Sensitivity analysis in optimized parametric curve fitting

Oscar E. Ruiz and Camilo Cortés
Laboratorio de CAD CAM CAE, Universidad EAFIT, Medellín, Colombia
Diego A. Acosta
Grupo de Investigación DDP, Universidad EAFIT, Medellín, Colombia, and
Mauricio Aristizabal
Laboratorio de CAD CAM CAE, Universidad EAFIT, Medellín, Colombia

Received 13 March 2013
Revised 5 August 2013
1 November 2013
Accepted 17 December 2013

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), norm degree (k), and point sample size (r) on the optimized curve reconstruction measured by a penalty function (f). The paper aims to discuss these issues.

Design/methodology/approach – A numerical sensitivity analysis of the effect of m, b, k and r 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. The authors were able to detect the presence of such spurious features with spectral analysis. Also, the authors found that the method for curve fitting is robust to significant decimation of the point sample.

Research limitations/implications – The authors have addressed important voids of previous works in this field. The authors determined, among the curve fitting parameters m, b and k, which of them influenced the most the results and how. Also, the authors performed a characterization of the curve fitting problem from the optimization perspective. And finally, the authors devised a method to detect spurious features in the fitting curve.

Practical implications – This paper provides a methodology to select the important tuning parameters in a formal manner.

Originality/value – Up to the best of the knowledge, no previous work has been conducted in the formal mathematical evaluation of the sensitivity of the goodness of the curve fit with respect to different possible tuning parameters (curve degree, number of control points, norm degree, etc.).

Keywords Sensitivity analysis, Minimization, Noisy point sample, Parametric curve fitting, Reverse engineering

Paper type Research paper

Nomenclature

C_0:	unknown C^1-differentiable simple planar curve	r:	number of sampled points in set **S**
$C(u)$:	parametric planar curve approaching C_0	$C(u_i)$:	point on $C(u)$ closest to point $\boldsymbol{p}_i$
S:	$\{\boldsymbol{p}_0,\boldsymbol{p}_1\ldots,\boldsymbol{p}_r\}$. Noisy unordered point sample of C_0	$d(\boldsymbol{p},\mathbf{S})$:	distance from point $\boldsymbol{p}$ to the point set **S**
		k:	degree of norm: $\left(\sum \lvert x_i \rvert^k\right)^{1/k}$

Engineering Computations: International Journal for Computer-Aided Engineering and Software
Vol. 32 No. 1, 2015
pp. 37-61

0264-4401
DOI 10.1108/EC-03-2013-0086

ℓ	length units	$O(g(n))$:	computational expense is a function $g(n)$ of the data size n
m:	number of control points of $C(u)$		
P:	$[\boldsymbol{P}_0, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_{m-1}]$. Control polygon of $C(u)$	**X**:	knot vector
		B:	sequence of parameter values. $\mathbf{B} = [u_0, u_1, \ldots, u_n]$
b:	degree of parametric curve $C(u)$		

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 (Piecewise Linear (PL)) or higher (C^1, C^2, …) continuity. In the second case, we talk about a smooth parametric free form (i.e. the subject of this paper). In either case, a usual goal is to obtain C statistically "centered" at 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 paper, to find the free form curve fitting a noisy point set, we used the heuristic of minimizing the 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 paper.

In this paper we will address planar Open Uniform B-splines, with knot vector such that $0 \leqslant u \leqslant 1$ (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 **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 f (Equation (1)) between C and **S**:

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

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

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

1.2 Context of the optimization problem

The terms in Equation (1) that can be changed to minimize *f( )* are: *m* (number of control points), **P** (control polygon), *b* (degree of parametric curve), **X** (knot vector), *k* (norm type, Equation (2)), and *w* (power of the distances, Equation (1)). In this paper we choose to analyze the sensitivity of the curve fitting with respect to *m*, *k*, and *b* (parameters) with the tuning variable being **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 signs.

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 paper, *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 paper we work under the assumption of compliance of the Nyquist-Shannon conditions by the sample.

2. Literature review

2.1 Objective function

Flöry and Hofer (2010) employs first order residuals ($w = 1$ in Equation (1)) while 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

Equation (2) calculates the distance or residual d_i of the objective function (Equation (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 ith cloud point $\boldsymbol{p}_i$ such that d_i is a minimum (perpendicular distance point-curve, Equation (7)). Namely:

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

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

Solving for u in $G(u)=0$ is achieved by using Newton's Method (Liu *et al.*, 2005; Piegl and Tiller, 1997), numerically minimizing $G(u)$ (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 distance. Liu and Wang (2008) present a review of the methods for solving (or approaching) Equations (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, avoiding curls and outliers.

2.3 Effect of curve fitting parameters

Number of control points m. Ueng *et al.* (2007) present 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. Yang *et al.* (2004) show 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. Heidrich *et al.* (1996) perform a comparison among L_1, L_2 and L_∞ norms in curve fitting applications with several data sets. 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 performed yet. 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 paper 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. Pekerman *et al.* (2008) present 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 paper, we introduce the use 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 paper

There are several open issues in optimized curve fitting to point clouds: effect of parameters such as the number of control points m, knot vector $\mathbf{X}$ and norm k; detection of peaks and curls in $C(u)$ to validate optimal parameter value identification and mathematical characterization of the curve fitting problem from the viewpoint of optimization.

The knot vector $\mathbf{X}$ 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 paper, to explore: 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 a quantitative analysis in the frequency domain of the curvature of $C(u)$ to detect peaks and curls. We aim at reconstructing 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 still 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 Equation (1), for the curve fitting algorithm implemented in this paper.

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

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

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).

To avoid the computational expenses of algebraic roots calculation, we sample the domain for u in [0,1], (i.e. [0, Δ_u, $2\Delta_u$,…, 1.0]) and approximate the curve C with

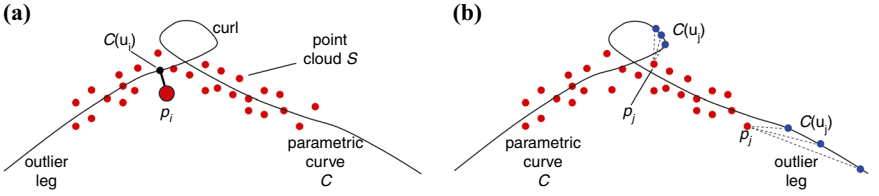


Figure 1.
Distances cloud points to/from curve

Notes: (a) Distances cloud point to curve; (b) distances curve to cloud point

a polyline [$C(0)$, $C(\Delta_u)$, $C(2\Delta_u)$,…, $C(1.0)$]. Approximating $C(u_i)$ in Equation (8) for a given $\boldsymbol{p}_i$ simply entails traversing [$C(0)$, $C(\Delta_u)$, $C(2\Delta_u)$,…, $C(1.0)$] to find the $C(\kappa\Delta_u)$ closest to $\boldsymbol{p}_i$. 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 (Equation (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 Equation (1) does not avoid spurious outlier legs and curls outside the boundaries of $\mathbf{S}$. 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 $\mathbf{S}$ 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:

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

$\mathbf{A_j}$, a partition of curve C, contains those points in the sequence [$C(0)$,$C(\Delta_u)$,$C(2\Delta_u)$ …, $C(1.0)$] that are closer to $\boldsymbol{p}_j\in\mathbf{S}$ than to any other point of $\mathbf{S}$. 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 $\mathbf{S}$.

From the previous discussion, residuals d_i for Equation (1) are defined by:

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

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

Figure 2 presents a simplified picture of the situation, with few cloud points, biased with respect to C curve. Calculations are depicted in Table I. 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.

