Sensitivity Analysis in Optimized Parametric Curve Fitting

Oscar E. Ruiz [a], Camilo Cortes [a], Diego A. Acosta [b], Mauricio Aristizabal [a]

[a] Laboratorio de CAD CAM CAE
[b] Grupo de Investigación DDP
Universidad EAFIT, COLOMBIA

Submitted to International Journal of Engineering Computations

Structured Abstract:

Purpose: Curve fitting from unordered noisy point samples is needed for surface reconstruction in many applications. In the literature, several approaches have been proposed to solve this problem. However, previous works lack formal characterization of the curve fitting problem and assessment on the effect of several parameters (i.e., scalars that remain constant in the optimization problem), such as control points number (m), curve degree (b), knot vector composition (U) and norm degree (k), on the optimized curve reconstruction measured by a penalty function (f).
Methodology: A numerical sensitivity analysis of the effect of m, b and k on f and a characterization of the fitting procedure from the mathematical viewpoint are performed. Also, the spectral (frequency) analysis of the derivative of the angle of the fitted curve with respect to u as a means to detect spurious curls and peaks is explored.
Findings: It is more effective to find optimum values for m than k or b in order to obtain good results because the topological faithfulness of the resulting curve strongly depends on m. Furthermore, when an exaggerate number of control points is used the resulting curve presents spurious curls and peaks. We were able to detect the presence of such spurious features with a spectral analysis.
Value: We have addressed important voids of previous works in this field. We determined, among the curve fitting parameters m, b and k, which of them influenced the most the results and how. Also, we performed a characterization of the curve fitting problem from the optimization perspective. And finally, we devised a method to detect spurious features in the fitting curve.

Keywords: parametric curve reconstruction, noisy point cloud, sensitivity analysis, penalty minimization

Nomenclature

C_0	: Unknown C^1-differentiable simple planar curve
$C(u)$	: Parametric planar curve approaching C_0
S	: $\{\boldsymbol{p}_0, \boldsymbol{p}_1 \ldots, \boldsymbol{p}_r\}$. Noisy unordered point sample of C_0
r	: Number of sampled points in set **S**
$C(u_i)$	: Point on $C(u)$ closest to point $\boldsymbol{p}_i$
$d(\boldsymbol{p}, \mathbf{S})$	: Distance from point $\boldsymbol{p}$ to the point set **S**
k	: Degree of norm: $(\Sigma\|x_i\|^k)^{1/k}$
ℓ	: Length units
m	: Number of control points of $C(u)$
P	: $[\boldsymbol{P}_0, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_{m-1}]$. Control polygon of $C(u)$
b	: Degree of parametric curve $C(u)$
$O(g(n))$	: computational expense is a function $g(n)$ of the data size n.
X	: knot vector
B	: sequence of parameter values. $\mathbf{B} = [u_0, u_1, \ldots, u_n]$

1. Introduction

Many engineering applications (e.g. terrain modeling, medical imaging, reverse engineering) require the recovery of a planar curve C from its unordered noisy point sample **S**. The curve C is usually an intermediate step in surface reconstruction. C may have C^0 (PL: Piecewise Linear) or higher (C^1, C^2,...) continuity. In the second case, we talk of a smooth parametric free form (i.e. the subject of this article). In either case, a usual goal is that C be statistically 'centered' in the point cloud **S,** assuming that the data has a uniform noise distribution, which is the case that we address in this work. It is customary to use heuristics to find the 'best curve' C fitting the given point set **S**. In the present article, to find the free form curve fitting a noisy point set, we used the heuristic of minimizing an accumulative distance function (discussed next) point cloud vs. curve, starting with an initial guess for **P**, the control polygon of C. This initial guess is based on Principal Component Analysis (PCA), by penalizing large curvatures, extreme curve excursions and curls. It must be clear, however, that even for very finely tuned heuristics, the curve C obtained must be double-checked by a human user in most of the cases, to avoid serious errors in sensitive applications.

The usual approach implies adjusting a parametric or implicit curve to the set of points by minimizing a cumulative unsigned distance function f between the points and their approximating curve. It is of interest to know which parameters are more effective to increase the goodness of the curve. The issue is important because an excess or deficit of the parameters produce equally disastrous results (curls, cusps, excursions, spurious self-intersections, etc.). The literature reviewed presents ad-hoc tests, which seem to favor a parameter over others, but not a systematic, quantified evaluation of the relative impact of the parameters in the goodness of the curve. This is the goal of the present article.

In this article we will address planar Open Uniform B-splines, with knot vector such that $0 \leq u \leq 1$ (Piegl and Tiller, 1997). Other curve types can be analyzed similarly. We address point samples with uniform sampling noise, leaving spatial-dependent noise for future work.

1.1. Objective function

Consider an unknown smooth finite curve C_0 and $\mathbf{S}$ a noisy point sample of C_0. The usual goal is to find a parametric curve C, which approximates C_0, by minimizing the distance function $\boldsymbol{f}$ (Eq. 1) between C and $\mathbf{S}$.

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

with the residual d_i being the minimal distance between the i-th cloud point $\boldsymbol{p}_i \in \mathbf{S}$ and the finite curve C (Eq. 2) , w being the order of the residual and k being the norm-degree to calculate the distance

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

1.2. Context of the optimization problem

The terms in Eq.1 that can be changed to minimize *f( )* are: m (number fo control points), $\mathbf{P}$ (control polygon), b (degree of parametric curve), $\mathbf{X}$ (knot vector), k (norm type, Eq. 2), and w (power of the distances, Eq. 1). In this article we choose to analyze the sensitivity of the curve fitting with respect to m, k, and b (parameters) and the tuning variable is $\mathbf{P}$ (i.e. $2m$ scalar values). This problem is non-linear, unconstrained with $2m$ degrees of freedom (Chong and Żak, 2008). The optimal solution is not a global one, because the eigenvalues of the Hessian matrix of *f( )* have mixed sign.

The relative (dimensionless) parametric sensitivity (Edgar *et al.*, 2001, Fiacco, 1983; Nocedal and Wright, 2006)) of f with respect to a parameter q (in this article, m, b and k) is given by:

$$S_q^f = \frac{q}{f}\frac{\partial f}{\partial q} = \frac{\partial ln(f)}{\partial ln(q)} \tag{3}$$

1.3 Nyquist – Shannon Compliance.

The Nyquist – Shannon principle states that the added value of sampling distance plus the sampling noise must be smaller than half of the minimal geometric feature to be reconstructed from the sample. In this article we work under the assumption of compliance of the Nyquist – Shannon conditions by the sample.

2. Literature Review

.

2.1. Objective function

Reference (Flöry and Hofer, 2010) employs first order residuals ($w = 1$ in Eq. 1) while references (Gálvez *et al.*, 2007; Liu and Wang, 2008; Liu *et al.*, 2005; Wang *et al.*, 2006) use second order residuals ($w = 2$).

Some references (Flöry and Hofer, 2010, 2008; Flöry, 2009; Liu *et al.,* 2005; Wang *et al.*, 2006) add a smoothing term f_c to the objective function in order to adjust the roughness of the curve:

$$f = \sum_{i=1}^{r} d_i^w + \lambda f_c. \tag{4}$$

The term f_c contains information on the curve's first and/or second derivatives and λ penalizes large curvatures and therefore it prevents reconstructing curves with sharp corners.

Constrained approaches (Flöry, 2009) and (Flöry and Hofer, 2008) are used to restrict the fitting to or outside entire regions (e.g. manifolds in R3), as opposed to high fidelity curve or surface fitting. Because of this reason, we do not consider them here.

2.2. Distance measurement

Eq.2 calculates distance or residual d_i of the objective function (Eq. 1). In curve fitting algorithms norm k is usually chosen to be $k = 2$ (i.e., Euclidean distance) as in (Liu *et al.*, 2005; Wang *et al.*, 2006).

The exact calculation of d_i is expensive because it requires calculating the roots of polynomial systems. It implies finding the parameter u_i which associates a point on the curve $C(u_i)$ with the i-th cloud point $\boldsymbol{p}_i$ such that d_i is a minimum (perpendicular distance point – curve, Eq. 7). Namely,

$$\|C(u_i) - \boldsymbol{p}_i\|^k = \min_{C(u)\in C} \|C(u) - \boldsymbol{p}_i\|^k \tag{5}$$

$$G(u) = |C'(u) \cdot (C(u) - \boldsymbol{p}_i)| \tag{6}$$

Solving for u in $G(u) = 0$ is achieved by usins Newton's Method (Liu *et al.*, 2005; Piegl and Tiller, 1997), numerically minimization of $G(u)$ (Flöry and Hofer, 2010; Liu and Wang, 2008; Saux and Daniel, 2003; Wang *et al.*, 2006) or using genetic algorithms (Gálvez *et al.*, 2007). All methods require the usual effort for finding initial guess for the solution (quadtree (Wang *et al.*, 2006), k-D tree (Liu and Wang, 2008)) and Euclidean minimum spanning tree (Liu *et al.*, 2005). Other methods avoid the actual calculation of point – curve perpendicular. Reference (Liu and Wang, 2008) presents a review of methods for solving (or approaching) Eqs 6 and 7.

It must also be remarked that using point-to-curve distance does not avoid curls formed outside the **S** boundaries and outliers in the final curve C. Therefore, we have included both (1) point-to-curve and (2) curve-to-point distance estimations, allowing for curls and outliers avoidance.

2.3. Effect of Curve fitting Parameters.

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. The experiments performed show that increasing m 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 removal of control points is part of their fitting strategy.

Norm Degree k. Reported research is oriented towards identifying which norm to use when 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. Ref. (Flöry and Hofer, 2010) concludes that the L_1 norm is less sensitive to outliers. In contrast, the problem of finding an adequate number of control points for correct geometry and topology reconstruction has not been discussed thoroughly.

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

2.4. Peaks and curls 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**, since spurious peaks and curls may appear. We show in this article that peaks and curls may be avoided by finding an optimal value for m, as opposed to the strategy of curvature penalization implemented in (Flöry and Hofer, 2010; Flöry, 2009; Liu *et al.,* 2005; Wang *et al.*, 2006), which presents the drawbacks discussed in section 2.1. Reference (Pekerman *et al.*, 2008) presents an algebraic approach to detect self-intersections solving $C(u) - C(v) = 0$, with $u \neq v$. In any case, peaks and curls detection is not trivial and it is an open problem. In this article, we introduce the usage of the frequency content of the $C(u)$'s curvature to detect peaks and curls.

2.5. Conclusions of the Literature review and contribution of this article

There are several open issues in optimized curve fitting to point clouds: (a) Effect of parameters such as the number of control points m, knot vector **X** and norm k, (b) Detection of peaks and curls in *C(u)* to validate optimal parameter value identification and (c) Mathematical characterization of the curve fitting problem from the viewpoint of optimization.

The knot vector **X** controls the variation of parameter velocity in the curves, and the adherence to specific control polygon vertices. The knot vector is, in itself, a whole area of research, given the large number of configurations and types that it admits. Because of this reason, we prefer, in this article, to explore: (1) A sensitivity analysis of the number of control points(m)*,* degree of curve (b), norm type (k) and size of point

sample (r) on $f(\ )$, and, (2) A quantitative analysis in the frequency domain of the curvature of $C(u)$ to detect peaks and curls. We aim to reconstruct curves which have sharp corners, and therefore, we do not consider a curvature penalty factor λ.
Nyquist-Shannon theory is well known in the domain of one-dimensional functions (i.e. signal processing). However, the quantification of Nyquist compliance in 2D/3D samples is an open problem in computational geometry. We do not know of any numerical estimation of Nyquist-compliance of 2D / 3D data. We do not, at this time, intend to undertake such a task.

