Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 876862, 9 pages http://dx.doi.org/10.1155/2015/876862



# Research Article

# Models and Algorithms for Optimal Piecewise-Linear Function Approximation

# Eduardo Camponogara<sup>1</sup> and Luiz Fernando Nazari<sup>2</sup>

<sup>1</sup>Department of Automation and Systems Engineering, Federal University of Santa Catarina, Cx.P. 476, 88040-900 Florianópolis, SC, Brazil

Correspondence should be addressed to Eduardo Camponogara; eduardo.camponogara@ufsc.br

Received 4 March 2015; Accepted 11 June 2015

Academic Editor: Jean-Christophe Ponsart

Copyright © 2015 E. Camponogara and L. F. Nazari. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Piecewise-linear functions can approximate nonlinear and unknown functions for which only sample points are available. This paper presents a range of piecewise-linear models and algorithms to aid engineers to find an approximation that fits best their applications. The models include piecewise-linear functions with a fixed and maximum number of linear segments, lower and upper envelopes, strategies to ensure continuity, and a generalization of these models for stochastic functions whose data points are random variables. Derived from recursive formulations, the algorithms are applied to the approximation of the production function of gas-lifted oil wells.

### 1. Introduction

Piecewise-linear functions are widely used to approximate functions for which only sample points are known and to model nonlinear functions. In petroleum engineering, the fluid flow from an oil well and the pressure drop in a pipeline can be approximated with a piecewise-linear function [1]. In optimization, nonlinear problems can be recast as a mixed-integer linear programming (MILP) problem, which is then solved with MILP algorithms [2–4].

Although many works in the literature are dedicated to piecewise linearization, few of them address the problem of generating piecewise-linear models which is often delegated to the scientist or engineer. Some techniques are based on point clustering, such as *K*-means, which has found applications in nonlinear dynamic systems and control [5]. This work attempts to mitigate undesirable effects of clustering, particularly the influence of outliers and the synthesis of subpotimal approximations. Akin to the previous technique, a fuzzy clustering strategy is developed in [6] to represent a nonlinear system with a piecewise-linear model. However, fuzzy clustering is dependent on the initial parameters and

can reach a local minimum. In order to reduce the chance of yielding a local minimum, genetic algorithms have been proposed for clustering [7]. Other works rely on local searches that introduce linear segments iteratively until satisfying a stopping criterion [8, 9].

However, all of the cited approaches do not offer a certificate of optimality and do not allow the user to specify desirable properties for the resulting approximation. To this end, this paper develops formal models and algorithms for generating piecewise-linear functions that fit best the needs of the end user. The models consider a fixed and maximum number of linear segments, lower and upper envelope approximations, constraints to ensure continuity, and stochastic generalizations of these models in which the data points are random variables. The underlying approximation problems for these models are cast in a recursive form that can often be efficiently solved with dynamic programming strategies. To the best of the knowledge of the authors, this is the first work on the synthesis of piecewise-linear functions based on formal models and algorithms which considers an extensive range of conditions and approximation metrics. Software tools are obtained from these strategies to assist

<sup>&</sup>lt;sup>2</sup>Instituto Federal Catarinense, Estrada do Redentor, No. 5665, 89163-356 Rio do Sul, SC, Brazil

scientists and engineers in the synthesis of piecewise-linear approximations of nonlinear models and processes.

Sections 2 through 5 develop a range of models and algorithms for piecewise-linear approximation. Section 6 reports computational results on the approximation of well production in gas-lifted oil fields. Section 7 offers a summary and suggests directions for future research.

# 2. Basic Models and Algorithms

Let  $P = \{p_1, \dots, p_n\}$  be a set of points of a function  $\Phi: \mathbb{R} \to \mathbb{R}$  such that  $p_i = (x_i, y_i), y_i = \Phi(x_i)$ , and  $x_1 < x_2 < \dots < x_n$ .  $\Phi$  may represent any function, hereafter referred to as *generating function*, which can be known explicitly, implemented by a simulator, or obtained from a real-world process such as the output flow from an oil well as a function of the compressed gas injection. The problem is to find a set  $R = \{r_1, \dots, r_T\}$  of linear segments that best approximates  $\Phi$ . Assuming that the segment breakpoints are points in P, a piecewise-linear model is a sequence  $I = \langle i(1), i(2), \dots, i(T+1) \rangle$  of breakpoints such that

(i) 
$$i(1) = 1 < i(2) < \cdots < i(T+1) = n$$
;

(ii) each subset 
$$P(t) = \{i(t), i(t) + 1, ..., i(t+1)\} \subseteq P$$
 is approximated by the linear segment  $r_t = \{(x, y) : y = a_t x + b_t, x_{i(t)} \le x \le x_{i(t+1)}\}$  for  $t = 1, ..., T$ .

An approximation for P(t) is obtained by minimizing the squared error between the linear segment  $r_t$  and the points in P(t), for which an analytical solution yields the intercept  $b_t$  and the slope  $a_t$ . Table 1 presents a set of sample points (the sample points were obtained in an arbitrary manner by perturbing the oil production function  $q_{\rm oil}(q_{\rm inj})$  of an oil well induced by the lift-gas injection rate  $q_{\rm inj}$ ; the purpose of these sample points is only to illustrate the behavior of the models and algorithms to be developed in this work) that will serve to illustrate models and algorithms.