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

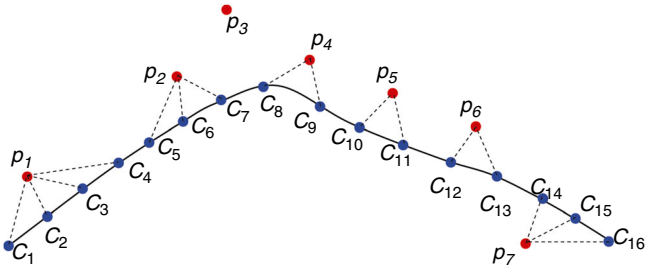


Figure 2.
Clusters of distances from curve to cloud points

anywhere in $\mathbb{R}^2$ the problem is unconstrained. The eigenvalues of f's Hessian matrix (Equation (10)) 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 ($0 \leqslant i < i_{\max}$) 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 Equation (11):

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

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 Ruiz *et al.,* 2011) on the whole point set $\mathbf{S}$. The number of sub-divisions of L is also 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 $\overrightarrow{\partial C / \partial u}$ that contain contributions of high frequencies. We computed the discrete Fourier transform

$\boldsymbol{p}_i$	$\mathbf{A_i}$	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 I. Calculations using curve to cloud-point distances for example in Figure 2

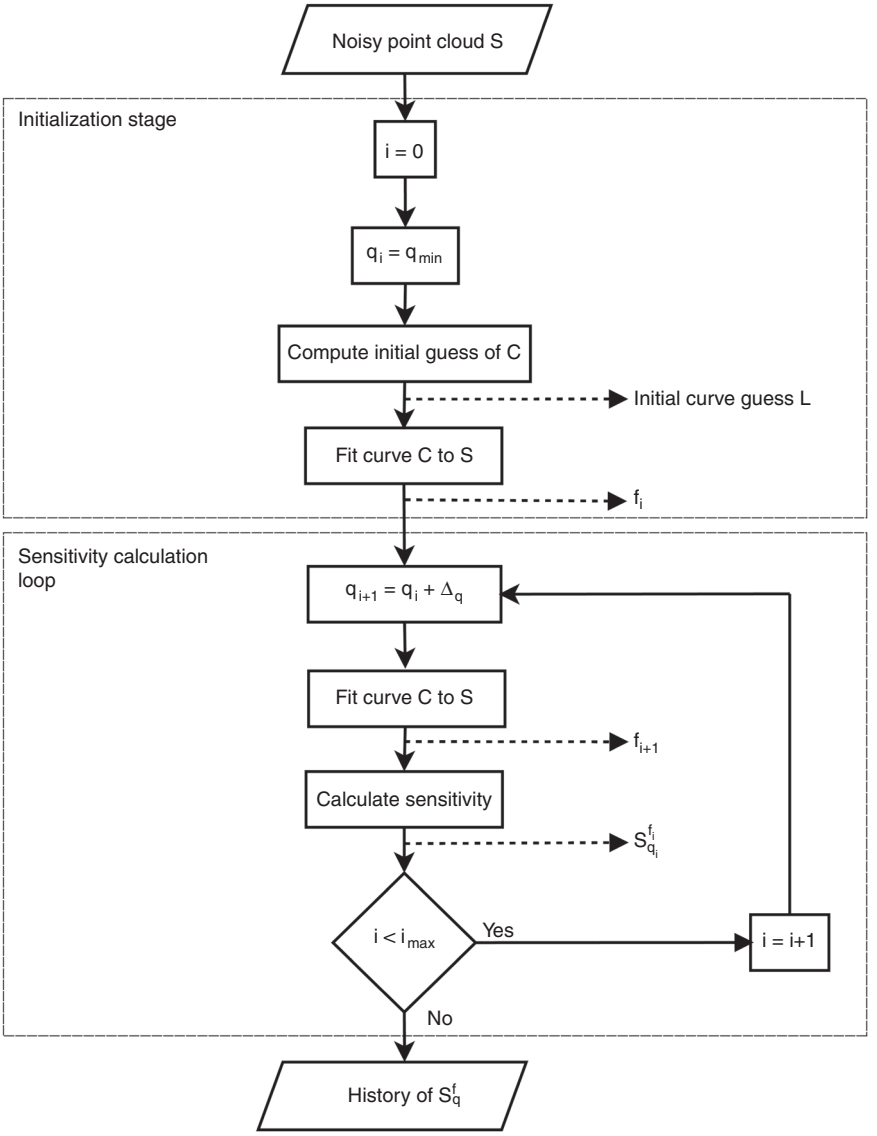


Figure 3.
Steps of the sensitivity calculation

(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 ($\mathbf{U_s} = \{u_0, \ldots, u_t\}$). Then we chose $d_s = 0.0001 \times l$, where l is a unit distance, and the sampling frequency f_s is 10,000/l.

Subsequently, the normalized tangent vectors of the curve were computed at all points of $\mathbf{U_s}$ yielding $\mathbf{V_s} = \{\hat{\boldsymbol{v}}_0, \ldots, \hat{\boldsymbol{v}}_t\}$. The dot product $\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{v}}_{i+1}$ is calculated, with $i = 0,1,2 \ldots, t-1$; and then the angle θ_i between $\hat{\boldsymbol{v}}_i$ and $\hat{\boldsymbol{v}}_{i+1}$ is obtained. Finally DFT is computed for $\theta = \{\theta_0, \ldots, \theta_{t-1}\}$, 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 five 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 $4 \leqslant m \leqslant 16$. 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 paper 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, if m overly grows, there appears a formation of curls and/or peaks and attraction among control points, while $f(\ )$ does not decrease significantly. 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.

