Proceedings of TMCE 2012, May 7–11, 2012, Karlsruhe, Germany, edited by I. Horváth and Z. Rusák

SENSITIVITY ANALYSIS OF OPTIMIZED CURVE FITTING TO UNIFORM-NOISE POINT SAMPLES

Oscar Ruiz
CAD CAM CAE laboratory
EAFIT University, Medellin, Colombia
oruiz@eafit.edu.co

Camilo Cortes
CAD CAM CAE laboratory
EAFIT University, Medellin, Colombia
ccortes@eafit.edu.co

Diego Acosta
DDP laboratory
EAFIT University, Medellin, Colombia
dacostam@eafit.edu.co

Mauricio Aristizabal
CAD CAM CAE laboratory
EAFIT University, Medellin, Colombia
maristi7@eafit.edu.co

ABSTRACT

Curve reconstruction from noisy point samples is needed for surface reconstruction in many applications (e.g. medical imaging, reverse engineering, etc.). Because of the sampling noise, curve reconstruction is conducted by minimizing the fitting error (f), for several degrees of continuity (usually C^0, C^1 and C^2). Previous works involving smooth curves lack the formal assessment of the effect on optimized curve reconstruction of several inputs such as number of control points (m), degree of the parametric curve (p), composition of the knot vector (U), and degree of the norm (k) to calculate the penalty function (f). In response to these voids, this article presents a sensitivity analysis of the effect of m and k on f. We found that the geometric goodness of the fitting (f) is much more sensitive to m than to k. Likewise, the topological faithfulness on the curve fit is strongly dependent on m. When an exaggerate number of control points is used, the resulting curve presents spurious loops, curls and peaks, not present in the input data. We introduce in this article the spectral (frequency) analysis of the derivative of the curve fit as a means to reject fitted curves with spurious curls and peaks. Large spikes in the derivative signal resemble Kronecker or Dirac Delta functions, which flatten the frequency content ad-infinitum. Ongoing work includes the assessment of the effect of curve degree p on f for non-Nyquist point samples.

KEYWORDS

parametric curve reconstruction, noisy point cloud, sensitivity analysis, minimization

NOMENCLATURE

C_0	=	Unknown C^1-derivable simple planar curve
$C(u)$	=	Parametric planar curve approaching C_0
$C(u_i)$	=	Point on C(u) closest to cloud point p_i
$d(p, S)$	=	Distance from point p to the point set S
k	=	Degree of norm: $(\Sigma\|x_i\|^k)^{1/k}$
l	=	Length unit
m	=	Number of control points of $C(u)$
P	=	$[P_0, P_1, ..., P_{m-1}]$. Control polygon of $C(u)$
PCA	=	Principal Component Analysis
PL	=	Piecewise Linear
S	=	$\{p_0, p_1 ..., p_n\}$ Noisy point sample of C_0

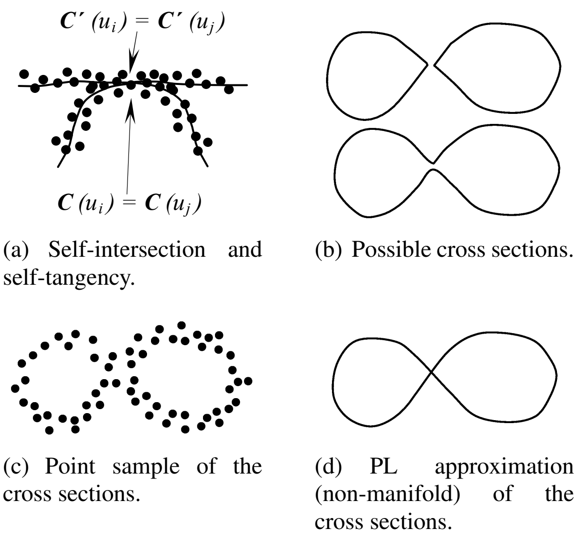
1. INTRODUCTION

Many engineering applications need to recover a planar curve from a noisy point sample. A possible approach is to fit a curve to the point set, recognizing the stochastic nature of the data. This approach consists in adjusting a parametric or implicit curve to the set of points, by minimizing the unsigned distance function between the points and their approximating curve. In the existing literature this approach is reported, in the form of heuristic - based experiments. The heuristics used affect the number of control points of the curve, its degree, the norm used to measure distances, the know vector for the parametric curve, etc. However, it must be noticed that a numerical systematic evaluation of the importance of these factors is not reported.

This present article presents a discussion of the optimality conditions of the curve fitting problem with b-splines (Piegl & Tiller, 1997) of the type Open, Uniform, degree 2, whose knot vectors were adjusted such that $0 \leq u \leq 1$. Our analysis and results apply for other curve types, although not with the same quantification. In this article we address point samples with uniform sampling noise, leaving spatial-dependent noise for future publications.

1.1. Curve self-intersection

Fig. 1 illustrates that, in the presence of sampling noise, it might be immaterial whether the sampled curve C_0 is self-intersecting or not. In the examples shown, the point sample will indicate a self intersecting curve in either case. We declare here that the issue of self-intersection of C_0 is outside the scope of this article and therefore we will consider open, simple (i.e. non self - intersecting) curves. We introduce the issue only for establishing a context for the term *self - intersection*.



(a) Self-intersection and self-tangency.

(b) Possible cross sections.

(c) Point sample of the cross sections.

(d) PL approximation (non-manifold) of the cross sections.

Figure 1 Ambiguous noise sample of near self-intersecting curves (Ruiz et al., 2011).

1.2. Objective function

In mathematical programming, an objective or cost function f is a function that represents how a dependent variable of a process (e.g., profit, cost, energy, etc) behaves in terms of a set of independent or decision variables. Depending on the nature of the optimization problem the objective function is maximized or minimized by tuning the independent variables.

In the context of reverse engineering the problem of parametric curve reconstruction from noisy point samples can be stated as follows:

Given an unknown target curve to reconstruct C_0, whose sampling (possibly noisy) constitutes a point set S, find a parametric curve C, which approximates C_0, by minimizing the distance between the curve and the elements of S. In general, the following expression is used to measure the fitting error and therefore is the objective function to be minimized:

$$f = \sum_{i=1}^{n} d_i^w \tag{1}$$

Where the residual d_i represents the minimum distance between the i-th cloud point and the curve C and w indicates the order of the residual. Then d_i is given by:

$$d_i = \min_{C(u) \in C} \|C(u) - S_i\|^k \tag{2}$$

k is the norm-degree to calculate the distance.

Sections 2.1 and 2.2 discuss the objective functions used in curve fitting and the strategies to calculate the residuals d_i.

1.3. Decision variables and parameters of the optimization problem

As seen in 1.2 there are plenty of terms involved in the calculation of Eq.1 that can be tunned to minimize it.

An option is to fit variables inherent to the definition of the parametric curve C.

If C is a b-spline curve, then m, P, the curve degree p and knot vector U can be adjusted to improve the fitting to S. In the literature the most common approach is the optimization of P, sometimes supplemented with the adjustment of U as found in (Ueng et al., 2007). On the other hand, the terms involved in the calculation of the objective function f such as k and w can be set to produce the desired results.