3. Methodology

3.1. Dual distance calculation

In addition to point-to-curve distance (section 2.2), curve-to-point distance is used to calculate d_i, used in Eq.1, for the curve fitting algorithm implemented in this article.

The distance from a given cloud point to the curve (point-to-curve, Figure 1(a))is defined as

$$d_i = ||\boldsymbol{p}_i - C(u_i)||^k \tag{7}$$

where u_i is the parameter in the domain of C which defines point $C(u_i)$ closest to $\boldsymbol{p}_i$. This calculation is computationally expensive because it seeks the common roots of a polynomial ideal (Kapur and Lakshman, 1992).

(a) Distances cloud point to curve. (b) Distances curve to cloud point.
Figure 1: Distances cloud points to/from curve.

To avoid the computational expenses of algebraic roots calculation, we sample the domain for u in [0,1], (i.e. $[0, \Delta_u, 2\Delta_u, \ldots, 1.0]$) and approximate the curve C with a poly-line $[C(0), C(\Delta_u), C(2\Delta_u), \ldots, C(1.0)]$. Approximating $C(u_i)$ in Eq. 8 for a given $\boldsymbol{p}_i$ simply entails traversing $[C(0), C(\Delta_u), C(2\Delta_u), \ldots, C(1.0)]$ to find the $C(\kappa\Delta_u)$ closest to $\boldsymbol{p}_i$. This PL approximation of C produces reasonable results if the sample of u in [0,1] is Nyquist-compliant.

Figure 1(a) displays the distance d_i (Eq. 1) from a particular cloud point $\boldsymbol{p}_i$ to its closest point $C(u_i)$ on the current curve C. Notice that $\boldsymbol{p}_i$ 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 does not avoid spurious outlier legs and curls outside the boundaries of **S**. To overcome this disadvantage, we also include in f the distances from *the curve points* C_i *to the cloud points* $\boldsymbol{p}_i$ (see Figure 1(b)).

For any point $\boldsymbol{p} \in \mathbb{R}^n$, the distance of this point to **S** is a well defined mathematical function: $d(\boldsymbol{p}, \mathbf{S}) = \min_{\boldsymbol{p}_j \in \mathbf{S}}(||\boldsymbol{p} - \boldsymbol{p}_j||^k)$. For the current discussion the points $\boldsymbol{p}$ belong to curve $C(u_i)$.

Notice that $d(\boldsymbol{p}, \mathbf{S}) = ||\boldsymbol{p}_j - \boldsymbol{p}||^k$ for some cloud point $\boldsymbol{p}_j \in \mathbf{S}$. Let the point set $\mathbf{A_j}$ (on curve C) be:

$$\mathbf{A_j} = \{C(u) | u \in \mathbf{B} \wedge d(C(u), \mathbf{S}) = ||\boldsymbol{p}_j - C(u)||^k\} \quad (8)$$

$\mathbf{A_j}$, a partition of curve C, contains those points in the sequence $[C(0), C(\Delta_u), C(2\Delta_u), \ldots, C(1.0)]$ that are closer to $\boldsymbol{p}_j \in \mathbf{S}$ than to any other point of **S**. We note with z_j the cardinality of $\mathbf{A_j}$. Observe that some z_j might be zero, since $\boldsymbol{p}_j$ could be far away from be curve C and no point on the curve would have $\boldsymbol{p}_j$ as its closest in **S**.

From the previous discussion, residuals d_i for Eq. 1 are defined by:

$$d_i = ||\boldsymbol{p}_i - C(u_i)||^k + \left(\frac{1}{z_i}\right) \sum_{C_v \in \mathbf{A_i}} ||C_v - \boldsymbol{p}_i||^k \quad (9)$$

where $||\boldsymbol{p}_i - C(u_i)||^k$ is the distance from cloud points in **S** to the curve C and $\left(\frac{1}{z_i}\right) \sum_{C_v \in \mathbf{A_i}} ||C_v - \boldsymbol{p}_i||^k$ expresses distances from the curve C to the cloud points in **S**, penalizing an increase on the curve length, by augmenting f.

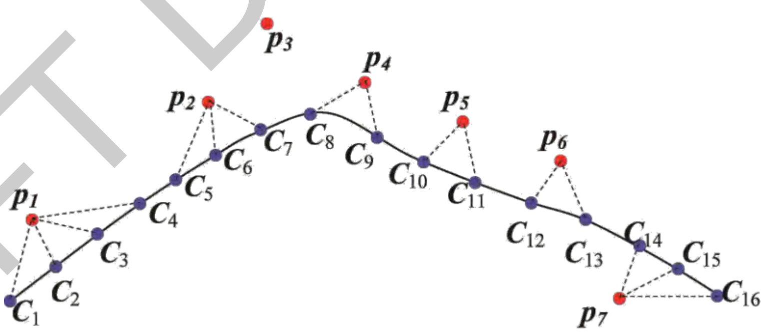


Figure 2: Clusters of distances from curve to cloud points.

Figure 2 presents a simplified picture of the situation, with few cloud points, biased with respect to C curve. Calculations are depicted in Table 1. Observe that $C_i = C(u_i)$, the point on C closest to $\boldsymbol{p}_i$, is not the exact one but an approximation using a tight PL approximation of C.

$\boldsymbol{p}_i$	$\mathbf{A_i}$	z_i	$C(u_i)$	d_i
$\boldsymbol{p}_1$	$\{C_1, C_2, C_3, C_4\}$	4	C_2	$\|\|\boldsymbol{p}_1 - C_2\|\| + \frac{1}{4}(\|\|\boldsymbol{p}_1 - C_1\|\| + \|\|\boldsymbol{p}_1 - C_2\|\| + \|\|\boldsymbol{p}_1 - C_3\|\| + \|\|\boldsymbol{p}_1 - C_4\|\|)$
$\boldsymbol{p}_2$	$\{C_5, C_6, C_7\}$	3	C_6	$\|\|\boldsymbol{p}_2 - C_6\|\| + \frac{1}{3}(\|\|\boldsymbol{p}_2 - C_5\|\| + \|\|\boldsymbol{p}_2 - C_6\|\| + \|\|\boldsymbol{p}_2 - C_7\|\|)$
$\boldsymbol{p}_3$	$\{\}$	0	C_8	$\|\|\boldsymbol{p}_3 - C_8\|\|$
$\boldsymbol{p}_4$	$\{C_8, C_9\}$	2	C_9	$\|\|\boldsymbol{p}_4 - C_9\|\| + \frac{1}{2}(\|\|\boldsymbol{p}_4 - C_8\|\| + \|\|\boldsymbol{p}_4 - C_9\|\|)$
$\boldsymbol{p}_5$	$\{C_{10}, C_{11}\}$	2	C_{10}	$\|\|\boldsymbol{p}_5 - C_{10}\|\| + \frac{1}{2}(\|\|\boldsymbol{p}_5 - C_{10}\|\| + \|\|\boldsymbol{p}_5 - C_{11}\|\|)$
$\boldsymbol{p}_6$	$\{C_{12}, C_{13}\}$	2	C_{12}	$\|\|\boldsymbol{p}_6 - C_{12}\|\| + \frac{1}{2}(\|\|\boldsymbol{p}_6 - C_{12}\|\| + \|\|\boldsymbol{p}_6 - C_{13}\|\|)$
$\boldsymbol{p}_7$	$\{C_{14}, C_{15}, C_{16}\}$	3	C_{14}	$\|\|\boldsymbol{p}_7 - C_{14}\|\| + \frac{1}{3}(\|\|\boldsymbol{p}_7 - C_{14}\|\| + \|\|\boldsymbol{p}_7 - C_{15}\|\| + \|\|\boldsymbol{p}_7 - C_{16}\|\|)$

Table 1: Calculations using *curve to cloud-point* distances for example in Figure 2.

3.2. Convexity

The *2m* variables to minimize f are the x and y coordinates of the vertices ($\boldsymbol{P}_j = (x_j, y_j)$) of the control polygon $\mathbf{P} = [\boldsymbol{P}_0, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_{m-1}]$. Because these points can be placed anywhere in $\mathbb{R}^2$ the problem is unconstrained. The eigenvalues of f's Hessian matrix (Eq. 11) indicate that the problem is non-convex.

$$H_f(\mathbf{P}) = \left[\frac{\partial^2 f}{\partial \boldsymbol{P}_i \, \partial \boldsymbol{P}_j}\right]_{ij} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_i \, \partial x_j} & \frac{\partial^2 f}{\partial x_i \, \partial y_j} \\ \frac{\partial^2 f}{\partial y_i \, \partial x_j} & \frac{\partial^2 f}{\partial y_i \, \partial y_j} \end{bmatrix}_{ij} \tag{10}$$

3.3. Sensitivity calculation

To calculate the relative sensitivity (S_q^f), of f with respect to a particular parameter q, the curve fitting algorithm is executed for specific values of q (i.e., q_i) to yield f_i, where i ($0 \leq i < i_{max}$) indicates the number of increments applied over an initial value q_{min}, bounded by a maximum number of increments i_{max}. S_q^f is calculated numerically as per Eq. 12.

$$S_{q_i}^{f_i} \approx \frac{(q_{i+1}+q_i)}{(f_{i+1}+f_i)} \quad \frac{(f_{i+1}-f_i)}{(q_{i+1}-q_i)} \tag{11}$$

The steps for the sensitivity calculation are shown in Figure 3. The initial guess L for this fitting process is a polyline with collinear intermediate vertices , which will eventually evolve as a result of the optimization process. L is calculated by using a global PCA (see reference (Ruiz *et al.*, 2011)) on the whole point set $\mathbf{S}$. The number of sub-divisions of L is a parameter of the sensitivity experiment and it is therefore chosen by the investigators. The optimized fitting of the curve C is performed with a penalized Gauss-Newton algorithm to adjust the control polygon $\mathbf{P}$ (whose first instance is L).

It must be noticed that a collinear polyline L is used for the specific purpose of the experiments related to the Sensitivity Analysis. In contrast, when the application of curve fitting arrives, a more sophisticated process is needed, given the fact that some 2D point clouds (e.g. stemming from closed curves) do not accept a simple collinear initial guess for L (or $\mathbf{P}$).

3.4. Peaks and curls detection

We perform an analysis of the frequency spectrum of the change of direction of the first derivative of $C(u)$ with respect to u that displays the presence of undesired features in $C(u)$. Peaks and curls produce large sudden changes in the direction of $\frac{\overrightarrow{\partial C}}{\partial u}$ that contain contributions of high frequencies. We computed the discrete Fourier transform (DFT) of $\frac{\overrightarrow{\partial C}}{\partial u}$. In order to sample this information according to the Nyquist criterion, we chose a series of u parameters located at equal distances d_s, on the curve ($\mathbf{U_s} = \{u_0, \dots, u_t\}$). Then we chose $d_s = 0.0001 * \ell$, where ℓ is a unit distance, and the sampling frequency f_s is $10000 / \ell$.