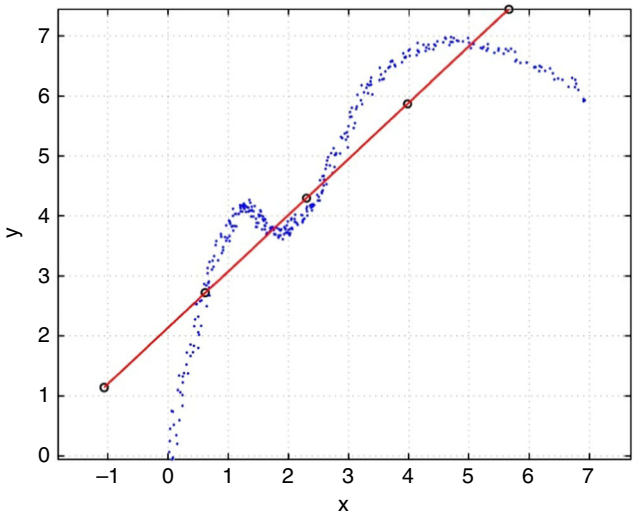


Figure 4.
Point cloud and initial curve guess with five control points

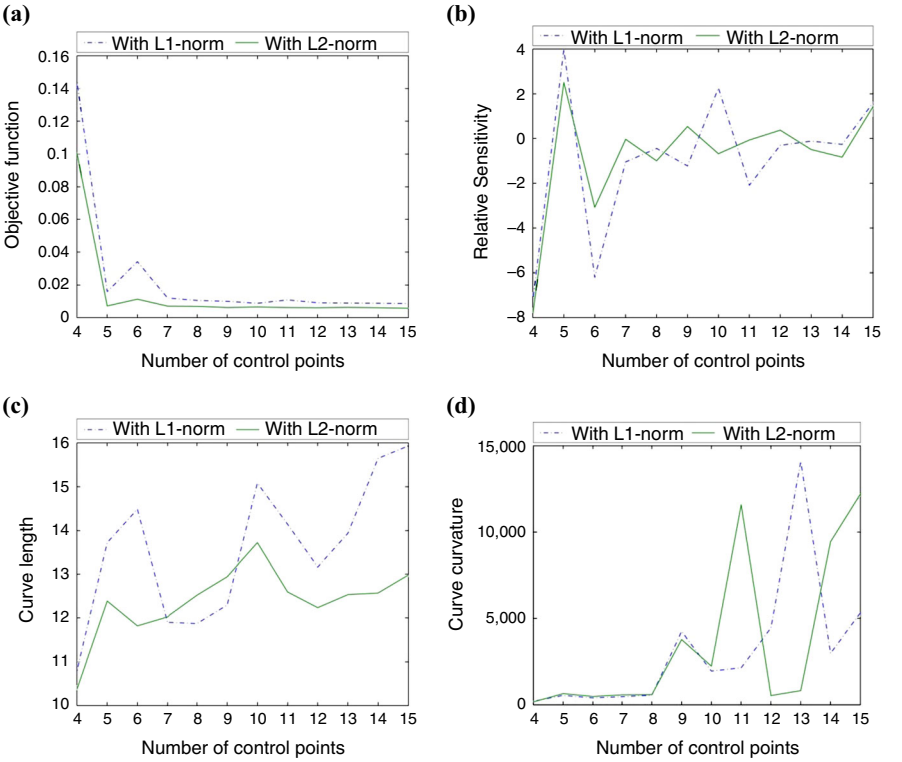


Figure 5.
Resulting metrics of the fitting curve with different number of control points using $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ norms

Notes: (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. Here the units of the length are ℓ and the units of the curvature are $1/\ell$

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 respect to k when $m = 5$ Curve length and curvature in Figure 7(c) and (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

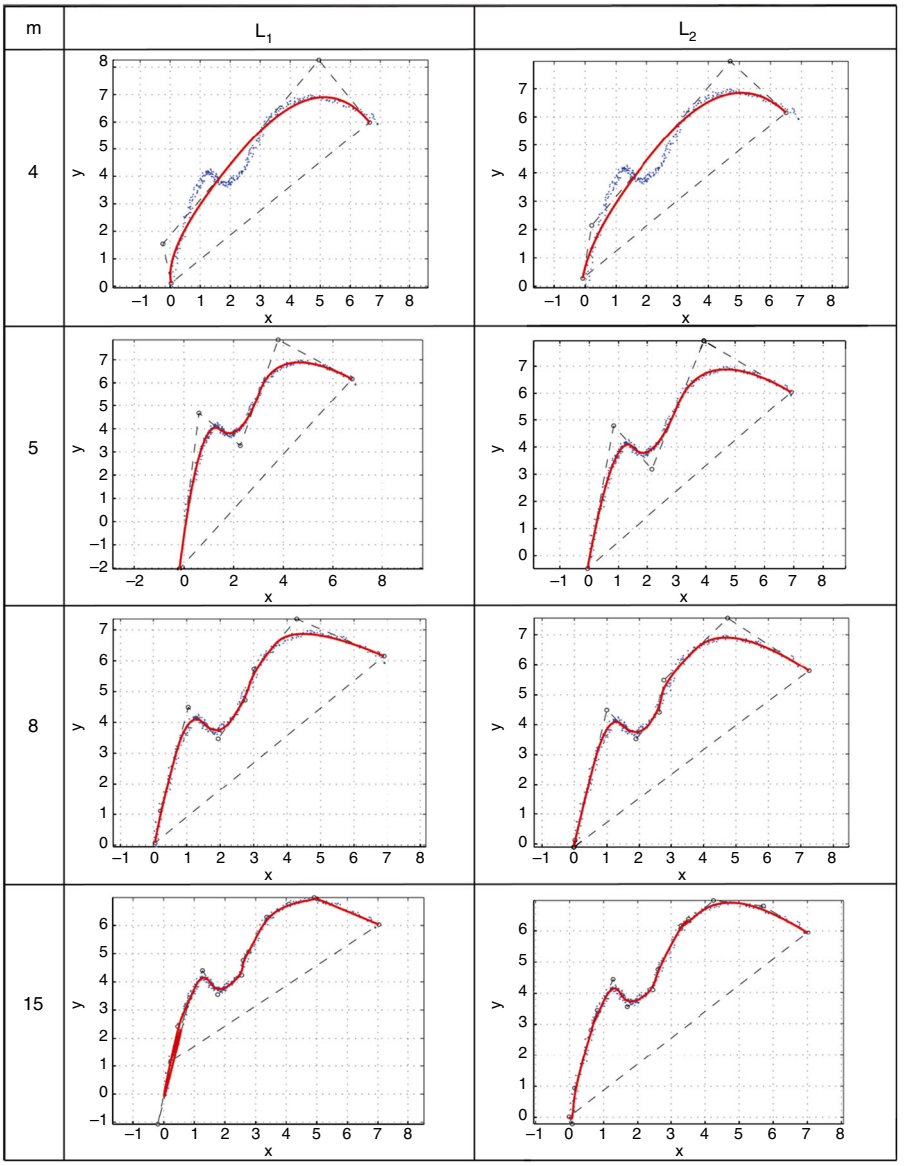


Figure 6.
Resulting curves of the fitting with different number of control points m, using $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ norms

would impede the fitting of curves to sharp corners, which is an advantage of our approach.

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 $\mathbf{S}$.

4.5 Sensitivity analysis for degree b

This experiment varies b, in the range $1 \leqslant b \leqslant m-1$. Figures 10 and 11 presents results with $m = 5$ and $m = 8$, and $k = 2$ (Euclidean norm).

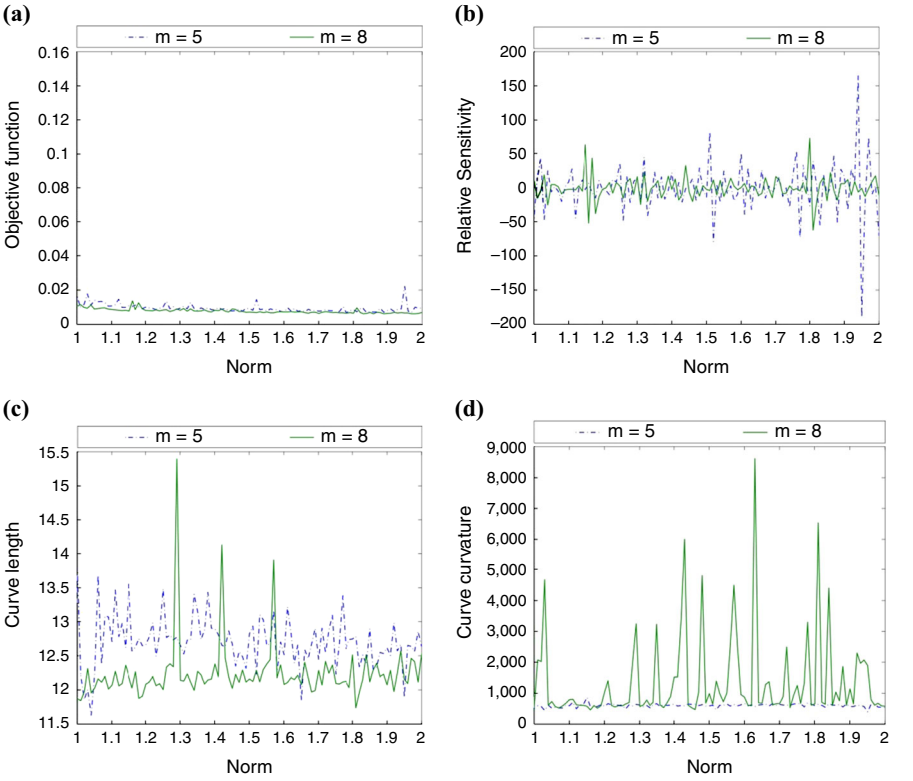


Figure 7.
Resulting metrics of the fitting curve with different norms using 5 and 8 control points

Notes: (a) Objective function vs norm; (b) S_k^f vs norm; (c) curve length vs norm; (d) curve curvature vs norm. Here the units of the length are ℓ and 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 (second 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 data set 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 data set is 334 points. At each iteration of this analysis, the data set 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 percent of the original points remained.