2.1. Fixed Number of Linear Segments. To find the piecewise-linear approximation with a fixed number T of linear segments that minimizes the squared error, the following problem is solved:

$$\min_{I} E = \sum_{t=1}^{T} E_{t} = \sum_{t=1}^{T} \sum_{k=i(t)}^{i(t+1)} (y_{k} - a_{t} x_{k} - b_{t})^{2}.$$
 (1)

This problem can be recast in a recursive form and solved with a dynamic programming (DP) algorithm. For an introduction to dynamic programming in discrete domains, the interested reader can refer to the text books by Kleinberg and Tardos [10] and Cormen et al. [11]. The book by Bertsekas [12] is recommended for an advanced treatment of DP with particular emphasis on continuous domains. Let F(j,t) be the minimum cost to approximate the points  $\{p_1, p_2, \ldots, p_j\}$  with exactly t segments. Then,

$$F(j,t) = \min_{t \le i \le j} F(i,t-1) + E(i,j), \qquad (2)$$

TABLE 1: Set of sample points.

i	$x_i$	$y_i$
1	0.0114	0.1857
2	0.0861	0.4502
3	0.1721	0.5584
4	0.2928	0.5857
5	0.3130	0.7020
6	0.3212	0.8338
7	0.4899	0.8713
8	0.8807	0.9104

where E(i, j) is the minimum squared error of approximating  $P_{i,j} = \{p_i, p_{i+1}, \dots, p_j\}$  with just one segment. Notice that  $E(i, i) = \infty$ ,  $F(j, t) = \infty$  for all  $j \le t$ , and F(j, 1) = E(1, j) for all j > 1. From recursive equation (2), clearly  $F(n, T) = \min E$ .

**Theorem 1.** Recursive equation (2) yields the piecewise-linear approximation of P of minimum squared error having exactly t linear segments.

*Proof.* The proof is by induction in j and t. For the basis with j=1,  $F(j,t)=\infty$  is the optimal cost since the approximation of one point is undefined for all t, and, with t=1, F(j,t)=E(1,j) for all j>1. For the induction, take any j>1 and t>1. Any approximation of  $\{p_1,\ldots,p_j\}$  with t segments must approximate  $\{p_1,\ldots,p_i\}$  with t-1 segments and  $\{p_i,\ldots,p_j\}$  with one segment. The point  $p_i$  is chosen to minimize the aggregated cost F(i,t-1) and the cost E(i,j) to approximate  $\{p_i,\ldots,p_j\}$  with one segment. Since F(i,t-1) is by induction the minimum cost to approximate  $\{p_1,\ldots,p_i\}$  with t-1 segments, F(j,t) is the minimum cost. □

The optimal sequence  $I = \langle i(1), i(2), \dots, i(T+1) \rangle$  is obtained recursively: i(T+1) = n; i(T) is the index i that minimizes F(i(T+1), T) in (2); i(T-1) is the index i that minimizes F(i(T), T-1); and so forth until F(i(2), 1) is reached, at which point i(1) = 1. Algorithm 1 outlines the DP algorithm PL-Fixed for computing F(j, t). The algorithm records in table I(i, t) the indices i that minimize (2), which are the optimal breakpoints.

PL-Fixed is correct because it solves recursive equation (2). It runs in  $\Theta(n^3)$  time since T < n. The notation  $\Theta(h(n)) = \{g(n) : \exists n_0, c_1, c_2 > 0$ , such that  $c_1h(n) \le g(n) \le c_2h(n)$  for all  $n \ge n_0\}$  represents the set of all nondecreasing functions g(n) that grow at the same asymptotic rate of h(n). Therefore, the time taken by algorithm PL-Fixed to solve an instance of size n is approximately  $n^3$ . A precise and extensive discussion on the running-time complexity of algorithms is found in [11].

Running the algorithm on the sample instance with T=3 yields the values F(i,j) and I(i,j) depicted in Table 2. The optimal sequence is recursively obtained from the table as follows: i(4)=8; i(3)=I(8,3)=5; i(2)=I(5,2)=2; and i(1)=I(2,1)=1, as indicated in boldface. The optimal sequence is  $\langle 1,2,5,8 \rangle$  according to which the first segment approximates the points  $\{p_1,p_2\}$ , the second segment approximates  $\{p_2,\ldots,p_5\}$ , and the third segment approximates

```
(1) for i = 1 to n do
         E(i,i) := \infty;
(2)
         for j = i + 1 to n do
              a_{i,j} := \frac{(j-i+1)\sum_{k=i}^{k=j} x_k y_k - (\sum_{k=i}^{k=j} x_k)(\sum_{k=i}^{k=j} y_k)}{(j-i+1)\sum_{k=i}^{k=j} x_k^2 - (\sum_{k=i}^{k=j} x_k)^2}
(4)
             \begin{aligned} b_{i,j} &:= \frac{\sum_{k=i}^{k=j} y_k - a_{i,j} \sum_{k=i}^{k=j} x_k}{(j-i+1)} \\ A_{i,j} &:= a_{i,j}, B_{i,j} := b_{i,j}; \end{aligned}
(5)
(6)
             E(i, j) := \sum_{j=1}^{J} (y_k - a_{i,j} x_k - b_{i,j})^2;
(8) for j = 1 to n do
(9)
          for t = j to T do
(10)
              F(j,t) := \infty; I(j,t) := 0;
          F(j, 1) := E(1, j); I(j, 1) := 1;
(11)
          for t = 2 to min\{j - 1, T\} do
(12)
              F(j,t) := \infty; I(j,t) := 0;
(13)
(14)
              for i = t to i do
(15)
                  if F(j,t) > F(i,t-1) + E(i,j) then
(16)
                       F(j,t) := F(i,t-1) + E(i,j);
(17)
                      I(j,t) := i;
```

ALGORITHM 1: PL-Fixed(P, T).

TABLE 2: Tables with values F(j, t) and I(j, t) for the sample instance.

$j \setminus t$	F(j,t): $I(j,t)$		
	1	2	3
1	∞:1	$\infty:0$	∞:0
2	0:1	$\infty:0$	$\infty:0$
3	5.5459e - 03:1	0:2	$\infty:0$
4	2.2611 <i>e</i> - 02:1	1.7718e - 03:2	0:3
5	2.3016e - 02:1	5.5550e - 03:2	1.7718e - 03:4
6	3.5038e - 02:1	2.3016e - 02:5	4.0758e - 03:4
7	4.4083e - 02:1	2.4617e - 02:2	1.3596e - 02:5
8	1.2685e - 01:1	3.4109e - 02:5	1.6648e - 02:5

 $\{p_5, ..., p_8\}$ . The approximation cost is  $F(8,3) = 1.6648 \times 10^{-2}$ . Figure 1 depicts this approximation.

