

Fitting the Cusp Catastrophe in R: A cusp-Package Primer.

Raoul P. P. P. Grasman Han L. J. van der Maas Eric-Jan Wagenmakers
Universiteit van Amsterdam Universiteit van Amsterdam Universiteit van Amsterdam

Abstract

This vignette for the **cusp** package for R is a altered version of (Grasman, van der Maas, and Wagenmakers 2009) published in the *Journal of Statistical Software*.

Of the seven elementary catastrophes in catastrophe theory, the “cusp” model is the most widely applied. Most applications are however qualitative. Quantitative techniques for catastrophe modeling have been developed, but so far the limited availability of flexible software has hindered quantitative assessment. We present a package that implements and extends the method of (Cobb and Watson 1980; Cobb, Koppstein, and Chen 1983), and makes it easy to quantitatively fit and compare different cusp catastrophe models in a statistically principled way. After a short introduction to the cusp catastrophe, we demonstrate the package with two instructive examples.

Keywords: Catastrophe theory, cusp, modeling.

Catastrophe theory studies qualitative changes in behavior of systems under smooth gradual changes of control factors that determine their behavioral state. In doing so, catastrophe theory has been able to explain for example, how sudden abrupt changes in behavior can result from minute changes in the controlling factors, and why such changes occur at different control factor configurations depending on the past states of the system. Specifically, catastrophe theory predicts all the types of qualitative behavioral changes that *can* occur in the class of so called gradient systems, under influence of gradual changes in controlling factors (Poston and Stewart 1996).

Catastrophe theory was popularized in the 1970’s (Thom 1973; Thom and Fowler 1975; Gilmore 1993), and was suggested as a strategy for modeling in various disciplines, like physics, biology, psychology, and economics, by Zeeman (1971 as cited in Stewart and Peregoy 1983; Zeeman 1973, 1974; Isnard and Zeeman 1976; Poston and Stewart 1996). Although catastrophe theory is well established and applied in the physical sciences, its applications in the biological sciences, and especially in the social and behavioral sciences, has been criticized (Sussmann and Zahler 1978; Rosser 2007). Applications in the latter fields range from modeling of exchange market crashes in economics (Zeeman 1974), to perception of bistable figures (Stewart and Peregoy 1983; Ta’eed, Ta’eed, and Wright 1988; Fürstenau 2006) in psychology. The major points of criticism concerned the overly qualitative methodology used in applications in the latter sciences, as well as the ad hoc nature of the choice of variables used as control variables. The first of these seems to stem from the fact that catastrophe theory concerned deterministic dynamical systems, whereas applications in these sciences concern data that are more appropriately considered stochastic. For an historical summary of the developments and critiques of catastrophe theory we refer the reader to Rosser (2007).

Stochastic formulations of catastrophe theory have been found, and statistical methods have been developed that allow quantitative comparison of catastrophe models with data (Cobb and Ragade 1978; Cobb and Watson 1980; Cobb 1981; Cobb *et al.* 1983; Guastello 1982; Oliva, Desarbo, Day, and Jedidi 1987; Lange, Oliva, and McDade 2000; Wagenmakers, Molenaar, Grasman, Hartelman, and van der Maas 2005a; Guastello 1992). Of these methods, the maximum likelihood approach of Cobb and Watson (1980); Cobb *et al.* (1983) is arguably most appealing for reasons that we discuss in the Appendix A.

In this paper we present an add-on package for the statistical computing environment R (R Development Core Team 2009) which implements the method of (Cobb *et al.* 1983), and extends it in a number of ways. In particular, the approach of Oliva *et al.* (1987) is adopted to allow for a behavioral variable that is embedded in multivariate response space. Furthermore, the implementation incorporates many of the suggestions of Hartelman (1997), and indeed intends to succeed and update the “*cusplfit*” FORTRAN program of that author. The package is available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=cuspl>

1. Cusp catastrophe

Consider a (deterministic) dynamical system that obeys equations of motion of the form

$$\frac{\partial y}{\partial t} = -\frac{\partial V(y; c)}{\partial y}, \quad y \in R^k, c \in R^p. \quad (1)$$

In this equation $y(t)$ represents the system’s state variable(s), and c represents one or multiple (control) parameter(s) whose value(s) determine the specific structure of the system. If the system state y is at a point where $\partial V(y; c)/\partial y = 0$ the system is in equilibrium. If the system is at a non-equilibrium point, the system will move towards an equilibrium point where the function $V(y; c)$ acquires a minimum with respect to y . These equilibrium points are stable equilibrium points because the system will return to such a point after a small perturbation to the system’s state. Equilibrium points that correspond to maxima of $V(y; c)$ are unstable equilibrium points because a perturbation of the system’s state will cause the system to move away from the equilibrium point towards a stable equilibrium point. In the physical world no system is fully isolated and there are always forces impinging on a system. It is therefore highly unlikely to encounter a system in an unstable equilibrium state. Equilibrium points that correspond neither to maxima nor to minima of $V(y; c)$, at which the Hessian matrix $(\partial^2 V(y)/\partial y_i \partial y_j)$ has eigenvalues equal to zero, are called degenerate equilibrium points, and these are the points at which a system can give rise to unexpected bifurcations in its equilibrium states when the control variables of the system are changed.

1.1. Canonical form

Catastrophe theory (Thom 1973; Thom and Fowler 1975; Gilmore 1993; Poston and Stewart 1996) classifies the behavior of deterministic dynamical systems in the neighborhood of degenerate critical points of the potential function $V(y; c)$. One of the remarkable findings of catastrophe theory is its discovery that degenerate equilibrium points of systems of the form described above, which have an arbitrary numbers of state variables and are controlled by no more than four control variables, can be characterized by a set of only seven canonical forms,

or “universal unfoldings”, in only one or two canonical state variables. These universal unfoldings are called the “elementary catastrophes”. The canonical state variables are obtained by smooth transformations of the original state variables. The elementary catastrophes constitute the different families of catastrophe models.

In the biological and behavioral sciences, the so-called cusp catastrophe model has been applied most frequently, as it is the simplest of catastrophe models that exhibits discontinuous transitions in equilibrium states as control parameters are varied. The canonical form of the potential function for the cusp catastrophe is

$$-V(y; \alpha, \beta) = \alpha y + \frac{1}{2}\beta y^2 - \frac{1}{4}y^4. \quad (2)$$

Its equilibrium points, as a function of the control parameters α and β , are solutions to the equation

$$\alpha + \beta y - y^3 = 0. \quad (3)$$

This equation has one solution if $\delta = 27\alpha - 4\beta^3$, which is known as *Cardan’s discriminant*, is greater than zero, and has three solutions if $\delta < 0$. These solutions are depicted in Figure 1 as a two dimensional surface living in three dimensional space, the floor of which is the two dimensional (α, β) coordinate system called the “control plane”. The set of values of α and β for which $\delta = 0$ demarcates the *bifurcation set*, the cross hatched cusp shaped region on the floor in Figure 1. From a regression perspective, the cusp equilibrium surface may be conceived of as response surface, the height of which predicts the value of the dependent variable y given the values of the control variables. This response surface has the peculiar property however, that for some values of the control variables α and β the surface predicts two possible values instead of one. In addition, this response surface has the unusual characteristic that it “anti-predicts” an intermediate value for these values of the control variables—that is, the surface predicts that certain state values, viz. unstable equilibrium states, should not occur (Cobb 1980). As indicated, the dependent variable y is not necessarily an observed quantity that characterizes the system under study, but is in fact a canonical variable that in general depends on a number of actually measurable dependent variables. Similarly, the control coordinates α and β represent canonical coordinates that are dependent on actual measured or controlled independent variables. The α coordinate is called the “normal” or “asymmetry” coordinate, while the β coordinate is called the “bifurcation” or “splitting” coordinate (Stewart and Peregoy 1983).

1.2. Cusp catastrophe as a model

In evaluating the cusp as a model for data there are two complementary approaches. The first approach evaluates whether certain qualitative phenomena occur in the system under consideration. In the second approach, a parameterized cusp is fitted to the data.

Gilmore (1993) derived a number of qualitative behavioral characteristics of the cusp model; the so-called *catastrophe flags*. Among the more prominent are sudden jumps in the value of the (canonical) state variables; hysteresis—i.e., memory for the path through the phase space of the system; and multi-modality—i.e., the simultaneously presence of multiple preferred states. Verification of the presence of these flags constitute one important stage in gathering evidence for the presence of a cusp catastrophe in the system under scrutiny. For extensive discussions of these qualitative flags we refer to Gilmore (1993) (see also Stewart and Peregoy 1983; van der Maas and Molenaar 1992; van der Maas, Kolstein, and van der Pligt 2003).

