

A novel feasibility analysis approach based on dimensionality reduction and shape reconstruction

Ipsita Banerjee, and Marianthi G. Ierapetritou*
Department of Chemical and Biochemical Engineering
Rutgers University, Piscataway, NJ 08854

Abstract

Optimal and feasible operation of process plants demand accurate knowledge of the effect of parameter uncertainty on process design and operation. There has been considerable effort towards accurate representation of the feasible operation range and different metrics have been proposed in literature to quantify the operational flexibility. While these methods are largely successful in addressing convex problems, their applicability becomes restricted for general nonconvex problems. The feasibility analysis technique proposed in this paper considers the feasible region as an object, and applies surface reconstruction ideas to capture and define the shape of the object. The procedure starts by first sampling the feasible region to have a representation of the feasible space, an α shape is then constructed with the sampled points, thus generating a polygonal representation of the feasible parameter space. With this information at hand, any point can be checked for its feasibility by applying the point-in-polygon algorithm. The proposed method is general, and can be applied to any convex, non-convex even disjoint problems without any further modifications.

Keywords: Uncertainty, Feasibility analysis, α -shape reconstruction.

1. Introduction

The determination of the range of feasible operation for a given design has been the subject of extensive research for the last two decades. Grossmann and his coworkers have pioneered this field by introducing the flexibility and feasibility analysis approach (Swaney and Grossmann 1985). This was followed by a series of articles focussing on improving and extending the use of such a metric in the design process. A detailed review of feasibility analysis can be found in Ierapetritou (2001). The main shortcoming of all these approaches is the assumption of convexity required to guarantee that the determined feasible region does not include infeasible points. That makes these approaches inapplicable to a wide range of problems appearing in chemical engineering, which are non convex in nature.

* Author to whom correspondence should be addressed: marianth@soemail.rutgers.edu

Recently Goyal and Ierapetritou, (2001) introduced a new approach to quantify process feasibility based on the description of the feasible region by an approximation of the convex hull. Their approach results in accurate representation of process feasibility. However, it also relies on the utilization of process model and specific convexity assumptions. This approach has been further extended in the present work to capture the detailed shape of the feasible region by utilizing the ideas of shape reconstruction used in the field of computer graphics. The main problem definition for surface reconstruction is that given a set of points, reconstruct a manifold that closely approximates the surface of the original model. Surface reconstruction finds extensive applications in the areas of automatic mesh generation and geometric modelling, molecular structure and protein folding analysis. The problem of feasibility analysis is analogous to the ideas of surface reconstruction, since the main effort of feasibility analysis lies in identifying and accurately estimating the boundary of the feasible region. The main advantage of the surface reconstruction scheme is that it can successfully describe both nonconvex and disjoint regions defining the bounding surface by piecewise linear functions.

2. α -Shape Approach

There are various approaches described in literature for determining the shape of a pattern class from sampled points. Many of these approaches are concerned with efficient construction of convex hulls for a set of points in the plane (Jarvis, 2001; Akl and Toussaint, 1978; Fairfield, 1979). A mathematically rigorous definition of α shape was introduced by Edelsbrunner et al., (1983). They proposed a natural generalization of the convex hulls, which are referred to as α -hulls. The α -hull of a point set is based on the notion of generalised discs in the plane. The family of α -hulls includes the smallest enclosing circle, the set itself, and an essentially continuous set of enclosing regions in between these two extremes.

Edelsbrunner et al., (1983) also define a combinatorial variant of the α -hull called the α -shape of a planar set, which can be viewed as the boundary of the α -hull with curved edges replaced by straight edges. Conceptually, α -shapes are a generalization of the convex hull of a point set S , with α varying from 0 to ∞ . The α -shape of S is a polytope that is neither necessarily convex nor connected. For $\alpha \rightarrow \infty$ the α -shape is identical to the convex hull of S . However, as α decreases, the α -shape shrinks by gradually developing cavities. When α becomes small enough, the polytope disappears and reduces to the data set itself.