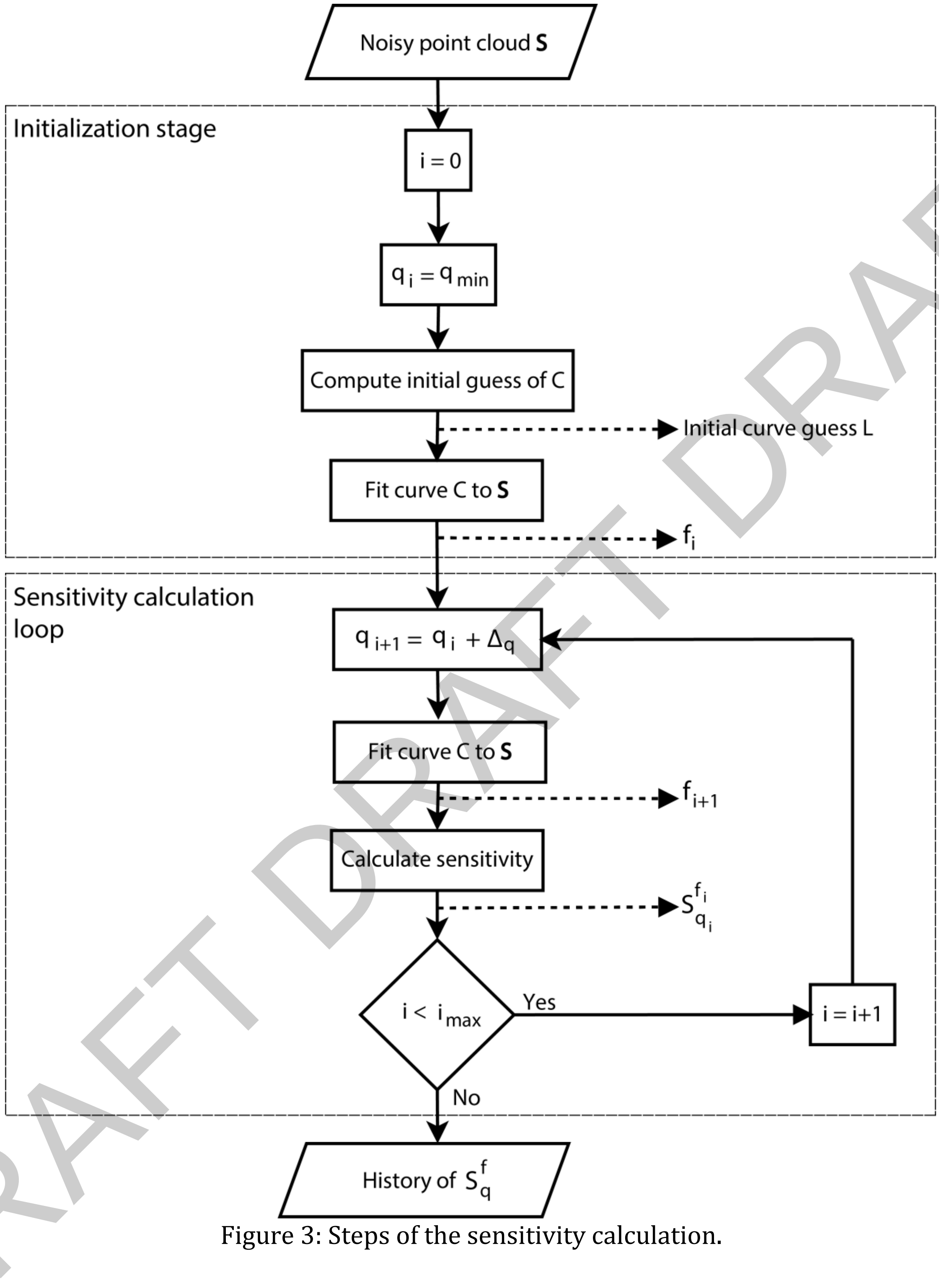


Figure 3: Steps of the sensitivity calculation.

Next, the normalized tangent vectors of the curve were computed at all points of $\mathbf{U_s}$ yielding $\mathbf{V_s} = \{\widehat{\boldsymbol{v}}_0, \dots, \widehat{\boldsymbol{v}}_t\}$. The dot product $\widehat{\boldsymbol{v}}_i \cdot \widehat{\boldsymbol{v}}_{i+1}$ is calculated, with $i = 0,1,2\dots, t-1$; and then the angle θ_i between $\widehat{\boldsymbol{v}}_i$ and $\widehat{\boldsymbol{v}}_{i+1}$ is obtained. Finally DFT is computed for $\boldsymbol{\theta} = \{\theta_0, \dots, \theta_{t-1}\}$, and properly scaled to achieve a single-sided spectrum of power vs.

frequencies.

4. Results and Discussion

4.1. Test point set

The point sample S appears in Figure 4. It is sampled on a curve generated with 5 control points. The initial guess for curve fitting is a straight line obtained from a naive PCA on the complete point cloud. The Hessian matrix $H_f(\mathbf{P})$ and its eigenvalues e were calculated at each iteration of the optimization procedure using m=5, 8, 9, 15 control points.

4.2. Convexity

In certain iterations, the eigenvalues of the Hessian matrix $H_f(\mathbf{P})$ are not all positive. Therefore, the solutions found are local optima (non-convex domain). This is an inherent challenge of the curve fit problem..

4.3. Sensitivity Analysis for Number of Control Points (m).

The number of control points m satisfies $4 \leq m \leq 16$. The runs used the norms L_1 and L_2 (k=1, 2). The results for S_m^f appear in Figure 5(b) showing that, as m increases f becomes less sensitive to it, specially when using L_2 norm.

In addition to f and S_m^f, the curve length and curvature were used to obtain information about curls, peaks, long legs, etc., of the fitting curve. In this article the word *curvature* corresponds to the summation of the local curvatures at the PL samples on the curve.

Figure 5(a) shows that, as the number of control points (m) increases, the objective function decreases. However, a m overly grows, there appears a formation of curls and/or peaks and attraction among control points, while *f( )* does not decrease in significant manner. This behavior may be traced back to over-abundance of degrees of freedom. Figure 6 shows the resulting curves with different number of control points (for L_1 and L_2 norms).

By using the dual distance in *f( ),* peaks and curls are discouraged. In particular, excursions of the curve *away* from the point sample **S** disappear.

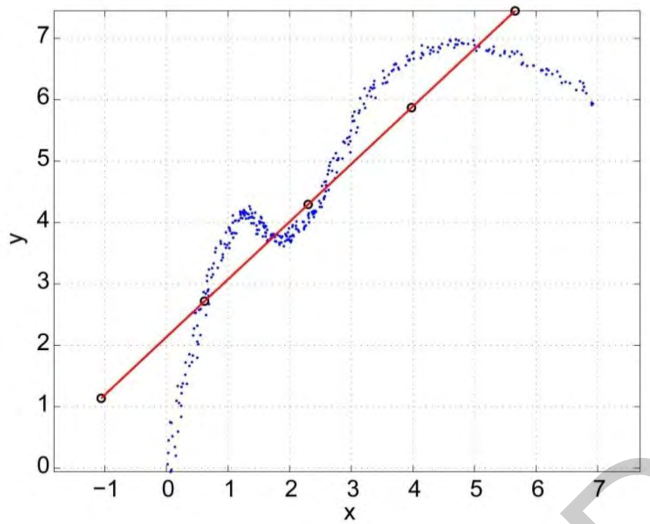
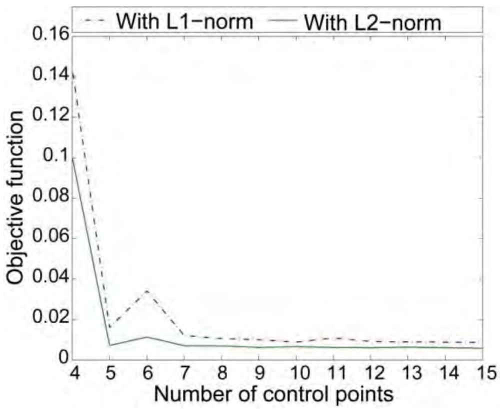


Figure 4: Point cloud and initial curve guess with five control points.

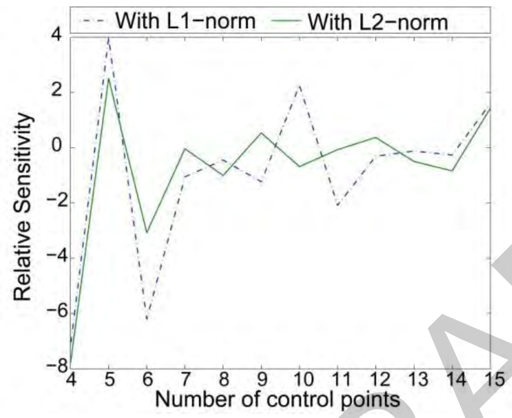
4.4. Sensitivity Analysis for Norm Type (k)

Figure 7 shows the influence of k on several goodness measures (*f( )*, length, curvature, sensitivity S_k^f), calculated for norms $k = 1$ to $k = 2$ with increments $\Delta k = 0.01$, for $m = 5$ and $m = 8$ control points. There are less oscilations of these measures with respecto to k when m=5 Curve length and curvature in Figure 7(c) and Figure 7(d) reflect a very stable behavior as k changes using $m = 5$ control points, in opposition to the results obtained when $m = 8$. Likewise, regarding the curve topology and geometry obtained, Figure 8 shows that when the number of control points m is properly chosen, the influence of the norm type k is negligible.

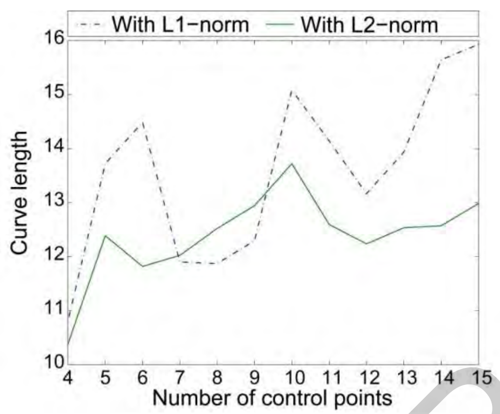
Figure 9 expands on Figure 7(d). We have revisited the case in which an exaggerate number of control points (*m*=8) is used. We ran a significant number of curve fitting tests, and found that the maximal curvature value in the curve fit behaves erratically and curls appear for high *m*. The reason is that the superfluous control points drift in the 2D plane, reaching positions very near to each other. When this jamming of control points occurs, the curve is squeezed between them, being forced to adopt local high-curvature detours. If these control points drift apart, the high curvature disappears. Notice that, penalizing curvature itself would impede the fitting of curves to sharp corners, which is an advantage of our approach.



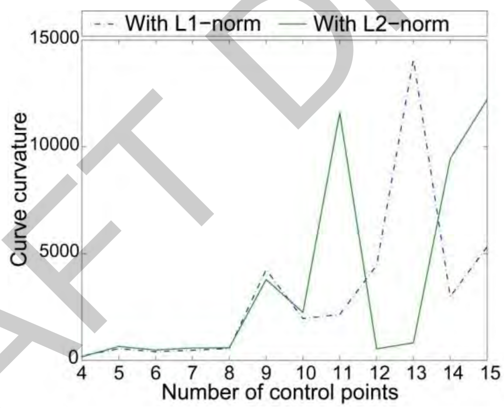
(a) Objective function vs. number of control points.