This experiment (Figures 12 and 13) was performed twice, with $m=5$ and $m=8$ control points, norm degree $k=2$ and curve degree $b=2$.

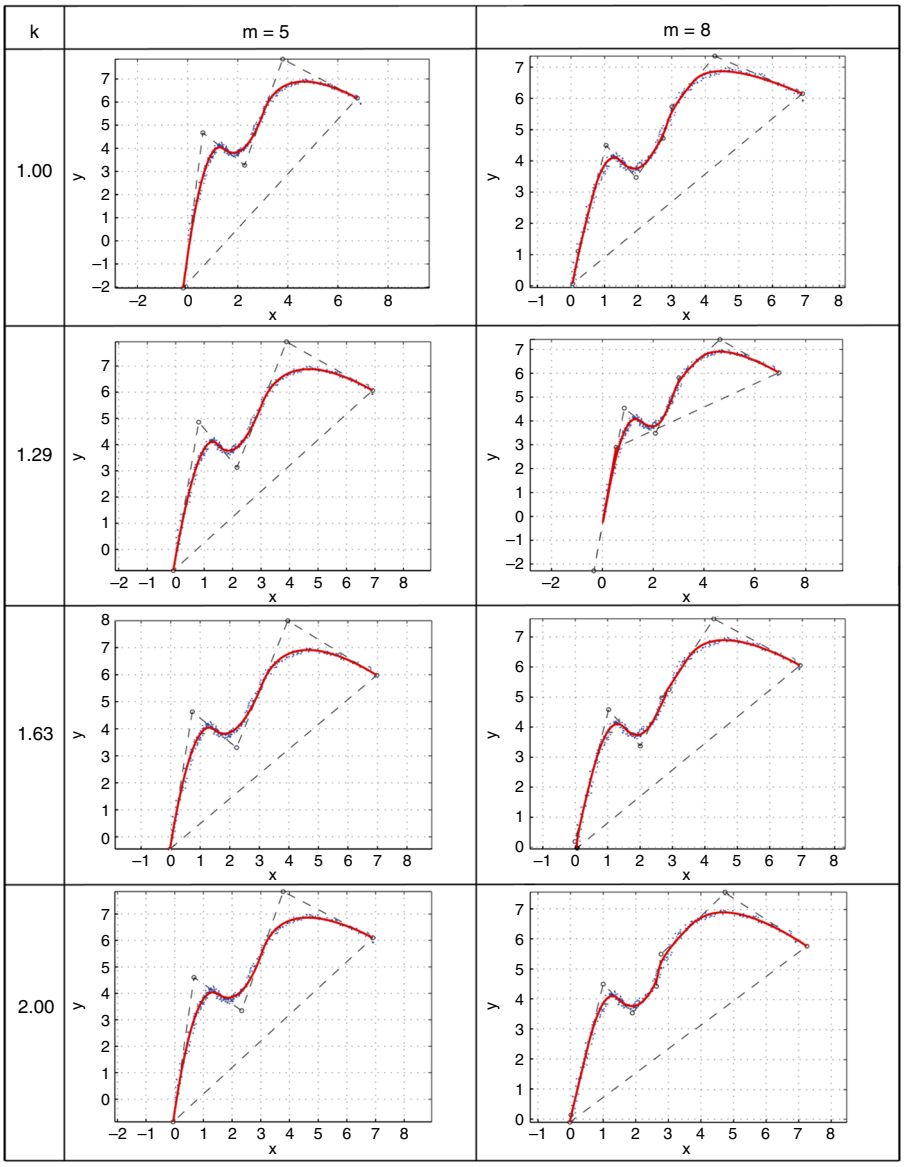


Figure 8.
Resulting curves of the fitting with different norms k, using 5 and 8 control points

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**.

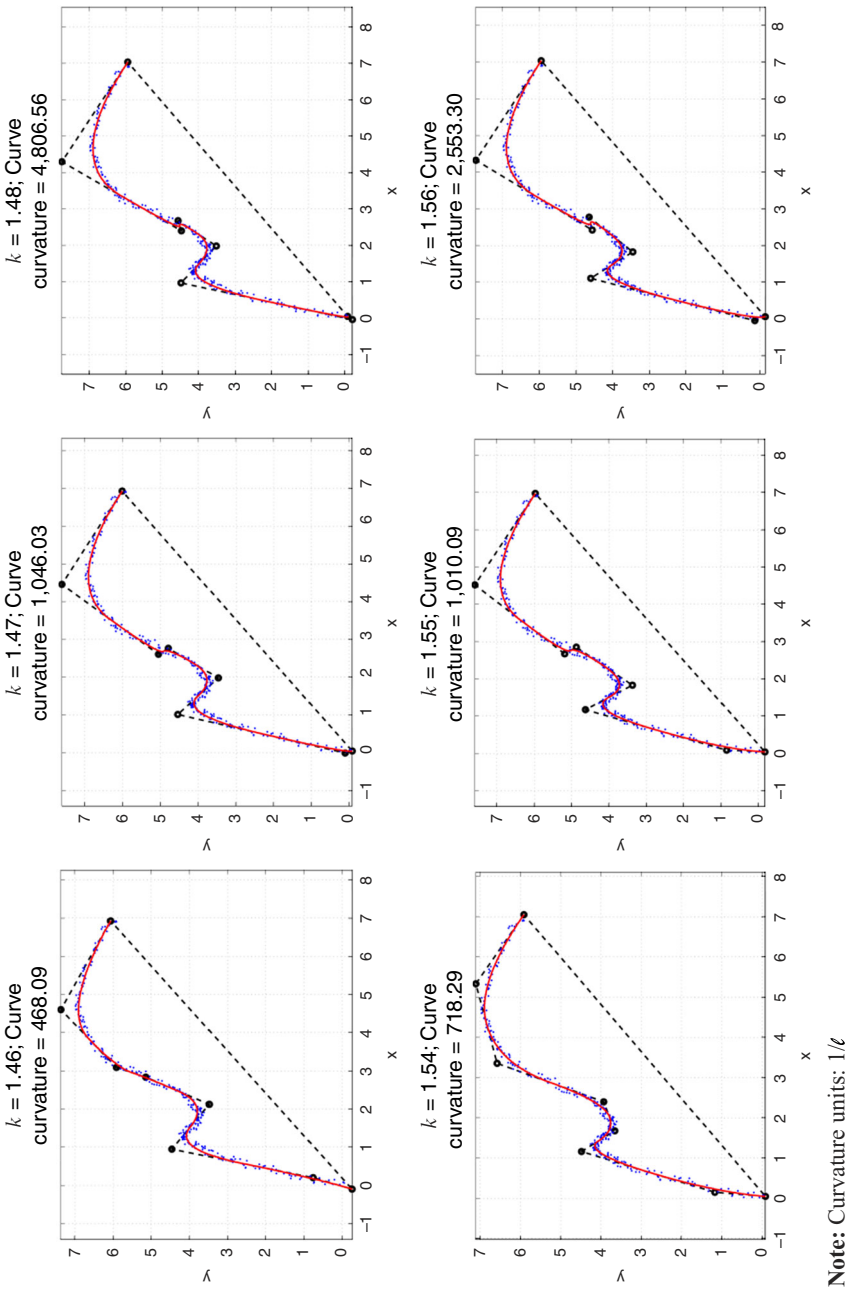


Figure 9.
Detail of curvature variation with norm type (k) for cases with over-population of control points