In our approach the decision variables are the control points P. All other terms remain constant and are considered parameters of the problem (i.e, norm k, number of control points m, knot vector U and curve degree p).

1.4. Constrains and degrees of freedom

Most optimization problems include some constrains on the decision variables, which bound the region of search of an optimal solution. In the context of curve fitting to noisy data sets some constrained approaches have been developed, in section 2.1 a brief reference to them is performed. In our implementation there are no constrains on the decision variables. As is discussed in the following sections this fact is decisive in the determination of the uniqueness of the solution and its global scope. This optimization problem is classified as Non-linear unconstrained.

The degrees of freedom G of an optimization problem are given by the subtraction of the number of the number of equality constrains E, from the number of decision variables V $(G = V - E)$. Optimization techniques are used to solve underdetermined systems, this means $G > 0$. Notice that for curve fitting problems in which only the control points are adjusted, the number of decision variables is $2m$ when implementing planar curves and $3m$ for the case of curves in the euclidean three-dimensional space. As our development is performed using planar curves and there are no equality constrains $G = 2m$.

1.5. Sensitivity analysis

This analysis consists of studying how the objective function behaves when the parameters are perturbed. The calculation of the relative sensitivity allows to determine which parameter influences f the most .

Let $F(K, Q)$ be the objective function of an optimization problem where K is a decision variable and Q is a parameter, then the relative sensitivity of $F(K, Q)$ with respect to Q, S_Q^F, as can be found in reference (Edgar et al., 2001), is given by:

$$S_Q^F = \frac{\partial F/F}{\partial Q/Q} = \frac{\partial ln(F)}{\partial ln(Q)} \tag{3}$$

The value of S_Q^F is the ratio between the percent change in F and the percent change in Q, which is dimensionless. For this reason it is possible to compare the relative effect of each parameter on the objective function.

When the required derivatives are difficult to calculate the sensitivity must be calculated numerically as shown in (Nocedal & Wright, 2006; Fiacco, 1983). In this paper the sensitivity analysis is performed to determine the influence of the number of control points m and the norm k on f.

1.6. Convexity

The objective function and search region convexity determine the classification and scope of an optimization problem solution. Let $\vec{x}$ be the decision variables vector of an optimization problem with objective function f. Let $\vec{x^*}$ be such that $\nabla f(\vec{x^*}) = 0$, then the assessment of the convexity of f at $\vec{x^*}$ allows to determine if $f(\vec{x^*})$ is a local extremum. If the optimization problem includes equality and inequality constrains, the convexity of the bounded region must be verified in order to conclude the uniqueness of the extremum, as discussed in (Edgar et al., 2001).

In the case of the objective function, its convexity is evaluated using the eigenvalues of its Hessian matrix $H_f(\vec{x^*})$, which is defined in (Papadimitriou & Steiglitz, 1998) as:

$$H_f(\vec{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{ij} \tag{4}$$

The eigenvalues e obtained from solving $det\left[H_f(\vec{x^*}) - eI\right] = 0$ indicate whether the function is convex, non-convex or a saddle point at $\vec{x^*}$. Furthermore, if $f(\vec{x^*})$ is optimized on a closed convex region, global maximum and minimum will be calculated as shown in (Edgar et al., 2001).

2. LITERATURE REVIEW

Very few discussions about some important concepts inherent to mathematical optimization can be found in the curve fitting literature because most of the research is focused on exploring better algorithms to perform the optimization of the decision variables.

2.1. Objective function

As discussed in section 1.2 Eq.1 is the general representation of the objective function in curve fitting problems. Reference (Flöry & Hofer, 2010) employs first order residuals ($w = 1$) and references (Wang et al., 2006; Liu et al., 2005; Gálvez et al., 2007; Liu & Wang, 2008) use second order residuals ($w = 2$).

Some references add a smoothing term f_c to the objective function, in order to adjust the final roughness of the curve:

$$f = \sum_{i=1}^{n} d_i^w + \lambda f_c. \tag{5}$$

The term f_c may contain information on the curve's first and second derivatives as in (Wang et al., 2006; Liu et al., 2005; Flöry & Hofer, 2008) or only information of the former as in (Flöry & Hofer, 2010; Flöry, 2009) and λ determines its influence, penalizing large curvatures. Notice that penalizing the curvature prevents the curve fitting for non-Nyquist samples.

Some authors have explored constrained approaches, reference (Flöry, 2009) presents constrained curve and surface fitting to a set of noisy points in the presence of obstacles, which are regions that the curve or surface must avoid. Reference (Flöry & Hofer, 2008) considers the problem of curves that must lie on a 2-manifold (surface), also with forbidden regions. These procedures are implemented using a constrained non-linear optimization strategy.

2.2. Distance measurement

As seen in section 1.2 Eq.2 corresponds to the calculation of the distance d_i, which represents the residuals of the objective function in Eq. 1. In Curve Fitting algorithms the norm k is usually chosen to employ the Euclidean distance (Wang et al., 2006; Liu et al., 2005) ($k = 2$).

The exact calculation of d_i is expensive, since it is obtained by a minimization procedure at each fitting iteration. The procedure consists of finding the parameter u_i which associates a point on the curve $C(u_i)$ with the i-th cloud point p_i such that d_i is a minimum, namely

$$\|C(u_i) - p_i\|^k = \min_{C(u)\in C} \|C(u) - p_i\|^k \qquad (6)$$

The minimum distance is obtained performing an orthogonal projection of the point p_i to the curve C, which occurs when the dot product between the tangent vector at $C(u_i)$ and the distance vector d_i is null (see Eq. 7). Therefore, an alternative to face this problem is to solve for u in $g(u) = 0$ using the Newton's Method, as implemented in references (Piegl & Tiller, 1997; Liu et al., 2005). Other approach is to minimize $g(u)$, references (Wang et al., 2006; Liu & Wang, 2008), for curve fitting, and (Flöry & Hofer, 2010), for surface fitting, propose the use of Newton-like iterative schemes, while (Saux & Daniel, 2003) employs a gradient method. On the other hand, reference (Gálvez et al., 2007) implement a Genetic Algorithm to obtain the parameter u_i at each iteration of the minimization procedure.

$$g(u) = \left|C'(u) \cdot (C(u) - p_i)\right| \qquad (7)$$

The approaches previously mentioned have drawbacks inherent to numerical methods, such as the need of a good initial guess, poor convergence and stagnation at local minima. These issues may lead to poor approximations of the distance d_i yielding unsatisfactory results of the fitting procedure.

Methodologies to automatically calculate an adequate initial guess for u in $g(u) = 0$ have been presented based on the point cloud subdivision using quadtree (Wang et al., 2006), k-D tree (for general dimensional fitting (Liu & Wang, 2008)) and Euclidean minimum spanning tree (Liu et al., 2005) strategies.