(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 $1/l$.

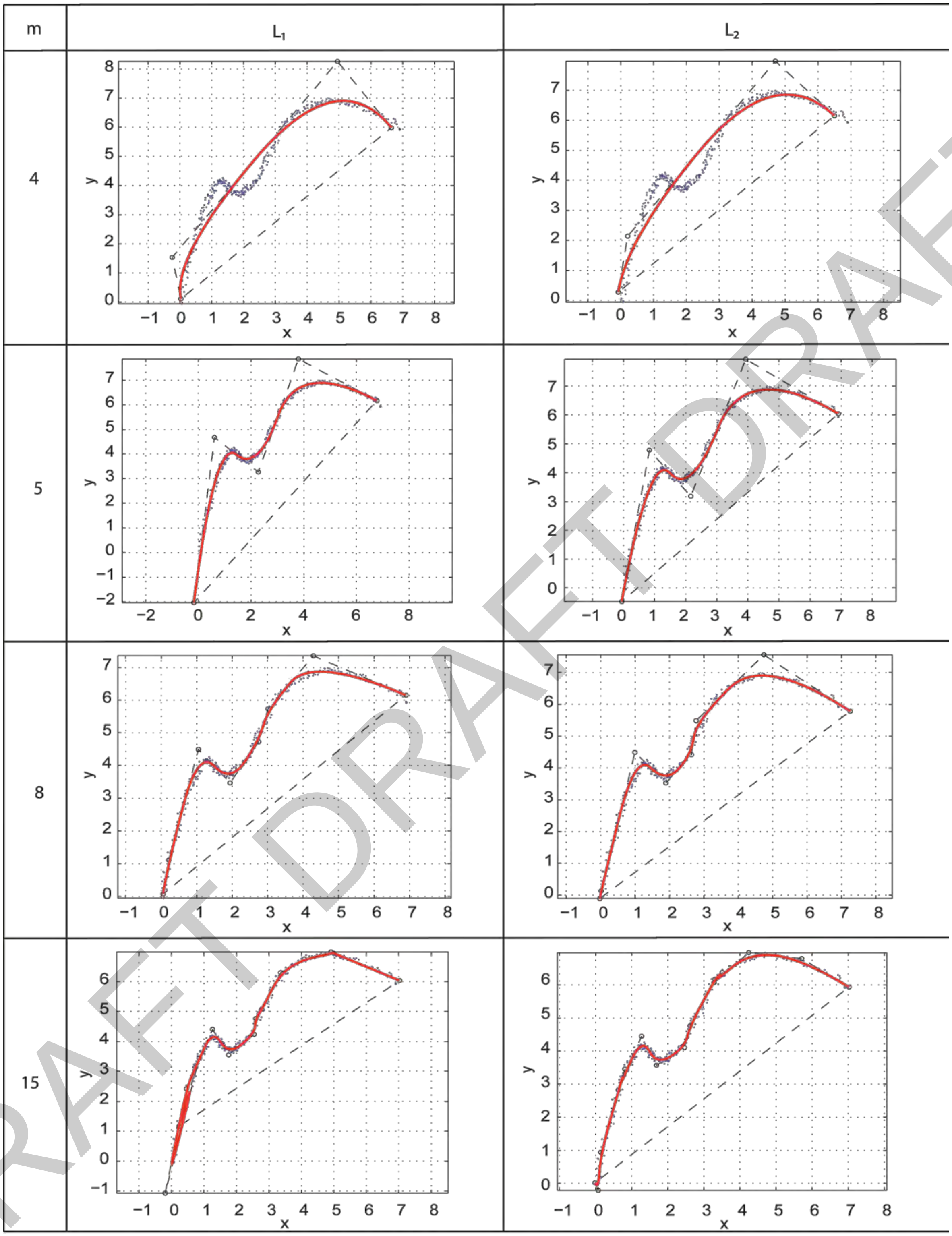
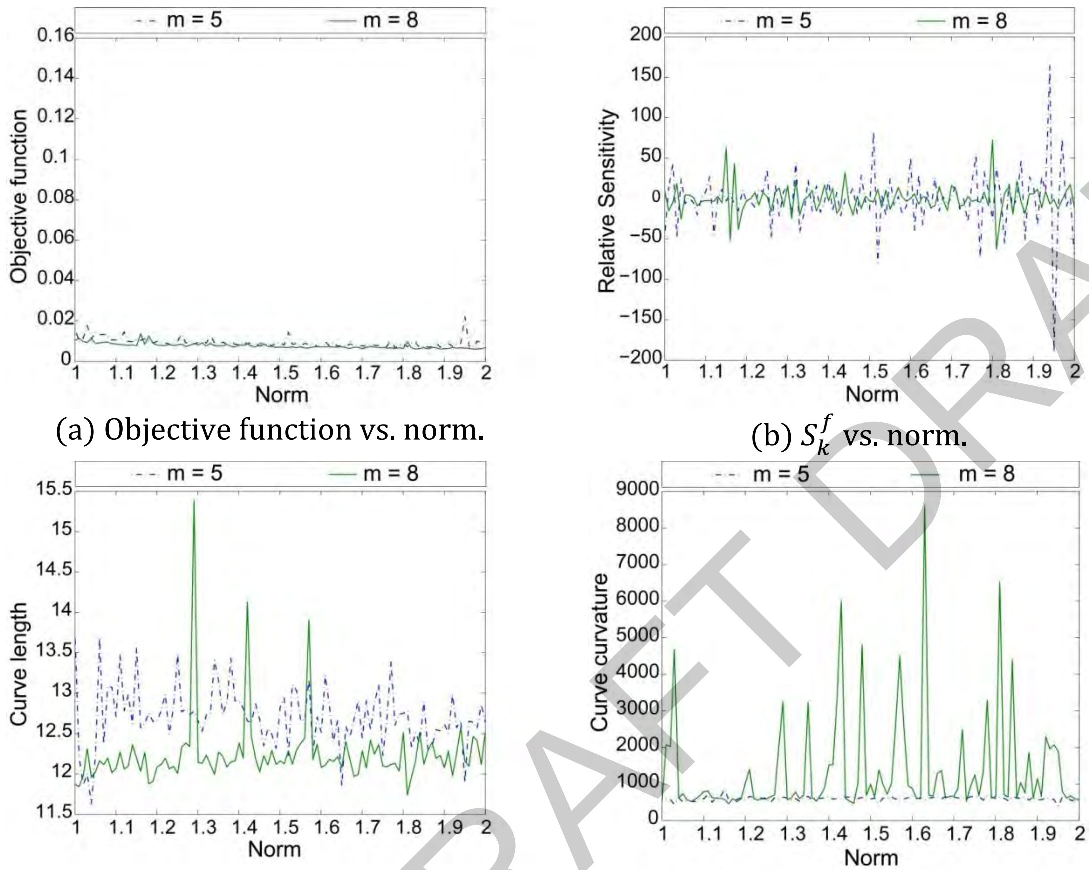


Figure 6: Resulting curves of the fitting with different number of control points m, using L_1 and L_2 norms.

As a result, 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**.



(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 $1/l$.

4.5. Sensitivity Analysis for Degree b

This experiment varies b, in the range $1 \leq b \leq m - 1$. Figure 10 presents results with $m = 5$ and $m = 8$, and k=2 (euclidean norm).

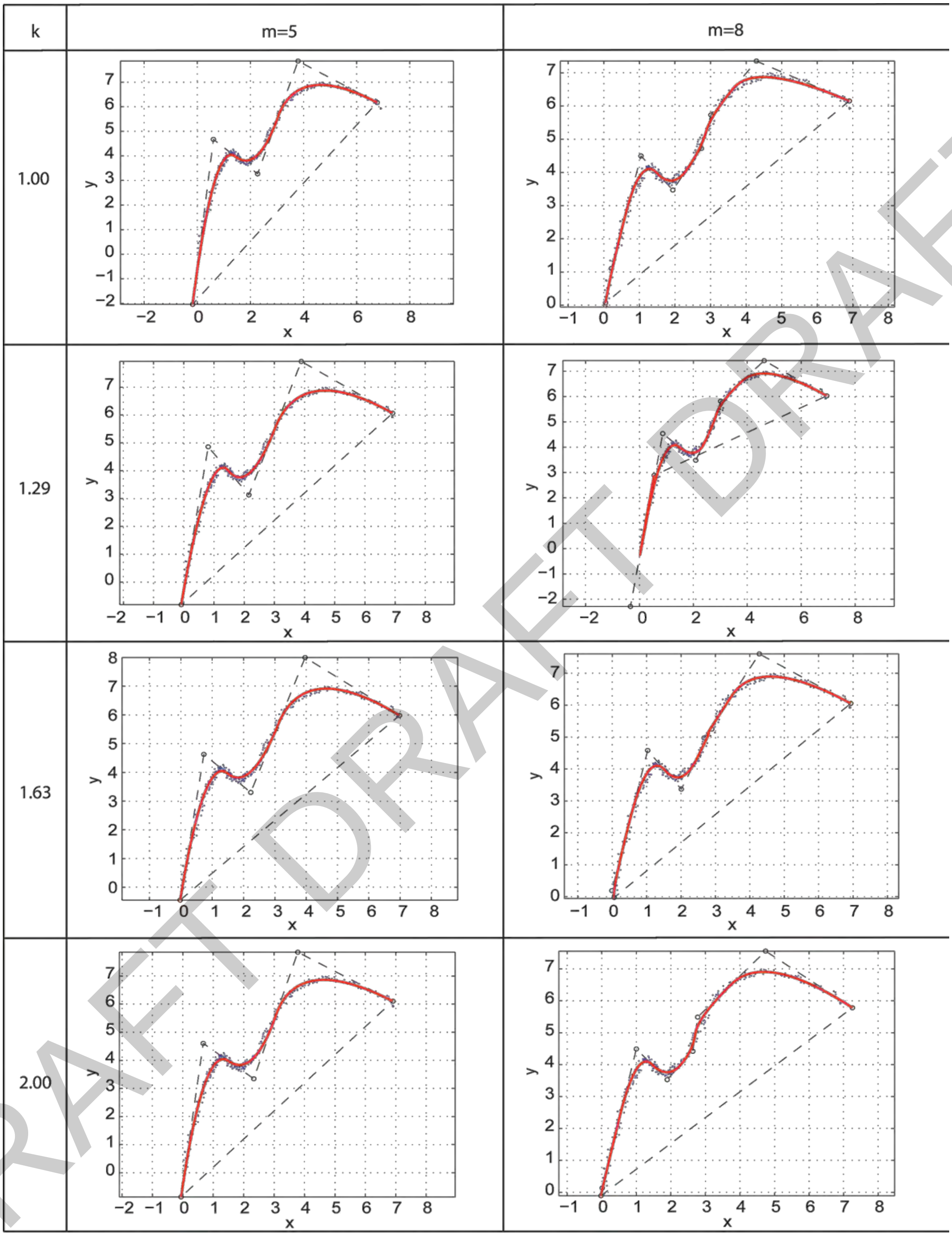
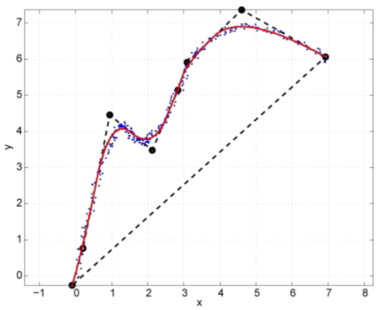
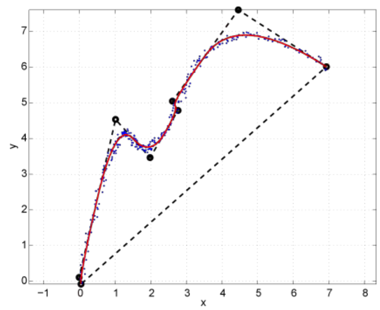
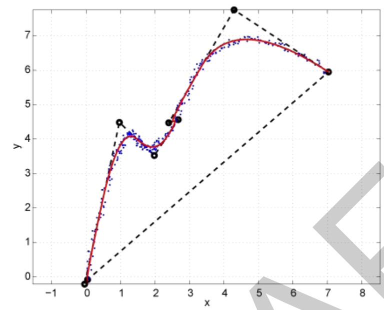


Figure 8: Resulting curves of the fitting with different norms k, using 5 and 8 control points.