To provide an intuitive notion of the concept, Edelsbrunner describes the space R^3 to be filled with styrofoam and the point set S to be made of more solid material, such as rock. Now if a spherical eraser with radius α curves out the styrofoam at all positions where it does not enclose any of the sprinkled rocks (the point set S), the resulting object formed will be called an α -hull. The surface of the object can be straightened by substituting straight edges for the circular ones and triangles for spherical caps. The

obtained object is the α shape of S . It is a polytope in a fairly general sense: it can be concave and even disconnected; it can contain two-dimensional patches of triangles and one-dimensional strings of edges, and its components can be as small as single points. The parameter α controls the degree of details captured by the α shape.

2.1 Selection of α

The computed α shape of a given set of sample points explicitly depends on the chosen value of α , which controls the level of details of the constructed surface. Mandal et al. (1979) presents a systematic methodology for selecting the value of α in R^2 . They visualize the problem of obtaining the shape of S as a set estimation problem where an unknown set $A \in B$ is to be estimated on the basis of finite number of points $X_1, X_2, \dots, X_n \in A$. As n increases $S(n)$ will cover many parts of A and hence the value of α for $S(n)$ should depend on the sample size (n), thus α is a function of n . Additionally α should also be a function of the inter-point distance of the sampled n points of $S(n)$. To account for the dependence on inter-point distance, the authors have constructed the Minimum Spanning Tree (MST) of the sampled data points. If l_n represents the sum of edge weights of the MST, where the edge weight is taken to be the Euclidean distance between the points, then the appropriate value of α for the construction of α shape is given by:

$$h_n = \sqrt{\frac{l_n}{n}}$$

where n is the total number of sample points.

2.2 Feasibility analysis using α shape

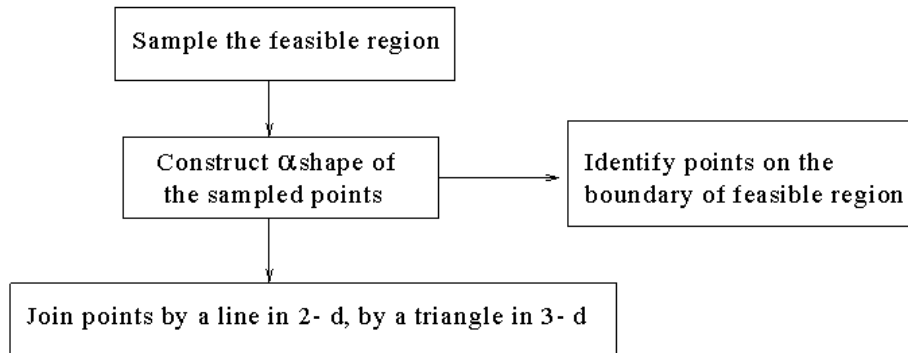


Figure 1. Algorithm of feasibility analysis using α shape

The overall aim of feasibility analysis is the determination of the range of parameters over which a particular process is feasible. A formal definition of this problem is to obtain mathematical description of the region in parameter space bounded by the process constraints. This region can be considered analogous to an object, the shape or surface of which can be estimated using the α shape technique. The input to any surface reconstruction algorithm needs to be a set of points representing the object, whose surface needs to be determined. Hence the first step is the generation of good sample data points to represent the feasible region under consideration. The α -shape can then

be constructed for the sampled data, using the α estimate obtained from the minimum-spanning tree of the data set. The α shape essentially identifies points from the input data set that lie on the surface of the object. These points are then connected by a line in two dimensions and triangle in three dimensions, giving rise to a polygon enclosing the feasible region. Figure (1) illustrates the steps of the proposed approach.

Having defined the surface or shape of the feasible region, the next step involves the determination of whether a particular point belongs to the feasible region. Since the feasible region has been approximated by a polygon, a simple way to check if a point is inside the polygon is by using one of the point-in-polygon tests (Haines, 1994). One method to determine whether a point is inside a region is the Jordan Curve Theorem, which states that a point is inside a polygon if, for any semi-infinite ray from this point, there is an odd number of intersections of the ray with the polygon's edges. Conversely, a point is outside a polygon if the ray intersects the polygon's edges an even number of times, or does not intersect at all. Following this idea, whenever a parameter needs to be checked for feasibility in a polygon estimated feasible region, a semi-infinite ray is drawn from the point in any direction, and number of intersections is noted, which decides on whether the point is feasible or not.

