Non-convex onion peeling using a shape hull algorithm

Jalal M. Fadili, Mahmoud Melkemi, Abderrahim Elmoataz

Non-convex onion-peeling using a shape hull algorithm

M.J. Fadili [a,*], M. Melkemi [b], A. ElMoataz [a,1]

[a] GREYC CNRS UMR 6072, ENSICAEN 6, Bd du Maréchal Juin, 14050 Caen, France
[b] LIRIS FRE 2672 CNRS, Université Claude Bernard Lyon 1, 43 boulevard du 11 Novembre 1918, Bat. 710, 69622 Villeurbanne, France

Abstract

The convex onion-peeling of a set of points is the organization of these points into a sequence of interpolating convex polygons. This method is adequate to detect the shape of the "center" of a set of points when this shape is convex. However it reveals inadequate to detect non-convex shapes. Alternatively, we propose an extension of the convex onion-peeling method. It consists in representing a set of points with a sequence of non-convex polylines which are computed using the $\mathscr{A}$-shape descriptor. This method is applied to robust statistical estimation. It is shown that it makes the estimators robust to the presence of outliers by removing suspect samples from the available population.

Keywords: Onion-peeling; Non-convex shapes; $\mathscr{A}$-shape; Robust statistics

1. Introduction

The convex onion-peeling method is a popular tool of computational geometry organizing a finite non-organized set of points in a sequence of strips (Chazelle, 1985; Preparata and Shamos, 1985; Abellanas et al., 1992; Okabe et al., 1992; Boissonnat and Yvinec, 1995). The first strip is the convex hull of the set of points, the second one is the convex hull of the set of points minus the points on the first strip. The remaining strips are computed similarly. The process continues until the empty set is reached. Building this method is motivated by the need of efficient methods allowing to organize a set of points and to extract the convex shape embedded in their center. This problem is of prime importance in computational geometry and has several applications, e.g. in statistics (Preparata and Shamos, 1985).

An important problem in statistics is the estimation of a population parameter, such as the mean, by observing only a sample of size n drawn randomly from the population. While the sample mean is an unbiased estimator of the population mean, it is extremely sensitive to outliers, observations that lie abnormally far from most of the

* Corresponding author. Tel.: +33-231452920; fax: +33-231452698.

E-mail addresses: jalal.fadili@greyc.ismra.fr (M.J. Fadili), melkemi@bat710.univ-lyon1.fr (M. Melkemi), abder.elmoataz@greyc.ismra.fr (A. ElMoataz).

[1] Present address: LUSAC EA 2607, Site Universitaire BP 78, 50130 Cherbourg-Octeville, France.

others. It is adequate to reduce the effects of outliers because they often represent spurious data that would otherwise introduce errors in the analysis. A related property that a good estimator should enjoy is that of robustness, i.e. insensitivity to deviations from the idealized assumptions for which the sample estimator is optimized.

Statisticians have developed various sorts of robust statistical estimators. An important class of such estimators are those known as the Gastwirth estimators (Gastwirth, 1966) (also called the *L*- or α-trimmed estimators). They are based on the fact that we tend to trust observations more the closer they are to the center of the samples. In higher dimensions, a convex-peeling method is used (see Huber, 1979, 1981; Abellanas et al., 1992; Chazelle, 1985; Preparata and Shamos, 1985). It consists on stripping away the convex hull of the set, then removing the convex hull of the remainder, and continuing until $(1-2\alpha)$ fraction of the n points remains. Here, α is the proportion of outliers among available samples. However, the convex hull-based method is inadequate when the points are on a concave region.

In this article, we propose a generalization of the convex onion-peeling algorithm. We organize a finite points set as a sequence of strips, where each strip is a closed non-convex interpolating polyline. The novelty of this approach is that the strips are not constrained to be convex and the method makes possible the extraction of a non-convex shape embedded in the center of a cloud of points. These properties are essential for the estimation of parameters with α-trimmed estimators.

In our method the strips are computed using the $\mathscr{A}$-shape concept (Melkemi and Djebali, 2000, 2001). The first strip corresponds to an $\mathscr{A}$-shape of the set of points where $\mathscr{A}$ is the set of vertices of elongated Voronoi polygons. The points interpolated by the first strip are added to $\mathscr{A}$ and deleted from the original set of points, the new $\mathscr{A}$-shape of the remaining set of points corresponds to the second strip. The other strips are computed iteratively according to the same process.