Different methodologies to measure the point-to-curve distance have been proposed, which are summarized as follows: (i) Point distance, which preserves the Euclidean distance between the cloud point and the paired point of the curve, discussed in (Wang et al., 2006; Flöry & Hofer, 2008), (ii) Tangent distance, which only preserves the distance between the cloud point and the tangential line projected at the paired point (Blake & Isard, 1998), and (iii) Squared distance, which is a curvature-based quadratic approximation of d_i^2 (Wang et al., 2006). Reference (Liu & Wang, 2008) presents a deep comparison between these methodologies.

It must be also remarked that using the point-to-curve distance does not make the method sensitive to curls and loops (formed outside the S boundaries) and outliers in the final curve C. Because of these reasons, in our work we have included both the point-to-curve and curve-to-point distance calculation (in approximate manner), giving emphasis instead to curl and outlier avoidance.

2.3. Optimality conditions and sensitivity analysis

Regarding the number of control points m, reference (Ueng et al., 2007) presents unconstrained and constrained approaches to solve the curve fitting problem to a set of low-noise organized data points using different values of m. The experiments performed show that increasing the number of control points helps, in general, to diminish f, although with the collateral effect of obtaining a more erratic curve.

Reference (Yang et al., 2004) shows similar results to (Ueng et al., 2007), with the difference that the insertion and removal of control points is part of their fitting strategy. If the local approximation of a parametric curve segment is poor, a new control point is added to it. On the other hand, when redundant control points are detected in a curve segment, the control points of that segment are removed one by one taking care of not producing fitting errors above a defined threshold.

In this approach the curve to reconstruct is comprised by ordered dense data points. When working with highly noisy unordered data new challenges arise. In particular, the problem of finding and adequate number of control points for correct geometry and topology reconstruction has not been discussed throughly. For other parameters such as the norm k, the reported researches are oriented to identify which norm to use when certain features such as outliers and particular noise distributions are present in the point data set.

Reference (Heidrich et al., 1996) performs a comparison amongst L_1, L_2 and L_∞ norms in curve fitting applications with several data sets. In reference (Flöry & Hofer, 2010) curve and surface fitting case studies are presented using the L_1 and L_2 norm when the data set contains outliers. It is concluded that L_1 norm is less sensitive to outliers, therefore better results are obtained.

In summary, few discussions are presented about the influence of m and k on the behavior of f. Furthermore, a formal sensitivity analysis for these parameters has not been performed yet, to the best of our knowledge. In addition, some features of the optimization problem have not been discussed, such as the objective function convexity, and its role in extrema characterization.

2.4. Peaks, curls and closed loops detection

A mathematically optimal solution for the fitting curve problem does not necessarily imply a correct topological and geometrical reconstruction of the curve C_0 represented by the point cloud S. Some strong oscillations may appear during the fitting process, such as peaks, curls and closed loops. When pursuing the reconstruction of smooth simple curves (i.e., non-self-intersecting) these features are undesirable and may be avoided by finding an optimal value for m, as shown in this paper, as opposed to the strategy of curvature penalization implemented in (Flöry, 2009; Liu et al., 2005; Wang et al., 2006; Flöry & Hofer, 2010).

The main drawback of the curvature penalization is that it is difficult to properly establish the weight λ with respect to the contributions of the distance residuals d_i in Eq.5, for each case study. Additionally, optimizing m results in an efficient use of the decision variables. Therefore, detection of peaks, curls and closed loops is necessary to find a reasonable number of control points.

In the literature, efficient methods to detect self-intersections can be found, covering the closed loops detection case. Reference (Pekerman et al., 2008) presents an algebraic approach to detect self-intersections solving $C(u) - C(v) = 0$, with u being different from v, and proposing a new function that does not contain zeros in this diagonal. In any case, peaks and curls detection is not a trivial and a method to detect all undesired features is necessary. In this paper we open the discussion of the use of the $C(u)$'s curvature information in the frequency domain to detect the presence of peaks, curls and closed loops.

2.5. Literature review conclusions and contribution of this article

According to the taxonomy conducted in this literature review, there are several issues that remain open in optimized curve fitting to point clouds. These subjects include: (a) Identification of the effect of the parameters such as the number of control points m, knot vector U and norm k in the curve fitting problem, (b) Detection of the presence of peaks, curls and closed loops in C to support the parameter optimal value identification and (c) Characterization of the curve fitting problem from the viewpoint of mathematical optimization.

In response to these issues, this article reports, in addition to formulating the optimization problem, the implementation of: (i) Sensitivity analysis of the number of control points m and norm k on f and (ii) Quantitative analysis of $C(u)$ curvature information in the frequency domain to detect the presence of peaks, curls and closed loops.

3. METHODOLOGY

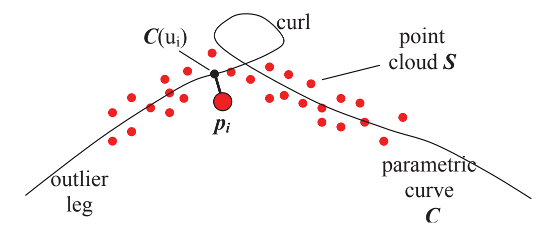
3.1. Dual distance calculation

In addition to the point-to-curve distance introduced in section 2.2 the curve-to-point distance is used to calculate the distance d_i used in Eq.1, for the implementation of the curve fitting algorithm used in this research.

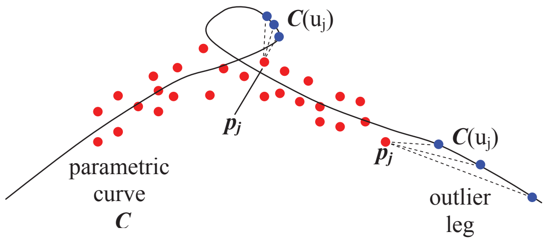
When implementing the point-to-curve distance we define the residuals as

$$d_i = ||p_i - C(u_i)||^k \tag{8}$$

where u_i is the parameter in the domain of C which



(a) Distances cloud point to curve.



(b) Distances curve to cloud point.

Figure 2 Distances cloud points to/from curve.

defines the point $C(u_i)$ closest to p_i. The term d_i represents the distance measured from each cloud point to the curve C (see Fig. 2(a)). This calculation of the distance between a point and an algebraic curve is a very expensive proposition because it implies the calculation of common roots of a polynomial ideal (see (Ruiz & Ferreira, 1996), (Kapur & Lakshman, 1992)).

Notice that the vector $p_i - C(u_i)$ is normal to the curve C at the point $C(u_i)$. To avoid the computational expenses of algebraic root calculation, we approximate $C(u)$ in PL manner and calculate d_i simply by an iterative process. We sample the domain for $C(u)$, ($[0,1]$) getting $U = [0, \Delta_u, 2\Delta_u, ..., 1.0]$ and approximate the current C curve with the poly-line $[C(0), C(\Delta_u), C(2\Delta_u), ..., C(1.0)]$. Calculating an approximation of $C(u_i)$ for a given p_i simply entails traversing $[C(0), C(\Delta_u), C(2\Delta_u), ..., C(1.0)]$ to find the $C(\kappa\Delta_u)$ closest to p_i.