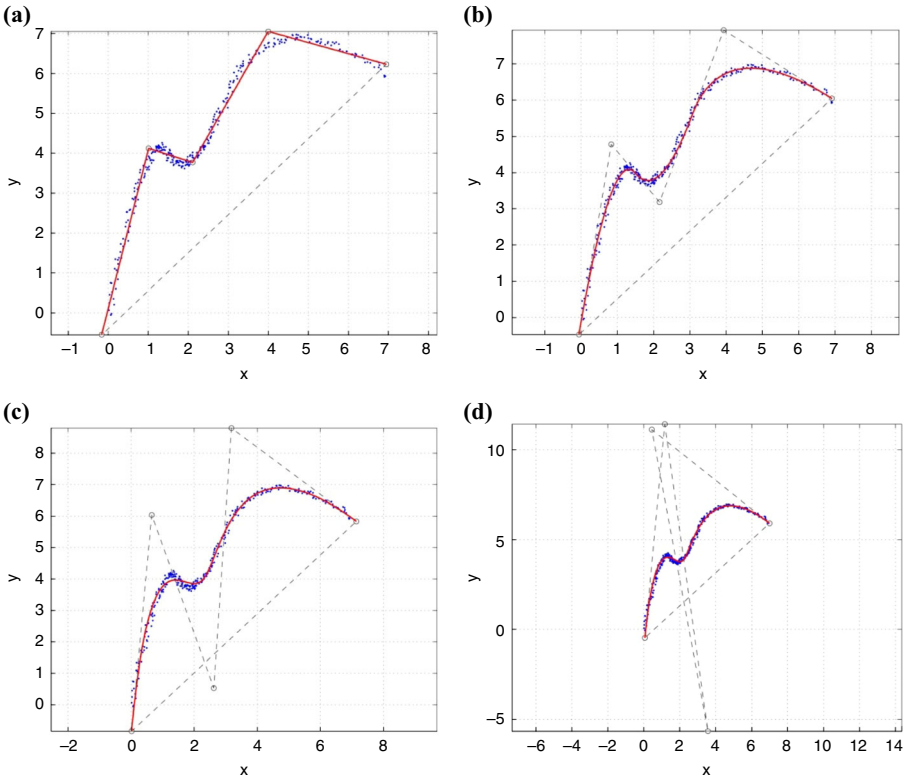


Figure 10.
Fitting results with different b, using 5 control points and $\mathbf{L}_2$ norm

Notes: (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 13 shows that using $m = 5$ control points, the resulting $C(u)$ is very robust with respect to the decimation of $\mathbf{S}$ (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 $\mathbf{S}$ remains Nyquist-compliant, our algorithm is able to reconstruct C_0, even if $\mathbf{S}$ is decimated significantly.

However, it is worth mentioning that if the samples in $\mathbf{S}$ 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.

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

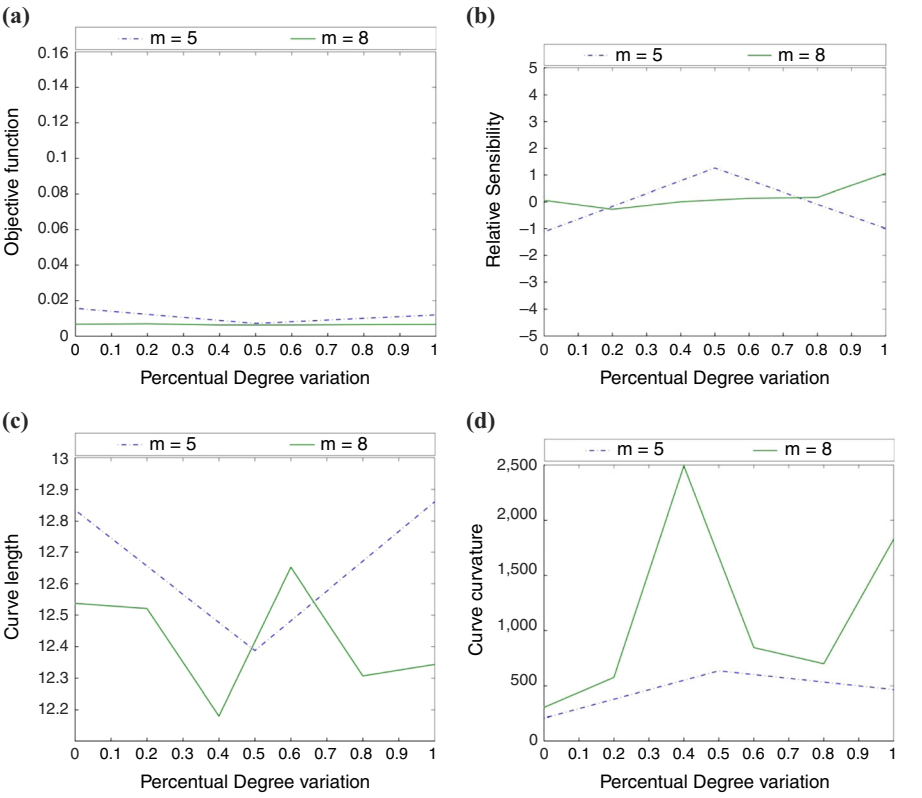


Figure 11.
Resulting metrics of the fitting curve with different degrees using 5 and 8 control points

Notes: (a) Objective function vs degree; (b) S_k^f vs norm; (c) curve length vs degree; (d) curve curvature vs degree. The units of the length are ℓ. The units of the curvature are $1/\ell$

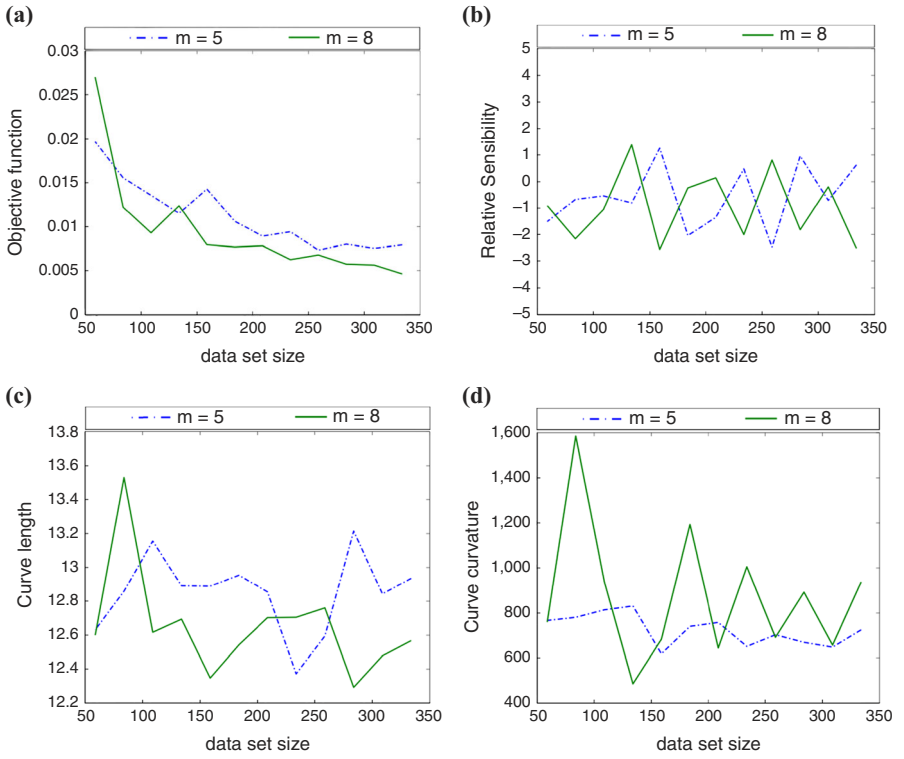
magnitude of Θ remains small as C is traversed. For $m = 9$ and $m = 15$ control points, large peaks were obtained for θ, indicating large oscillations in C.

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 5,000/l. The cases $m = 8$, 9, 15 have a frecuency signature characteristic of a spike or Dirac δ function (high frequency content at infinite). The case $m = 5$ presents low frequencies only, showing that curls and cusps are avoided.