This article is organized as follows: in Section 2, we present geometrical concepts used to build the onion-peeling algorithm. Section 3 is devoted to the non-convex onion-peeling algorithm based on the $\mathscr{A}$-shape concept. Section 4 presents the application of non-convex onion-peeling to robust statistical estimation.

2. Voronoi diagram and Delaunay triangulation

Let $S=\{p_1,p_2,\ldots,p_n\}$ be a set of n distinct points of the plane, $d(p,q)$ is the Euclidean distance between two points p and q. We define a Voronoi polygon of a point p_i by:

$$R(S,p_i)=\{p\in\mathbb{R}^2; d(p,p_i) < d(p,p_j)\forall j \text{ such that } i\neq j\} \quad (1)$$

The set $V(S)=\{R(S,p_1),\ldots,R(S,p_n)\}$ defines a partition of the plane. This set is called Voronoi diagram of S. The Delaunay triangulation of S, denoted $\mathrm{DT}(S)$, is a decomposition of the convex hull of S by triangles. For each triangle having its vertices in T (T is subset of S and contains three points), there exists an open disc b such that $b\cap S=\emptyset$ and $\partial b\cap S=T$, where ∂b is the circle bounding b (∂b is called the *Delaunay circle*).

3. $\mathscr{A}$-shape onion-peeling

3.1. $\mathscr{A}$-shape definition

Consider again a finite set of points $S=\{p_1,p_2,\ldots,p_n\}$, and $\mathscr{A}$ is a finite set of points. The following definitions introduce the concept of $\mathscr{A}$-shape of the points set S:

Definition 1

- An edge $[p_ip_j]$ of the Delaunay triangulation $\mathrm{DT}(S\cup\mathscr{A})$ is $\mathscr{A}$-exposed, if there exists a triangle (P_iP_ja) in $\mathrm{DT}(S\cup\mathscr{A})$ where a belongs to $\mathscr{A}$.
- $\mathscr{A}$-shape of S, $F(S,\mathscr{A})$, is the set of $\mathscr{A}$-exposed edges extracted from $\mathrm{DT}(S\cup\mathscr{A})$.
- $VF(S,\mathscr{A})$ denotes the vertices set of $F(S,\mathscr{A})$.

Varying the positions of the points of $\mathscr{A}$, $\mathscr{A}$-shapes of a set S describe a family of graphs which represents S with different levels of details. Some examples of $\mathscr{A}$-shape spectrum are presented in Fig. 1. Figures in the top row correspond to the Voronoi diagrams of the union of the points set S

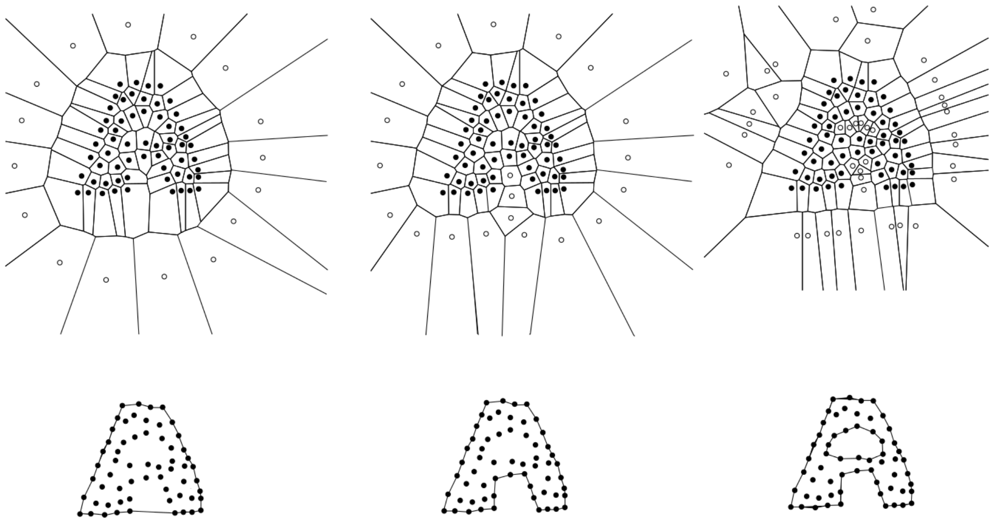


Fig. 1. Examples of $\mathscr{A}$-shape spectrum obtained with different choices of $\mathscr{A}$.