Fig. 2(a) displays the distance from a particular (emphasized) cloud point p_i to its closest point $C(u_i)$ on the current curve C. Such a distance has influence in f as per Equation 1. Notice, however, that p_i and $C(u_i)$ (and hence f) do not change if large legs and curls appear in the synthesized C. Therefore, considering only the distance *from cloud points to the curve* in Eq.1 allows the incorrect formation of outlier legs and curls outside the boundaries of S.

If one can make spurious legs and curls to inflate the objective function f, the minimization of f avoids them. This is achieved by including the distances from *the curve points C_i to the cloud points p_i* (see Fig. 2(b)) to penalize in f.

For any point $p \in \mathbb{R}^n$, the distance of this point to S is a well defined mathematical function: $d(p, S) = \underset{p_j \in S}{min}(||p - p_j||^k)$. For the current discussion the points p are of the type $C(u_i)$ (i.e. they are points of curve C). The u_i parameters to use are the sequence $U = [0, \Delta_u, 2\Delta_u, ..., 1.0]$, already mentioned.

Notice that $d(p, S) = ||p_j - p||^k$ for some cloud point $p_j \in S$. Let us define the point set A_j (on the curve C) as:

$$A_j = \{C(u) | u \in U \wedge d(C(u), S) = ||p_j - C(u)||^k\} \quad (9)$$

The set A_j contains those points in the sequence $[C(0), C(\Delta_u), C(2\Delta_u), ..., C(1.0)]$ that are closer to the point $p_j \in S$ than to any other point of S. We note with M_j the cardinality of A_j. Observe that some M_j might be zero, since p_j could be far away from be curve C and no point on the curve would have p_j as its closest in S. The set of all A_js could also be understood as a partition of the curve C.

With the previous discussion, a new definition of the residuals d_i, to be used in Eq. 1, is possible:

$$d_i = ||p_i - C(u_i)||^k + (\frac{1}{M_i}) \sum_{C_v \in A_i} ||C_v - p_i||^k \quad (10)$$

The $||p_i - C(u_i)||^k$ in Eq. 10 considers the distance from cloud points in S to the curve C. The term $(\frac{1}{M_i}) \sum_{C_v \in A_i} ||C_v - p_i||^k$ expresses distances from the curve C to the cloud points in S. This term penalizes the length of the curve, by increasing f.

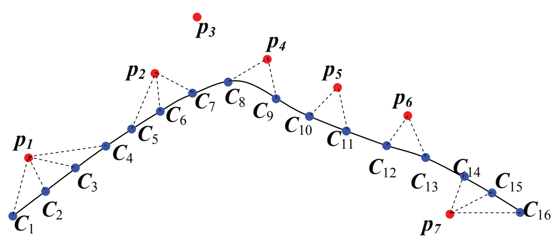


Figure 3 Clusters of distances from curve to cloud points.

Fig. 3 presents a rather simplified materialization of the situation, with very few cloud points, moreover biased with respect to the instantaneous C curve.

However, it serves the purpose of illustrating the algorithm.

3.2. Convexity

Since the optimization problem that is attacked in this article has no constrains, only the convexity of the objective function is analyzed. The variables to minimize f are the x and y coordinates of the control points ($P_j = (x_j, y_j)$) contained in the control polygon $P = [P_0, P_1, ..., P_{m-1}]$, therefore, the corresponding Hessian matrix is given by:

$$H_f(P) = \left[\frac{\partial^2 f}{\partial P_i \partial P_j}\right]_{ij} \tag{11}$$

with

$$\frac{\partial^2 f}{\partial P_i \partial P_j} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_i \partial x_j} & \frac{\partial^2 f}{\partial x_i \partial y_j} \\ \frac{\partial^2 f}{\partial y_i \partial x_j} & \frac{\partial^2 f}{\partial y_i \partial y_j} \end{bmatrix} \tag{12}$$

Numerical differentiation is implemented for the calculation of $H_f(P)$, using approximations for the second order and mixed derivatives. Next the eigenvalues of $H_f(P)$ are computed and the convexity or concavity of f is evaluated as discussed in 1.6.

3.3. Sensitivity calculation

In order to calculate the relative sensitivity of the objective function with respect to the parameters on a defined domain of each one of them, the curve fitting problem is solved for a set of parameter values. For a given value of the parameter, m_i or k_i (when computing S_m^f or S_k^f respectively) the curve fitting problem is solved obtaining a value of the objective function f_i. Notice that i indicates the number of increments applied over an initial value of the parameter of interest and goes from 0 to a defined number of maximum increments Max_{inc}.

Next, the sensitivity at an increment i ($0 \leq i < Max_{inc} - 1$) is numerically calculated by using the values of the parameter and objective function at i and $i + 1$ as discussed in 1.5. While calculating S_m^f, k is kept constant. Similarly, when computing S_k^f, m is kept constant.

The steps of this procedure are summarized as follows:

1. *initialization*: The point cloud S, to be fitted is loaded. Depending on the sensitivity to be computed m_0 or k_0 is assigned with an initial value.
2. *initial guess calculation*: A straight line L is used as an initial guess for the b-spline curve C. This allow us to provide an initial guess whose topology and shape is not affected by the number of control points used to build it. L is obtained performing a principal component analysis (PCA) (see reference (Ruiz et al., 2011)) over the whole S set. In the case of computing S_m^f the m_i control points are placed, equally-spaced, along L. When calculating S_k^f, the number of control points is kept constant, so that the initial guess will be the same during the whole procedure.
3. *curve fitting*: A penalized Gauss-Newton algorithm is used to perform the adjustment of P. Once the stopping criteria is met, the value of f is saved as f_i.
4. *parameter value increment*: The parameter value increment is defined as: m_{i+1}=m_i+1, for the number of control points and $k_{i+1} = k_i + \Delta k$, for the norm. Where Δk is an arbitrary small constant value.
   Only when $i < Max_{inc}$ the increment is performed and the process goes back to step 2, otherwise the procedure finishes.

In this article we analyze the sensitivity only with simple curves, ignoring for the time being self-intersecting curves.

3.4. Peaks, curls and closed loops detection approach

When solving the curve fitting problem some undesired features, as closed loops, curls and peaks, may appear in the adjusting curve for certain configurations of parameters. The length and curvature of C can provide some information about the topology and geometry of the curve, but without knowing what are the expected values for this metrics, it is difficult to conclude from them the correct topological reconstruction of S and therefore to establish an optimum value of the parameters of the problem. These reference values may be expensive to obtain in the preprocessing stage from the cloud point.

Adding information about the first and/or second derivatives of the curve to f helps to obtain smooth fitting curves, however weighting factors between the contributions of the distance residuals (2) and curve derivatives to f must be defined, interactively, for every case of study (see references (Liu et al., 2005) and (Flöry, 2009)) as a consequence of not having benchmark values for these measurements. Therefore it is desirable to devise a method to determine the presence of curls and peaks in the fitting curve, without the high overheads derived form extracting reference values form the point cloud.