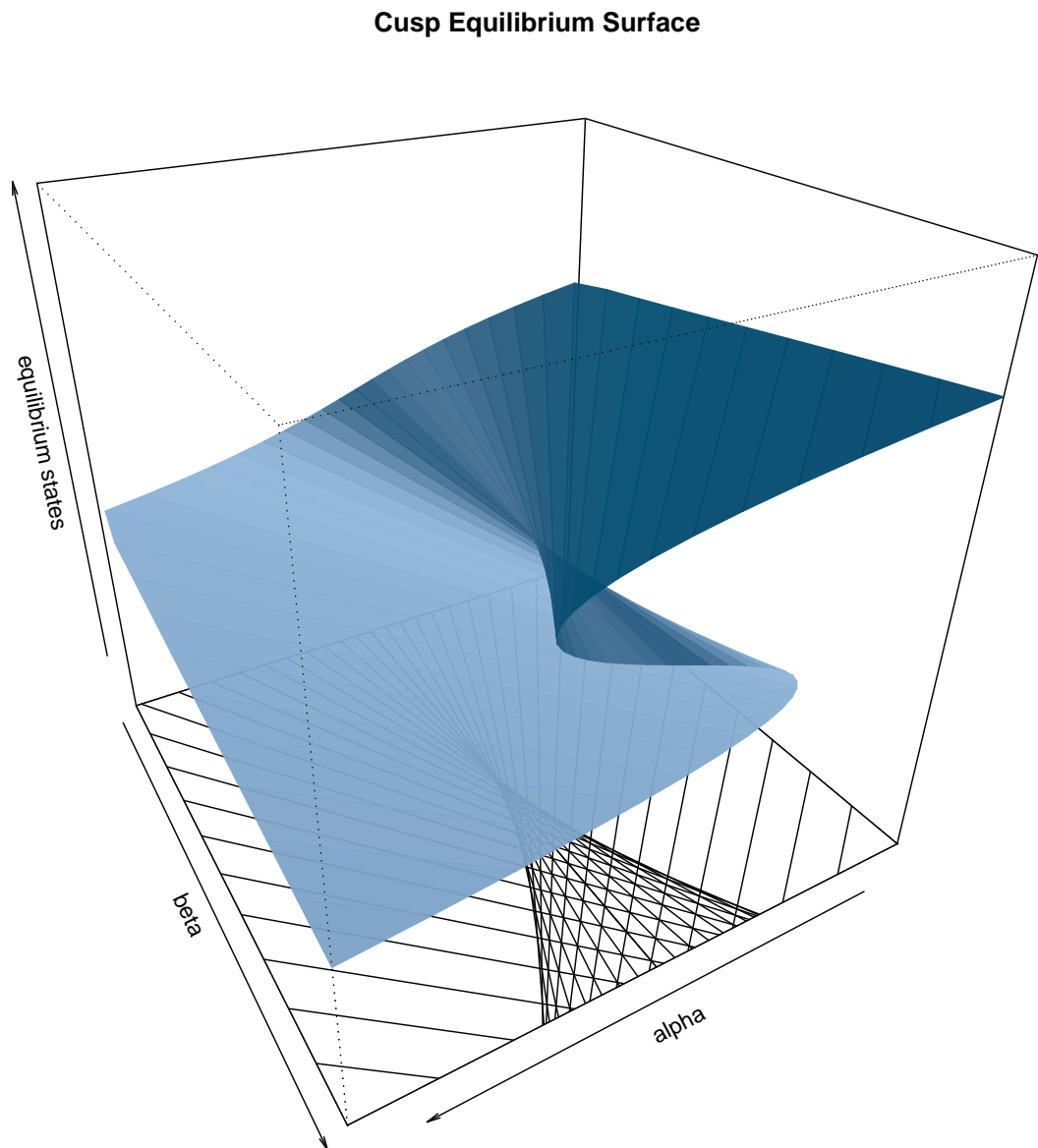


Figure 1: Cusp surface.

Most applications of catastrophe theory in general, and the cusp catastrophe in particular, have focused entirely on this qualitative verification. An alternative approach is constituted by a quantitative evaluation of an actual match between a cusp catastrophe model and the data using statistical fitting procedures. This is indispensable for sound verification that a cusp catastrophe does a better job at describing the data than other conceivable models for the data.

A difficulty that arises when constructing empirically testable catastrophe models is the fact that catastrophe theory applies to deterministic systems as described by equation (1). Being inherently deterministic, catastrophe theory cannot be applied directly to systems that are subject to random influences which is commonly the case for real physical systems, especially in the biological and behavioral sciences.

To bridge the gap between determinism of catastrophe theory and applications in stochastic environments, Loren Cobb and his colleagues (Cobb and Ragade 1978; Cobb 1980; Cobb and Watson 1980; Cobb and Zacks 1985) proposed to turn catastrophe theory into a stochastic catastrophe theory by adding to equation (1) a (white noise) Wiener process, $dW(t)$, with variance σ^2 , and to treat the resulting equation as a stochastic differential equation (SDE):

$$dY = \frac{\partial V(Y; \alpha, \beta)}{\partial Y} dt + dW(t).$$

This SDE is then associated with a probability density that describes the distribution of the system's states on any moment in time, which may be expressed as

$$f(y) = \frac{\psi}{\sigma^2} \exp \left[\frac{\alpha (y - \lambda) + \frac{1}{2}\beta (y - \lambda)^2 - \frac{1}{4}(y - \lambda)^4}{\sigma^2} \right]. \quad (4)$$

Here ψ is a normalizing constant, and λ merely determines the origin of scale of the state variable. In this stochastic context β is called the bifurcation factor, as it determines the number of modes of the density function, while α is called “asymmetry” factor as it determines the direction of the skew of the density (the density is symmetric if $\alpha = 0$ and becomes left or right skewed depending on the sign of α ; Cobb 1980). The function is implemented in the R package described below as `dcusp()`. Figure 2 displays the density for different regions of the control plane.

2. Estimation methods

As indicated earlier, the state variable y of the cusp is a “canonical variable”. This means that it is a (generally unknown) smooth transformation of the actual state variables of the system. If we have a set of measured dependent variables Y_1, Y_2, \dots, Y_p , to a first order approximation we may say

$$y = w_0 + w_1 Y_1 + w_2 Y_2 + \dots + w_p Y_p, \quad (5)$$

where w_0, w_1, \dots, w_p are the first order coefficients of a polynomial approximation to the “true” smooth transformation. Similarly, the parameters α and β are “canonical variates” in the sense that they are (generally unknown) smooth transformations of actual control variates. Again to first approximation, for experimental parameters or measured independent variables X_1, \dots, X_q , we may write

$$\alpha = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_q X_q, \quad (6)$$

$$\beta = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_q X_q. \quad (7)$$

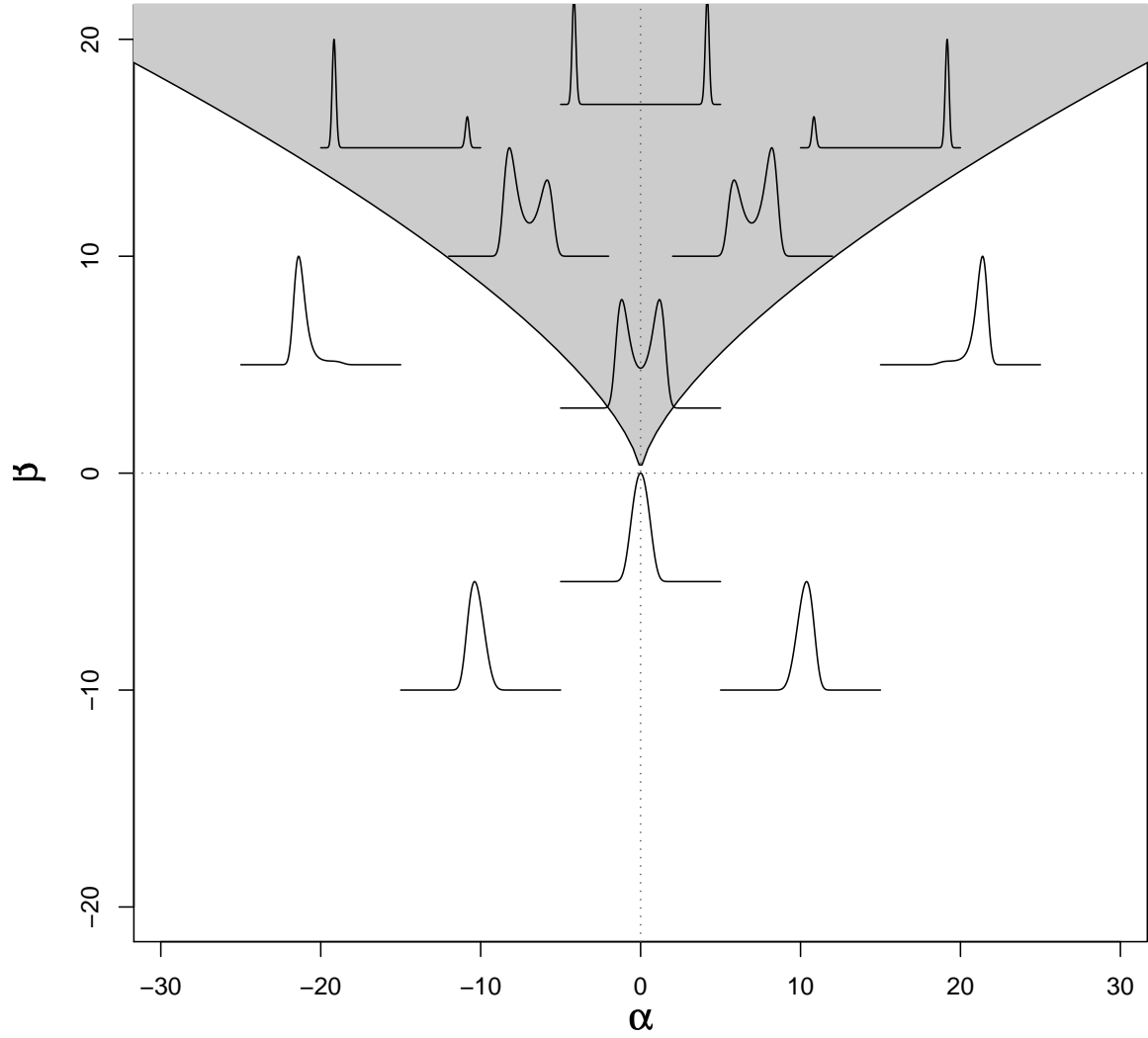


Figure 2: Shape of cusp density in different regions of the control plane. Bimodality of the density only occurs in the “bifurcation set”, the cusp shaped shaded area of the control plane.