(black discs) and three different examples of the set $\mathscr{A}$ (white discs). The $\mathscr{A}$-shapes associated to these choices of $\mathscr{A}$ are shown respectively in the bottom row of Fig. 1. When the points of $\mathscr{A}$ are outside the union of the Delaunay discs circumscribing the triangles of the Delaunay triangulation of S, and when the set of vertices of the convex hull of $S \cup \mathscr{A}$ is equal to the set of $\mathscr{A}$, then $\mathscr{A}$-shape of S simply reduces to the convex hull of S.

3.2. Computing the first strip: computing the set $\mathscr{A}$

In this section, we present the method to compute the points of the set $\mathscr{A}$, which is at the heart of the non-convex onion-peeling algorithm. Let us consider the bounded Voronoi polygons, $R(S,x)$ $x \in S$, adjacent to the unbounded ones.

Definition 2. (*Definition of $\mathscr{A}$*)
Let $r_{\max}(x)$ be the maximum distance from x to the boundary of the polygon $R(S,x)$, $r_{\min}(x)$ denotes the minimal distance, and $g(x)$ is the center of gravity of $R(S,x)$.

- The bounded polygon $R(S,x)$, adjacent to an unbounded polygon, is said to be an *elongated polygon* if and only if $r_{\max}(x)/r_{\min}(x) > t$ (t is a fixed real number).
- $A(x)$ is the set of the Voronoi vertices v of the elongated Voronoi region $R(S,x)$ which verify the relationship:

$$d(v,g(x))^2 \leqslant d(v,x)^2 + \frac{1}{2}d(x,g(x))^2 \tag{2}$$

- Let $\mathscr{A}'$ be a finite set of points which are outside the union of the Delaunay discs circumscribing the triangles of the Delaunay triangulation of S. The vertices of the convex hull of $S \cup \mathscr{A}'$ are the points of the set $\mathscr{A}'$.
- $\mathscr{A}$ is the union set of $\mathscr{A}'$ and the sets $A(x)$, for all $x \in S$ with $R(S,x)$ bounded and adjacent to an unbounded polygon.
- The $\mathscr{A}$-shape of S is the first strip of the $\mathscr{A}$-shape onion-peeling sequence.

An example of non-elongated and elongated polygons are shown in Fig. 2. When $t = 3$, the polygon of Fig. 2(a) is non-elongated ($r_{\max}(x)/r_{\min}(x) = 1.33 < t = 3$) and the polygon of Fig. 2(b) is elongated ($r_{\max}(x)/r_{\min}(x) = 10.01 > t = 3$).

The vertices of $A(x)$ are located in the half-space containing the gravity center $g(x)$ and defined by the radial axis of the weighted points $(g(x), \sqrt{\frac{1}{2}}d(x,g(x)))$ and $(x,0)$ (see Fig. 3 for illustration).

Referring to Definition 2, the set $\mathscr{A}$ depends on a single parameter t; a positive real number greater

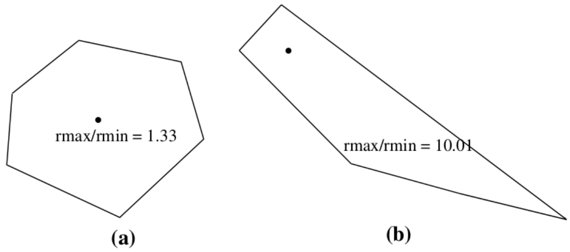


Fig. 2. (a) A non-elongated polygon, (b) an elongated polygon.

than 1. When t is sufficiently large such that all the polygons are considered non-elongated, then the sets $A(x)$ are empty. Thus $\mathscr{A}$ is the set $\mathscr{A}'$. In this case, the $\mathscr{A}$-shape of a set of points S is the convex hull of S (Fig. 4(b)). Otherwise, $\mathscr{A}$-shape reflects concave shapes. The finer shape reflected by $\mathscr{A}$-shape is achieved when t is equal 1. In this case, all the polygons adjacent to the unbounded ones

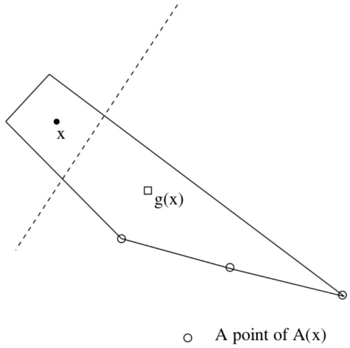


Fig. 3. The points of the set $A(x)$.

are considered elongated. Two examples of finer shapes are illustrated in Fig. 4(c) and (d). In our experiments, we employ the threshold $t = 3$. In Figs. 5(b) and 6, we depict an example of the points set $\mathscr{A}$ used to compute the first strip.