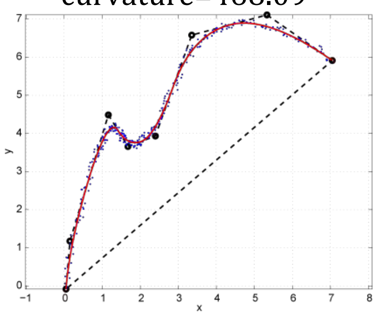


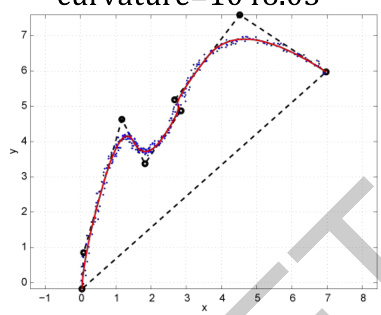


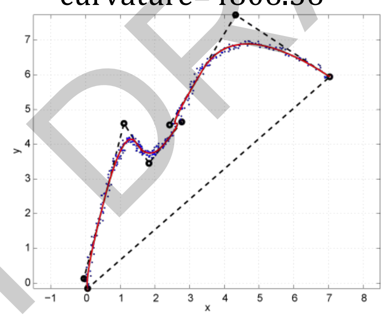
k =1.46; Curve curvature=468.09

k =1.47; Curve curvature=1046.03

k =1.48; Curve curvature=4806.56





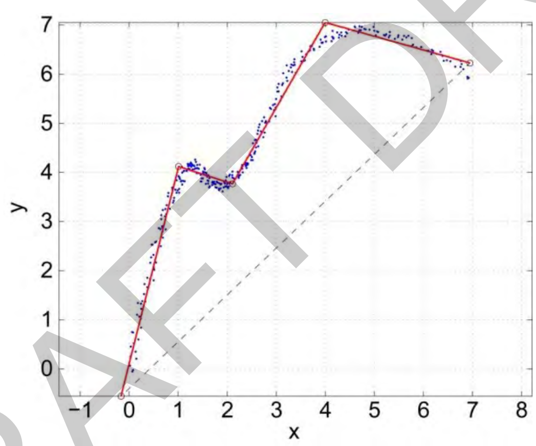


k =1.54; Curve curvature=718.29

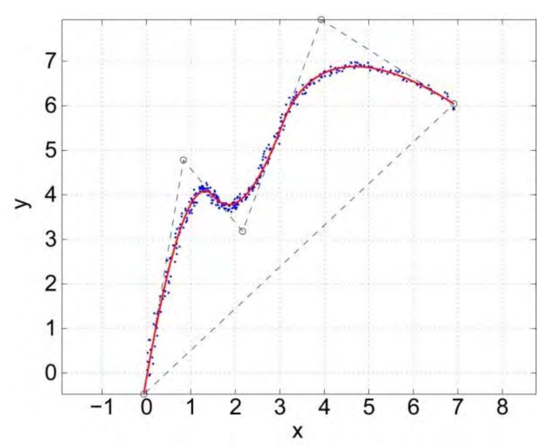
k =1.55; Curve curvature=1010.09

k =1.56; Curve curvature=2553.30

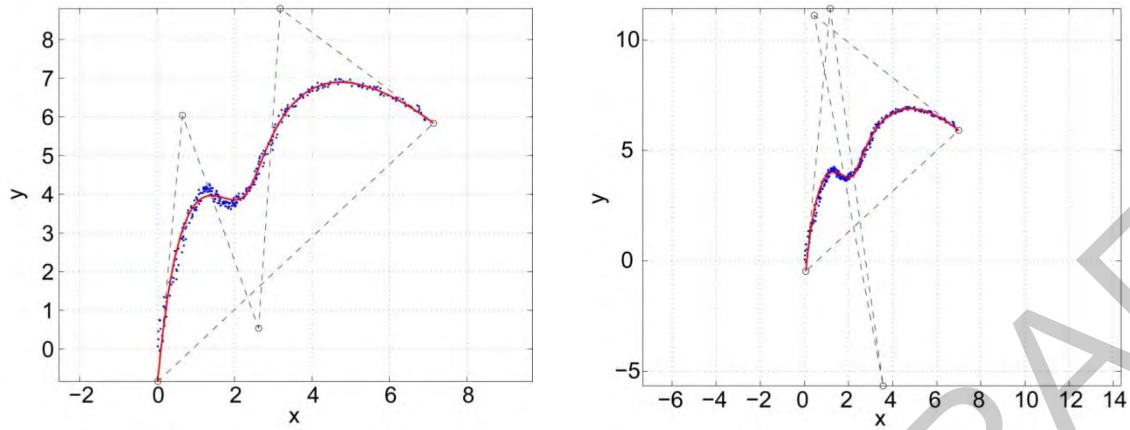
Figure 9: Detail of Curvature variation with norm type (k) for cases with over-population of control points. (curvature units : $1/\ell$)



(a) Fitted curve with $b = 1$.

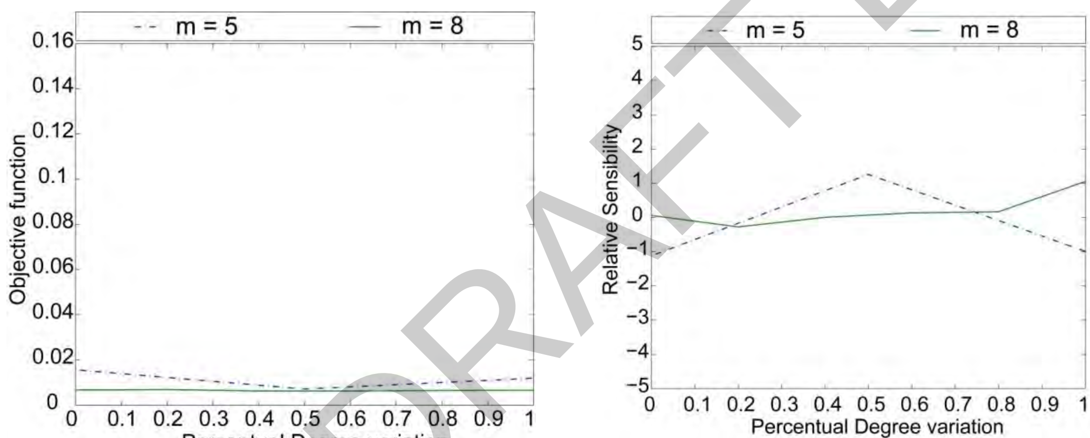


(b) Fitted curve with $b = 2$.

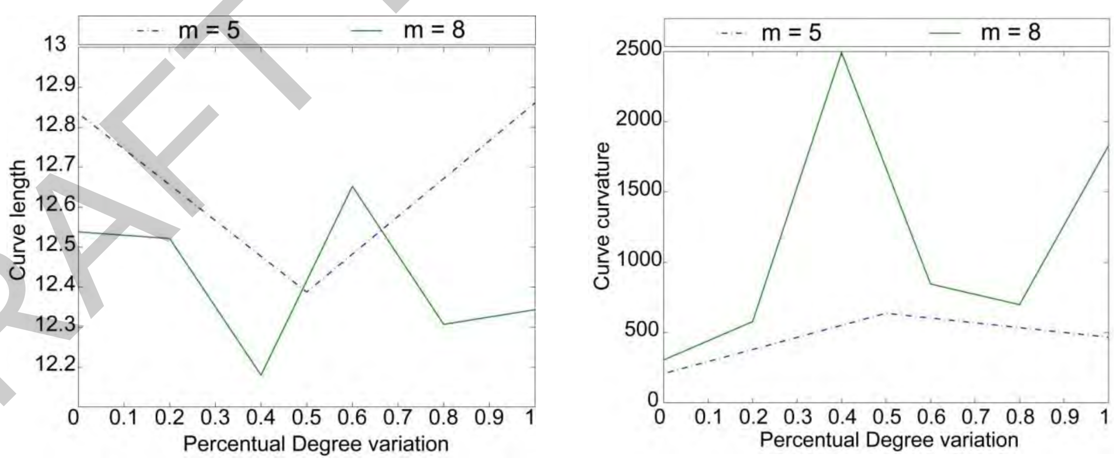


(c) Fitted curve with $b = 3$. (d) Fitted curve with $b = 4$.

Figure 10: Fitting results with different b, using 5 control points and L_2 norm.



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



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

Figure 11: Resulting metrics of the fitting curve with different degrees using 5 and 8 control points. The units of the length are ℓ. The units of the curvature are $1 / \ell$.

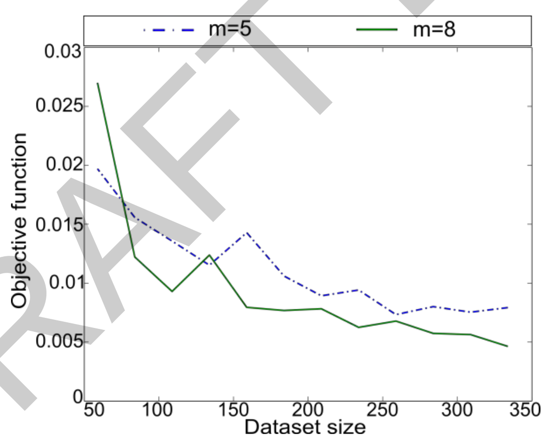
f results to be less sensitive to *b* than to *m* and to *k*. Increasing *b* improves *f( )* only for small b values. A curve degree *b*=3 maintains continutiy in curvature (2[nd] derivative of *C*(u)). It is hard (vis-à-vis the point cloud) to find justification for degrees larger than 3. . Notice that varying *b* affects *f( )* without changing the number of loops or curls in *C*(u).