Fitting the cusp model to empirical data then reduces to estimating the parameters $w_0, w_1, \dots, w_p, a_0, \dots, a_q, b_0, \dots, b_q$.

On the basis of equation (4) and the associated stochastic differential equation discussed earlier, Cobb (1981) and Cobb *et al.* (1983) respectively developed method of moments estimators and maximum likelihood estimators. The maximum likelihood method of Cobb (Cobb and Watson 1980; Cobb and Zacks 1985) has not been commonly used however. Two important reasons for this are instability of Cobb’s software for fitting the cusp density, and difficulties in its use (see van der Maas *et al.* 2003 for a discussion; see the appendix for a discussion of other estimation methods that have been proposed in the literature).

Cobb’s methods assume that the state variable is directly accessible to measurement. As argued by Oliva *et al.* (1987), in the behavioral sciences both dependent (state) as well as independent (control) variables are more often than not constructs which cannot be easily measured directly. It is therefore important to incorporate equation (5) so that the state variable that adheres to the cusp catastrophe may be embedded in the linear space spanned by a set of dependent variables.

In the **cusp** package that we describe below, we use the maximum likelihood approach of Cobb and Watson (1980), augmented with the subspace fitting method (Equation 5) proposed by (Oliva *et al.* 1987)

3. Package description

The **cusp** package is a package for the statistical computing language and environment R (R Development Core Team 2009). The package contains a number of functions for use with catastrophe modeling, including utility functions to generate observations from the cusp density, to evaluate the cusp density and cumulative distribution function, to fit the cusp catastrophe, to evaluate the model fit, and to display the results. The core user interface functions of the package are listed in Table 1.

3.1. Model specification

As is the with many model fitting routines R, the fitting routines in the **cusp** package allow the user to specify models in terms of dependent and independent variables in a compact symbolic form. Basic to the formation of such models is the \sim (tilde) operator. An expression of the form $y \sim \text{model}$ signifies that the response y is modeled by a linear predictor that is specified in `model`. In the **cusp** package, the dependent variables are however always y , α , and β , in accordance with Equations (8–10) discussed below. Model formulas are used to generate an appropriate design matrix. It should be mentioned that R makes a strict distinction between “factors” and variables. The former are study design factors that have an enumerable number of levels (e.g., education level, treatment level, etc.), whereas the latter are continuous variates (e.g., age, heart rate, response time). Factors and variates are (appropriately) treated differently in the constructing the design matrix.

3.2. Cost function (likelihood function)

At the heart of the package is the fitting routine that performs maximum likelihood estimation of all the parameters in Equations (5–7). That is, for observed dependent variables

Function name	Description and example call
cusp	Fits cusp catastrophe to data. <code>fit <- cusp(y~z, alpha~x1+x2, beta~x1+x2, data)</code>
summary method	Computes statistics on parameter estimates and model fit (by default this function compares the cusp model to a linear regression model). For model comparison a pseudo- R^2 statistic, AIC and BIC are computed. <code>summary(fit)</code>
confint method	Computes confidence intervals for the parameter estimates. <code>confint(fit)</code>
vcov method	Computes an approximation to the parameter estimator variance-covariance matrix. <code>vcov(fit)</code>
logLik method	Returns the optimized value of the log-likelihood. <code>logLik(fit)</code>
plot method	Generates a graphical display of the fit, including the estimated locations on the cusp control surface for each observation, non-parametric density estimates for different regions of the control surface, and a residuals plot. An example is provided in Figure 3. The graph is fully customizable. <code>plot(fit)</code>
cusp3d	Generates a three dimensional display of the cusp equilibrium surface on which the estimated states are displayed. The graphic is fully customizable. Figure 5 displays an example. <code>cusp3d(fit)</code>
cusp3d.surface	Generates a three dimensional display of the cusp equilibrium surface. Figure 1 was generated with it. <code>cusp3d.surface()</code>
rcusp	Generates a random sample from the cusp distribution. <code>y <- rcusp(n=100, alpha=-1/2, beta=2)</code>
dcusp , pcusp , qcusp	The cusp density, cumulative distribution and quantile functions. <code>dcusp(y=1.0, alpha=-1/2, beta=2)</code>
cusp.logist	Fits an logistic curve to the data. Is used in summary.cusp if optional parameter logist=TRUE is specified for testing the presence of observations in the bifurcation set. <code>fit <- cusp.logist(y~z, alpha~x1+x2, beta~x1+x2, data)</code>

Table 1: Summary, including example calls, of most important functions in **cusp** package. A full description of these functions and others are given in the package documentation.

$Y_{i1}, Y_{i2}, \dots, Y_{ip}$, and independent variables $X_{i1}, X_{i2}, \dots, X_{iq}$, for subjects $i = 1, \dots, n$, the distribution of

$$y_i = w_0 + w_1 Y_{i1} + w_2 Y_{i2} + \dots + w_p Y_{ip} \quad (8)$$

is modeled by the density (4), with $\alpha \mapsto \alpha_i$, $\beta \mapsto \beta_i$, where

$$\alpha_i = a_0 + a_1 X_{i1} + a_2 X_{i2} + \dots + a_q X_{iq} \quad (9)$$

$$\beta_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_q X_{iq}. \quad (10)$$

If we collect the observed dependent variables in the $n \times (p+1)$ matrix $\mathbf{Y} = [\mathbf{1}_n | (Y_{ij})]$, the independent variables in the $n \times (q+1)$ matrix $\mathbf{X} = [\mathbf{1}_n | (X_{ik})]$, where $\mathbf{1}_n$ is an n -vector whose entries all equal 1, and furthermore collect the coefficients in the vectors $\mathbf{w} = (w_0, w_1, \dots, w_p)'$, $\mathbf{a} = (a_0, a_1, \dots, a_q)'$, and $\mathbf{b} = (b_0, b_1, \dots, b_q)'$, where $'$ denotes transpose, we can write these equations succinctly as

$$\mathbf{y} = \mathbf{Y}\mathbf{w}, \quad \boldsymbol{\alpha} = \mathbf{X}\mathbf{a}, \quad \boldsymbol{\beta} = \mathbf{X}\mathbf{b},$$

where $\mathbf{y} = (y_1, \dots, y_n)'$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$. Therefore, the negative log-likelihood for a sample of observed values $(x_{i1}, \dots, x_{iq}, y_{i1}, \dots, y_{ip})$, $i = 1, \dots, n$, is

$$L(\mathbf{a}, \mathbf{b}, \mathbf{w}; \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n \log \psi_i - \sum_{i=0}^n \left[\alpha_i y_i + \frac{1}{2} \beta_i y_i^2 - \frac{1}{4} y_i^4 \right]. \quad (11)$$

Note that compared to equation (4), here we have absorbed the location and scale parameters λ and σ into the coefficients w_0, w_1, \dots, w_p . It should be noted furthermore that there is an ambiguity in the equations above: The signs of the a_j and w_j may all be switched without affecting the value negative log-likelihood function. This is the case, because the quadratic and quartic terms in y_i do not determine the sign of y_i (i.e., $y_i^2 = (-y_i)^2$ and $y_i^4 = (-y_i)^4$), while for the linear term $\alpha_i y_i = (-\alpha_i)(-y_i)$ holds. Thus the sign of y_i (and hence, the sign for the w_j 's) can be switched without affecting the value of (11) if the sign of α_i , and hence the signs of the a_j 's, is switched as well. The estimates of $a_1, \dots, a_q, w_1, \dots, w_p$ are therefore identifiable up to a change in sign.

The `cusp` routine of the package minimizes L with respect to the parameters $w_0, \dots, w_p, a_0, \dots, a_p, b_0, \dots, b_q$. The computational load of evaluating L is mainly burdened by the calculation of the normalization constants ψ_i , which has to be carried out numerically. To this end, we use an adaptive quadrature routine to minimize the computational cost. To speed up the calculations substantively, this is carried out in linked C code. Speed is especially important from the user experience perspective, if different models are to be tried and compared.