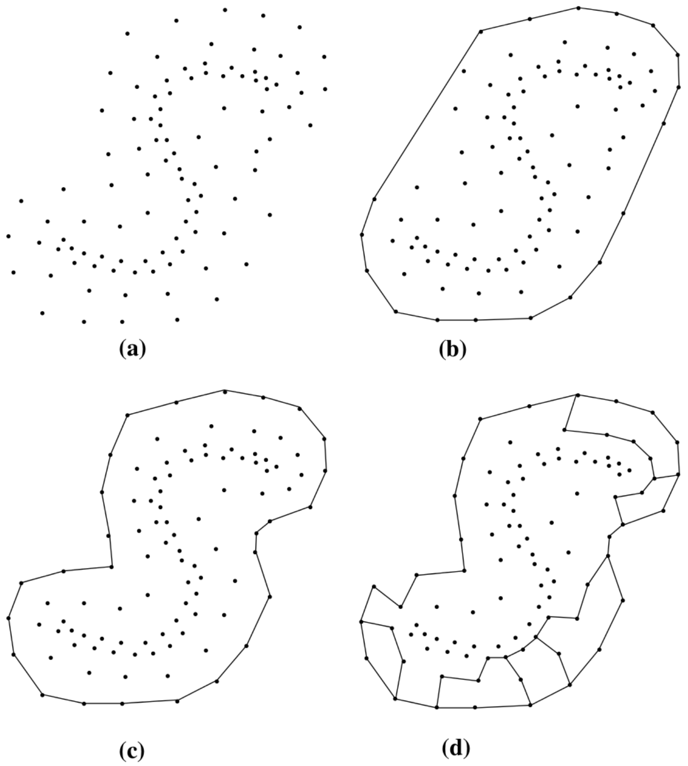
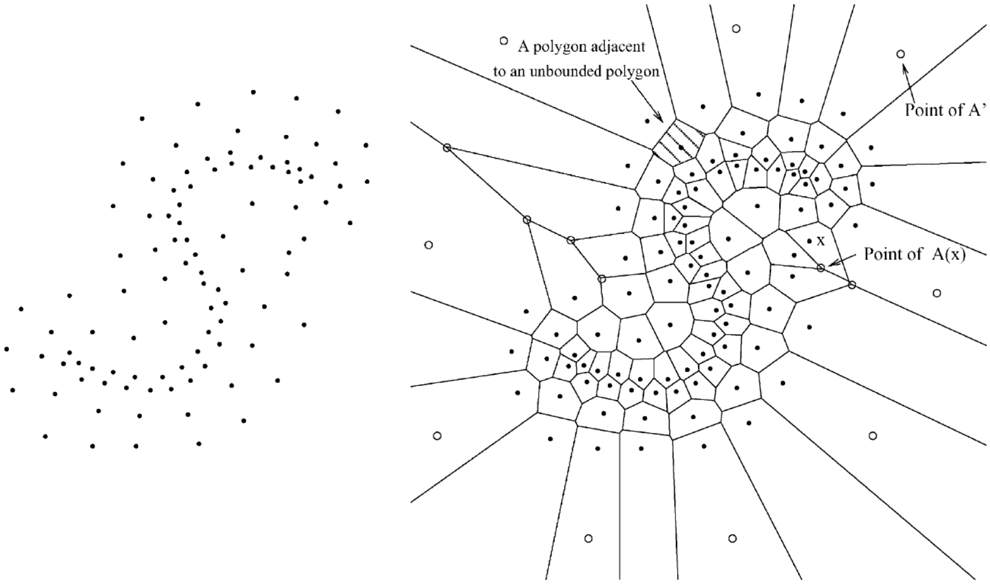
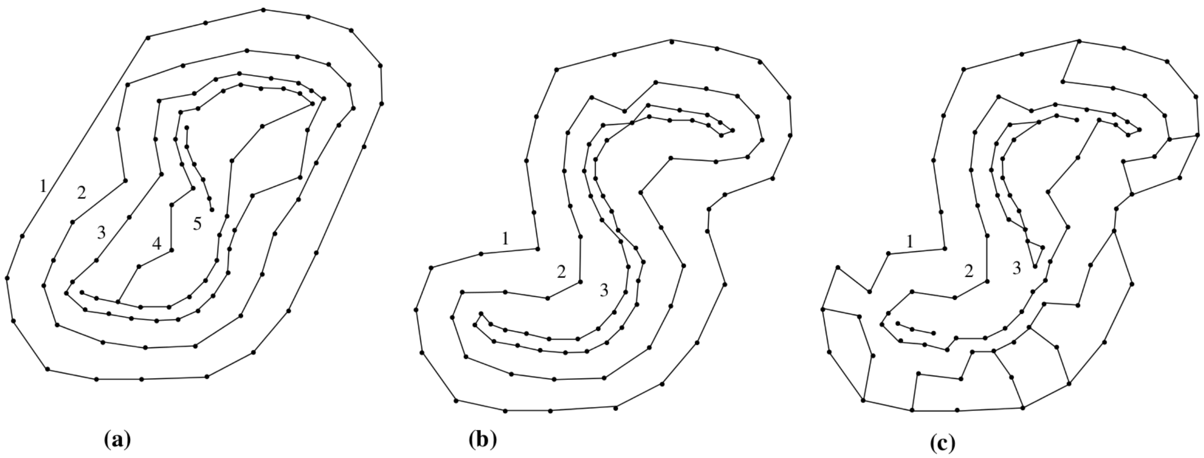


