderón Contreras for his useful help in the preparation of this paper.

## 2 Concepts and notation

Before we introduce this classifying method, we present a brief introduction to concepts and notation common in both disciplines, dynamical systems and statistics.

A vector field is a vector valued function with a vector domain in the  $n$  real vector space  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , in this context we consider differential vector fields which define a one parameter group of diffeomorphisms  $\Phi : A \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,

where  $A \subset \mathbb{R}$  that contains 0 in such a way that the following properties hold for any  $t, s \in A, \bar{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  such that  $t + s \in A$ :

1.  $\Phi(t + s, \bar{x}) = \Phi(t, \Phi(s, \bar{x})) = \Phi(s + t, \bar{x})$ .
2.  $\Phi(0, \bar{x}) = \bar{x}$ .
3. if  $\Phi_t(\bar{x}) = \Phi(t, \bar{x})$  then  $\Phi'_t(\bar{x}) = \frac{\partial \Phi(t, \bar{x})}{\partial t} = F(\bar{x})$ .

The curves  $\Phi_t(\bar{x})$  are the integral curves or orbits of  $F$  and their importance in this paper is crucial since they are the classifying vehicle.

Given a differentiable real valued function,  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , we construct the gradient vector field

$$\nabla f(x_1, \dots, x_n) = \left( \frac{\partial f}{\partial x_1}(x_1, \dots, x_n), \dots, \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) \right)$$

in this context, we call  $f$  the potential of  $\nabla f$ . When possible, a phase portrait will be produced where the integral curves of the vector field will make apparent the past and future of the evolution that the vector field yields for every point  $x$  in  $\mathbb{R}^n$ .

A usual way to determine how the phase portrait may be is to find the isolated zeros of the vector field, linearize it around each zero and determine its quality by spectral analysis, that is, its eigenvalues will determine whether it is a sink (attractor), a source (repeller), a hyperbolic point or none of them.

Given the vector field  $F = (F_1, \dots, F_n)$  where each  $F_i$  is a differentiable real valued function that vanishes only at  $\bar{x}_0$  in a neighborhood of it, then the  $n$  by  $n$  matrix  $L_{\bar{x}_0} = \left( a_{ij} = \frac{\partial F_i}{\partial x_j}(\bar{x}_0) \right)$  determines the linear transformation that approximates  $F = (F_1, \dots, F_n)$  near  $\bar{x}_0$  and the eigenvalues of  $L_{\bar{x}_0}$  determine qualitatively the decomposition of a neighborhood of  $\bar{x}_0$  in  $\mathbb{R}^n$  in orbits that are attracted by  $\bar{x}_0$ , orbits that are repelled by  $\bar{x}_0$  or none. The number, counting multiplicity, of negative eigenvalues and those whose real part is negative, in case they are complex, determine the dimension of the subset of  $\mathbb{R}^n$  whose orbits, close to  $\bar{x}_0$ , are attracted by  $\bar{x}_0$ ; the number of positive eigenvalues and, those whose real part is positive, determine the dimension of the orbits repelled by  $\bar{x}_0$ .

Since in this paper we will only consider the one and two dimensions, phase portrait of vector fields with isolated zeros are reasonable easy to achieve.

Consider two populations and suppose that they distribute as two normal random variables  $\bar{X}_i = (X, Y) \in \mathbb{R}^2$  with means  $\bar{\mu}_i = (\mu_{iX}, \mu_{iY})$ , correlation coefficients  $\rho_i$ , as well as covariances matrices

$$V_i = \begin{pmatrix} \sigma_{iXX} & \sigma_{iXY} \\ \sigma_{iXY} & \sigma_{iYY} \end{pmatrix} = \begin{pmatrix} \sigma_{iXX} & \rho_i \sqrt{\sigma_{iXX} \sigma_{iYY}} \\ \rho_i \sqrt{\sigma_{iXX} \sigma_{iYY}} & \sigma_{iYY} \end{pmatrix}.$$

For  $i = 1, 2$  in such a way that their probability density functions are

$$f_i(\bar{x}_i) = \frac{1}{2\pi\sqrt{\det(V_i)}} \exp\left(\frac{-(\bar{x}_i - \bar{\mu}_i)^t V_i^{-1} (\bar{x}_i - \bar{\mu}_i)}{2}\right) \quad (1)$$

for every  $\bar{x}^t = (x, y) \in \mathbb{R}^2$  where  $\bar{x}^t$  is the transponse of  $\bar{x}$ .