Internally, the data are standardized using a QR decomposition. This is both for stability of the estimation algorithm, as well as handling collinear predictors in the design matrix. Only coefficients for each dimension of the column space of the design matrix are estimated. The estimated coefficients are related back to the actual dependent variables. If the design matrix does not have full column rank, coefficients for some independent variables—those that are fully explained by others—are not determined. For which of the independent variables coefficients *are* estimated in such collinear cases depends partly on the order of their occurrence in the formula.

Note that, because the models for α_i 's and β_i 's are specified separately, different sets of independent variables can figure in each, hence allowing for confirmatory data analysis. This has been considered a deficiency the method of Cobb and Watson (Stewart and Peregoy

1983; Alexander, Herbert, DeShon, and Hanges 1992), although this option has been present in Hartelman’s (1997) program.

3.3. Optimization algorithm

The negative log-likelihood function is minimized using one of the built-in optimization routines of R. By default the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm with bounds (L-BFGS-B; Zhu, Byrd, Lu, and Nocedal 1997) is used by the *culp* function. Other methods provided by R can be used, but in our experience L-BFGS-B works best. The main problem is that the *culp* density quickly decreases beyond numerical precision for $|\alpha|, |\beta| > 5$, and hence some boundary restrictions are helpful. Boundaries (both lower and upper) default to -10 and 10 , which is sufficiently large in combination with the internal standardization of the data in all the cases we have encountered. Different boundaries may be specified, but our experience with earlier applications of the method is that the values of $|\alpha|$ and $|\beta|$ rarely exceed 3.

3.4. Starting point

The starting values used by default turn out to often render convergence of the optimization algorithm without problems. When convergence problems arise a warning is issued, and this may be interpreted as an invitation to the user to provide alternative starting values. In our experience convergence to a proper minimum of the negative log-likelihood can be achieved most of the time by providing alternative starting values. A strong indication of improper convergence is a warning about NaN’s that is issued when summary statistics are computed. If this happens, the model should be refit using a different set of starting values.

3.5. Statistical evaluation of model fit

To assess the correspondence of data with the predictions made by the *culp* catastrophe, a number of diagnostic tools have been suggested.

Maxwell vs delay convention and R^2

First of all, Cobb (1998; see also Stewart and Peregoy 1983; Hartelman 1997) have suggested a pseudo- R^2 as a measure of explained variance defined by

$$1 - \frac{\text{error variance}}{\text{Var}(y)}. \quad (12)$$

This is clearly inspired by the familiar relation between the squared (multiple) correlation coefficient and the “explained variance” from ordinary regression. However, the term “error variance” needs to be defined, which is not trivial for the *culp* catastrophe model.

As indicated previously, the *culp* catastrophe is not an ordinary regression model. Rather, it is an implicit regression model of an irregular type.¹ As such, unlike ordinary regression models, for a given set of values of the independent variables the model may predict multiple values for the dependent variable. In ordinary regression the predicted value is usually the

¹The *culp* catastrophe is an irregular implicit regression model in the sense that the available statistical theory for implicit regression models cannot be applied to catastrophe models (Hartelman 1997).

expected value of the dependent variable given the values of the independent variables. In the case of the cusp density, for certain values of α and β the cusp density is bimodal, and the expected value of this density lies in a region of low probability between the two modi. That is, the mathematical expectation of the cusp density is a value that in itself is relatively unlikely to be observed. Two alternatives for the expected value as the predicted value can be used, which are closely related to a similar problem regarding interpretation conventions in deterministic catastrophe theory: One can choose the mode of the density closest to the state values as the predicted value, or one can use the mode at which the density is highest. The former is known as the *delay convention*, the latter as the *Maxwell convention*. Although in the physical sciences both have their uses, Cobb and Watson (1980, Cobb 1998) suggest to use the delay convention (Stewart and Peregoy 1983). Both conventions are provided by the package, the default being the delay convention.

Hence, in the pseudo- R^2 statistic defined above, the error variance is defined as the variance of the differences between the observed (or estimated) states and the mode of the distribution that is closest to this value. It should be noted however that this pseudo- R^2 can become negative if many of the α_i 's deviate from 0 in the same direction—in that case the distribution is strongly skewed, and deviation from the mode is on average larger than deviation from the mean. Negative pseudo- R^2 's are thus perfectly legitimate for the cusp density, which limits its value for model fit assessment. Alternatives to the pseudo- R^2 are discussed below.

AIC, BIC and logistic curve

In addition to the pseudo- R^2 statistic, to establish convincingly the presence of a cusp catastrophe, Cobb gives three guidelines for evaluating the model fit (Cobb 1998; Hartelman 1997): First, the fit of the cusp should be substantially better than multiple linear regression—i.e., its likelihood should be significantly higher than that of the ordinary regression model. Second, any of the coefficients w_1, \dots, w_p should deviate significantly from zero (w_0 does not have to), as well as at least one of the a_j 's or b_j 's. Thirdly, at least 10% of the (α_i, β_i) pairs should lie within the bifurcation region. The former two guidelines can be assessed with the summary function of the packages; the latter guideline can be assessed with the plot function; both of these are detailed in the Examples section below.

A problem arises when the general case of equation (8) with more than one dependent variable is used: The linear regression model with which to compare the cusp model is not uniquely defined. The most natural approach seems to be linear subspace regression: The first canonical variate between the X_{ji} 's and the Y_{ki} 's is used as the predicted value, and the first canonical correlation is used for calculating the explained and residual variance. For models in which (8) contains only one dependent variable this automatically reduces to standard univariate linear regression.

The 10% guideline of Cobb is somewhat arbitrary, and Hartelman (1997); Hartelman, van der Maas, and Molenaar (1998); van der Maas *et al.* (2003) propose a more stringent test of the presence of bifurcation points: They suggest to compare the model fit to the non-linear least squares regression to the logistic curve

$$y_i = \frac{1}{1 + e^{-\alpha_i/\beta_i^2}} + \epsilon_i, \quad i = 1, \dots, n, \quad (13)$$

where y_i , α_i , β_i are defined in Equations (8–10), and the ϵ_i 's are zero mean random disturbances. (One may assume ϵ_i to be normally distributed, but this is not necessary; see e.g.,

Seber and Wild 1989.) The rationale for the logistic function is that this function does not possess degenerate critical points, while it does have the possibility to model arbitrarily steep changes in the (canonical) state variable as a function of minute changes in an independent variable, thus mimicking “sudden” transitions of the cusp (Hartelman 1997). The summary function in the package provides an option to fit this logistic curve to the data. Because the cusp density and the logistic functions are not nested models, the fit cannot be assessed on the basis of differences in the likelihood, and one has to resort to other indicators such as AIC and BIC. These fit indices are both computed by the summary function, as well as an AIC corrected for small sample sizes (AICc; Burnham and Anderson 2004).

Instead of the 10% guideline, Hartelman (1997) proposes to require that the AIC and BIC indicate a better fit for the cusp density than for the logistic curve. Wagenmakers, van der Maas, and Molenaar (2005b) suggest to use the BIC of both models to compute approximation of the posterior odds for the cusp relative to the logistic curve, assuming equal prior probabilities for both models.

4. Examples

For illustration purposes, we provide three example analyses with the package. The first two examples entail data that have been analyzed with cusp catastrophe methods before and have been published elsewhere.

Example I

The first example is taken from van der Maas *et al.* (2003), and concerns attitudinal response transitions with respect to the statement “The government must force companies to let their workers benefit from the profit as much as the shareholders do”. Some 3000 Dutch respondents indicated their level of agreement with this statement on a 5 point scale (1 = *totally agree*, 5 = *totally disagree*). As a normal factor political orientation (measures on a 10 point scale from 1 = *left wing* to 10 = *right wing*) was used. As a bifurcation factor the total score on a 12 item “political involvement” scale was used. The theoretical social psychological details are discussed in (van der Maas *et al.* 2003). The data thus consist of a table with three columns: `Orient`, `Involv`, and `Attitude`.