Fig. 4. Examples of $\mathscr{A}$-shape computed according to different choices of the parameter t. (a) The set of points S, (b) $\mathscr{A}$-shape of S ($t = 20$), (c) $\mathscr{A}$-shape of S ($t = 3$), (d) $\mathscr{A}$-shape of S ($t = 1.3$).



(a) (b)

Fig. 5. (a) The points set S and (b) construction of the points of the set $\mathscr{A}$.



(a) **(b)** **(c)**

Fig. 6. Continuation of Fig. 4. $\mathscr{A}$-shape onion-peeling sequences associated to the thresholds (a) $t = 20$, (b) $t = 3$ and (c) $t = 1.3$.

3.3. $\mathscr{A}$-shape onion-peeling algorithm

The $\mathscr{A}$-shape onion-peeling algorithm computes iteratively a sequence of strips. The first one is computed following the method presented in Section 3.2. The vertices of the first strip are added to the initial set $\mathscr{A}$ to obtain the set $\mathscr{A}_1$, and they are removed from S to obtain S_1. The $\mathscr{A}_1$-shape of S_1 represents the second strip, the following strips are computed according to the same process. Formally, this iterative process is as follows:

We start with $S_0 = S$ and $\mathscr{A}_0 = \mathscr{A}$, and we compute $S_{k+1} = S_k - VF(S_k, \mathscr{A}_k)$, and

$\mathscr{A}_{k+1} = VF(S_k, \mathscr{A}_k) \cup \mathscr{A}_k$, for all $k = 0, \ldots,$ until the empty set is reached. The graphs $F(S_k, \mathscr{A}_k)$ are the strips of the set and the process of iteratively removing the strips is the $\mathscr{A}$-shape onion-peeling. Examples of different $\mathscr{A}$-shape onion-peeling sequences, obtained with different choices of the threshold t, are shown in Fig. 4. In our experiments, it seems that the adequate value of t is in the range $(3, 4)$. Specifically we here used the value $t = 3$. Another illustrative example of the $\mathscr{A}$-shape onion-peeling sequence is shown in Fig. 7(d). For comparison purposes, an example of the convex hull onion-peeling is depicted in Fig. 7(b). In the one hand, one can legitimately conclude that the obtained convex hull sequence is inadequate in this case as it fails to detect the "cross"-like shape embedded in the center of the set of points. In the other hand, this shape is properly recovered using the $\mathscr{A}$-shape onion-peeling.

The following steps summarize our proposed $\mathscr{A}$-shape-based onion-peeling algorithm:

Algorithm 1. $\mathscr{A}$-shape onion-peeling
Inputs: The set of $S_0 = S$, the threshold t.
Output: The sequence $F(S_0, \mathscr{A}_0)$, $F(S_1, \mathscr{A}_1), \ldots$