Since  $\sqrt{\det(V_i)} = \sqrt{\sigma_{iXX}\sigma_{iYY}(1 - \rho_i^2)}$  we can also express the density.

In order to get a more compact expression we will denote:

$$a_i = \frac{1}{2\pi\sigma_{iXX}\sigma_{iYY}\sqrt{1-\rho_i^2}},$$

$$q_i = \frac{1}{(1-\rho_i^2)} \left[ \left( \frac{x-\mu_{iX}}{\sqrt{\sigma_{iXX}}} \right)^2 - 2\rho_i \left( \frac{x-\mu_{iX}}{\sqrt{\sigma_{iXX}}} \right) \left( \frac{y-\mu_{iY}}{\sqrt{\sigma_{iYY}}} \right) + \left( \frac{y-\mu_{iY}}{\sqrt{\sigma_{iYY}}} \right)^2 \right] \text{ hence}$$

$f_i(x, y) = a_i \exp\left(-\frac{q_i}{2}\right)$  and their partial derivatives are

$$\begin{aligned} f_{ix}(x, y) &= \frac{-a_i}{1-\rho_i^2} \left( \frac{x-\mu_{iX}}{\sigma_{iXX}} - \rho_i \frac{y-\mu_{iY}}{\sqrt{\sigma_{iXX}\sigma_{iYY}}} \right) \exp\left(-\frac{q_i}{2}\right) \\ &= \frac{-1}{1-\rho_i^2} \left( \frac{x-\mu_{iX}}{\sigma_{iXX}} - \rho_i \frac{y-\mu_{iY}}{\sqrt{\sigma_{iXX}\sigma_{iYY}}} \right) f_i(x, y), \end{aligned}$$

$$\begin{aligned} f_{iy}(x, y) &= \frac{-a_i}{1-\rho_i^2} \left( \frac{y-\mu_{iY}}{\sigma_{iYY}} - \rho_i \frac{x-\mu_{iX}}{\sqrt{\sigma_{iXX}\sigma_{iYY}}} \right) \exp\left(-\frac{q_i}{2}\right) \\ &= \frac{-1}{1-\rho_i^2} \left( \frac{y-\mu_{iY}}{\sigma_{iYY}} - \rho_i \frac{x-\mu_{iX}}{\sqrt{\sigma_{iXX}\sigma_{iYY}}} \right) f_i(x, y). \end{aligned}$$

Hence, we can compute the gradients  $\nabla f_i = (f_{ix}, f_{iy})$  with the formulas above.

### 3 Main results

We will define the classifying vector field as

$$CF(\bar{x}) = \left( \frac{f_1(\bar{x})}{f_1(\bar{x}) + f_2(\bar{x})} \right) \nabla f_1(\bar{x}) + \left( \frac{f_2(\bar{x})}{f_1(\bar{x}) + f_2(\bar{x})} \right) \nabla f_2(\bar{x}). \quad (2)$$

We will use it below to classify from data in a two class setting, but before that, we will make an analysis of the behavior of this vector field in the one dimensional space, as well as in the two dimensional space, for some theoretical specific examples in order to get an insight of the classifying process.