We use the same subset of 1387 cases that was also analyzed in van der Maas *et al.* (2003). In that paper an extensive comparison of models was done. These data are made available in the package and can be loaded using the `data("attitudes")` command. Here we only present the analysis for the best fitting model (model 12 in van der Maas *et al.* 2003, according to AIC and BIC). In that model, the bifurcation factor (β_i) is determined exclusively by `Involv`, while the normal factor (α_i) is determined by both `Orient` and `Involv`. This model can be fitted with the command

```
R> fit <- culp(y ~ Attitude,
+   alpha ~ Orient + Involv,
+   beta ~ Involv,
+   data = attitudes, start=attitudeStartingValues)
```

Here `attitudes` is the `data.frame` in which the table with data is stored, and the vector `attitudeStartingValues` is a set starting values that is included in the package to obtain a

	Estimate	Std. Error	z value	P(> z)
a[(Intercept)]	0.15271	0.56218	0.272	.7859
a[Orient]	0.46210	0.06414	7.204	< .0000
a[Involv]	0.09736	0.05445	1.788	.0738
b[(Intercept)]	0.12397	0.38643	0.321	.7484
b[Involv]	0.22738	0.02993	7.597	< .0000
w[(Intercept)]	0.10543	0.24485	0.431	.6668
w[Attitude]	0.87758	0.06682	13.134	< .0000

Table 2: Coefficient summary table for attitudes example obtained with the `summary` method.

solution quickly for this example. Both were loaded with a `data` call. Note the use of formula's discussed earlier: The first formula specifies that the state variable y_i is determined by the variable `Attitude` for which a location and scaling parameter (w_0 and w_1 in Equation 8) have to be estimated. The second formula states that the normal factor α_i is determined by the variables `Orient` and `Involv` for which regression coefficients and an intercept must be estimated. Similarly the third formula which specifies that the bifurcation factor is determined by the variable `Involv` for which an intercept and a regression coefficient have to be estimated. When the `culp` function returns, the estimates can be printed to the console window of R by typing `print(fit)`. More informative however is a summary of the parameters and the associated statistics, which is obtained with the statement

```
R> summary(fit, logist = TRUE)
```

where `logist` is set to `TRUE` to compare the cusp to the logistic curve fit, in addition to a comparison with a linear model. Part of the result is display in Tables 2 and 3. There are only very small differences between the fit presented here and the fit presented in (van der Maas *et al.* 2003), which are entirely due to differences in optimization algorithm. It should be mentioned that in (van der Maas *et al.* 2003) the parameter estimates are the coefficients with respect to standardized data. To standardize the data the `scale` function of R can be used. The column headed " R^2 " in Table 3 lists the squared multiple correlation for the linear regression model and logistic curve model, and the pseudo- R^2 for the cusp catastrophe model. A visual display of the data fit and diagnostic plots is generated with the command `plot(fit)`; the result is displayed in Figure 3. The figure display the control plane along with the estimated (α_i, β_i) , $i = 1, \dots, n$, for each of the observations. Furthermore, it (optionally) displays for each of four regions in the control plane a density estimate of the estimated

	R^2	AIC	AICc	BIC
Linear model	0.1309	3997.6	3997.6	4018.5
Logist model	0.1599	3954.5	3954.5	3985.9
Cusp model	-0.0962	3623.8	3623.9	3660.4

Table 3: Model fit statistics for synthetic data example as obtained with `summary`. The column labeled " R^2 " gives conventional R^2 for the linear and logist model, and the pseudo- R^2 statistic for the cusp model.

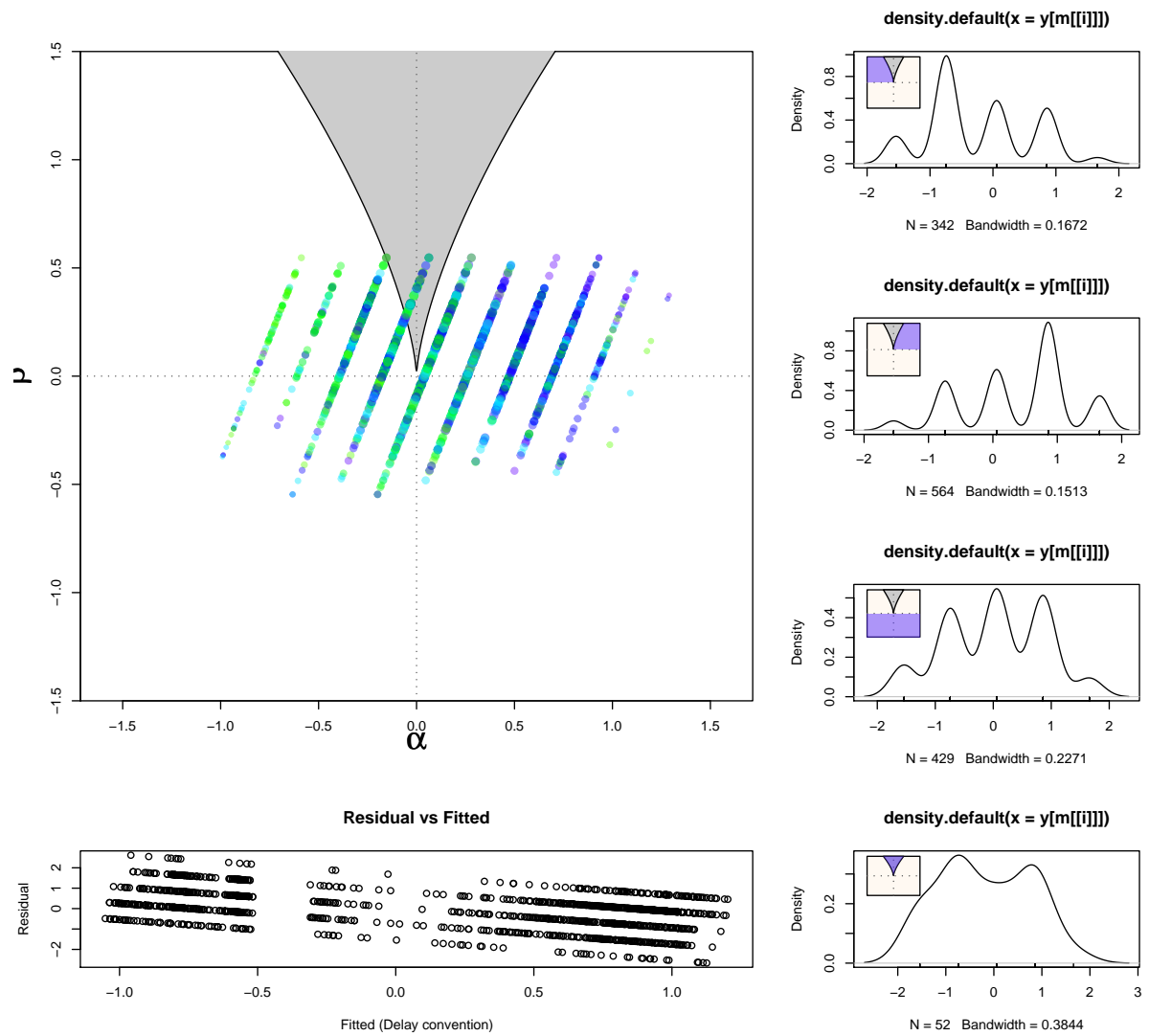


Figure 3: Diagnostic plot of fit of attitude data obtained with the `plot` method that is available in the package.

state variable in that region. These are displayed on the right in the figure: The top two panels reflect the densities in the region left and right of the bifurcation region respectively. These should be positively skewed for the left side and negatively skewed for the right side (compare Figure 2). The next panel display the density estimate for the state estimates below the bifurcation region. Here the density estimate should be approximately uni-modal and less skewed than in the other two regions. The last panel displays the density estimate for estimated state values within the bifurcation region. In this region the density should be bimodal. For the data displayed here the first three densities seem heavily multi-modal. This is due to the fact that the bifurcation factor “political orientation” was measured on a discrete ordinal 10 points scale, thus artifactually introducing modes around these values. The residual versus fitted plot displays the estimated errors as a function of predicted states. As indicated earlier, by default the errors are computed using the *delay* convention. Although with a good model fit one would expect to see no systematic relationship between the estimated errors and the predicted states, we have observed a consistent negative trend between the two—even when the data were generated from the cusp density in simulations. This may simply result from cases of which the density is strongly skewed. Hence, a slight negative trend should not be taken too seriously as an indication of model misfit.

Note from Figure 3 that the overwhelming majority of cases lie in regions where the cusp density is moderate to strongly skewed. This fact accounts for the negative pseudo- R^2 . Lange *et al.* (2000) recommend to reject the cusp model entirely when negative pseudo R^2 result. This seems to be an unwarranted recommendation because negative pseudo- R^2 values are—because of the definition of this measure of fit—entirely possible for the cusp density. In fact, negative pseudo- R^2 statistics are possible for all (strongly) skewed distribution such as for instance a chi-square distribution. Rather than rejecting the cusp catastrophe model on the basis of a negative pseudo- R^2 , one might consider forgetting about this measure of fit entirely, as it clearly is not well-suited for its intended purposes in non-symmetrical distributions. Instead, one can use AIC and BIC to compare the cusp model with competitor models such as the linear regression model and the logistic curve.

Example II

In the second example we use the model specified by Oliva *et al.* (1987) to demonstrate the use of multiple state variables. The model corresponds to the data in Table 2 of that paper, which displays synthetic data for 50 cases with scores on two (dependent) state variables (Z_1 and Z_2), four (independent) bifurcation variables (Y_1 , Y_2 , Y_3 , and Y_4), and three (independent) normal variables (X_1 , X_2 , and X_3). The data can be loaded from the package with `data(oliva)`. The “true” model for their synthetic data is

$$\begin{aligned}\alpha_i &= X_1 - .969 X_2 - .201 X_3, \\ \beta_i &= .44 Y_1 + 0.08 Y_2 + .67 Y_3 + .19 Y_4, \\ y_i &= -0.52 Z_1 - 1.60 Z_2.\end{aligned}\tag{14}$$

For testing purposes, Oliva *et al.* did not add noise to any of the variables, and hence, the data perfectly comply with their implicit regression equation (3).