In this paper we propose to perform an analysis of the frequency spectrum of certain information of the fitting curve that reflects the presence of the undesired features previously mentioned. The representation of the data in the frequency domain indicates how it is composed of low and high frequency waves, making easier to establish whether the curve follows the desired behavior or not. This task is accomplished by studying the changes of direction of the first derivative of the curve with respect to its parameter u. Peaks and curls produce large sudden changes in the direction of $\frac{\partial \vec{C}}{\partial u}$ that are represented in the frequency domain with a considerable presence of high frequency sinusoidal curves.

We have obtained the frequency spectrum computing the discrete Fourier transform (DFT) of the previously mentioned data. To guarantee that the desired information is sampled according to the Nyquist criterion, for this process, is of prime importance. To achieve this, we have chosen a series of u parameters that are located at equal distances d_s, of each other, on the curve, constituting $u_s = \{u_0, \ldots, u_g\}$. We have chosen $d_s = 0.0001l$, where l is the unit of distance, thus the sampling frequency is $f_s = 10000l^{-1}$.

Next, the normalized tangent vectors of the curve were computed at all the points given by the parameters u_s obtaining $V_s = \left\{\hat{V}_0, \ldots, \hat{V}_g\right\}$. The dot product of every $\hat{V}_i$ and $\hat{V}_{i+1}$ pair is computed, where $0 \leq i \leq g-1$, and therefore the angle θ_i between them is obtained. Finally the magnitude of the DFT is obtained, and properly scaled to achieve a single-sided spectrum of power vs. frequencies of the obtained history of θ.

4. RESULTS AND DISCUSSION

4.1. Test point set

The point cloud shown in Fig.4 was used to run the procedures discussed here and in the following sections. As in the sensitivity experiments, the fitting curve initial guess used here was a straight line obtained from a PCA of the complete point cloud. The Hessian matrix $H_f(P)$ and its eigenvalues e were calculated in every iteration of the optimization procedure using 5, 8, 9 and 15 control points.

4.2. Convexity

As discussed before, the region of search and objective function convexity is a necessary condition to claim the global scope of a solution of a minimization problem. Since the problem we deal with is unconstrained, the region of search is unbounded and its convexity can not even be verified. Therefore, by definition, the solutions found from the minimization procedure can only be classified as local. However the behavior of f is still of interest, since it hints to possible better results to be obtained .

For all these cases of study (5, 8, 9 15 control points) the convexity of f depends on the location of the control points P used to calculate $H_f(P)$, given that the eigenvalues obtained did not comply with the condition $e_j \geq 0\ \forall j$, where $1 \leq j \leq 2m$, at certain iterations of such tests. Therefore, no unique extremum exists and only convergence to a local minimum can be guaranteed.

Because of the behavior of f it is of prime importance to provide an initial guess for C close to a satisfactory solution avoiding large optimization times and stagnation in local minimum with poor topological and geometrical reconstruction.

4.3. Number of control points sensitivity calculation

The process was run twice with a number of control points ranging between 4 and 16, using both norms, L_1 and L_2. The results of S_m^f are summarized in Fig.5(b), where can be noticed that as m increased f became less sensitive to it, specially when using L_2 norm.

In addition to the value of f and S_m^f, the curve length and curvature were calculated to obtain information about the topology of the fitting curve (i.e, curls, peaks, long legs, etc). In this paper what is presented as curvature corresponds to the sum of the curvature at a determined number of samples along the curve.

The results show a general trend in which as the number of control points increases the objective function decreases (see Fig.5(a)). However, with exaggerated number of control points given a topological situation, some undesirable features begin to appear, such as curve roughness, curls and/or peaks and attraction among control points. These outcomes were obtained with both of the norms tested (i.e., $k = 1$ and $k = 2$). In Fig.6 the resulting curves of the fitting with different number of control points are shown using L_1 and L_2 norms.

For the particular point cloud used in these tests the minimum number of control points to reconstruct its topology is 5. Increasing the number of control points does not yield in a considerable diminution of f and more importantly, a better topological reconstruction of the curve is not necessarily obtained.

With the usage of the dual distance in f (see 1.2), the peaks and curls that appear are located inside the boundaries of the point cloud S, and what they produce is a reduction of f. The excessive amount of degrees of freedom of the curve allows the appearance of these curls and peaks, as the optimization algorithm place the control points minimizing the objective function.

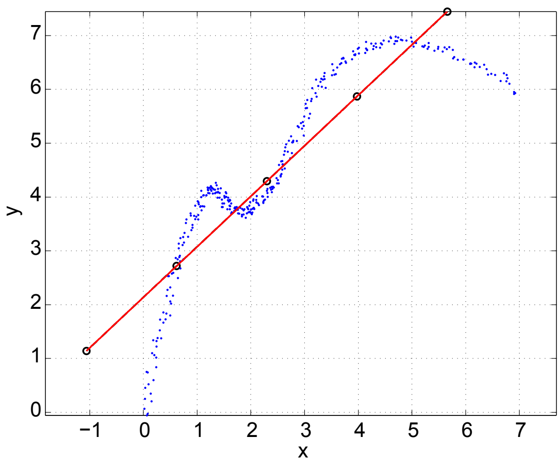


Figure 4 Point cloud and initial curve guess with five control points.

4.4. Norm sensitivity calculation

The relative sensitivity S_k^f was calculated from norm $k = 1$ to $k = 2$ with $\Delta k = 0.01$. The test was run twice, initially, with 5 control points, which are the minimum number of control points to successfully fit S, as determined in section 4.3, and then with 8 control points. The results show that small changes in the norm k may produce large relative changes in the values of f obtained in the domain defined by $1 \leq k \leq 2$ for both of the m tested. Figures 7(a) and 7(b) show an irregular behavior from which no particular value or range of k derive a remarkable improvement of the fitting.

It must be considered that even if the same curve topology and geometry are obtained implementing two different norms, the value of f will be distinct for each case, due to the modification in the residuals calculation in Equation 2. This fact magnifies the effect that k has on f and is the reason of the variability observed. However with respect to the quality (e.i., topology and geometry) of the curves obtained along the procedure (see Fig.8), the influence of k is almost imperceptible, when m is chosen properly.

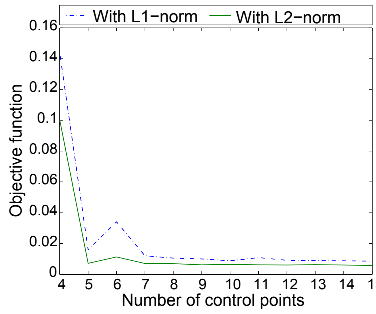
The curve length and curvature in figures 7(c) and 7(d) reflect a very stable behavior as k changes using 5 control points. The outlier curve segments observed in Fig.8 when $m = 5$ can be adjusted changing the stopping criteria of the optimization algorithm, so a few more iterations are performed. On the other hand, using 8 control points some peaks and curls appear at certain values of k. This can be identified in a large increment in the curve length and curvature, with respect to the values obtained for other norms implemented. Therefore it is concluded that it is more effective to optimize m than k, in the pursuit of high topology and geometry fidelity in the reconstruction of S.