2.2. Maximum Number of Linear Segments. A variation of the previous model is obtained by limiting rather than fixing the number of linear segments. Each segment incurs a cost c to trade off model simplicity and quality of approximation. This problem is also cast in a recursive form, with F(j,t) being the minimum cost to approximate  $\{p_1, p_2, \ldots, p_j\}$  with at most  $t \ge 1$  segments. Formally,

$$F(j,t) = \min_{1 \le i \le j} F(i,t-1) + E(i,j) + c, \tag{3}$$

where E(i, j) is defined as above,  $F(1, t) = \infty$ , and F(2, t) = E(1, 2) + c for all t.

**Theorem 2.** Recursive equation (3) yields the piecewise-linear model of P with at most t segments, whose objective is

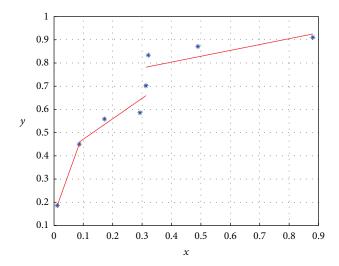


FIGURE 1: Piecewise-linear approximation for the sample instance with 3 segments.

the combination of the approximation error and a cost c per segment.

Recursive equation (3) leads to a DP algorithm PL-Max, which is omitted here in the interest of brevity. PL-Max(P,T,c) runs in  $\Theta(n^3)$  provided that  $T \leq n$ . This algorithm records the index i corresponding to the optimal breakpoint of (3) in a table I(j,t). This table is key to determine the optimal number  $T^*$  of segments and construct the optimal sequence  $\langle i(1), i(2), \ldots, i(T^*+1) \rangle$ .

The optimal number of segments and sequence of breakpoints are determined from the tables produced by PL-Max(P,T,c) as follows: i(T+1)=n; i(T)=I(n,T) is the index that minimizes F(n,T) in (3); i(T-1)=I(i(T),T-1) is the index that minimizes F(i(T),T-1); and so forth until reaching the first breakpoint where i(T+1-k)=1 for some k. Then,  $T^*=k$  and the optimal breakpoint sequence is  $\langle i(T-k+1), i(T-k+2), \ldots, i(T+1) \rangle$ .

With a limit T=5 and cost c=0.01, PL-Max produces an optimal approximation that uses 4 out of the 5 segments available, inducing a total cost of F(8,5)=0.044208. The optimal sequence is  $\langle i(1),\ldots,i(5)\rangle=\langle 1,2,4,6,8\rangle$  with four segments approximating the point sets  $\{p_1,p_2\},\{p_2,p_3,p_4\},\{p_4,p_5,p_6\}$ , and  $\{p_6,p_7,p_8\}$ .

#### 3. Continuous Approximations

The preceding models may produce discontinuous piecewiselinear functions. To enforce continuity either because the actual function is known to be continuous or because continuity is desirable for numerical optimization, some adaptations are proposed for the basic models.

3.1. Ensuring Continuity at the Breakpoints. Continuity is enforced by approximating the set  $P_{i,j} = \{p_i, p_{i+1}, \dots, p_j\}$  with segments going through  $p_i$  and  $p_j$ . The recursive formulations ensuring continuity with a fixed and maximum

```
(1) Same as lines (1) through (7) of PL-Fixed
(2) for j = 1 to n do
(3)
      for t = j to T do
          F(j,t) := \infty; I(j,t) := 0; X(j,t) := 0;
(4)
(5)
      F(j, 1) := E(1, j); I(j, 1) := 1; X(j, 1) := 0;
      for t = 2 to min{j - 1, T} do
(6)
(7)
          F(j,t) := \infty; I(j,t) := 0; X(j,1) := 0;
(8)
           for i = t to j - 1 do
              k := I(i, t - 1);
(9)
              if k \neq 0 and A(k, i) \neq A(i, j) then
(10)
                        (B(k,i)-B(i,j))
(11)
                       (A(i, j) - A(k, i))
                  if x_{\max\{1,i-1\}} \le x \le x_{\min\{n,i+1\}} and F(j,t) > F(i,t-1) + E(i,j) then
(12)
                     F(j,t) := F(i,t-1) + E(i,j); I(j,t) := i; X(j,t) := x;
(13)
```

ALGORITHM 2: PL-Fixed-Cont-Reject(P, T).

number of segments are identical to the preceding ones, except that  $P_{i,j}$  is approximated with the segment passing through  $p_i$  and  $p_j$ . The theorem below is a consequence of this observation.

**Theorem 3.** Suppose that the coefficients of the linear segment approximating  $P_{i,j}$  are  $a_{i,j} = (y_j - y_i)/(x_j - x_i)$  and  $b_{i,j} = y_j - a_{i,j}x_j$ . Then, recursive formulations (2) and (3) produce optimal piecewise-linear approximation functions for P that are continuous at the breakpoints using exactly and at most t linear segments, respectively.

Applying PL-Max to the sample instance with a maximum of T=5 segments and cost c=0.01, but forcing the segments approximating  $P_{i,j}$  to go through the end points, yields the piecewise-linear function with  $T^*=4$  segments, which is illustrated in Figure 2. This approximation is given by the sequence  $\langle i(1),\ldots,i(5)\rangle = \langle 1,2,4,6,8\rangle$  which, coincidentally, is identical to the one obtained without ensuring continuity. However, its approximation cost F(8,5)=0.046556 exceeds the optimal cost 0.044208 of the discontinuous approximation, which is expected because a discontinuous model encompasses its continuous counterpart. This property is formally stated by the proposition below.

**Proposition 4.** The approximation cost induced by the schemes that enforce continuity at the breakpoints, with a fixed and maximum number of segments, is not lower than the cost induced by the schemes that do not enforce continuity.