Because we are using the stochastic approach of Cobb, we cannot use these deterministic data. We therefore generated data in accordance with the model in equation (14), where X_1 , X_2 , and X_3 were uniformly distributed on the interval $(-2, 2)$, Y_1 , Y_2 and Z_1 were uniformly

Model I				
	Estimate	Std. Error	z value	P(> z)
a[x1]*	0.82896	0.21994	3.769	.0002
a[x2]*	-0.67698	0.19928	-3.397	.0007
a[x3]	-0.17317	0.15706	-1.103	.2702
b[y1]*	0.44729	0.13899	3.218	.0013
b[y2]	0.24059	0.16021	1.502	.1332
b[y3]*	0.81809	0.12873	6.355	< .000
b[y4]	-0.14654	0.14315	-1.024	.3060
w[z1]*	-0.49822	0.04633	-10.753	< .000
w[z2]*	-1.50097	0.08555	-17.545	< .000
Model II				
	Estimate	Std. Error	z value	P(> z)
a[(Intercept)]	-0.8868	0.6240	-1.4213	0.1552
a[x1]	-0.9109	0.2564	-3.5527	0.0004
a[x2]	0.7251	0.2191	3.3100	0.0009
a[x3]	0.1584	0.1735	0.9130	0.3613
a[y1]	0.0653	0.0933	0.6998	0.4841
a[y2]	-0.0341	0.1099	-0.3108	0.7560
a[y3]	0.1474	0.1318	1.1182	0.2635
a[y4]	0.0590	0.1149	0.5131	0.6079
b[(Intercept)]	0.1572	0.9575	0.1642	0.8696
b[x1]	0.0317	0.3053	0.1039	0.9173
b[x2]	-0.3197	0.2690	-1.1883	0.2347
b[x3]	-0.2083	0.2428	-0.8581	0.3909
b[y1]	0.4379	0.1389	3.1528	0.0016
b[y2]	0.2794	0.1721	1.6235	0.1045
b[y3]	0.7788	0.1983	3.9266	0.0001
b[y4]	-0.1432	0.1701	-0.8419	0.3999
w[(Intercept)]	-0.0976	0.1046	-0.9328	0.3509
w[z1]	0.4963	0.0493	10.0720	< .000
w[z2]	1.5186	0.0897	16.9330	< .000

Table 4: Coefficient summary table for synthetic data example obtained with `summary`. Significant parameters are indicated with an asteriks (*). *Model I*: $y \sim z1 + z2 - 1$, $\alpha \sim x1 + x2 + x3 - 1$, $\beta \sim y1 + y2 + y3 + y4 - 1$. *Model II*: $y \sim z1 + z2$, α , $\beta \sim x1 + x2 + x3 + y1 + y2 + y3 + y4$.

	R^2	AIC	AICc	BIC
Linear model	0.5381	131.72	137.37	150.84
Logist model	0.8485	92.17	114.23	126.58
Cusp model	0.7821	80.60	105.93	116.93

Table 5: Model fit statistics for [Oliva et al. \(1987\)](#) synthetic data example obtained with `summary` for the less informed model (*Model II*; see text for details). *Note*: “ R^2 ” value for cusp is pseudo- R^2 .

distributed on $(-3, 3)$, and Y_3 and Y_4 were uniformly distributed on $(-5, 5)$. The states y_i were then generated from the cusp density with their respective α and β as normal and splitting factors, and then Z_2 was computed as $Z_{2i} = (y_i + 0.52 Z_{1i})/(-1.60)$.

The call for fitting the model of [Oliva et al.](#) for the resulting data set is

```
R> oliva.fit <- cusp(y ~ z1 + z2 - 1,
+   alpha ~ x1 + x2 + x3 - 1,
+   beta ~ y1 + y2 + y3 + y4 - 1,
+   data = oliva)
```

Note that in the model there are no intercept coefficients; this is signified in the `cusp` call by the `-1` added to the formula’s (see the formula section of the R Manual, [R Development Core Team 2009](#) for details). The data are stored in the data frame `oliva` in this case. The result from `summary(oliva.fit, logist = TRUE)` is displayed in the upper part of Table 4.

Some of the estimates differ substantially from the true values. This should however not be too surprising given that there are 9 estimated parameters and only 50 observations, yielding a ratio of $55/9 \approx 5.5$ observations per parameter. Six coefficients differ significantly from zero: a_{x1} , a_{x2} , b_{y1} , b_{y3} , and both w_{z1} and w_{z2} . Using `confint` to calculate confidence intervals, we obtain for a_{x1} , a_{x2} , b_{y1} , b_{y3} , w_{z1} and w_{z2} 95% confidence intervals of respectively $(-1.26, -0.3979)$, $(0.2864, 1.0676)$, $(0.1749, 0.7197)$, $(0.5658, 1.07)$, $(0.4074, 0.589)$ and $(1.333, 1.669)$, which all neatly cover the true values. The same confidence interval for b_{y4} however, is $(-0.4271, 0.134)$, which does not contain the true value. This may indicate that the estimator is biased. Indeed [Hartelman \(1997\)](#) showed in simulations that the estimators are in fact biased. The control factors (α_i ’s and β_i ’s) were however recovered quite decently: Correlations between the actual α_i ’s and those predicted by the model fit (which can be obtained with the function `predict`) was .996, and the correlation between actual β_i ’s and those predicted by the model fit was .924.

In contrast to the model specified in this fit, which is rather informed on the variables and their role in the cusp catastrophe, it is often the case that one doesn’t know which variables determine the splitting factor, and which variables determine the normal factor. We therefore also fitted a less informed model, in which both `alpha` and `beta` are specified as `x1 + x2 + x3 + y1 + y2 + y3 + y4`. That is, all independent variables are used to model both the splitting as well as the normal factor. The resulting estimates are given in the second part of Table 4. The same set of estimates differs significantly from zero, demonstrating the usefulness of having significance tests for each estimate separately. The confidence intervals all covered the true parameter value, except for the same parameter as in the informed model.

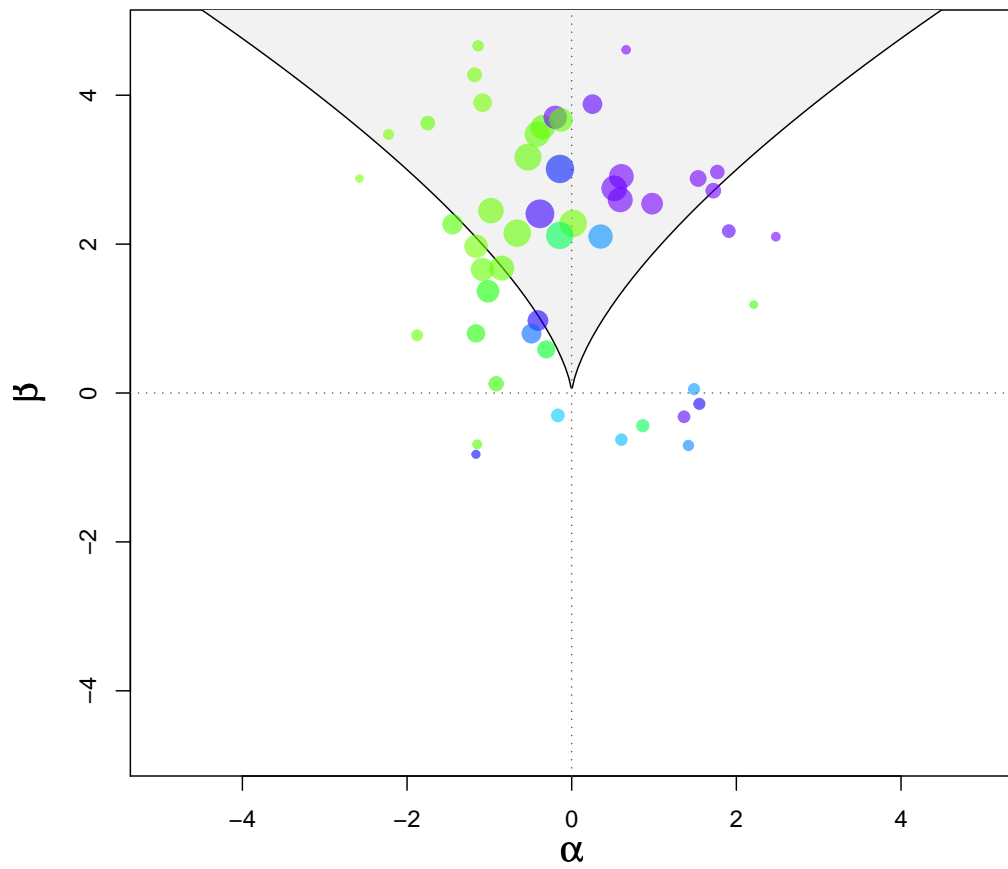


Figure 4: Fit of simulated data generated conform the [Oliva *et al.* \(1987\)](#) synthetic data example.