1. Compute the Voronoi diagram of S.
2. Compute the points of the set $\mathscr{A}_0$.
3. Compute the Voronoi diagram of $S \cup \mathscr{A}_0$.
4. Extract $F(S_0, \mathscr{A}_0)$ from $V(S_0 \cup \mathscr{A}_0)$.
5. $S_1 = S_0 - VF(S_0, \mathscr{A}_0)$, $\mathscr{A}_1 = VF(S_0, \mathscr{A}_0) \cup \mathscr{A}_0$ and $k = 1$.
6. **While** $S_k \neq \emptyset$ **do**.
7. Compute $F(S_k, \mathscr{A}_k)$ as the set of the edges $[pq]$, $(p, q) \in S_k \times S_k$ such that p and q have a neighbor $x \in \mathscr{A}_k$ in the Voronoi diagram of $S \cup \mathscr{A}_0$.
8. $S_{k+1} = S_k - VF(S_k, \mathscr{A}_k)$ and $\mathscr{A}_{k+1} = VF(S_k, \mathscr{A}_k) \cup \mathscr{A}_k$.

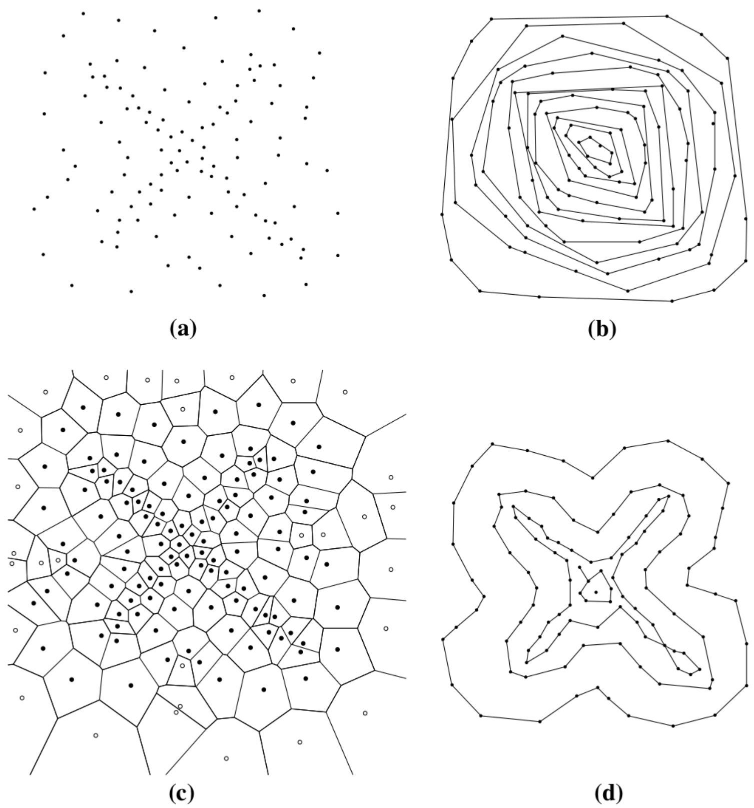


Fig. 7. (a) A point set, (b) onion-peeling using convex hull algorithm, (c) the Voronoi diagram of $S \cup \mathscr{A}$, (d) onion-peeling using $\mathscr{A}$-shape concept.

9. $k = k + 1$.
10. **end while**

This algorithm is based on the computation of the Voronoi diagram, which is efficiently computed in computational geometry (Okabe et al., 1992; Boissonnat and Yvinec, 1995). A large fraction of time is spent in computing the Voronoi diagram of $S \cup \mathscr{A}_0$. The resulting complexity of this step is $O(n \log n)$. The rest of the running-time is employed to extract the strips from the Voronoi diagram $S \cup \mathscr{A}_0$ (the complexity of this sub-step is $O(n)$). Therefore, the overall complexity of this non-convex onion-peeling algorithm is $O(n \log n)$.

We also point out that the presented algorithm can be easily extended to any d-dimensional space ($d \geqslant 3$). We employ the d-dimensional $\mathscr{A}$-shape and Voronoi diagram. The definition of the onion-peeling sequence does not change. Only the computation of the first strip (the first $\mathscr{A}$-shape of the sequence) should be slightly modified. Particularly, the notion of elongated polygons should be adapted to the shape of polytopes in higher dimensions. The complexity of the algorithm is the complexity of computing the Voronoi diagram in a d-dimensional space which is $O(n \log n + n^{d/2})$.

4. Application to robust statistical estimation

The connection between statistics and geometry is a close one as multivariate samples can be viewed as points in Euclidean space. Problems in statistics such as robust estimation are then equivalent to purely geometric ones. Outliers can be stripped away using the convex hull concept. However, this approach is clearly inadequate and shows serious limitations in the case of non-convex points sets, which occurs very often in real life datasets. Motivated by the $\mathscr{A}$-shape peeling concept which is well adapted to the non-convex case, we now turn to the robust statistical estimation problem.