### Analysis in dimension one

Let us suppose that we have two random variables in  $\mathbb{R}$ , that are  $N(2, \sigma^2)$  and  $N(0, 1)$  respectively, so that the corresponding density functions are

$$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-2)^2}{2\sigma^2}\right)$$

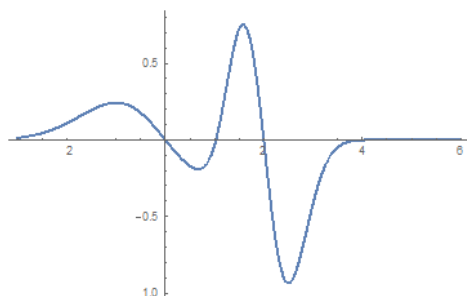
$$f_2(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right).$$

The gradients are, in this case, the respective derivatives:

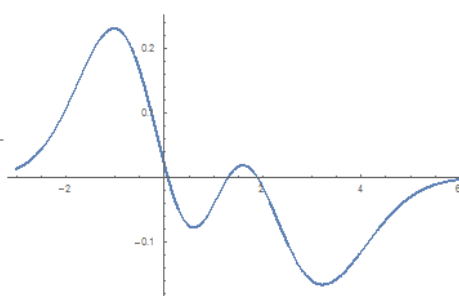
$$f_1'(x) = \frac{-(x-2)}{\sigma^3\sqrt{2\pi}} \exp\left(\frac{-(x-2)^2}{2\sigma^2}\right)$$

$$f_2'(x) = \frac{-x}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right).$$

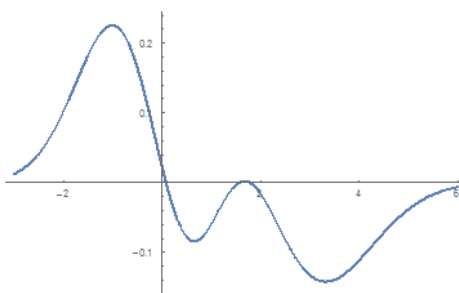
Our goal will be to find the isolated zeros of  $CF = \left(\frac{f_1}{f_1+f_2}\right) f_1' + \left(\frac{f_2}{f_1+f_2}\right) f_2'$  and classify them as sinks, sources or hyperbolic by the linearization method discussed in the last section; suppose that  $CF(x_0) = 0$ , the eigenvalue of the linear part in this case is  $CF'(x_0)$ , hence,  $x_0$  is a source if  $CF'(x_0) > 0$  and a sink  $CF'(x_0) < 0$ ; when it vanishes, it attracts on one side and repels on the other. When we vary  $\sigma^2$ , starting with small values for it and keep increasing them, we can do a graphic assisted analysis with images obtained with Mathematica. We chose  $\sigma = 0.25, 1.2, 1.3$  and obtained the following graphs 1,2,3.



**Figure 1:** Graph of  $CF$  for  $\sigma = 0.25$ .

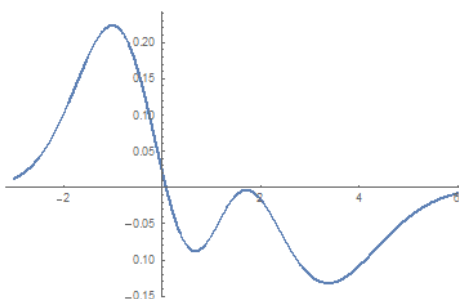


**Figure 2:** Graph of  $CF$  for  $\sigma = 1.2$ .



**Figure 3:** Graph of  $CF$  for  $\sigma = 1.3$ .

We can see that we get three zeros being, from left to right, sink, source, sink; these are: a point close to 0, a point between 0 and 2, and a point close to 2. The last two zeros approach one to each other as  $\sigma$  increases. Finally, see Figure 4, we graphed  $CF$  when  $\sigma = 1.35$ , in this case, the only zero is close to 0 and it is a source.



**Figure 4:** Graph of  $CF$  for  $\sigma = 1.35$ .

This suggests that the source and the sink coalesce as the graph becomes tangent to the  $x$  axis and as  $\sigma$  increases, only the sink at 0 persists.