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),



Notes: (a) Objective function vs data set S size r; (b) S_r^f vs data set S size r; (c) curve length vs data set S size r; (d) curve curvature vs data set S size r. The units of the length are ℓ. The units of the curvature are $1/\ell$

Figure 12. Resulting metrics of the fitting curve with different sizes of **S** using 5 and 8 control points

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.

Pre-processing: first, the initial guess for **P** is obtained by using the methodology presented in Ruiz *et al.* (2013) with Marching Ellipsoid PCA used to obtain a PL approximation of the curve; second, 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 pre-processing 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)).

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.

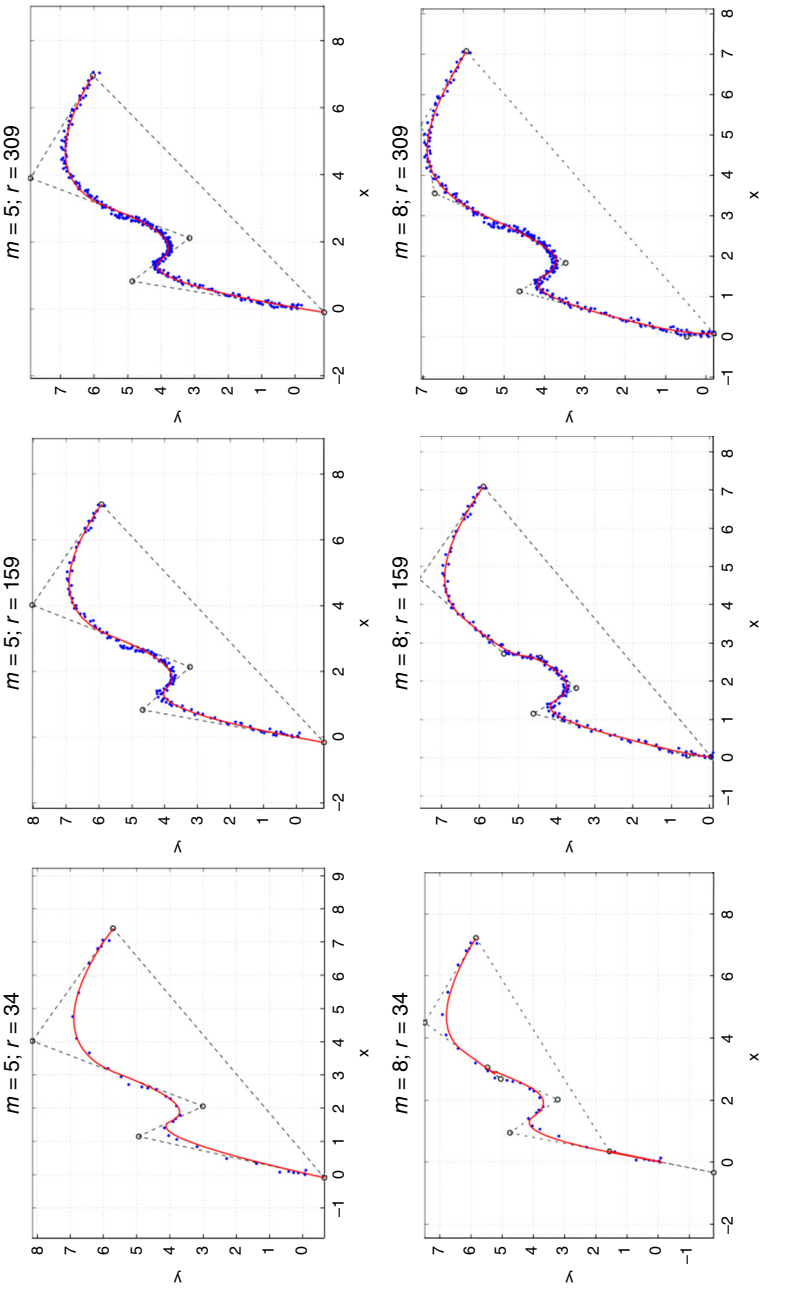


Figure 13.
Resulting curves of the fitting with different sizes of S, using 5 and 8 control points

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.

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 paper. The information is inserted here for the sake of enhancing the understanding of the material.

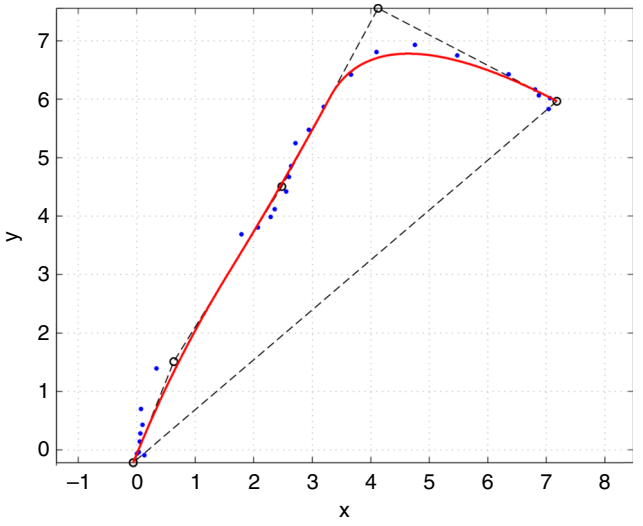


Figure 14.
An excesive decimation ($r = 28$ from an initial $r = 334$) results in non-nyquist samples and in $C(u)$ not resembling the original curve C_0 (test with $m = 5$ control points)

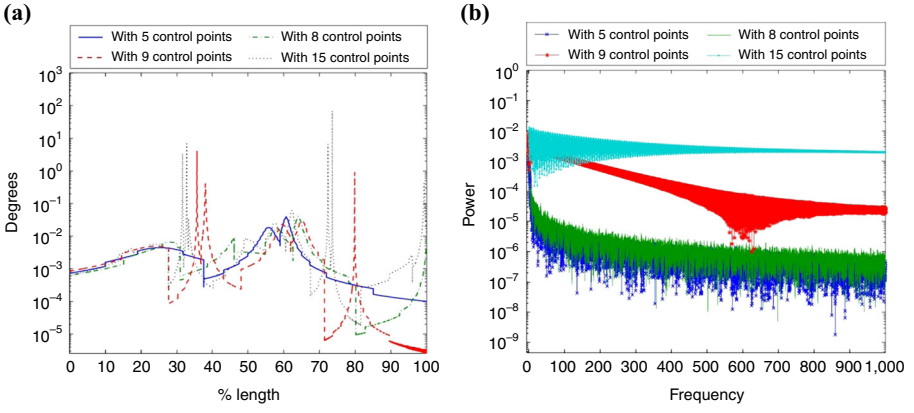


Figure 15.
Changes in direction of curve's first derivative and frequency spectrum

Notes: (a) Changes in direction of curve's first derivative (degrees) vs length (percentage); (b) power vs frequency

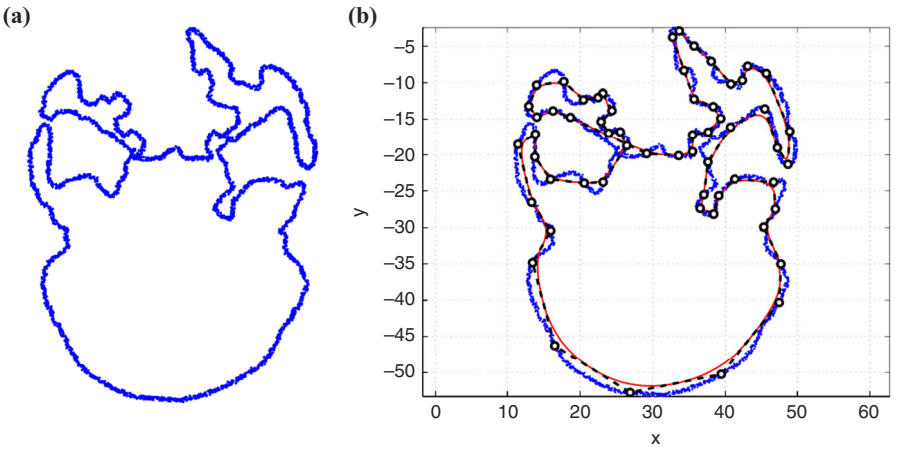


Figure 16.
Complete cross-section sample of a skull and initial guess used to perform the fitting procedure