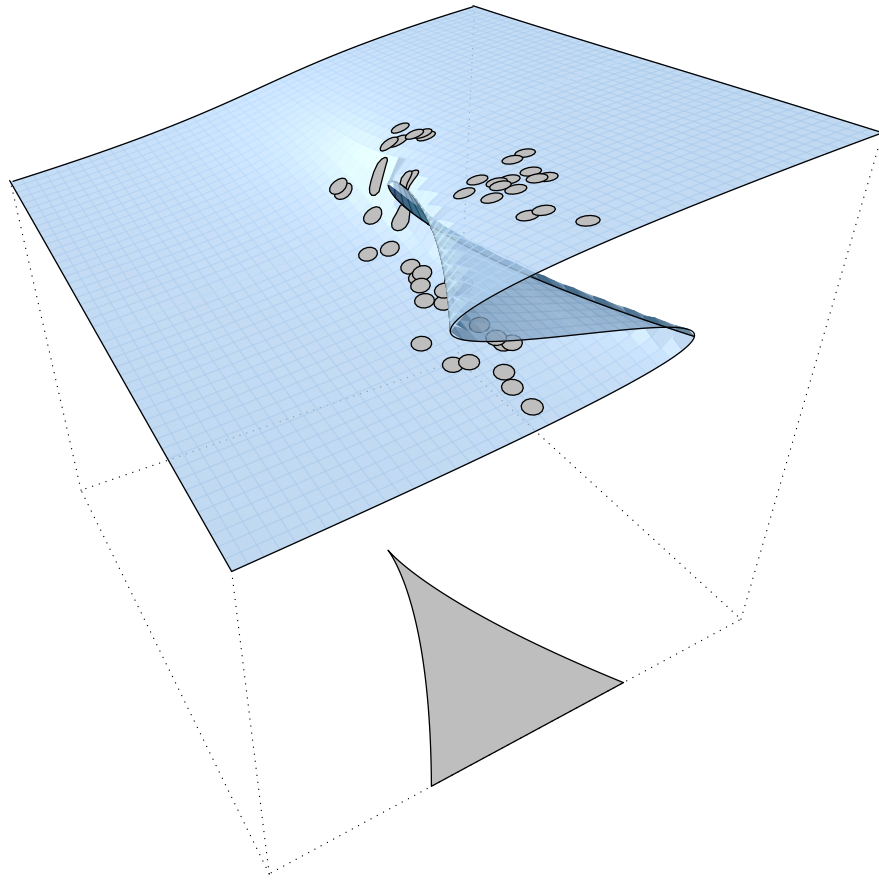


Figure 5: Three dimensional display of the fit of the [Oliva *et al.* \(1987\)](#) synthetic data example generated with the `cusp3d` function.

The fit indices for this less informed model are displayed in Table 5. Note that the pseudo- R^2 statistics indicates that the fit does not differ substantially between the cusp model and the logistic curve model. In fact, it even indicates that the logistic curve model gives a slightly better fit! Thus clearly, the pseudo R^2 is not in all cases a trustworthy guide in selecting a model, and since we are usually unaware of what the “correct” model is, this makes it difficult to rely on it at all. The AIC, AICc, and BIC on the other hand all (correctly) indicate that the cusp is the best model (of the ones compared) for these data. The chi-square Likelihood Ratio test given in the program output indicated that the cusp model fit was significantly better than the linear model with normal errors ($\chi^2 = 68.71$, $df = 9$, $p < .000$).

A control plane scatter plot of the synthetic data for the [Oliva *et al.*](#) model fit is presented in Figure 4. A three dimensional display of the model fit as generated with `cusp3d` is presented in Figure 5. A couple of things may be noted from the scatter plot: First of all, the sizes of the dots differ. In fact the size of the dots, each of which corresponds to a single case, varies according to the observed bivariate density of the control factor values at the location of the point. A second observation is that cases that lie inside the bifurcation set are plotted indiscriminately of whether they lie on the upper surface or on the lower surface. To make it possible to visually distinguish these cases, the color of the points vary according to the value of the state variable; higher values are associate with more intense purple, lower values with more intense green.

5. Discussion

In this paper we have presented an add-on package for cusp catastrophe modeling in R. The core user interface functions allow the user to easily specify and fit a broad range of models using the maximum likelihood approach of [Cobb \(1980; Cobb and Watson 1980\)](#). It also provides the user a number of tools for the assessment of the cusp catastrophe model fit.

Although the focus of this paper was entirely on quantitative methodology, by no means do we consider the statistically satisfactory fit of a cusp catastrophe model—or any catastrophe model for that matter—definite evidence for the for the presence of dynamical phase transitions. Without concurrent qualitative assessment of the presence of catastrophe flags implied by the model, and without a sound theoretical framework that implicates an underlying gradient system near equilibrium states, any catastrophe model fit remains unconvincing.

[Cobb’s](#) method is not without it limitations. Deterministic catastrophe classifies systems up to a set of smooth nonlinear scalings of the state variables. [Hartelman \(1997\)](#) points out that this poses a problem for [Cobb’s](#) approach. The alternative approach offered in [Hartelman \(1997; see also Wagenmakers *et al.* 2005a\)](#) however, only works for time series. Practical experience with applications to cross-sectional data indicates that [Cobb’s](#) method is suitable for these cases.

A number of improvement of the package can be thought of. The current package tests for the presence of bifurcation points by fitting a logistic curve that accommodates arbitrarily sudden changes in the state variable as a function of smooth gradual changes in the control variables. More ideally, to test for the presence of bifurcation points one should minimize the negative log-likelihood L of equation (11) subject to the n constraints $\delta_i = \alpha_i^2/4 - \beta_i^3/27 \leq 0$. The parameter δ_i is known as Cardan’s Discriminant, named after the 16th century Italian mathematician who first published it ([Cobb and Zacks 1985](#)). It is positive, only if the

cuspid equation (3) has three solutions—i.e., only if there are multiple equilibrium states. One then can use likelihood ratio chi-square testing to compare this model with the unconstrained model. This could be a viable alternative approach to the logistic curve fit, but it would require a different optimization routine. Unfortunately, R currently has no optimization routine that allows for arbitrary nonlinear inequality constraints.

One might further consider a Bayesian approach to estimation. The Bayesian approach however, comes of course at the cost of having to specify a prior belief about the likelihood of parameter values. Given the relations between parameters $w_0, w_1, \dots, b_{q-1}, b_q$ and the data in equations (8–10) the latter seems rather laborious in the most general case, and a far from intuitive enterprise. Only (relatively) uninformative priors seem straightforward, hence leaving out the heart of a true Bayesian approach.

These considerations will be explored in future improvements of the package.

Acknowledgments

Preparation of this article was sponsored in part by a VENI-grant and a COF-grant from the Netherland Organisation for Scientific research (NWO).

References

- Alexander RA, Herbert GR, DeShon RP, Hanges PJ (1992). “An Examination of Least-Squares Regression Modeling of Catastrophe-Theory.” *Psychological Bulletin*, **111**(2), 366–374.
- Burnham K, Anderson D (2004). “Multimodel Inference: Understanding AIC and BIC in Model Selection.” *Sociological Methods & Research*, **33**, 261–304.
- Cobb L (1980). “Estimation Theory for the Cusp Catastrophe Model.” *Proceedings of the Section on Survey Research Methods*, pp. 772–776.
- Cobb L (1981). “Parameter Estimation for the Cusp Catastrophe Model.” *Behavioral Science*, **26**(1), 75–78.
- Cobb L (1998). “An Introduction to Cusp Surface Analysis.” *Technical report*, Aetheling Consultants, Louisville, CO, USA. URL <http://www.aetheling.com/models/cusp/Intro.htm>.
- Cobb L, Koppstein P, Chen N (1983). “Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family.” *Journal of the American Statistical Association*, **78**(381), 124–130.
- Cobb L, Ragade RK (1978). “Applications of Catastrophe Theory in the Behavioral and Life Sciences.” *Behavioral Science*, **23**, 291–419.
- Cobb L, Watson B (1980). “Statistical Catastrophe Theory: An Overview.” *Mathematical Modelling*, **1**(4), 311–317.