3.2. Rejecting Discontinuous Approximations. The basic models with fixed and maximum number of segments can be modified to ensure continuity by rejecting discontinuous approximations. The recursive equation for generating a piecewise-linear function with a fixed number t > 1 of segments is

$$F(j,t) = \min_{i \in S_{j,t}} F(i,t-1) + E(i,j), \qquad (4)$$

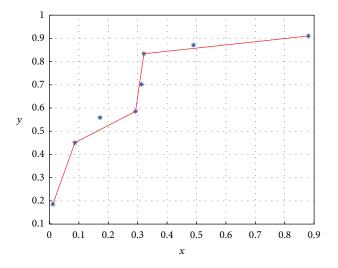


FIGURE 2: Continuous piecewise-linear approximation of the sample instance with a maximum of 5 segments.

where  $S_{j,t} = \{i : t \le i \le j, \text{ the lines approximating } P(t-1) \text{ and } P(t) \text{ intercept at point } x \in [x_{\max\{1,i-1\}}, x_{\min\{n,i+1\}}]\}, \text{ and } F(j,t) = \infty \text{ if } S_{j,t} = \emptyset. \text{ Further, } E(i,i) = \infty, F(j,t) = \infty \text{ for all } j \le t, \text{ and } F(j,1) = E(1,j).$ 

Algorithm 2 gives the dynamic programming solution of recursive equation (4), for a fixed number T of segments. The algorithm records in table X(j,t) the points where the linear segments intercept, which are used to determine the breakpoints of the piecewise-linear function. The algorithm runs in  $\mathcal{O}(n^3)$  time provided that  $T = \mathcal{O}(n)$ .

Algorithm 2 applied to the sample instance with T=3 yields  $\langle i(1),\ldots,i(4)\rangle=\langle 1,5,6,8\rangle$  as the optimal sequence. However, the breakpoints are not those of the sequence but defined by the intercept of the segments: the first segment is defined for  $[x_1,0.3120]=[0.0114,0.3120]$ , the second for [0.3120,0.3216], and the third for  $[0.3216,x_8]=[0.3216,0.9104]$ . Notice that  $x_4\leq 0.3112\leq x_6$  and  $x_5\leq 0.3216\leq x_7$  as expected. The approximation cost is

 $F(8,3) = 2.3148 \times 10^{-2}$ , which exceeds the cost  $1.6648 \times 10^{-2}$  of the discontinuous approximation.

3.3. Generating Continuous Approximations. Continuity may be ensured by solving a least squares problem at the moment that the recursive equation is evaluated. For a fixed number of segments t > 1, recursive equation (2) becomes

$$F(j,t) = \min_{t \le i \le j} F(i,t-1) + E(i,j)^*, \qquad (5)$$

while for a maximum number of segments (3) becomes

$$F(j,t) = \min_{1 \le i \le j} F(i,t-1) + E(i,j)^* + c, \tag{6}$$

where  $E(i, j)^*$  is obtained by solving the following problem:

$$E(i, j)^* = \min_{a_{i,j}, b_{i,j}, x} \sum_{k=i}^{j} (a_{i,j} x_k + b_{i,j} - y_k)^2$$
 (7a)

s.t.: 
$$A_{i,t-1}x + B_{i,t-1} = a_{i,j}x + b_{i,j}$$
 (7b)

$$x_{\max\{1,i-1\}} \le x \le x_{\min\{i+1,n\}},$$
 (7c)

where  $A_{i,t-1}$  and  $B_{i,t-1}$  are the slope and intercept of the rightmost segment of the optimal approximation of  $P_{1,i}$  with exactly or at most t-1 segments, depending on which recursion is being solved.  $E(i,j)^* = \infty$  if the problem is infeasible, for example, when j=i.

The computation of  $E(i, j)^*$  entails minimizing the squared error of the approximation of  $\{p_i, p_{i+1}, \ldots, p_j\}$  with a linear segment, while ensuring that this segment intercepts the preceding one, the rightmost segment of the piecewise-linear approximation of  $\{p_1, \ldots, p_i\}$ , in the interval  $[x_{\max\{1,i-1\}}, x_{\min\{i+1,n\}}]$ . The chief difficulty lies in the nonlinearity of (7b). However, the optimal solution is obtained by solving two quadratic programs. First, notice that variable x is eliminated by replacing the constraints (7b)-(7c) with the following nonlinear inequality:

$$x_{\max\{1,i-1\}} \le \frac{b_{i,j} - B_{i,t-1}}{A_{i,t-1} - a_{i,j}} \le x_{\min\{i+1,n\}}$$
(8)

which, for  $a_{i,j} < A_{i,t-1}$ , is transformed into the linear inequalities

$$\begin{split} x_{\max\{1,i-1\}} a_{i,j} + b_{i,j} &\geq B_{i,t-1} + x_{\max\{1,i-1\}} A_{i,t-1}, \\ x_{\min\{i+1,n\}} a_{i,j} + b_{i,j} &\leq B_{i,t-1} + x_{\min\{i+1,n\}} A_{i,t-1}, \\ a_{i,j} &< A_{i,t-1} \end{split} \tag{9a}$$

and the following ones when  $a_{i,j} > A_{i,t-1}$ :

$$\begin{aligned} x_{\max\{1,i-1\}} a_{i,j} + b_{i,j} &\leq B_{i,t-1} + x_{\max\{1,i-1\}} A_{i,t-1}, \\ x_{\min\{i+1,n\}} a_{i,j} + b_{i,j} &\geq B_{i,t-1} + x_{\min\{i+1,n\}} A_{i,t-1}, \\ a_{i,j} &> A_{i,t-1}. \end{aligned} \tag{9b}$$

The end result is that  $E(i, j)^*$  can be computed by solving two convex, quadratic programs (QPs) and choosing the least costly solution. The QPs minimize the objective (7a), with the first being subject to constraint (9a), whereas the second is constrained by (9b).