4.5. Peaks, curls and closed loops detection approach

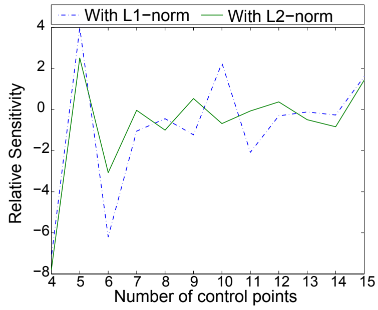
The cloud point S to reconstruct and the procedure to obtain the fitting curve initial guess is the same implemented in previous sections. The methodology described in section 3.4 was applied for the fitting curves that resulted from the optimization procedure, using 5, 8, 9 and 15 control points, using only L_2 norm. The change of direction of $\frac{\partial \vec{C}}{\partial u}$, represented by θ, in degrees, is consigned in Fig.9(a) for all the cases of study. In this figure it is shown how for the curves generated with 5 and 8 control points the magnitude of θ was kept small as C is traversed. When 9 and 15 control points were used, very large peaks were obtained in θ, in agreement with the presence of strong oscillations in the fitting curves.

In the frequency spectrum representation (see figure 9(b)), the θ data obtained when using 5 and 8 control points, consist of very low frequency waves (i.e, near zero), while for the 9 and 15 control points cases the large peaks are represented by a considerable and stable presence of high frequency sinusoidal waves that go up to $5000l^{-1}$. Notice that this is the maximum frequency that can be resolved according to the sampling rate implemented, but it is not the higher frequency of the waves that comprise θ for the later cases.

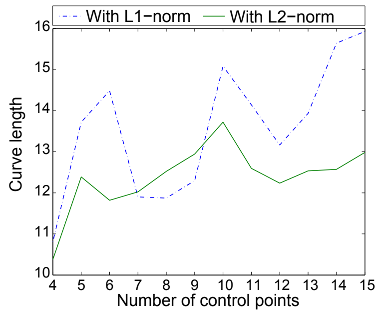
The information obtained from the frequency spec-



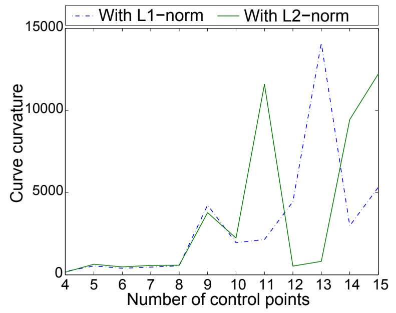
(a) Objective function vs. number of control points.



(b) S_m^f vs. number of control points.



(c) Curve length vs. number of control points.



(d) Curve curvature vs. number of control points.

Figure 5 Resulting metrics of the fitting curve with different number of control points using L_1 and L_2 norms. Here the units of the length are l and the units of the curvature are $1/l$.

trum can be used to conclude the presence of peaks or curls in the fitting curve, by comparing the contributions of low and high frequencies waves to θ. With the usage of the dual distance penalization, the curls an peaks in the fitting curve produce large values in θ because of their sharp shape, due to the fact that they are formed within the S boundaries. Therefore the low and high frequencies waves contributions to θ are very similar.

When optimizing the number of control points, the peaks and curls detection is useful to determine its upper limit, so the curve is not provided with an excessive degrees of freedom. If the optimization technique includes information about the curvature in f, the information obtained from the frequency spectrum can be processed to establish the weight of the curvature penalization in f dynamically.

5. CONCLUSIONS AND FUTURE WORK

This article presented a sensitivity analysis of the number of control points m and norm k on the objective function f. It has been found that using an adequate number of control points the formation of peaks, curls and closed loops in C is prevented, making unnecessary to add a curvature penalization term to f in order to avoid them. Finding proper values of m also reduces the number of decision variables of the problem, which results in a more efficient process since redundancy of control points is avoided.

Changes in the values of k do not influence significantly the result of the reconstruction process when m is chosen properly. Although k produces larger percent changes in f than m, the optimization of m produce better results in terms of topology and geometry of the reconstructed curve. The analysis of the C curvature information in the frequency domain allows to identify the presence of peaks, curls and closed loops as they map into high frequency components in the frequency spectrum.

Ongoing studies are being undertaken to determine the influence of knot vector U and curve degree p on the minimization of the penalty function f, in case studies that include non-Nyquist and self-intersecting point samples. A remaining open issue is the implementation of a method that uses information provided by the DFT of the curvature of C to find appropriate values for parameters such as m.

Notice that the complexity of the fast Fourier transform (FFT) and related efforts depends on the num-