2.3 Sampling Technique

Obtaining a good representation of the feasible region is the first and very important step in the α -shape technique of feasibility analysis. In many problems, typically, the feasible region covers only a very restricted region of the entire parameter space. Hence sampling techniques covering the entire range of uncertain parameter proves to be inefficient, particularly when evaluation of the process constraints is an expensive operation. Most of the common sampling techniques sample the parameter space based on the distribution of the uncertain parameter, which are considered to be uniform for the cases considered here for simplicity of presentation. Under this condition it will lead to uniform sampling of the entire parameter space, irrespective of whether the sampled points are feasible or not. A new sampling technique is thus introduced here, which takes advantage of the fact that typically a small section of the entire parameter space is feasible. The sampling problem is formulated as an optimization problem and solved using Genetic Algorithm (GA) in order to take advantage of the inherent property of the algorithm to focus on the feasible regions.

The purpose of the sampling step is to have a good approximation of the feasible space by reducing the number of the expensive feasibility function evaluation. The formulation of the sampling problem as an optimization problem is given by:

$$\begin{aligned} & \max \sum \theta_{feas} \\ & \text{subject to: } (f_1)_{\theta_{feas}} \leq 0 \\ & \quad \vdots \\ & \quad (f_n)_{\theta_{feas}} \leq 0 \end{aligned}$$

where θ_{feas} is a feasible point in the uncertain space, and $f_1, f_2, \dots, f_n$ are the constraints of the feasibility problem evaluated at θ_{feas} . The idea is to maximize the sampled feasible space, hence whenever a chosen value of θ satisfies the constraint

functions, it improves the objective function, but when the value is not feasible it is ignored. By solving this problem using GA reduces the required number of function evaluations by minimizing the unnecessary evaluation of infeasible parameter space. In order to solve the optimization problem by GA, the optimization variables are encoded as a string of bits, and these strings are appended together to form a chromosome. The solution procedure starts by generating a large population of solutions, where each individual in the population has a particular chromosome value, which can be decoded to evaluate the parameter values and the objective function, also called the fitness function. The populations are evolved through several generations, following rules like reproduction, crossover, mutation, until the objective function cannot be improved any further.

The main feature of GA is that it concentrates on regions where the feasible solution can be obtained. This is of particular importance for the present problem since only a limited region of the variable space is feasible, and exploring the entire variable space to have an idea of the feasible region becomes computationally expensive. To further enhance the efficiency of GA, the feasible solutions are stored and the objective function is updated only for new solutions.

3. Case Studies

In the example considered here, the feasible region is defined by the following sets of constraints:

$$f_1 = \theta_2 - 2\theta_1 - 15 \leq 0$$

$$f_2 = \frac{\theta_1^2}{2} + 4\theta_1 - \theta_2 - 5 \leq 0$$

$$f_3 = \theta_2(6 + \theta_1) - 80 \leq 0$$

$$f_4 = 10 - \frac{(\theta_1 - 4)^2}{5} - 2\theta_2^2 \leq 0$$

where θ_1 and θ_2 are the uncertain parameters. In order to estimate this region using the technique of α shape described before, first one needs to generate points representing the shape of the region by sampling the feasible space. The technique described in section (2.3) is used to sample the feasible space. Both uncertain parameters are considered to vary within the range of (-20,20). In order to solve the problem using Genetic Algorithm, the parameters θ_1 and θ_2 are encoded as bits, with 7 bits for each parameter, giving rise to a 14-bit chromosome. A population size of 10 is chosen for this problem and the chromosomes (solutions) are evolved for 1300 generations, hence requiring a total of 13000 function calls. This gave rise to 950 feasible points as illustrated in Figure (2). However, a large number of these function calls are for the same value of the parameters, and are not unique. Hence by storing the parameter values and its corresponding solutions, it is possible to avoid repeating function evaluations, and hence increases the efficiency of the algorithm by evaluating the constraints only for unique parameter values. By following this procedure it was observed that out of 13000 required function calls only 4835 were unique. Hence this procedure could