There are many, if not most, statistical methods where insensitivity to outlying points is required. Among these methods, one can cite examples such as population statistics estimators (e.g. moments, quantiles, cumulants, etc.), linear and non-linear regression, non-parametric function estimation, hypothesis testing, model selection, etc. (Staudte and Sheather, 1990).

Let us consider the problem of linear regression in the presence of outliers. Our goal in this paper is to illustrate the usefulness of the $\mathscr{A}$-shape onion-peeling for robust statistical estimation. To reach this goal, we will consider examples where the geometrical shape formed by the points (i.e. bivariate samples) in the 2D Euclidean space are not necessarily convex. Such an example is depicted in Fig. 8(a). The true curve, which is a superposition of two sine functions Eq. (3), is shown with a solid line:

$$y(t) = \sin 2\pi t + 0.75 \sin \pi t, t \in [0, 0.2] \tag{3}$$

This true curve has been contaminated with an additive Gaussian white noise with a standard deviation $\sigma = 0.3$. Finally, bivariate uniform deviates have been introduced to simulate outliers. The proportion of outliers was 23% ($n = 130$ data points with 30 outliers).

Fig. 8(b) and (c) show respectively the onion-peeling using the convex hull and the $\mathscr{A}$-shape concepts. The latter clearly outperforms the convex hull-based peeling reflecting the shape of the true points set. Furthermore, the convex hull onion-peeling is unable to track outliers when they are in the cavities.

A Fourier basis (at frequencies 1 and 0.5) has been fitted to these data. The corresponding design matrix X has five columns, corresponding to two sines, two cosines and a constant. This yields five coefficients to be estimated. The least squares estimator was used in three configurations: without onion-peeling (LS), with convex hull (CHP) and $\mathscr{A}$-shape (ASP) onion-peeling. The usual value of $\alpha = 5\%$ was used. The results are shown in Fig. 7(d) and Table 1. As expected, the classical least-squares estimator has poor performances. It yields biased estimates with variances higher than the expected variances shown in parentheses in the first column of Table 1. The variances have been estimated from the diagonal of the classical covariance matrix:

$$\widehat{\mathrm{Cov}}(\beta) = \hat{\sigma}^2 (X^{\mathrm{T}} X)^{-1} \tag{4}$$