Another interesting observation is that the source close to 0 starts as a mild attractor and becomes a stronger one since  $|CF'(x_0)|$  increases as  $\sigma$  increases. From the classifying point of view, this is a remarkable observation. We formulate and prove the following

**Proposition 1** *Suppose we have two real random variables  $N(\mu_i, \sigma_i^2)$  for  $i = 1, 2$  and  $\mu_1 < \mu_2$ . Each gradient,  $\nabla f_i$  of the density function  $f_i$ , has a sole zero at  $\mu_i$  and it is a sink. When  $\sigma_i$  is small, the classifying field  $CF$  has three zeros two of which are close to each  $\mu_i$  and are sinks, the third one is a source and lies in between the two others, as  $\sigma_2$  becomes relatively larger than  $\sigma_1$ , the sink close to  $\mu_2$  vanishes, after the coalescence of the sink, and the source.*

**Proof.** In order to prove the proposition we will, without loss of generality, suppose that  $\mu_1 = 0$  and  $\sigma_1 = 1$  since what is relevant is the ratio between the two variances. The equation  $CF = 0$  is transcendental; so we will do a qualitative analysis to prove the existence of solutions as well as the sign of their eigenvalues. Our first observation is that the densities intercept in two points

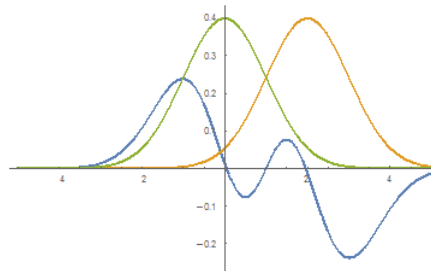
$$x_1 = \frac{-\mu_2 + \sigma_2 \sqrt{\mu_2^2 + 2(\sigma_2^2 - 1) \ln \sigma_2}}{\sigma_2^2 - 1} \quad (3)$$

$$x_2 = \frac{-\mu_2 - \sigma_2 \sqrt{\mu_2^2 + 2(\sigma_2^2 - 1) \ln \sigma_2}}{\sigma_2^2 - 1}. \quad (4)$$

The first one lies between 0 and  $\mu_2$ , the second one will be either greater than  $\mu_2$  or smaller than 0 according to whether  $\sigma_2$  is smaller or bigger than one respectively.

$CF = 0$  if and only if  $f_1 f'_1 + f_2 f'_2 = 0$ . We analyze this expression under the assumption that,  $\sigma_2 \leq 1$  and when  $\sigma_2 > 1$  afterwards.

Case  $\sigma_2 \leq 1$ . The signs of  $f_1, f'_1, f_2$  and  $f'_2$  are all positive when  $x < 0$  therefore  $f_1 f'_1 + f_2 f'_2 > 0$  nevertheless, when  $\mu_2 > x > 0, f'_1 < 0$ , where else, the signs of the other elements remain the same. Note that  $f_2$  and  $f'_2$  are very small relatively to  $f_1$  hence, as  $x$  increases  $f_1 f'_1 + f_2 f'_2 < 0$ . When  $x$  approaches  $x_1$ ,  $f_1$  and  $f_2$  become closer in value as well as the absolute values of  $f'_1$  and  $f'_2$ ; when  $x_1 + h^2 < x < \mu_2$ , for  $h$  small, we again have  $f_1 f'_1 + f_2 f'_2 > 0$ , by using similar arguments we will find that when  $\mu_2 + h^2 < x$ , for  $h$  small,  $f_1 f'_1 + f_2 f'_2 < 0$  and the sign will not change any more since both densities decrease when  $\mu_2 < x$ . When the variance is one the sole zero is  $\frac{\mu_2}{2}$  and the analysis is essentially the same. These all imply that we will end up with three zeros, from left to right, a sink, a source and a sink, see Figure 5.