- Cobb L, Zacks S (1985). “Applications of Catastrophe Theory for Statistical Modeling in the Biosciences.” *Journal of the American Statistical Association*, **80**(392), 793–802.
- Fürstenau N (2006). “Modelling and Simulation of Spontaneous Perception Switching with Ambiguous Visual Stimuli in Augmented Vision Systems.” *Perception and Interactive Technologies*, **4021**, 20–31.
- Gilmore R (1993). *Catastrophe Theory for Scientists and Engineers*. Dover, New York, NY, USA.
- Grasman RPPP, van der Maas HLJ, Wagenmakers EJ (2009). “Fitting the Cusp Catastrophe in R: A **cusp** Package Primer.” *Journal of Statistical Software*, **32**(8), 1–27. URL <http://www.jstatsoft.org/v32/i08/>.
- Guastello SJ (1982). “Moderator Regression and the Cusp Catastrophe – Application of 2–Stage Personnel-Selection, Training, Therapy, and Policy Evaluation.” *Behavioral Science*, **27**(3), 259–272.
- Guastello SJ (1988). “Catastrophe Modeling of the Accident Process: Organizational Subunit Size.” *Psychological Bulletin*, **103**(2), 246–255.
- Guastello SJ (1992). “Clash of the Paradigms – A Critique of an Examination of the Polynomial Regression Technique for Evaluating Catastrophe–Theory Hypotheses.” *Psychological Bulletin*, **111**(2), 375–379.
- Hartelman PAI (1997). *Stochastic Catastrophe Theory*. Ph.D. thesis, University of Amsterdam, Amsterdam, the Netherlands.
- Hartelman PAI, van der Maas HLJ, Molenaar PCM (1998). “Detecting and Modelling Developmental Transitions.” *British Journal of Developmental Psychology*, **16**, 97–122.
- Isnard C, Zeeman E (1976). “Some Models from Catastrophe Theory in the Social Sciences.” In L Collins (ed.), *The Use of Models in the Social Sciences*. Tavistock, London, UK.
- Lange R, Oliva TA, McDade S (2000). “An Algorithm for Estimating Multivariate Catastrophe Models: GEMCAT II.” *Studies in Nonlinear Dynamics & Econometrics*, **4**(3), 169–182.
- Oliva T, Desarbo W, Day D, Jedidi K (1987). “GEMCAT: A General Multivariate Methodology for Estimating Catastrophe Models.” *Behavioral Science*, **32**(2), 121–137.
- Poston T, Stewart I (1996). *Catastrophe Theory and Its Applications*. Dover, New York.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rosser J (2007). “The Rise and Fall of Catastrophe Theory Applications in Economics: Was the Baby Thrown Out with the Bathwater?” *Journal of Economic Dynamics and Control*, **31**(10), 3255–3280.
- Seber GAF, Wild CJ (1989). *Nonlinear regression*. Wiley, New York, NY, USA.

- Stewart IN, Peregoy PL (1983). “Catastrophe–Theory Modeling in Psychology.” *Psychological Bulletin*, **94**(2), 336–362.
- Sussmann HJ, Zahler RS (1978). “Catastrophe Theory as Applied To the Social and Biological Sciences: A Critique.” *Synthese*, **37**(2), 117–216.
- Ta’eed LK, Ta’eed O, Wright JE (1988). “Determinants Involved in the Perception of the Necker Cube: An Application of Catastrophe Theory.” *Behavioral Science*, **33**(2), 97–115.
- Thom R (1973). *Structural Stability and Morphogenesis: Essai D’une Théorie Générale Des Modèles*. W. A. Benjamin, California.
- Thom R, Fowler DH (1975). *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*. W. A. Benjamin, Michigan.
- van der Maas HLJ, Kolstein R, van der Pligt J (2003). “Sudden Transitions in Attitudes.” *Sociological Methods & Research*, **23**(2), 125–152.
- van der Maas HLJ, Molenaar PCM (1992). “Stagewise Cognitive Development: An Application of Catastrophe Theory.” *Psychological Review*, **99**(3), 395–417.
- Wagenmakers EJ, Molenaar PCM, Grasman RPPP, Hartelman PAI, van der Maas HLJ (2005a). “Transformation Invariant Stochastic Catastrophe Theory.” *Physica D*, **211**, 263–276.
- Wagenmakers EJ, van der Maas HLJ, Molenaar PCM (2005b). “Fitting the Cusp Catastrophe Model.” In BS Everitt, DC Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*, volume 1, pp. 234–239. Wiley, Chichester, UK.
- Zeeman E (1973). “Catastrophe Theory in Brain Modelling.” *International Journal of Neuroscience*, **6**, 39–41.
- Zeeman E (1974). “On the Unstable Behavior of the Stock Exchanges.” *Journal of Mathematical Economics*, **1**, 39–44.
- Zeeman EC (1971). “The Geometry of Catastrophe.” *Times Literary Supplement*, pp. 1556–1557.
- Zhu C, Byrd R, Lu P, Nocedal J (1997). “L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization.” *ACM Transactions on Mathematical Software*, **23**(4), 550–560.

A. Other estimation methods

Different fitting techniques for the cusp catastrophe, and for catastrophe models in general, have been proposed in the literature. Two of the most prominent seem to suffer from a number of problems, one of which is that the “anti-predictions” of the cusp model are not taken into account. The polynomial regression technique of [Guastello \(1988\)](#) approximates Cobb’s stochastic form of equation (1) by a difference equation which essentially results in a polynomial regression equation. Least squares regression is then used to estimate the

parameters. In this regression procedure however, the occurrence of unstable equilibrium states is rewarded just as much as stable states, rather than punished (Hartelman 1997).

Alexander *et al.* (1992) have furthermore criticized this approach because the dependent variable in the regression is a difference score of two variables, one of which is also present as a predictor. As a consequence, the explained variance can be up to 50% for completely random data. Thus the method cannot distinguish between a (cusp) catastrophe model and a linear model. In a reply to Alexander *et al.* (1992), Guastello (1992) demonstrates that his polynomial regression technique can give an R^2 estimate of 0.55 for purely random data. By constructing the bifurcation and asymmetry factors to be “known”, as the author calls it, “a near-perfect [i.e., $R^2 = 0.99$] cusp model could be obtained” (Guastello 1992, p. 387). On the basis of an, in our view misguided, interpretation of chaos theory, the author argues that from these examples it can be concluded that these random numbers conform to a fold- and even a cusp-catastrophe. Clearly, this cannot be a “cusp” in the sense of Cobb’s stochastic version of equation (3). The polynomial regression method of Guastello (1992) estimates the coefficients correctly if the data set is a time series, in which case we can show however, at least for an Ornstein-Uhlenbeck SDE, that theoretically R^2 ranges from 0 for closely spaced samples to 0.5 for widely spaced samples (when samples are nearly uncorrelated). For the cusp SDE a similar result can be demonstrated using simulation. A scenario under which the method would yield correct coefficient estimates *and* a high R^2 , is when multiple independent systems are observed on two occasions in which the systems are perturbed on the first occasion by at least two standard deviations of the equilibrium noise level.

The multivariate “GEMCAT” methodology of Oliva *et al.* (Oliva *et al.* 1987; Lange *et al.* 2000) assumes that the measured states are at the equilibrium surface and that any discrepancy is due to additive noise (much like Cobb’s stochastic extension of catastrophe theory but, as we shall see, not quite the same). Hence, GEMCAT simply minimizes the square of the left hand side of equation (3), summed across all observed states. The qualification “multivariate” refers to the fact that GEMCAT allows for the use of equation (5), whereas the technique of Guastello (1988) presumes that the state variable y is accessible for direct measurement, and is not, as implicated by equation (5), a set of dependent variables that “predict” the state. As is true for the method of Guastello (1988), a problem with this approach is that stable and unstable equilibrium states are treated indiscriminately, and both types of states are “rewarded” for their presence in the model fit. In fact, in the multivariate synthetic data example, for eleven of the cases in Table 2 of (Oliva *et al.* 1987), the simulated observed states were unstable (i.e., inaccessible) equilibrium states. This is opposite to the predictions from the cusp catastrophe that the unstable equilibrium state is unlikely to be observed, and contrasts with Cobb’s conception of stochastic catastrophe theory. Furthermore, as astutely noted in Hartelman (1997), the approach for model evaluation proposed in Oliva *et al.* (1987) is not valid because their stochastic counterpart of equation (3) as a model, entails an *implicit equation* to which data should adhere—save for random disturbances. Unlike conventional non-linear regression, implicit equations allow for multiple predictions for each set of values of the predictor variables. Equation (3) and its stochastic counterpart thus do not constitute non-linear regression in the conventional sense. Even more importantly, the conditions under which (asymptotic) statistical inference theory is developed for such implicit models (Seber and Wild 1989) precisely excludes those cases that are central to catastrophe theory (Hartelman 1997). As a consequence, the suggested chi-square statistics initially used in GEMCAT for model comparison and model selection are rendered invalid. In a new version of GEM-

CAT, GEMCAT II (Lange *et al.* 2000), this situation was changed and inference is based on resampling techniques (jackknife and non-parametric bootstrap).

In contrast to the other techniques which are solely based on the derivative in equation (1). The methods developed by Cobb and his colleagues are based on the density in (4) and therefore take into account all characterizing aspects of the system's potential function V . For instance, where the two previous methods "reward" the presence of unstable equilibrium states, the maximum likelihood approach of Cobb *et al.* (1983) punishes for their presence, as these correspond to points in an area of the density function of low probability, that lies in between two high probability states.

Affiliation:

Raoul Grasman
Department of Psychology
University of Amsterdam
1018KP Amsterdam, the Netherlands
E-mail: rgrasman@uva.nl
URL: <http://users.fmg.uva.nl/rgrasman>