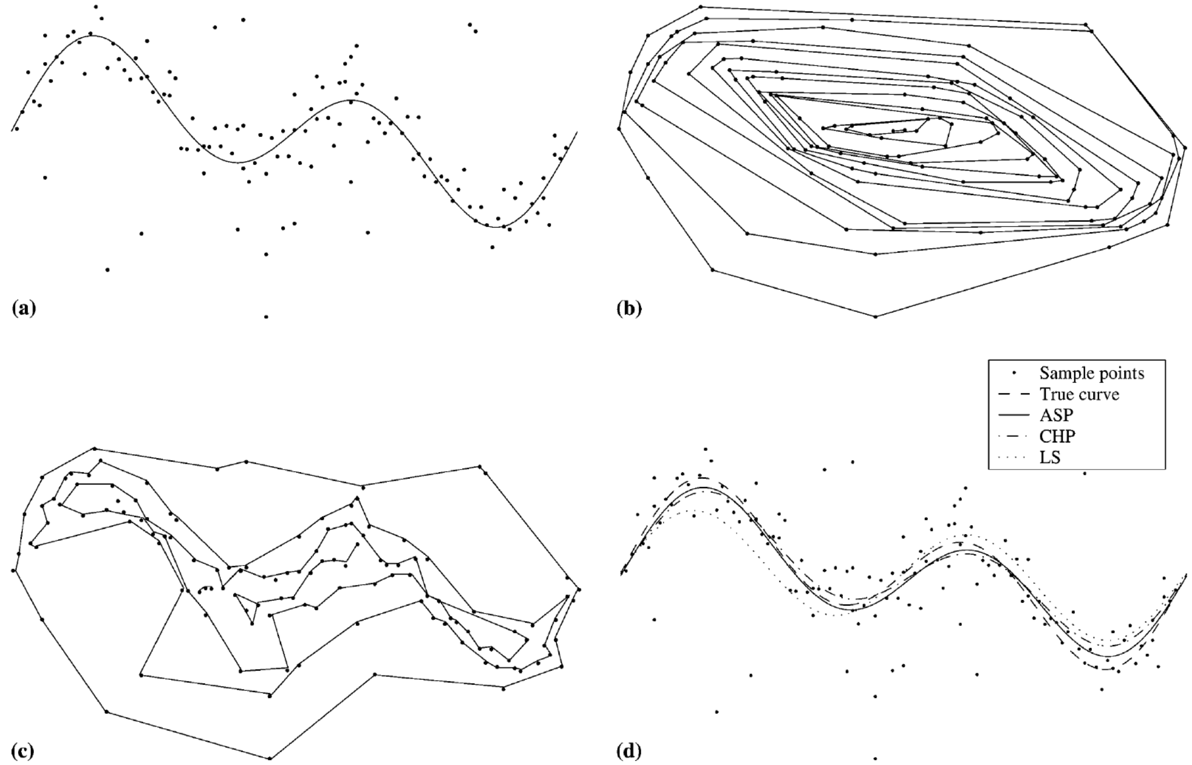


Fig. 8. (a) Points set obtained using superposition of two sine functions (shown with a solid line) contaminated with an additive Gaussian white noise with a standard deviation 0.3. Bivariate uniform deviates have been introduced to simulate outliers. (b) Onion-peeling using convex hull. (c) Onion-peeling using $\mathscr{A}$-shape. (d) A Fourier basis (at frequencies 1 and 0.5) fitted to the points with and without onion-peeling. For the onion-peeling α-trimmed estimators, the usual value of $\alpha = 5\%$ was used.

Table 1
True (first column) and estimated model parameters using the three approaches

True data	ASP estimates	CHP estimates	LS estimates
0 (0.0293)	−0.0293 (0.0495)	0.0154 (0.0747)	−0.0266 (0.1010)
1 (0.0417)	0.9012 (0.0673)	0.7854 (0.1039)	0.8501 (0.1426)
0 (0.0413)	−0.0734 (0.0705)	−0.1126 (0.1066)	0.0507 (0.1427)
0.75 (0.0417)	0.6370 (0.0671)	0.5260 (0.1072)	0.2846 (0.1468)
0 (0.0413)	0.0660 (0.0729)	0.0231 (0.1040)	0.0040 (0.1388)

Shown in parentheses are the standard deviations of the estimates calculated using Eq. (4).

where $\hat{\sigma}$ is the standard deviation of the residuals. One can also clearly see the higher performance of the ASP estimator yielding lower bias and variance levels. The variances given by the ASP are closer to the ideal Cramer–Rao bounds, shown in parentheses in the first column, than the CHP estimator.

5. Conclusion

We have presented an alternative algorithm to the convex onion-peeling method. The layers in the $\mathscr{A}$-shape onion-peeling method are not necessarily convex. The method takes into account non-convex shapes embedded in a points set. The algorithm is fast, simple to implement and is promising for robust statistical estimation. It can be used in a variety of situations in statistics to robustify the estimators against the presence of undesirable outliers. Furthermore, the presented technique can be easily applied to multivariate statistics even for higher dimensions than 2.

References

Abellanas, M., García, J., Hernández, G., Hurtado, F., Serra, O., 1992. Onion polygonizations. In: 4th Canadian Conference on Computational Geometry, Canada. pp. 127–131.

Boissonnat, J., Yvinec, M., 1995. Comput. Geomet. Springer, Cambridge.

Chazelle, B., 1985. On the convex layers of a planar set. IEEE Trans. Inform. Theory 31, 509–517.

Gastwirth, J., 1966. On robust procedures. J. Amer. Stat. Assn. 65, 929–948.

Huber, P., 1979. Robust statistical procedures. SIAM, Philadelphia.

Huber, P., 1981. Robust Statistics. Wiley, New York.

Melkemi, M., Djebali, M., 2000. Computing the shape of a planar points set. Pattern Recogn. 33, 1423–1436.

Melkemi, M., Djebali, M., 2001. Weighted *a*-shape: a descriptor of a finite points set. Pattern Recogn. 34, 1159–1170.

Okabe, A., Boots, B., Sugihara, K., 1992. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. Texts and Monographs in Computer Science. John Wiley & Sons, England.

Preparata, F., Shamos, M., 1985. Computational Geometry: An Introduction. Texts and Monographs in Computer Science. Springer-Verlag, New York.

Staudte, R., Sheather, S., 1990. Robust Estimation and Testing. Wiley, New York.