Dynamic programming algorithms are obtained to solve recursive equations (5) and (6) using a standard QP solver. The algorithms run in  $\mathcal{O}(n^3g)$  time, with g being an upper bound on the cost for solving QPs with 2 variables and 3 constraints. The application of the DP algorithm for the sample instance with a fixed number of T=3 segments yields the sequence  $\langle i(1),\ldots,i(4)\rangle=\langle 1,5,6,8\rangle$ . The breakpoints of the segments are defined by their intercepts, which are  $[x_1,0.3120]=[0.0114,0.3120]$  for the first segment, [0.3120,0.3215] for the second, and finally  $[0.3215,x_8]=[0.3215,0.9104]$  for the third. The approximation cost is  $F(8,3)=2.3147\times 10^{-2}$ .

# 4. Lower and Upper Envelopes

The preceding models can be modified to obtain piecewise-linear lower and upper envelopes of the generating function  $\Phi$ . For the upper envelope, the slope  $a_{i,j}$  and intercept  $b_{i,j}$  should minimize the squared error of the line  $y=a_{i,j}x+b_{i,j}$  approximating  $P_{i,j}=\{p_i,p_{i+1},\ldots,p_j\}$ , while keeping these points below the line. Formally, this is achieved by solving the QP:

$$L_{i,j}: \quad \min_{a_{i,j},b_{i,j}} \sum_{k=i}^{j} \left( a_{i,j} x_k + b_{i,j} - y_k \right)^2$$
 (10a)

s.t.: 
$$y_k \le a_{i,j} x_k + b_{i,j}, \ p_k \in P_{i,j}$$
. (10b)

The basic algorithms can compute the optimal upper envelope, with a fixed and maximum number of segments, by solving  $L_{i,j}$  to obtain the coefficients of the segment approximating  $P_{i,j}$ . The theorem below formalizes these results for the upper envelope approximation. Algorithms are developed similarly for the lower envelope by changing the direction of inequality (10b).

**Theorem 5.** Recursive equations (2) and (3) yield the optimal piecewise-linear upper envelope approximation of P with exactly and at most t linear segments, respectively, having the squared approximation error as the objective, provided that the coefficients for the linear segments are computed by solving problem  $L_{i,j}$  given by (10a) and (10b).

The running time of the resulting algorithm is  $\mathcal{O}(n^3g(n))$  where g(n) is an upper bound on the cost to solve the quadratic program  $L_{i,j}$  with 2 variables and n constraints.

Algorithm PL-Fixed applied to the sample instance, with the coefficients obtained by solving  $L_{i,j}$  and a fixed number of segments T=3, yields the optimal sequence  $\langle i(1),\ldots,i(T+1)\rangle=\langle 1,2,5,8\rangle$  with cost  $F(8,3)=2.8762\times 10^{-2}$ . The slopes and intercepts of the segments are  $(a_{ij},b_{ij})=(3.5413,0.1453)$  for (1,2); (1.0188,0.3830) for (2,5); and (0.2225,0.7623) for (5,8).

The models and algorithms presented in Section 3 can be modified to produce continuous lower and upper envelopes. For PL-Fixed-Cont-Reject, it suffices to use the slope  $a_{i,j}$  and intercept  $b_{i,j}$  obtained by solving  $L_{i,j}$ . This algorithm applied to the sample instance with T=3 yields the optimal sequence  $\langle i(1), \ldots, i(T+1) \rangle = \langle 1, 5, 6, 8 \rangle$  with cost  $F(8,3) = 4.609 \times 10^{-2}$ . The domains of the linear segments are  $[x_1, 0.3153] = [0.0114, 0.3153]$  for the first segment, [0.3153, 0.3222] for the second, and  $[0.3222, x_8] = [0.3222, 0.9104]$  for the third. This approximation is illustrated in Figure 3.

For the algorithm that computes linear segments by solving QP problem (7a), (7b), and (7c), we only need to introduce the linear inequalities that ensure bounding, such as the constraints (10b) for an upper envelope approximation.

## 5. Dealing with Random Data

Uncertainty is addressed by modeling each point  $p_i \in P$  as a random variable  $\mathbf{p}_i = (\mathbf{x}_i, \mathbf{y}_i)$  characterized by a probability density function  $f_i(x, y)$ . Two models are proposed, one minimizing expected squared error and the other using chance constraints [13].

5.1. Minimizing Expected Squared Error. Uncertainty may be treated by minimizing the expected squared error of the piecewise-linear approximation. For the points in the set  $\mathbf{P}_{i,j} = \{\mathbf{p}_i, \mathbf{p}_{i+1}, \dots, \mathbf{p}_j\}$ , this approach finds a linear segment given by the slope  $a_{i,j}$  and intercept  $b_{i,j}$  that solves the problem:

$$F_{i,j}: f_{i,j} = \min_{a_{i,j}, b_{i,j}} \sum_{k=i}^{J} f_{i,j}^{k} \left( a_{i,j}, b_{i,j} \right), \tag{11a}$$

where

$$f_{i,j}^{k}\left(a_{i,j}, b_{i,j}\right) = \int_{x \in X_{b}} \int_{y \in Y_{b}} \left(a_{i,j}x + b_{i,j} - y\right)^{2} f_{k}\left(x, y\right) dx dy,$$
(11b)

 $(X_k, Y_k)$  is the domain of  $\mathbf{p}_k = (\mathbf{x}_k, \mathbf{y}_k)$ ,  $f_k$  is the joint probability density function for  $(\mathbf{x}_k, \mathbf{y}_k)$ , and  $f_{i,j}^k$  is the expected error incurred by  $\mathbf{p}_k$  when approximating  $\mathbf{P}_{i,j}$  with a linear segment.

The function  $f_{i,j}^k(a_{i,j},b_{i,j})$  is convex on the variables  $(a_{i,j},b_{i,j})$  since the term  $(a_{i,j}x+b_{i,j}-y)^2$  is convex for fixed (x,y), generalizing an infinite weighted sum [14, Sec. 3.2.1]. This means that  $f_{i,j}$  is a convex function and so is problem  $F_{i,j}$ . Hence, an optimal solution to  $F_{i,j}$  can be found with numerical optimization methods, such as the interior-point method, using numerical differentiation procedures or directly when a closed form is known for the gradient.