generate 950 feasible points by 4835 function evaluations. The same problem was solved by drawing random samples in the range (-20,20) for both uncertain parameters, as illustrated in Figure (2). To generate 950 feasible points using random search required 9830 function calls. This clearly shows the advantage of the proposed approach of sampling the feasible space. However, this procedure is particularly advantageous when the feasible region is a small portion of the entire parameter range. Otherwise its performance becomes comparable to random sampling over the entire parameter range.

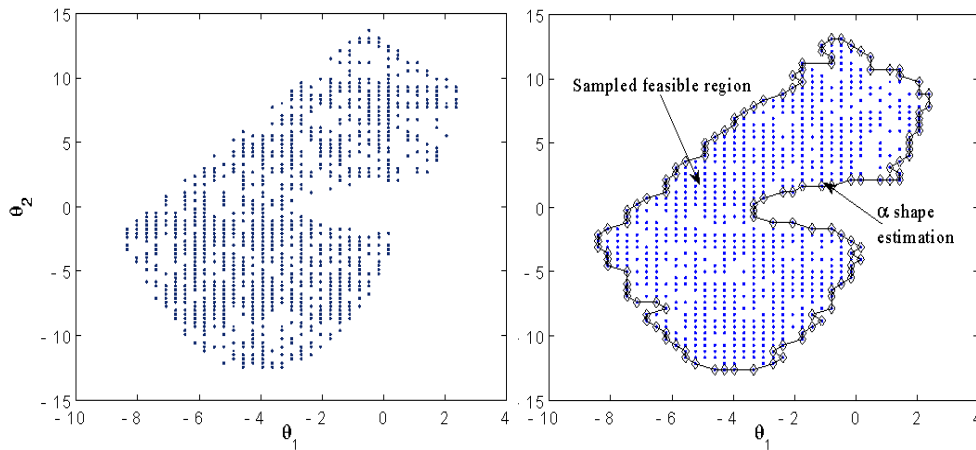


Figure 2. Predicting the feasible region using α shape

The next step is the generation of α shape from the sampled points, which identifies the points forming boundaries of the object, which are then joined by straight line. The value of α plays a crucial role in determining the degree of details captured by the α shape. The α value determined by the procedure outlined in section (2.1) is 25, which was found to capture the nonconvex nature of the object with adequate accuracy, as illustrated in Figure 2.

References

- Akl, S.K. and Toussaint, G.T., 1978, Proc. Fourth Int. Jr. Conf. Pattern Recognition, 483.
- Edelsbrunner, H. and Kirkpatrick, D.G. and Seidel, R., 1983, IEEE Trans. Inform. Theory IT-29, 551.
- Fairfield, J., 1979, Proc. IEEE Conf. Cybernet and Soc., 60.
- Goyal, V. and Ierapetritou, M.G., 2002, AIChE J, 48, 2902.
- Haines, E., 1994, Point in polygon strategies, ed. Paul Heckbert, Academic Press, 24.
- Ierapetritou, M.G. 2001, AIChE J, 47, 1407.
- Jarvis, R.A., 1977, Proc. IEEE Comp. Soc. Conf. Pattern Recognition and Image, 231.
- Mandal, D.B. and Murthy, C.A., 1997, Pattern Recognition, 30, 1759.
- Swaney, R.E. and Grossmann, I.E., 1985, AIChE J, 31, 621.

Acknowledgements

The authors gratefully acknowledge financial support from the Office of Naval Research under the contract N00014-03-1-0207 and National Science Foundation (CTS 0224745). The α -shape code used in this paper was developed by Dr. Ken Clarkson (<http://cm.bell-labs.com/netlib/voronoi/hull.html>).