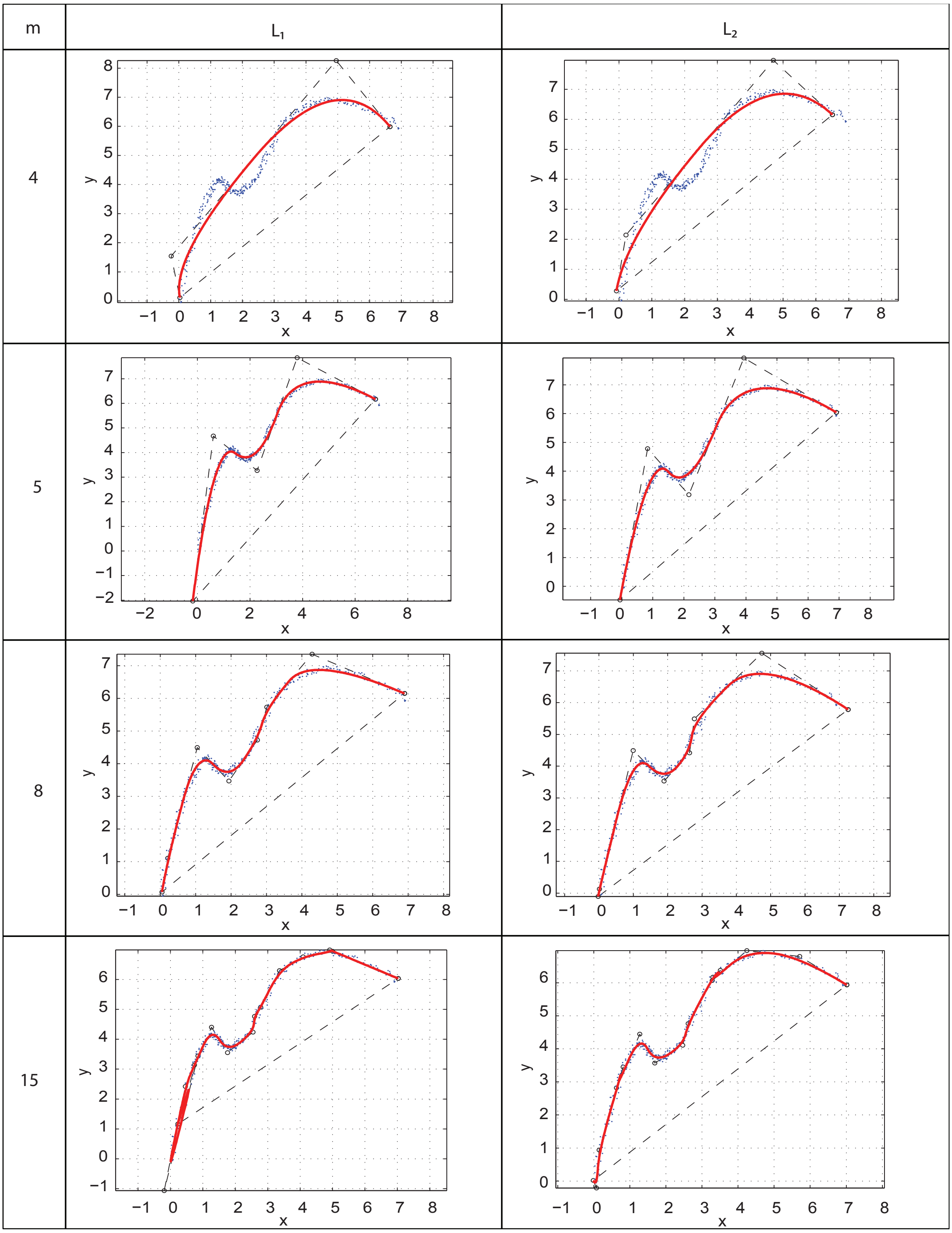
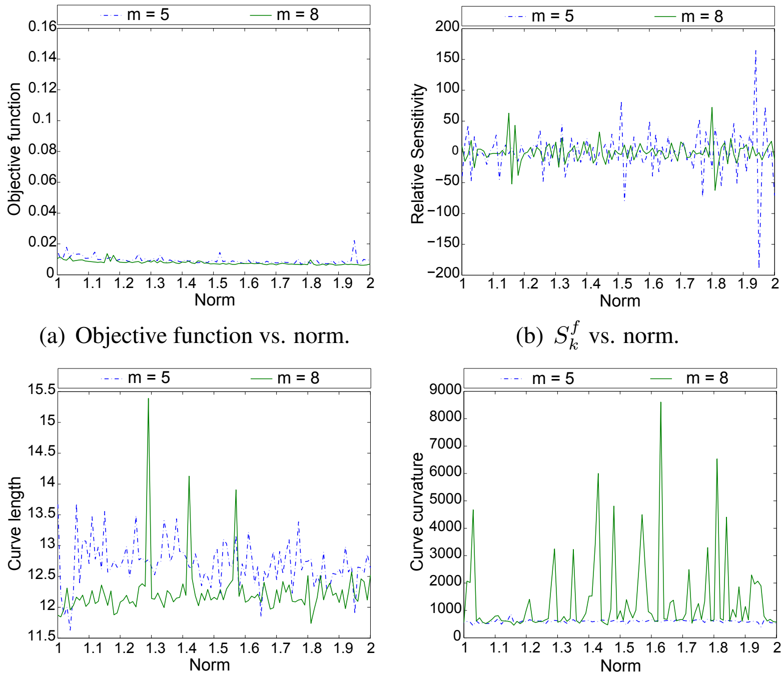


Figure 6 Resulting curves of the fitting with different number of control points m, using L_1 and L_2 norms.



(a) Objective function vs. norm. (b) S_k^f vs. norm.

(c) Curve length vs. norm. (d) Curve curvature vs. norm.

Figure 7 Resulting metrics of the fitting curve with different norms using 5 and 8 control points. Here the units of the length are l and the units of the curvature are $1/l$.

ber of PL segments of the curve $C(u)$ (n: length of the θ history). It does not depend on the number of cloud points. The frequency content of the θ signal is obtained by using the FFT, whose complexity is $O(n.log(n))$. Although FFT has very reasonable computational expenses, more work is required in lowering the expenses of automatically analyzing the results of the FFT to detect curls and cusps.

Stochastic noise vs. Sampling Density. For a proper curve reconstruction the quality of the digitalization is of prime importance. If the sampling density and/or stochastic noise violate the Nyquist criteria, an accurate reconstruction becomes impossible. Further elaboration of this topic is left for future publications.

REFERENCES

Blake, A. & Isard, M. (1998). Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion. Springer.

Edgar, T., Himmelblau, D., & Lasdon, L. (2001). Optimization of chemical processes. McGraw-Hill.

Fiacco, A. (1983). Introduction to sensitivity and stability analysis in nonlinear programming. Academic Press.

Flöry, S. (2009). Fitting curves and surfaces to point clouds in the presence of obstacles. Computer Aided Geometric Design, 26(2), pp. 192–202.

Flöry, S. & Hofer, M. (2008). Constrained curve fitting on manifolds. Computer-Aided Design, 40(1), pp. 25–34.

Flöry, S. & Hofer, M. (2010). Surface fitting and registration of point clouds using approximations of the unsigned distance function. Computer Aided Geometric Design, 27(1), pp. 60–77.

Gálvez, A., Iglesias, A., Cobo, A., Puig-Pey, J., & Espinola, J. (2007). Bézier curve and surface fitting of 3d point clouds through genetic algorithms, functional networks and least-squares approximation. Computational Science and Its Applications–ICCSA 2007, pp. 680–693.

Heidrich, W., Bartels, R., & Labahn, G. (1996). Fitting uncertain data with nurbs. In Proceedings of 3^{rd} Int. Conf. on Curves and Surfaces in Geometric Design: Vanderbilt University Pres pp. 1–8.

Kapur, D. & Lakshman, Y. (1992). Elimination Methods: An Introduction, pp. 45–88. Academic Press.

Liu, Y. & Wang, W. (2008). A revisit to least squares orthogonal distance fitting of parametric curves and surfaces. Advances in Geometric Modeling and Processing, pp. 384–397.