The basic algorithms of Section 2 yield the optimal approximation, for a fixed and maximum number of segments, by solving  $F_{i,j}$  and using the obtained coefficients for the linear segment approximating  $\mathbf{P}_{i,j}$ . This result is formalized in the theorem below, which can be easily proven.

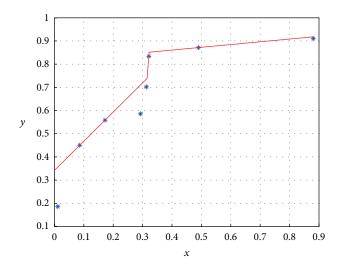


FIGURE 3: Continuous, upper envelope approximation with 3 linear segments obtained with the rejection strategy.

**Theorem 6.** Recursions (2) and (3) yield the optimal piecewise-linear approximation of a set  $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  of random points with exactly and at most t segments, respectively, having the expected squared approximation error as the objective if the coefficients for the linear segments are obtained by solving  $F_{i,j}$  given by (11a) and (11b).

The resulting algorithms run in  $\mathcal{O}(n^3g(n))$  time, with g(n) being an upper bound on the running time to solve  $F_{i,j}$  with 2 variables and n integration terms.

A sample instance is obtained by defining  $\mathbf{x}_i$  and  $\mathbf{y}_i$  as independent random variables, both characterized by Gaussian distributions with mean  $\mu_i^x = x_i$  and  $\mu_i^y = y_i$  and standard deviations  $\sigma_i^x = x_i/2$  and  $\sigma_i^y = y_i/2$ . Table 3 shows the statistics of the random points of the stochastic sample instance. Thus, the joint probability density function is

$$f_i(x_i, y_i) = \frac{e^{-(1/2)((x_i - \mu_i^x)/\sigma_i^x)^2}}{\sigma_i^x \sqrt{2\pi}} \times \frac{e^{-(1/2)((y_i - \mu_i^y)/\sigma_i^y)^2}}{\sigma_i^y \sqrt{2\pi}}.$$
 (12)

With a fixed number T=3 of segments and E(i,j)=f(i,j), PL-Fixed produced the optimal sequence  $\langle i(1),...,i(T+1)\rangle=\langle 1,2,4,8\rangle$  with an expected approximation cost F(8,3)=1.1311. This piecewise-linear function differs from the one with zero variance, for which the optimal sequence is  $\langle i(1),...,i(T+1)\rangle=\langle 1,2,5,8\rangle$  with cost  $F(8,3)=1.6648\times 10^{-2}$ 

5.2. Lower and Upper Envelopes. Instead of minimizing only the expected squared error, approximations for the lower and upper envelope can be obtained by introducing probabilistic constraints to problem  $F_{i,j}$ . An algorithm for the lower envelope is developed below, which is easily adapted for the upper envelope.

To find the lower envelope of  $\mathbf{P}_{i,j} = \{\mathbf{p}_i, \mathbf{p}_{i+1}, \dots, \mathbf{p}_j\}$ , the constraint that the random point  $\mathbf{p}_k = (\mathbf{x}_k, \mathbf{y}_k)$  is above the

TABLE 3: Statistics of the random points.

i	$\mu_i^x$	$\sigma_i^x$	$\mu_i^y$	$\sigma_i^y$
1	0.0114	0.0057	0.1857	0.0929
2	0.0861	0.0431	0.4502	0.2251
3	0.1721	0.0861	0.5584	0.2792
4	0.2928	0.1464	0.5857	0.2929
5	0.3130	0.1565	0.7020	0.3510
6	0.3212	0.1606	0.8338	0.4169
7	0.4899	0.2450	0.8713	0.4357
8	0.8807	0.4404	0.9104	0.4552

linear segment  $a_{i,j}x + b_{i,j}$ ,  $i \le k \le j$ , is imposed with a given probability  $\alpha$ . Mathematically,

$$\mathscr{P}\left(a_{i,j}x + b_{i,j} \le \mathbf{y}_k\right) \ge \alpha, \quad x \in X_k, \tag{13}$$

where  $\mathcal{P}(\cdot)$  denotes the probability operator. The DP algorithms can find the lower envelope by changing how the slope  $a_{i,j}$  and intercept  $b_{i,j}$  of the linear segments are computed. The problem of finding the segment coefficients that minimize the expected squared error, subject to the probabilistic constraints for the lower envelope, is

$$f_{i,j}^{le} = \min_{a_{i,j}, b_{i,j}} \sum_{k=i}^{j} f_{i,j}^{k} \left( a_{i,j}, b_{i,j} \right)$$
 (14a)

while being subject to

$$f_{i,j}^{k}\left(a_{i,j}, b_{i,j}\right) = \int_{x \in X_{k}} \int_{y \in Y_{k}} \left(a_{i,j}x + b_{i,j} - y\right)^{2} f_{k}\left(x, y\right) dx dy,$$
(14b)

$$\mathcal{P}\left(a_{i,j}x + b_{i,j} \le \mathbf{y}_k\right) \ge \alpha, \quad x \in X_k, \ k = i, \dots, j.$$
 (14c)

By expressing probabilistic constraint (14c) in terms of integrals and the joint probability density function  $f_k$ , a deterministic equivalent to this problem is obtained:

$$F_{i,j}^{le}: f_{i,j}^{le} = \min_{a_{i,j},b_{i,j}} \sum_{k=i}^{j} f_{i,j}^{k} \left( a_{i,j}, b_{i,j} \right)$$
 (15a)

while being subject to

$$f_{i,j}^{k}(a_{i,j},b_{i,j})$$

$$= \int_{x \in X_{k}} \int_{y \in Y_{k}} (a_{i,j}x + b_{i,j} - y)^{2} f_{k}(x,y) dx dy,$$

$$g_{i,j}^{k}(a_{i,j},b_{i,j}) = \int_{x \in X_{k}} \int_{y \in Y_{k}: y \leq a_{i,j}x + b_{i,j}} f_{k}(x,y) dx dy$$

$$\leq 1 - \alpha,$$
(15b)

$$k = i, \ldots, j,$$

where  $g_{i,j}^k$  is the probability of  $\mathbf{p}_k$  lying below the linear segment that approximates  $\mathbf{P}_{i,j}$ . This problem may be posed

in terms of the probability  $\overline{g}_{i,j}^k$  of  $\mathbf{p}_k$  being above the linear segment.