4.6. Sensitivity Analysis for Dataset Size r

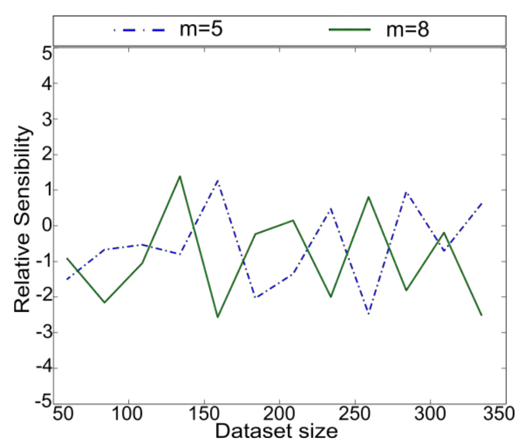
The sensitivity analysis of *f( )* with respect to *r* (population of point sample **S**) that follows assumes that **S** remains Nyquist – compliant in spite of being decimated. When **S** loses too many points and is not Nyquist compilant (e.g. Figure 14), no algorithm is able to recover the original sampled geometry.

We performed a sensitivity analysis of *f* with respect to *r*, the number of sampled points in **S**. The original size of the dataset is 334 points. For each iteration of this analysis, the dataset was reduced by eliminating 25 points. Each decimation of the input set eliminates points evenly along the original curve (i.e. as in lower quality samples). We conducted the process until 10% of the original points remained.
This experiment (Figure 12 and Figure 13) was performed twice, with *m*=5 and *m*=8 control points, norm degree *k*=2 and curve degree *b*=2.

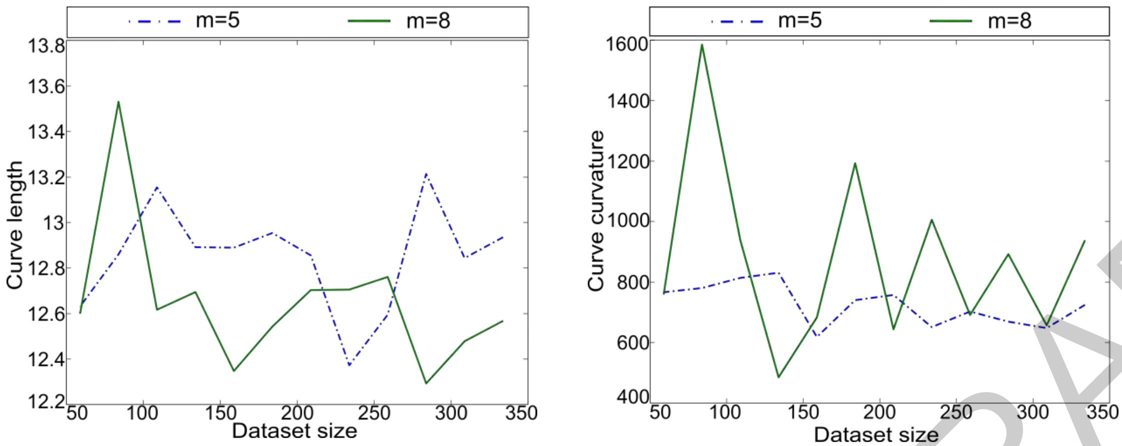
Figure 12(a) shows that the value of penalty function *f*( ) sharply increases as the sample size decreases (independent of *m*). Notice that *f* depends on the quality of the PL approximation of *C(u)* (i.e. size of set **B**, kept constant in all runs). When **S** becomes sparse, the curve-to-point distance component in *f* increases, as the distance for any point on $C(u)$ to its closest point in **S** augments, explaining why *f*( ) is negatively sensitive to *r* (Figure 12(b)). It follows that the dual distance penalization is sensitive to the level of dispersion of the point cloud **S**.



(a) Objective function vs. datset **S** size r.



(b) S_r^f vs. datset **S** size *r*.



(c) Curve length vs. datset **S** size *r* (d) Curve curvature vs. datset **S** size *r*

Figure 12: Resulting metrics of the fitting curve with different sizes of **S** using 5 and 8 control points. The units of the length are ℓ. The units of the curvature are $1 / \ell$.

Figure 13 shows that using *m*=5 control points, the resulting *C(u)* is very robust with respect to the decimation of **S** (i.e. decay in *r*). The length and curvature plots in Figure 12(c) and (d) confirm this observation. On the other hand, using *m*=8 control points produces curves *C(u)* more vulnerable to the point sample decimation. In this case, the curvature shows larger variations (Figure 12(d)). The curve *C(u)* has too many degrees of freedom available (as compared with constraints) and loops may appear in the fitted curve (Figure 13). These results show that if *m* is chosen properly and **S** remains Nyquist-compliant, our algorithm is able to reconstruct C_0, even if **S** is decimated significantly.

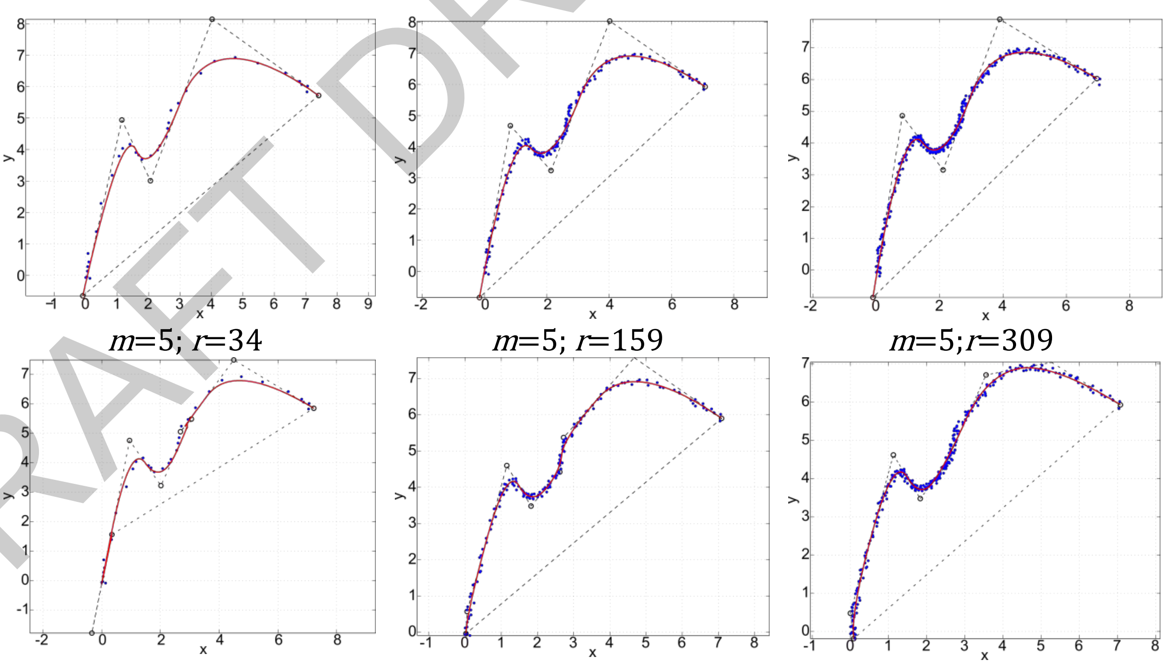


Figure 13: Resulting curves of the fitting with different sizes of **S**, using 5 and 8

control points.

However, it worth mentioning that if the samples in **S** are insufficient to describe or represent a feature in C_0, the geometry of $C(u)$ will not resemble that feature after the fitting process, as shown in Figure 14.

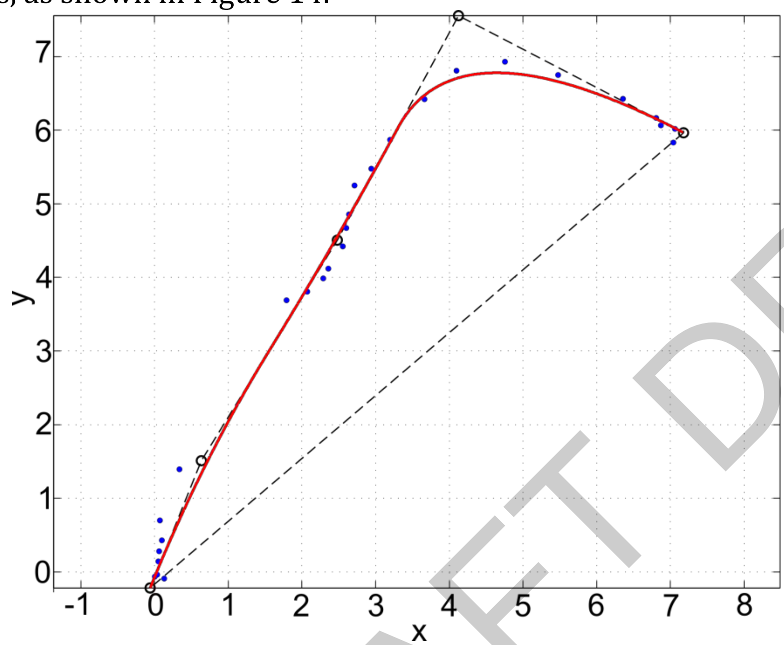
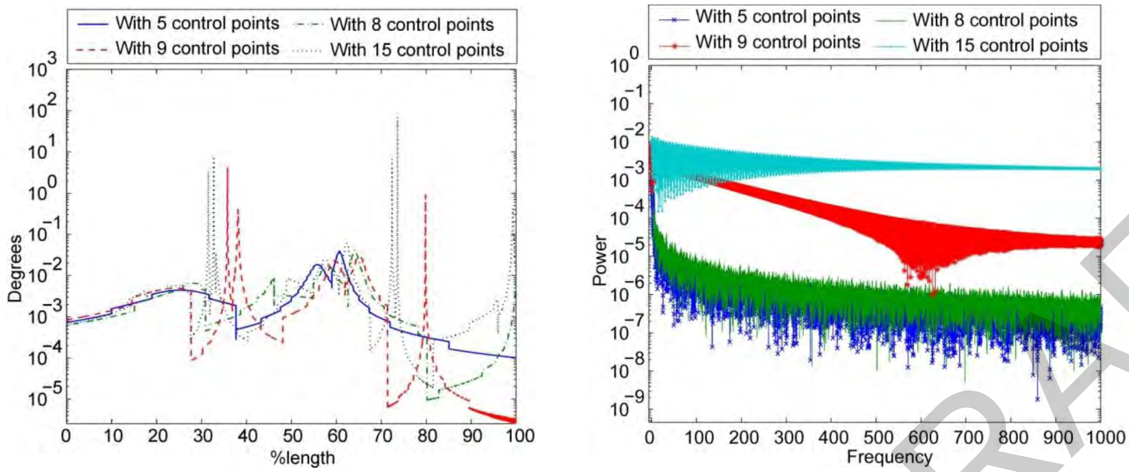


Figure 14: An excesive decimation (r=28 from an initial r=334) results in non-Nyquist samples and in $C(u)$ not resembling the original curve C_0 (test with m=5 control points).

4.7. Peaks and curls detection

The methodology described in section 3.4 was applied for the fitting curves that resulted from the optimization procedure, using m= 5, 8, 9, 15 control points, using L_2 norm. The change of direction of $\partial C/\partial u$, represented by Θ, appears in Figure 15(a) for all cases of study. For the curves generated with m=5 and m=8 control points the magnitude of Θ remains small as $\boldsymbol{C}$ is traversed. For m=9 and m=15 control points, large peaks were obtained for θ, indicating large oscillations in $\boldsymbol{C}$.