Liu, Y., Yang, H., & Wang, W. (2005). Reconstructing b-spline curves from point clouds–a tangential flow

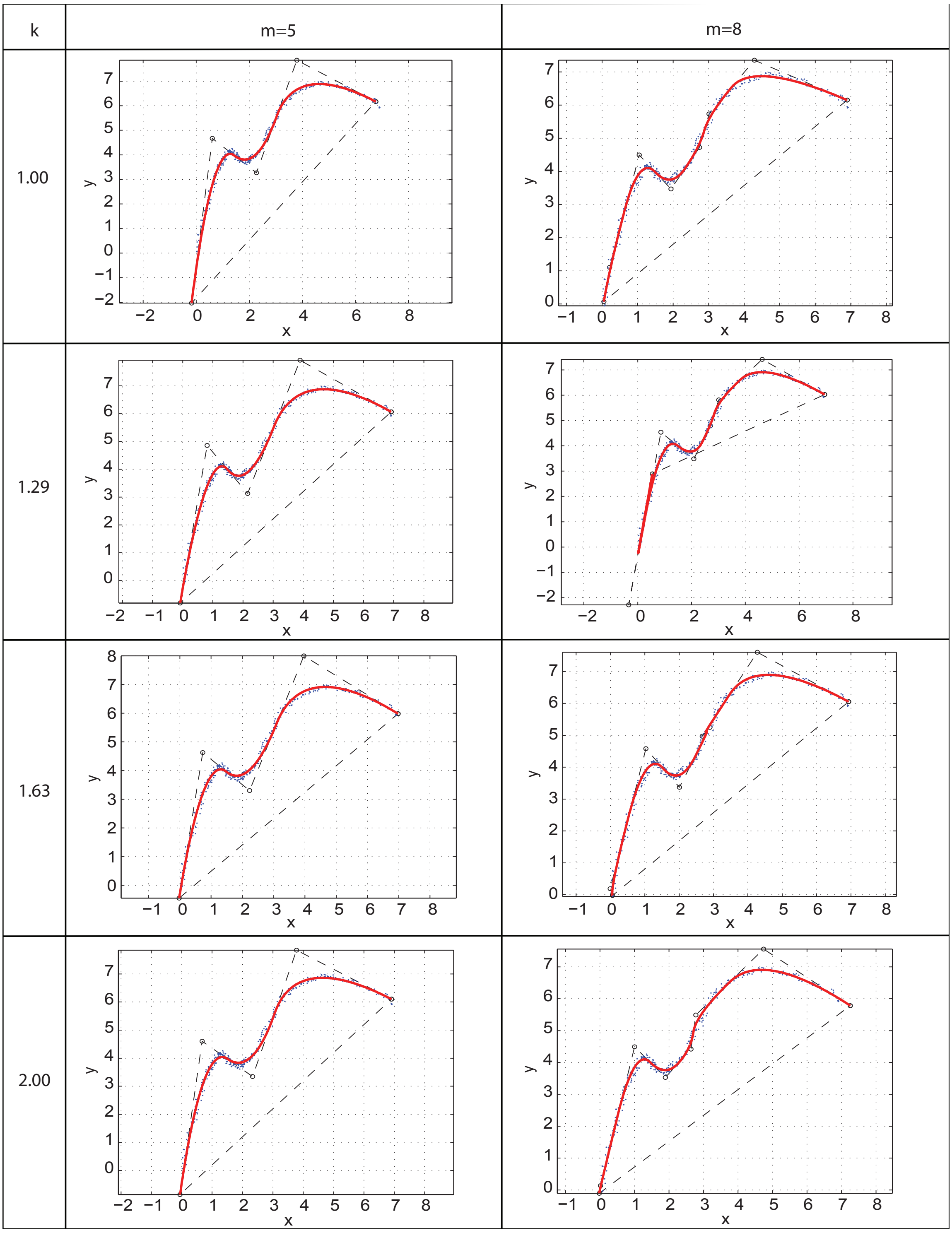
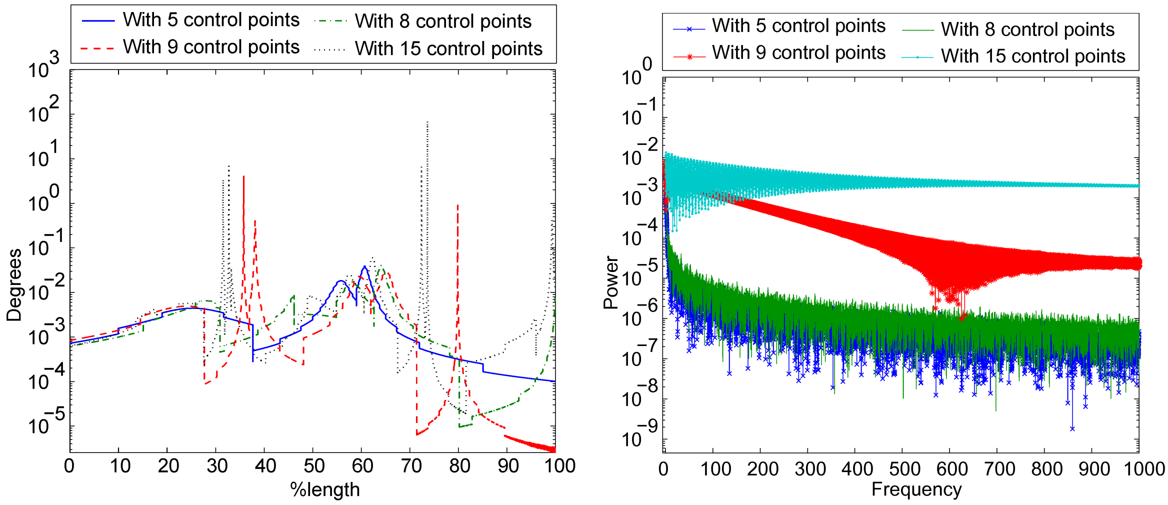


Figure 8 Resulting curves of the fitting with different norms k, using 5 and 8 control points.



(a) Changes in direction of curve's first derivative (degrees) vs. length (percentage).

(b) Power vs. frequency.

Figure 9 Changes in direction of curve's first derivative and frequency spectrum.

approach using least squares minimization. In Shape Modeling and Applications, 2005 International Conference: IEEE pp. 4–12.

Nocedal, J. & Wright, S. (2006). Numerical optimization, series in operations research and financial engineering.

Papadimitriou, C. & Steiglitz, K. (1998). Combinatorial optimization: algorithms and complexity. Dover Publications.

Pekerman, D., Elber, G., & Kim, M. (2008). Self-intersection detection and elimination in freeform curves and surfaces. Computer-Aided Design, 40(2), pp. 150–159.

Piegl, L. & Tiller, W. (1997). The NURBS book. Springer Verlag.

Ruiz, O. & Ferreira, P. (1996). Algebraic Geometry and Group Theory in Geometric Constraint Satisfaction for Computer Aided Design and Assembly Planning. IIE Transactions. Focussed Issue on Design and Manufacturing, 28(4), pp. 281–204.

Ruiz, O., Vanegas, C., & Cadavid, C. (2011). Ellipse-based principal component analysis for self-intersecting curve reconstruction from noisy point sets. The visual computer, pp. 1–16.

Saux, E. & Daniel, M. (2003). An improved hoschek intrinsic parametrization. Computer Aided Geometric Design, 20(8-9), pp. 513–521.

Ueng, W., Lai, J., & Tsai, Y. (2007). Unconstrained and constrained curve fitting for reverse engineering. The International Journal of Advanced Manufacturing Technology, 33(11), pp. 1189–1203.

Wang, W., Pottmann, H., & Liu, Y. (2006). Fitting b-spline curves to point clouds by curvature-based squared distance minimization. ACM Transactions on Graphics (TOG), 25(2), pp. 214–238.

Yang, H., Wang, W., & Sun, J. (2004). Control point adjustment for b-spline curve approximation. Computer-Aided Design, 36(7), pp. 639–652.