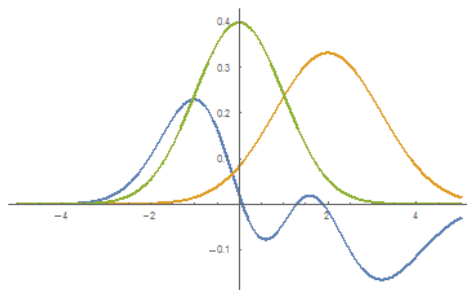


**Figure 5:** In orange the density of  $X_1$ , in green the density of  $X_2$  and CF in blue with  $\sigma$  small.

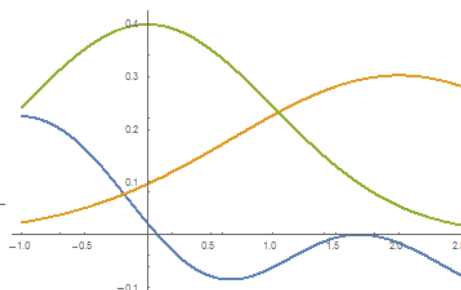
Case  $\sigma_2 > 1$ . As  $\sigma_2$  increases we get a similar situation as before but as  $f'_2$  becomes sufficiently small, the term  $f_2 f'_2$  is smaller, in absolute value, than



$f_1 f'_1$  so  $f_1 f'_1 + f_2 f'_2 < 0$ , this is a continuous process hence, at some point,  $f_1 f'_1 + f_2 f'_2 = 0$  but turn back to  $f_1 f'_1 + f_2 f'_2 < 0$ , that is the point where coalescence occurs, see Figure 7. ■



**Figure 6:** Graph as  $\sigma$  increases.



**Figure 7:** Graph when coalescence occurs.

This proposition is very important since it asserts that classification using the classifying field is not always possible due to the fact that one class will never be achieved by the orbits of  $CF$ , and this fact is intrinsic to the parameters relation between each other. This remark, we hope, will help to get a deeper insight to the one dimension classification problem.

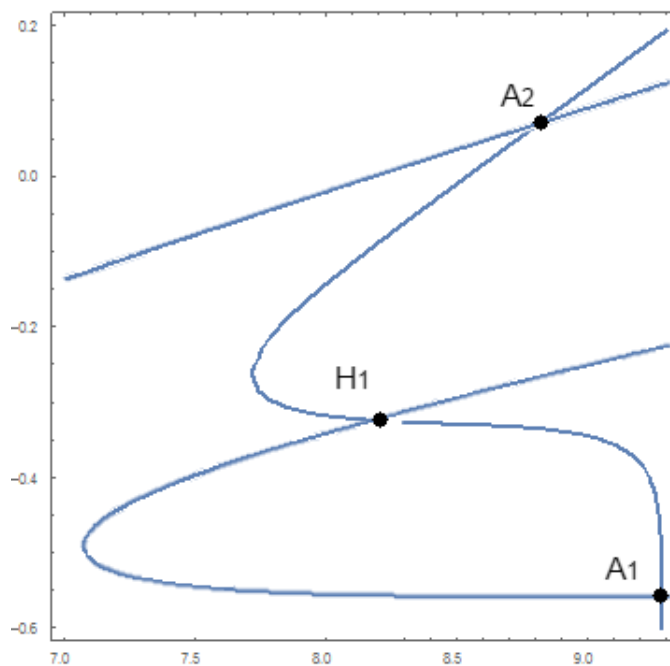
### Analysis in two dimensions

In order to deal with the two dimensional situation, we use actual data from the very well known problem of Leptograpsus crabs treated in Campbell and Mahon [1]. We chose two measurements of the carapace of the crabs, the rear width (RW) and the frontal width (FL) and classify, using the classifying field, for the crab sex. Let  $\Omega_1$  be the class of females, and  $\Omega_2$  the males' class, we used the estimated parameters, from the sample, to define the corresponding bivariate normal densities for each class. With that information, we constructed the corresponding gradients vector fields and the classifying field  $CF$ . The Confusion Matrix below was the result of classifying with  $CF$ , it is important to mention that we only used RW and FL, two of the five variables reported in the original paper, nevertheless the classification is fairly good.