Notes: (a) Cross-sectional sample of the skull to be fitted; (b) a curvature-based initial guess

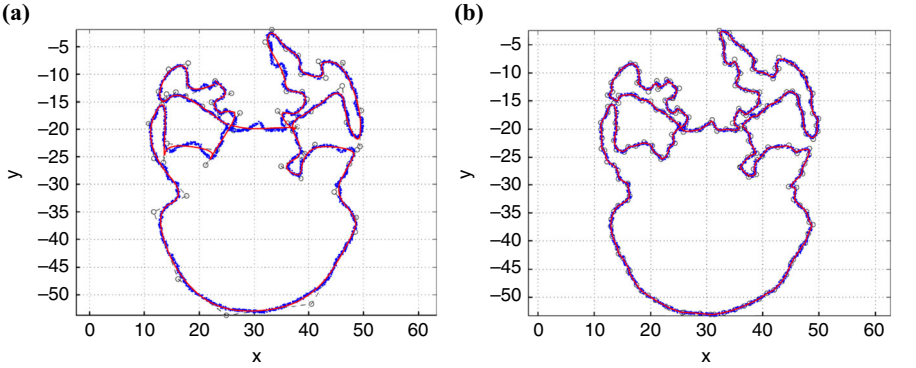


Figure 17.
Final fitting results of the complete cross sectional sample

Notes: (a) Final fitting result using 57 control points; (b) final fitting result using 201 control points

6. Conclusions and future work

This paper 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.

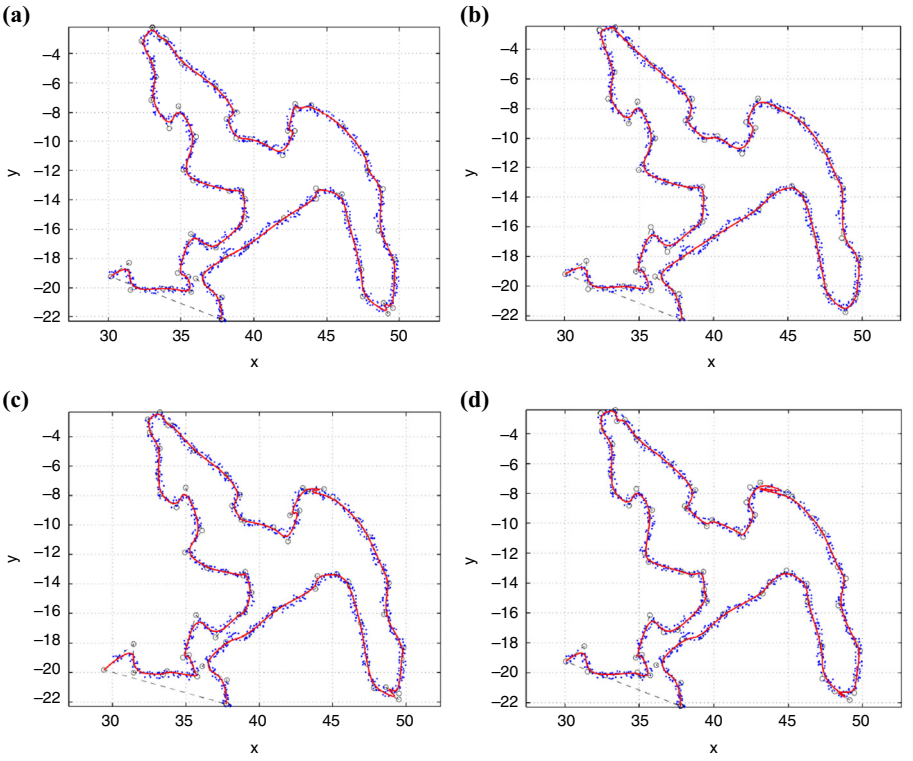


Figure 18.
Fitting results of the upper-right segment of the skull using different number of control points

Notes: (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

It is possible to decimate the point sample **S**, whenever 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.

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 $m = m_{opt}$.

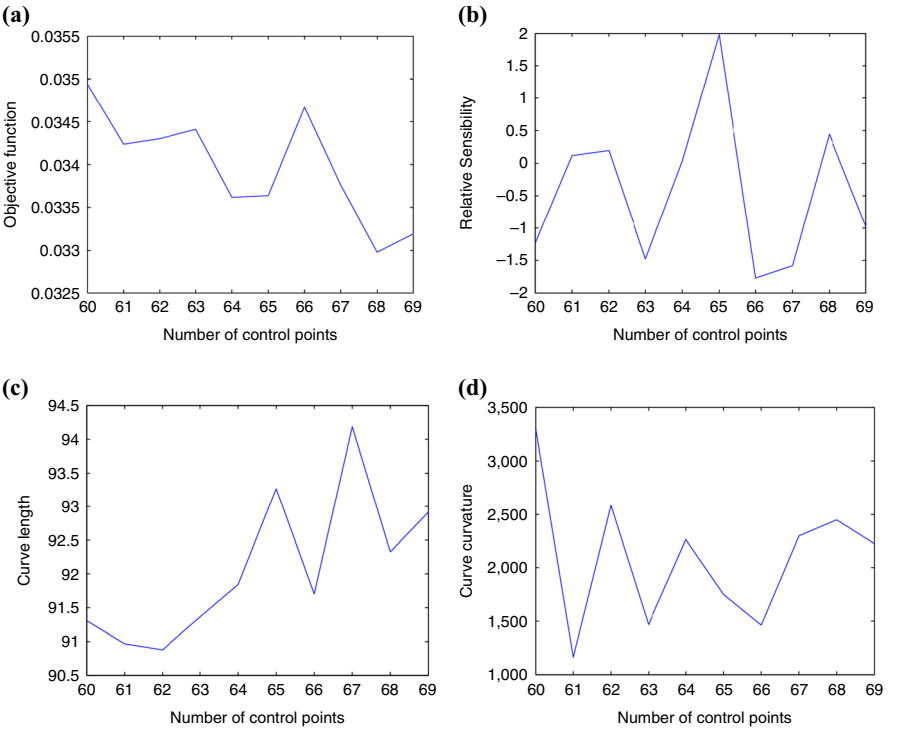


Figure 19. Quantitative properties of *C(u)* as function of the number of control points

Notes: (a) Objective function; (b) relative sensitivity S_m^f; (c) curve length; (d) curvature of C(u). The units of length are ℓ. The units of curvature are $1/\ell$

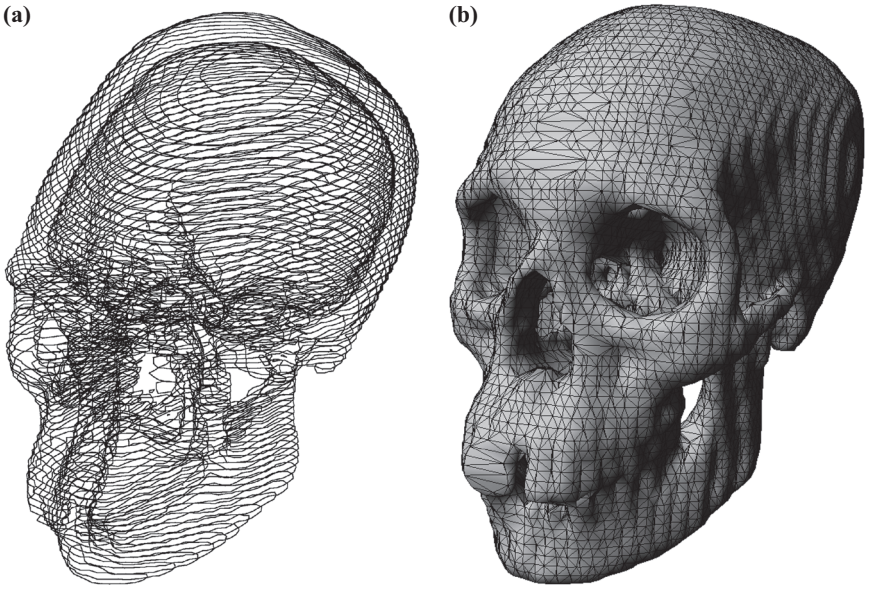
(2) With k constant and m_{opt}, vary b so that f is reduced. This results in an optimal $b = b_{opt}$.