Problem  $F_{i,j}^{le}$  is a nonlinear program for which algorithms reach a locally optimal solution, such as sequential quadratic programming (SQP).

The basic algorithms can find the optimal approximation, for a fixed and maximum number of segments, by solving  $F_{i,j}^{le}$  and using the obtained coefficients in the approximation of  $\mathbf{P}_{i,j}$ . This result is formalized in the theorem below, whose proof is similar to that of Theorem 1.

**Theorem 7.** Recursions (2) and (3) yield the optimal piecewise-linear approximation of a set  $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  of random points with exactly and at most t segments, respectively, having the expected squared approximation error as the objective and subject to chance constraints to probabilistically ensure the lower envelope, if the coefficients for the linear segments correspond to the optimal solution of problem  $F_{i,j}^{le}$ .

Algorithm PL-Fixed was applied to the stochastic sample instance given in Table 3, with a fixed number T=3 of segments and probabilities  $\alpha=60\%$  and 70%. The problems  $F_{i,j}^{le}$  were solved approximately using an SQP algorithm. For  $\alpha=60\%$ , the algorithm produced the sequence  $\langle i(1),\ldots,i(T+1)\rangle=\langle 1,2,5,8\rangle$  for the probabilistic lower envelope. The expected cost of this solution is F(8,3)=1.3280, above the expected cost of 1.1311 obtained for the solution without chance constraints. For  $\alpha=70\%$ , the algorithm produced the sequence  $\langle 1,3,5,8\rangle$  with an expected cost of F(8,3)=1.6369.

### 6. Computational Analysis

In order to keep the analysis brief, the experiments were restricted to the following approaches:

- (1) PL-Max, the basic model of Section 2.2 which enforces a maximum number of linear segments.
- (2) PL-Max-Cont, the version of the basic model that imposes continuity by solving quadratic programs according with the developments of Section 3.3.
- (3) PL-Max-Bound, the version of PL-Max which minimizes the number of linear segments while ensuring a bound on approximation error.

The algorithms were implemented in C++ in a workstation equipped with an Intel i7 CPU, 2.67 GHz, with 8 GB of RAM, and running the Linux operating system version Ubuntu 2.6.35-32.

The CFSQP software package and its embedded QP solver were used to solve optimization problems [15]. CFSQP follows the sequential quadratic programming (SQP) framework which solves a sequence of QPs approximating the original problem but having the Lagrangian as the objective to capture the curvature. The gradient of the objective and constraints were provided analytically. The Hessian of the objective was computed numerically using the BFGS strategy.

The approaches were put to the test in the piecewise-linear approximation of the production function of gas-lifted oil

TABLE 4: Running time of the algorithms (in seconds).

Well	Max	Max-Cont	Max-Bound
1	0.004	1.120	0.148
2	0.008	1.172	0.144
3	0.004	1.156	0.164
4	0.004	1.152	0.168
5	0.008	1.140	0.152
6	0.008	1.272	0.164
7	0.008	1.268	0.188
8	0.008	1.160	0.160
9	0.008	1.268	0.184
10	0.004	1.076	0.160

wells. As an oil field matures, the reservoir internal pressure reduces to a level that is not sufficiently high to sustain natural production, requiring the application of artificial lifting techniques. Among these techniques, gas-lift is widely used by the industry for its robustness and relatively low operating costs [16]. Gas-lift works by injecting high pressure gas at the bottom of the production tubing, reducing the fluid density due to gasification and thereby lifting the fluid to the surface.

For the computational analysis, the generation function  $\Phi$  corresponds to the total mass flow  $q^i(q^i_{\rm inj})$  of oil, water, and gas given as a function of the lift-gas rate  $q^i_{\rm inj}$  injected in an oil well i. The relation  $q^i(q^i_{\rm inj})$  is referred to as well production function. This work adopted the oil production system presented in [4], which consists of subsea gas-lifted wells, pipeline-risers, and topside processing equipment such as compressors and separators. The production system is modeled in Pipesim, a commercial multiphase flow simulator widely adopted by the oil industry.

A detailed piecewise-linear model was obtained by uniformly sampling the simulation model of  $q^i$  for a sufficiently high number of  $q^i_{\rm inj}$  values. In our experiments, 50 breakpoints produced a highly accurate and detailed description of  $q^i$ . However, simpler and sufficiently accurate piecewise-linear approximations can be obtained by a better distribution of the breakpoints within the function domain, a task to be performed by the algorithms proposed in this work.

Table 4 shows the running time of the three implemented strategies for piecewise-linear approximation of 10 oil-well production functions, each originally represented with 50 breakpoints. For PL-Max and PL-Max-Cont, the maximum number of segments was set to T=10 and the cost per segment to c=0. For PL-Max-Bound, T=10 and the approximation bound was  $\epsilon=10$ . The CPU times are consistent with the time complexity of the algorithms: PL-Max is faster than PL-Max-Bound due to the extra constraints of the latter; in turn, PL-Max-Bound is faster than PL-Max-Cont because he latter needs to solve QP problems.

The behavior of the PL-Max strategy is shown in Table 5 for a varying cost c and a maximum number of segments T = 15. The algorithm was applied to the production function of well #1.

TABLE 5: Behavior of the PL-Max strategy.