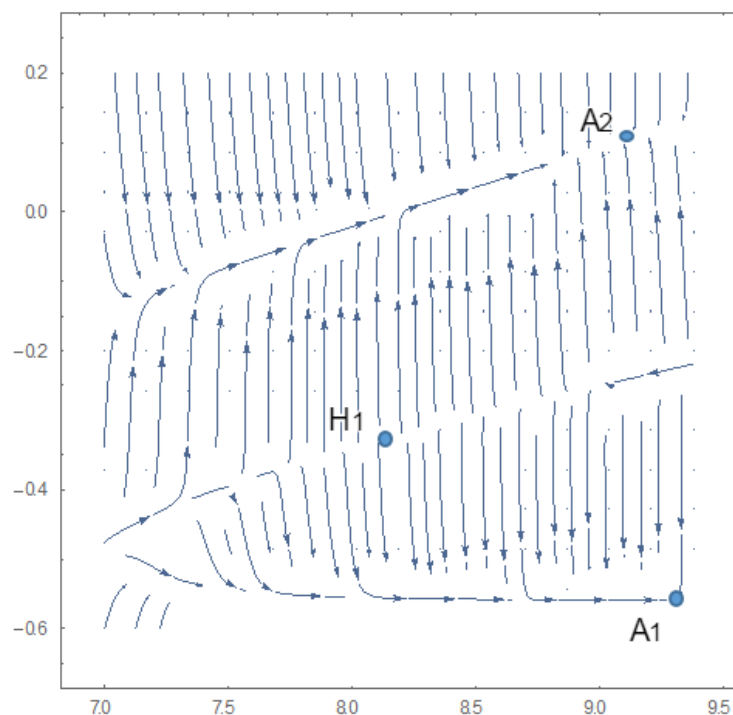
In order to make the analysis in a clearer way, we change coordinates to get correlation zero for the female population. Once we achieve that, we proceed to find the zeros for our classifying field, which now has two coordinates  $CF = (CF_1, CF_2)$ . We can see, in Figure 8, the graphs for the zero curves for each components where we depict the interceptions of both curves.

**Table 1:** Leptograpsus crabs: Confusion matrix.

	Belongs to $\Omega_1$	Belongs to $\Omega_2$
Classified into $\Omega_1$	0.85	0.90
Classified into $\Omega_2$	0.15	0.91

**Figure 8:**  $CF_1 = CF_2$  in the attractors  $A_1$  and  $A_2$  and in the hyperbolic point  $H_1$ .

We get, in the new coordinates, where both variables for the female population have correlation zero, two zeros,  $A_1 = (9.2828758, -0.5591295)$  and  $A_2 = (8.822026, 0.0724417)$ , close to the means of each class, that are attractors. A third zero,  $H_1 = (8.17003, -0.324096)$ , corresponds to a hyperbolic point whose stable manifold, in this case it is a pair of integral curves that are attracted by  $H_1$ , separates the corresponding basins of attraction of  $A_1$  and  $A_2$ . Due to this flow is that classification becomes possible. We obtained the Figure 9 with Mathematica.



**Figure 9:** Graph of the flow generated by  $CF$ .

The eigenvalues for the linearization were computed and are given in the following Table 2, a detailed analysis of their absolute value help to understand the phase portrait that we present above.

**Table 2:** Eigenvalues for the zeros.

Point	$\lambda_1$	$\lambda_2$
$A_1$	-4.80	-0.07
$A_2$	-8.21	-0.12
$H_1$	-0.05	4.93

We also applied this classifying method to the celebrated Fisher's Irises data for the Iris-setosa, which we denote by  $\Omega_1$ , and Iris versicolor, which we denote by  $\Omega_2$ , for this purpose we used as classifying variables, the length of the sepal and its width. For this exercise, we found the Confusion Matrix, which we include Table 3.