The frequency spectrum representation (see Figure 15(b)), of Θ (m=5 and m=8 control points), has low frequencies (i.e., near zero). For the cases with m=9 and m=15 control points there are high frequencies in Θ that go up to $5000/\ell$. The cases m=8, 9, 15 have a frecuency signature characteristic of a spike or Dirac delta function (high frequency content at infinite). The case m=5 presents low frequencies only, showing that curls and cusps are avoided.



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

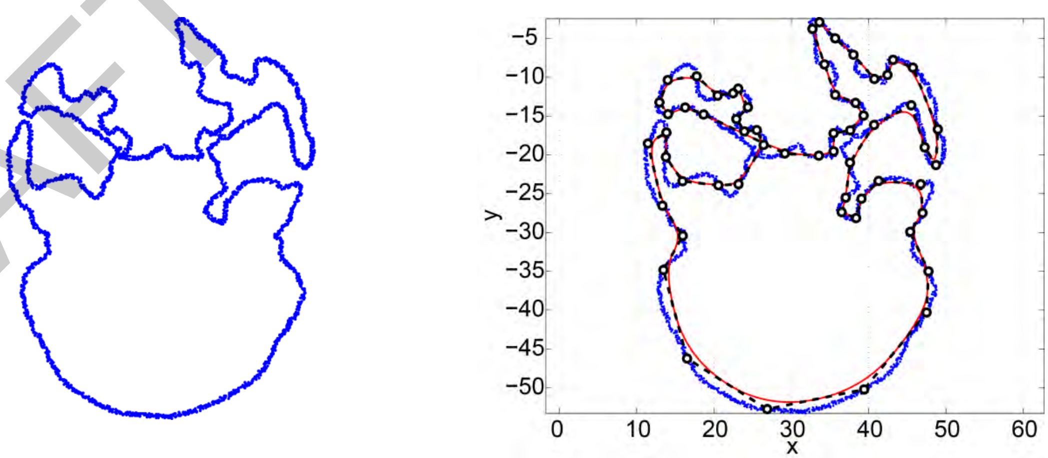
(b) Power vs. frequency.

Figure 15: Changes in direction of curve's first derivative and frequency spectrum.

The general domain of reverse engineering is closely related to the one of multi-dimensional stochastic signal processing. Therefore, it is only natural to import frequency content analysis from signal processing into the reverse engineering domain. We do recognize that our application of frequency domain methods in reverse engineering still requires a significant amount of work.

5. Real Case Scenario

The point cloud to be fitted is shown in Figure 16(a) where the principal challenges arise from the non-smoothness of the curve (high variation of curvature and concavity), as well as the presence of near self-intersecting regions (i.e. infinite Nyquist frequencies). The goal is to reconstruct the point cloud with only one parametric curve, in order to simplify its representation.



(a) Cross sectional sample of the skull to be fitted.

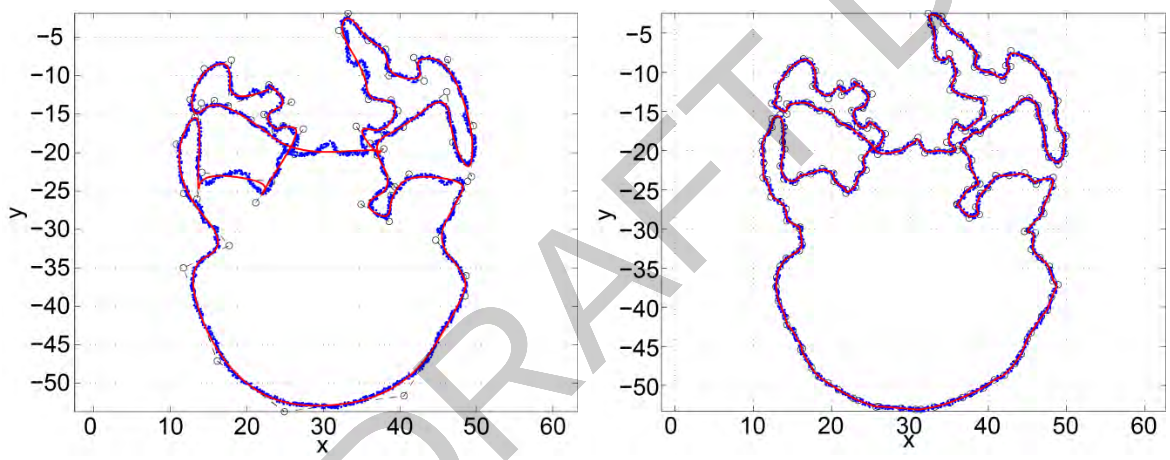
(b) A curvature-based initial guess.

Figure 16. Complete cross-section sample of a skull and initial guess used to

perform the fitting procedure.

Pre-processing: (1) The initial guess for **P** is obtained by using the methodology presented in reference (Ruiz *et al.,* 2013) with Marching Ellipsoid PCA used to obtain a PL approximation of the curve. (2) A curvature-based resampling is applied to diminish and optimize the number of control points on linear regions and to increase it in regions of high curvature.. The result of this preprocessing is shown in Figure 16(b).

As stated, m is the most critical parameter. When using insufficient control points, the resulting curve does not reproduce correctly the initial point cloud (Figure 17(a), m=57 control points, degree b=2, norm k=2), leading to a poor result. A successful fitting was obtained using m=201 control points (Figure 17(b)).

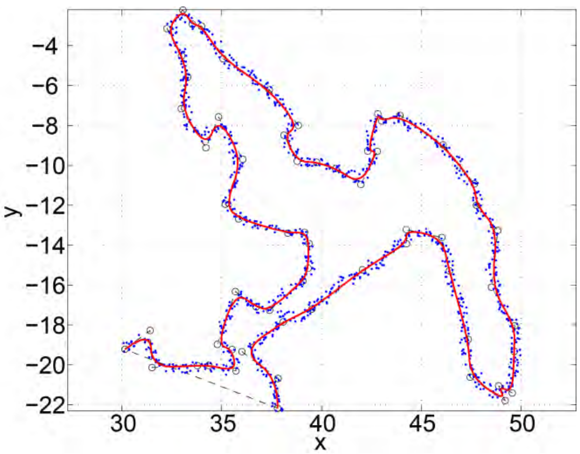


(a) Final fitting result using 57 control points.

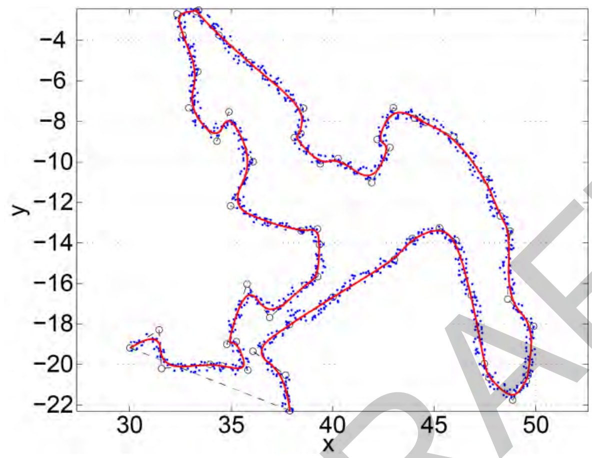
(b) Final fitting result using 201 control points.

Figure 17. Final fitting results of the complete cross sectional sample.

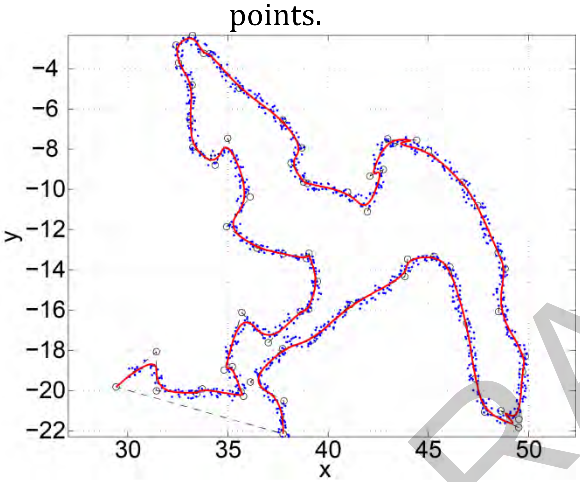
To observe the behavior when an excessive number of control points is used, one neighborhood was chosen, placing in it m=60, 63, 67, 70 control points (Figure 18). As expected, the cases m=67 and m=70 control points present curls and cusps in high frequency zones.



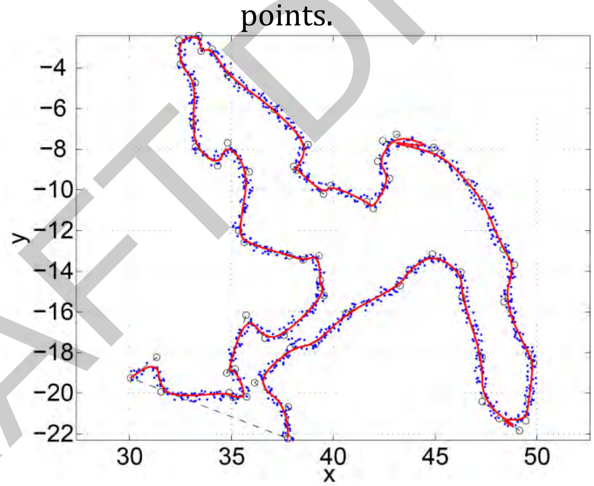
(a) Final fitting result using 60 control points.



(b) Final fitting result using 63 control points.



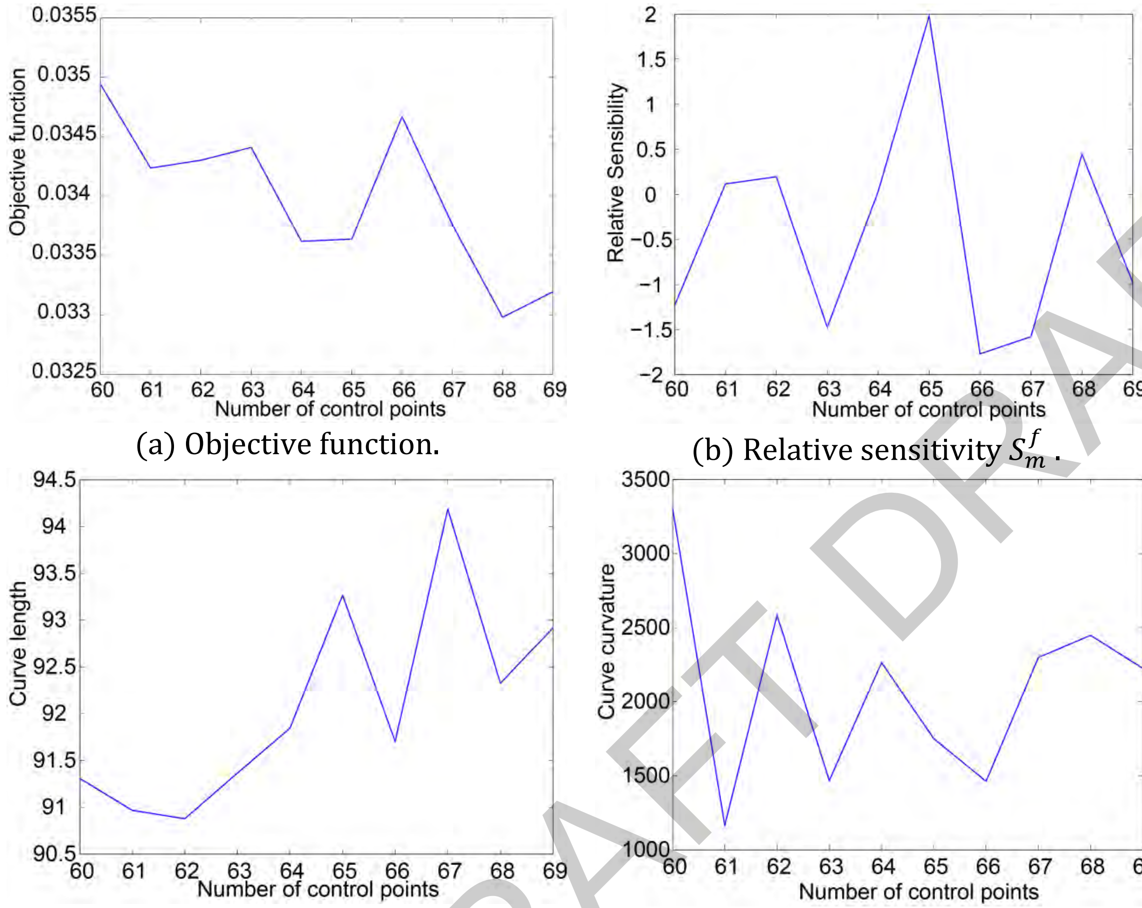
(c) Final fitting result using 67 control points.