С	Segments	Cost
0	15	29.42
10	15	179.42
40	12	565.28
50	11	677.97
100	10	1197.42
200	8	2 113.22
1000	6	7 639.54

TABLE 6: Behavior of the PL-Max-Cont strategy.

T	Cost
1	784 752.00
5	3 346.49
10	197.42
15	29.42
20	6.44
25	1.68
30	0.47

The squared error grows as c increases, corroborating the theoretical behavior.

The behavior of PL-Max-Cont strategy is illustrated in Table 6 for a varying number of maximum segments. The algorithm was also applied to the production function of well #1. As expected, the squared error decreases as the limit *T* increases.

The computational experiments show that piecewise-linear functions respecting the desired criteria can be obtained with fewer breakpoints than the original representation. Such simpler piecewise-linear models can have a major impact on the production optimization of large oil fields [17].

#### 7. Summary and Conclusion

Applications of piecewise-linear models abound in science and engineering. In petroleum engineering, the performance curve of an oil well can be approximated with a piecewise-linear function. Owing to their importance, this paper developed models for piecewise linearization of nonlinear functions that trade off model complexity and approximation quality. The models and algorithms serve as decision-support tools for engineers to find piecewise-linear models that fit best their needs.

Several models were developed for the synthesis of piecewise-linear approximations under a variety of conditions. Given a potentially large set  $P = \{p_1, \ldots, p_n\}$  of graph points generated by a function  $\Phi$ , not necessarily known explicitly, the problem consists in finding a reduced subset of points that induces a more compact, yet sufficiently accurate piecewise-linear approximation of  $\Phi$ . To this end, a number of criteria were proposed in order to formally state the approximation problems for which algorithms have been designed. The main contributions of this paper consist of

the following formulations and algorithms for piecewiselinear approximation:

- (1) A model and dynamic programming algorithm for finding the optimal approximation using a fixed number of linear segments, which minimizes a quadratic error with respect to the point set *P*.
- (2) A model and DP algorithm to find the optimal approximation using a maximum number of segments, which minimizes the quadratic approximation error and a fixed cost per segment.
- (3) Variations of the two preceding models and algorithms, referred to as basic approaches, to ensure a continuous piecewise-linear approximation.
- (4) Variations of the basic approaches that yield optimal upper and lower envelopes for  $\Phi$ .
- (5) Extensions of the preceding approaches to handle uncertain data, whereby the points of  $\Phi$  are random variables.

The developed approaches can be combined or extended to create models and algorithms that are tailored to the user's needs, such as the synthesis of a piecewise-linear function approximating the upper envelope, with continuity and in the presence of random data points. The proposed models measured error with the  $l_2$ -norm; however other norms can be easily adopted such as the  $l_1$ - and  $l_\infty$ -norm by solving suitable linear and quadratic programs.

#### **Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

# Acknowledgment

This work was supported in part by CNPq under Grant 302972/2013-7.

## References

- [1] V. Gunnerud and B. Foss, "Oil production optimization a piecewise linear model, solved with two decomposition strategies," *Computers & Chemical Engineering*, vol. 34, no. 11, pp. 1803–1812, 2010.
- [2] H. Lee, J. M. Pinto, I. E. Grossmann, and S. Park, "Mixed-integer linear programming model for refinery short-term scheduling of crude oil unloading with inventory management," *Industrial & Engineering Chemistry Research*, vol. 35, no. 5, pp. 1630–1641, 1996.
- [3] R. Rejowski Jr. and J. M. Pinto, "An MILP formulation for the scheduling of multiproduct pipeline systems," *Brazilian Journal of Chemical Engineering*, vol. 19, no. 4, pp. 467–474, 2002.
- [4] A. Codas and E. Camponogara, "Mixed-integer linear optimization for optimal lift-gas allocation with well-separator routing," *European Journal of Operational Research*, vol. 217, no. 1, pp. 222–231, 2012.

- [5] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.
- [6] S. Ghosh and S. Maka, "A fuzzy clustering based technique for piecewise affine approximation of a class of nonlinear systems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 9, pp. 2235–2244, 2010.
- [7] S. Ghosh, A. Ray, D. Yadav, and B. M. Karan, "A genetic algorithm based clustering approach for piecewise linearization of nonlinear functions," in *Proceedings of the International Conference on Devices and Communications (ICDeCom '11)*, pp. 1–4, February 2011.
- [8] T. L. Silva and E. Camponogara, "A computational analysis of multidimensional piecewise-linear models with applications to oil production optimization," *European Journal of Operational Research*, vol. 232, no. 3, pp. 630–642, 2014.
- [9] T. L. Silva, E. Camponogara, A. F. Teixeira, and S. Sunjerga, "Modeling of flow splitting for production optimization in offshore gas-lifted oil fields: simulation validation and applications," *Journal of Petroleum Science & Engineering*, vol. 128, pp. 86–97, 2015.
- [10] J. Kleinberg and E. Tardos, Algorithm Design, Addison Wesley, 2005.
- [11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, 2001.
- [12] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, Athena Scientific, 2000.
- [13] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Management Science*, vol. 6, pp. 73–79, 1959.
- [14] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
- [15] C. T. Lawrence, J. L. Zhou, and A. L. Tits, "User's guide for cfsqp version 2.4: a c code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints," Tech. Rep., Institute for Systems Research, University of Maryland, College Park, Md, USA, 1996.
- [16] E. Camponogara and A. M. de Conto, "Lift-gas allocation under precedence constraints: MILP formulation and computational analysis," *IEEE Transactions on Automation Science and Engi*neering, vol. 6, no. 3, pp. 544–551, 2009.
- [17] A. Codas, S. R. V. Campos, E. Camponogara, V. Gunnerud, and S. Sunjerga, "Integrated production optimization of oil fields with pressure and routing constraints: the Urucu field," *Computers & Chemical Engineering*, vol. 46, pp. 178–189, 2012.

















Submit your manuscripts at http://www.hindawi.com