**Table 3:** Fisher's Irises: confusion matrix.

	Belongs to $\Omega_1$	Belongs to $\Omega_2$
Classified in $\Omega_1$	0.98	0
Classified in $\Omega_2$	0.02	1

## 4 Discussion

Classical discriminant analysis assumes the observation of a  $p$ -dimensional random vector  $X : \Omega \rightarrow \mathbb{R}^p$ , defined on a sample space  $\Omega$  that is partitioned into two subsets  $\Omega_1$  and  $\Omega_2$ . Given  $\omega \in \Omega$  such that  $X(\omega) = x$ , the objective is to infer whether  $\omega \in \Omega_1$  or  $\omega \in \Omega_2$ . Let  $\pi_i = \Pr[\Omega_i]$  and let  $f_i$  be the class conditional density of  $X$  given  $\Omega_i$ ; then, Bayes Theorem defines the posterior probability for each category as

$$\Pr[\Omega_i|X = x] = \frac{\pi_i f_i(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)}.$$

It is worth noting that the Classifying Vector Field  $CF(x)$  as given in eq. (2), is a convex linear combination of two gradient vectors, which are weighted by its corresponding posterior probabilities. In this sense,  $CF(x)$  can be described as a dynamic discriminant score, that combines relevant information from Bayes Theorem and from the gradient vectors.

An important contribution of this dynamic approach is the description of how classification in terms of the Classifying Field is not possible when relative heteroscedasticity tends to infinity. This is an important assertion that also makes sense from the classical statistical perspective, since

$$\lim_{\sigma \rightarrow \infty} \frac{1}{\sigma} \exp\left[-\frac{x^2}{2\sigma}\right] = 0, \forall x \in \mathbb{R}.$$

Therefore, the posterior probabilities for one category tend to one, absorbing as a direct consequence, every single observation to be classified. The discrimination of two multivariate normal densities given in terms of the Bayes Classification rule, always generates a separatrix border that is always linear or quadratic. Contrasting with classical discriminant analysis, dynamic classification seems to generate basins which boundaries are not necessarily algebraic varieties. This feature is informative, since it offers a complementary insight of the geometry of the system, that is not obvious from a decision theoretic approach. Supervised classification is a fundamental problem in multivariate statistics. Classical

textbooks generally dedicate at least one chapter to analyze this topic. The discussion is usually presented from a decision theoretic approach (see Mardia *et al.* [5] and Duda *et al.* [2]). Here, we discussed it from an alternative perspective based on the theory of dynamical systems. The results are interesting and clarifying.

## 5 Conclusion

This new approach to statistical classification encourages to a deeper use of the method in more general settings, either in the supervised and the none supervised situation, as well as in the parametric and nonparametric cases. More than two classes in any finite dimension must be consider and the dynamical system theory should be exploited.

## References

- [1] Campbell, N.A.; Mahon, R.J. (1974) “A multivariate study of variation in two species of rock crab of the genus *Leptograpsus*”, *Australian Journal of Zoology* **22**: 417–425.
- [2] Duda, R.O.; Hart, P.E.; Stork, D.G. (2001) *Pattern Classification*, 2nd edition. Wiley-Interscience, New York.
- [3] Fisher, R.A. (1936) “The use of multiple measurements in taxonomic problems”, *Annals of Eugenics* **7**(2): 179–188.
- [4] Hirsch, M.W.; Smale, S.; Devaney, R. (2012) *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, Waltham MA.
- [5] Mardia, K.A.; Kent, J.T.; Bibby, J.M. (2000) *Multivariate Statistics*. Academic Press, London.
- [6] Milnor, J.W. (1985) “On the concept of attractor”, *Communications in Mathematical Physics* **99**: 177–195.
- [7] Thom, R. (1975) *Structural Stability and Morphogenesis*. W.A. Benjamin, Inc., Reading MA.
- [8] Trejos, J.; Castillo, W.; González, J. (2014) *Análisis Multivariado de Datos. Métodos y Aplicaciones*. Editorial UCR, San José.