(d) Final fitting result using 70 control points.

Figure 18. Fitting results of the upper-right segment of the skull using different number of control points.

Although these spurious features reduce *f( )* as appreciated in Figure 19(a), the topological and geometrical reconstruction is not the desired one. The observation agrees with the cases in section 4.3.



(a) Objective function. (b) Relative sensitivity S_m^f.

(c) Curve length. (d) Curvature of C(u).

Figure 19. Quantitative properties of $C(u)$ as function of the number of control points. The units of length are ℓ. The units of curvature are $1/\ell$.

Figure 20 displays the subsequent results using the data displayed in Figure 17. The cross sections in Figure 20(a) are fed to widely known surface reconstruction algorithms, resulting in Figure 20(b). These are not contributions of the present article. The information is inserted here for the sake of enhancing the understanding of the material.

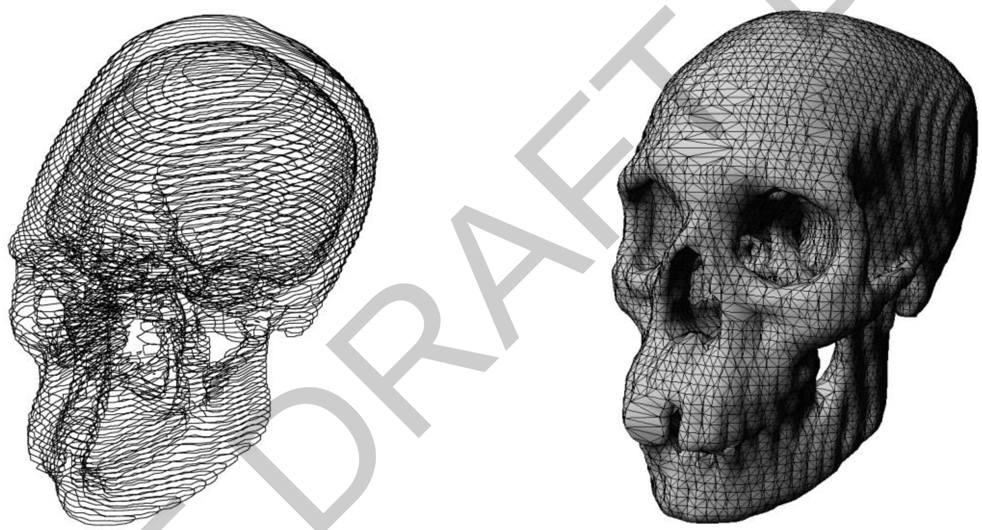
6. Conclusions and Future Work

This article presents a sensitivity analysis of the number of control points m, curve degree b and norm k on the objective function f. It has been found that using an adequate number of control points the formation of peaks and curls in C is prevented, making it unnecessary to add a curvature penalization term to f. 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 varying of m produce better results in terms of topology and geometry of the reconstructed curve. The results obtained from varying b show that the modification of this parameter is useful to perform a fine tuning of $C(u)$, without causing undesired effects on the final geometry of the curve.

It is possible to decimate the point sample **S**, whenever the it remains being Nyquist – compliant. When this boundary is crossed, **S** is simply insufficient, in terms of information content on ***C***, to reconstruct it (independent of the algorithm). Within the Nyquist compliant region, however, lowering the size of the point sample (r) leads to more difficulty in minimizing *f( )*. The dual – distance (point cloud to/from curve) proves to be robust in front of the sample decimation, in contrast with the point-to-curve distance traditional approach.

The frequency-domain analysis (DFT, FFT) of the curvature of $C(u)$ helps to detect peaks and curls, since they produce hight frequencies in the spectrum. Notice that the computational complexity of the FFT is $O(n \log n)$ (n being the number of segments approximating $C(u)$. The complexity of DFT / FFT is not a function of the size of **S**. More work is required in lowering the expense of the subsequent processing of FFT or DFT to detect curls and cusps.



(a) Cross section contours. (b) Surfaces Reconstructed from cross-section contours.

Figure 20. Usage of cross section contours in surface reconstruction, using Nuages (Boissonat and Geiger, 1993) and Contour-Mapped Nuages (Ruiz *et al.*, 2005)..

We conclude that a reasonable procedure to fit a parametric curve to a set of noisy 2D points is:

(1) With b and k constant, vary m so that the next three conditions are met: f is reduced, its variation reaches a threshold and no loops are detected. This results in an optimal $m = m_{opt}$.
(2) With k constant and m_{opt}, vary b so that f is reduced. This results in an optimal $b = b_{opt}$.
(3) With b_{opt} and m_{opt}, vary k so that f is reduced. This results in an optimal value of

$k = k_{opt}$ that leads to a curve with a better approximation to $\mathbf{S}$.

At the present time, our work is the first to address curve-fitting sensitivity within the existing literature. This article addresses the sensitivity of the goal function with respect to individual parameters. Future work includes the evaluation of cross-correlations among parameters as they influence the objective function.

The presented methodology can be applied to analyze the effect of parameters involved in a fitting process using other types of curves. Notice that for parameters that are independent of the curve type (e.g. degree of the norm k), their effect is determined according to the definition of f.

Additional work is required to study (1) the influence of the knot vector $\mathbf{X}$, (2) the systematic usage of the frequency content (DFT) of C to optimize it, (3) the quality of the digitalization and its noise distribution, since insufficient sampling density and/or stochastic noise put at risk the compliance of Nyquist criteria, and (4) different noise distributions which may dictate different strategies for the fitting process.

References

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

Boissonat, D. and Geiger, B. (1993), "Three Dimensional Reconstruction of Complex Shapes based on Delaunay Triangulation", in *Biomedical Image Processing and Biomedical Visualization*, SPIE, Vol. 1905, pp. 964-975.

Chong, E.K.P. and Żak, S.H. (2008), *An introduction to optimization, Wiley-Interscience series in discrete mathematics and optimization*, Wiley-Interscience.

Edgar, T.F., Himmelblau, D.M. and Lasdon, L.S. (2001), *Optimization of chemical processes*, McGraw-Hill.

Fiacco, A.V. (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*, Vol. 26 No. 2, pp. 192–202.

Flöry, S. and Hofer, M. (2008), "Constrained curve fitting on manifolds", *Computer-Aided Design*, Vol. 40 No. 1, pp. 25–34.

Flöry, S. and Hofer, M. (2010), "Surface fitting and registration of point clouds using approximations of the unsigned distance function", *Computer Aided Geometric Design*, Vol. 27 No. 1, pp. 60–77.

Gálvez, A., Iglesias, A., Cobo, A., Puig-Pey J. and 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*, Springer, pp. 680–693.

Heidrich, W., Bartels, R. and Labahn, G. (1996), "Fitting uncertain data with NURBS", *Proceedings of 3rd Int. Conf. on Curves and Surfaces in Geometric Design*, Vanderbilt University Pres., pp. 1–8.

Kapur, D. and Lakshman, Y. (1992), "Elimination Methods: An Introduction", in Donald B., Kapur D., Mundy J. (Ed.), *Symbolic and Numerical Computation for Artificial Intelligence*, Academic Press, pp. 45–88.

Liu, Y. and Wang, W. (2008), "A revisit to least squares orthogonal distance fitting of parametric curves and surfaces", *Advances in Geometric Modeling and Processing*, Vol. 4975, pp. 384–397.

Liu, Y., Yang, H. and Wang, W. (2005), "Reconstructing B-spline Curves from Point Clouds–A Tangential Flow Approach Using Least Squares Minimization", *2005 International Conference on Shape Modeling and Applications*, IEEE, pp. 4–12.

Nocedal, J. and Wright, S. (2006), *Numerical Optimization, Series in Operations Research and Financial Engineering*, Springer.

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

Pekerman, D., Elber, G. and Kim, M.S. (2008), "Self-intersection detection and elimination in freeform curves and surfaces", *Computer-Aided Design,* Vol. 40 No. 2, pp. 150–159.

Piegl, L.A. and Tiller, W. (1997), *The NURBS book*, Springer-Verlag.

Ruiz, O., Cadavid, C. Granados, M., Vasquez, E. and Peña, S. (2005), "2D shape similarity as a complement for Voronoi-Delone methods in shape reconstruction", *Elsevier Journal Computers & Graphics*, Vol. 29 No. 1, pp. 81-94.

Ruiz, O., Cortés, C., Aristizábal, M. and Acosta, D. (2013), "Parametric Curve Reconstruction from Point Clouds using Minimization Techniques", paper presented at the International Conference on Computer Graphics Theory and Applications GRAPP 2013, 21-24 February, Barcelona, Spain. pp 35-48.

Ruiz, O., Vanegas, C. and Cadavid, C. (2011), "Ellipse-based principal component analysis for self-intersecting curve reconstruction from noisy point sets", *The visual computer*, Vol. 27 No. 3, pp. 1–16.

Saux, E. and Daniel, M. (2003), "An improved Hoschek intrinsic parametrization", *Computer Aided Geometric Design*, Vol. 20 No. 8-9, pp. 513–521.

Ueng, W.D., Lai, J.Y. and Tsai, Y.C. (2007), "Unconstrained and constrained curve fitting for reverse engineering", *The International Journal of Advanced Manufacturing Technology*, Vol. 33 No. 11-12, pp. 1189–1203.

Wang, W., Pottmann, H. and Liu, Y. (2006), "Fitting B-spline curves to point clouds by curvature-based squared distance minimization", *ACM Transactions on Graphics (TOG)*, Vol. 25 No. 2, pp. 214–238.

Yang, H., Wang, W. and Sun, J. (2004), "Control point adjustment for B-spline curve approximation", *Computer-Aided Design*, Vol. 36 No. 7, pp. 639–652.