(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 paper 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 the influence of the knot vector $\mathbf{X}$, the systematic usage of the frequency content (DFT) of C to optimize it, 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 different noise distributions which may dictate different strategies for the fitting process.



Notes: (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)

References

Boissonat, D. and Geiger, B. (1993), "Three dimensional reconstruction of complex shapes based on Delaunay triangulation", in Acharya, R.S. and Goldgof, D.B. (Eds), *Proc. SPIE 1905, Biomedical Image Processing and Biomedical Visualization*, Vol. 1905, International Society for Optical Engineering, SPIE, San Jose, CA, pp. 964-975, available at: http://proceedings.spiedigitallibrary.org/volume.aspx?volumeid=11727

Chong, E.K.P. and Żak, S.H. (2008), *An Introduction to Optimization*, Wiley-Interscience Series in Discrete Mathematics and Optimization, ISBN, 9781118033340, John Wiley & Sons, Inc.,Hoboken, NJ, pp. 585-586, available at: http://dx.doi.org/10.1002/9781118033340.scard

Edgar, T.F., Himmelblau, D.M. and Lasdon, L.S. (2001), *Optimization of Chemical Processes*, ISBN, 9780070393592, 2nd ed., McGraw-Hill, New York, NY, available at: http://trove.nla.gov.au/work/12178671

Fiacco, A.V. (1983), *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Series Monographs and Textbooks Mathematics in Science and Engineering, Vol. 165, Academic Press, New York, NY.

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", in Gervasi, O. and Marina, L.G. (Eds), *Proceedings of the 2007 International Conference on Computational Science and its Applications – Volume Part II*, Lecture Notes in Computer Science ICCSA' 07, ISBN, 3-540-74475-4; 978-3-540-74475-7, Kuala Lumpur, Springer-Verlag, Berlin, Heidelberg, pp. 680-693, available at: http://dl.acm.org/citation.cfm?id=1802954.1803023,acmid = {1803023}

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

Kapur, D. and Lakshman, Y. (1992), "Elimination methods: an introduction", in Donald, B., Kapur, D. and Mundy, J. (Eds), *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", in Chen, F. and Jüttler, B. (Eds), *Advances in Geometric Modeling and Processing*, Lecture Notes in Computer Science, ISBN, 978-3-540-79245-1, Vol. 4975, Springer Verlag, Berlin, Heidelberg, pp. 384-397, available at: http://dx.doi.org/10.1007/978-3-540-79246-8_29

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, ISBN, 978-0-387-40065-5, Springer Verlag, New York, NY.

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*, Monographs in Visual Communication, ISBN, 978-3-642-59223-2, 2nd ed., Springer-Verlag, New York, NY.

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.

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, Barcelona, February 21-24, pp. 35-48.

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.

Saux, E. and Daniel, M. (2003), "An improved Hoschek intrinsic parametrization", *Computer Aided Geometric Design*, Vol. 20 Nos 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 Nos 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.

Further reading

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*, ISBN, 3540762175, 1st ed., Springer Verlag, Inc., New York, NY.

Papadimitriou, C.H. and Steiglitz, K. (1998), *Combinatorial Optimization: Algorithms and Complexity*, Series Dover Books on Computer Science, ISBN-13: 978-0486402581, ISBN-10: 0486402584 Dover Publications Inc., Mineola, New York, NY.

About the authors

Associate Professor Oscar E. Ruiz was born in 1961 in Tunja, Colombia. He obtained BSc Degrees in Mechanical Engineering (1983) and Computer Science (1987) at the Los Andes University, Bogota, Colombia, an MSc Degree with emphasis in CAM (1991) and a PhD with emphasis in CAD (1995) from the Mechanical & Industrial Engineering Department of University of Illinois at Urbana – Champaign, USA. Dr Ruiz has held a Visiting Researcher positions at Ford Motor Co (Dearborn, USA 1993 and 1995), Fraunhofer Inst. Graphische Datenverarbeitung (Darsmstad, Germany 1999 and 2001), University of Vigo (1999 and 2002), Max Planck Institute for Informatik (2004) and Purdue University (2009). In 1996 Dr Ruiz was appointed as Faculty of the Mechanical Engineering and Computer Science Departments at the EAFIT University, Medellin, Colombia, and has ever since the Coordinator of the Laboratory for Interdisciplinary Research on CAD/CAM/CAE. Dr Ruiz's interests are Computer Aided Geometric Design, Geometric Reasoning and Applied Computational Geometry. Associate Professor Oscar E. Ruiz is the corresponding author and can be contacted at: oruiz@eafit.edu.co

Dr Camilo Cortés was born in 1986 in Bogota, Colombia. Camilo obtained the Diploma in Mechatronics from Escuela de Ingenieria de Antioquia (Colombia) in 2010. He immediately joined the MSc program of Universidad EAFIT in the Laboratory of CAD CAM CAE under the supervision of Professor Dr Oscar Ruiz. In 11-2011, Camilo Cortés started an ongoing research internship in the Institute for Visual Communication Technologies-Vicomtech in San Sebastian, Spain as part of his MSc Thesis. His research interests are robotics (emphasis in medical and health applications), mechatronics, applied computational geometry and optimization of geometrical setups.

Professor Diego A. Acosta was born in 1969 at Spartanburg, SC (USA). Dr Acosta earned a BS in Chemical Engineering from the Universidad Pontificia Bolivariana (UPB) in Medellín, Colombia and MS and PhD Degrees in Chemical Engineering from the University of Oklahoma at Norman, Oklahoma (USA). Dr Acosta has worked at Ashland Chemicals, Xerox and Smurfit Kappa as a Process Engineer. Dr Acosta is a Tenured Professor of the Process Engineering Department at the Universidad EAFIT in Medellín, Colombia. His research interests include chemical process design, mathematical programming, advanced mathematics and statistics and design of experiments. Professor Acosta directs the research group Design and Development of Processes and Products (DDP) and lectures courses in Statistics, Design of Experiments, Process Engineering Design, Mass Transfer Laboratory, and Process Engineering Optimization.

Mauricio Aristizabal was born in 1989 in Medellín, Colombia. He obtained BSc Degree in Mechanical Engineering at the EAFIT University in 2012. He was a Teaching Assistant of the Statics course in 2009 and in the Basic-Sciences Department in 2010. He joined the CAD CAM CAE Laboratory as a Research Assistant since June 2010, under the supervision of Professor Dr Eng. Oscar Ruiz. Mauricio has developed code for optimized fitting of parametric curves to noisy point samples, calculation of singular points in parallel mechanisms, automated reuse of traditional SW for web rendering and accelerated rendering of vector fields (oceanic streams) on web platforms. His research interests are geometric modeling, CAD, compliant mechanisms and machine design.