

# Stochastic U-Curve Branch and Bound

**Student:** Gustavo Estrela de Matos

**Advisor:** Dr. Ulisses Braga-Neto

## 1 Input Generation

To generate the input for the problem we created two functions. The first creates a vector of floating points that simulates the values of a chain of a boolean lattice that respects the U-Curve assumption; and the second one adds a random noise to the values of the vector.

### 1.1 Chain Generation

The chain is generated based on a polynomial of second degree with positive leading coefficient. This polynomial has the minimum value of 0.1 evaluated at some point  $x^* \in [0, n]$  (where  $n$  the size of the chain) which is determined by the user. The algorithm receives two arguments: the first one,  $n$ , represents the size of the chain; and the second one,  $center \in [0, 1]$ , specifies where is the minimum of the polynom. After creating a random polynomial that holds our constraints, we generate  $n$  uniformly distributed points in the interval  $[0, n]$ , and evaluate the polynomial at these points, generating the curve.

### 1.2 Noise

The noise is applied to the vector created by *GeneratePoints*, by adding to every point  $v_i$  a Gaussian error,  $\epsilon_i$  with mean zero and standard deviation *sigma*. We also assume that every pair  $\epsilon_i, \epsilon_j$  are independent and identically distributed for  $i \neq j$ .

---

**Algorithm 1** U-Curve Input Creator

---

```
1: procedure GENERATEPOINTS( $n, max\_distance, center$ )
2:    $points \leftarrow \{0, \dots, 0\}$ 
3:    $p.a \leftarrow random()$ 
4:    $p.b \leftarrow -2 * p.a * center$ 
5:    $p.c = (p.a^2 - 0.1)/(4 * p.a)$ 
6:    $dx \leftarrow 1/n$ 
7:    $x \leftarrow 0$ 
8:   for  $i \in \{0, \dots, n\}$  do
9:      $points_i \leftarrow p(x)$ 
10:     $x \leftarrow x + dx$ 
11:  end for
12:   $j \leftarrow n * random()$ 
13:   $plain\_size \leftarrow (n - j) * random()$ 
14:  for  $k \in \{1, \dots, plain\_size\}$  do ▷ Creates a plain area in the chain
15:     $points[j + k] \leftarrow points[j]$ 
16:  end for
17:  return  $points$ 
18: end procedure
```

---

Figure 1: Example of a curve generated with  $n = 100$ ,  $center = 0.5$  and  $\sigma = 0$

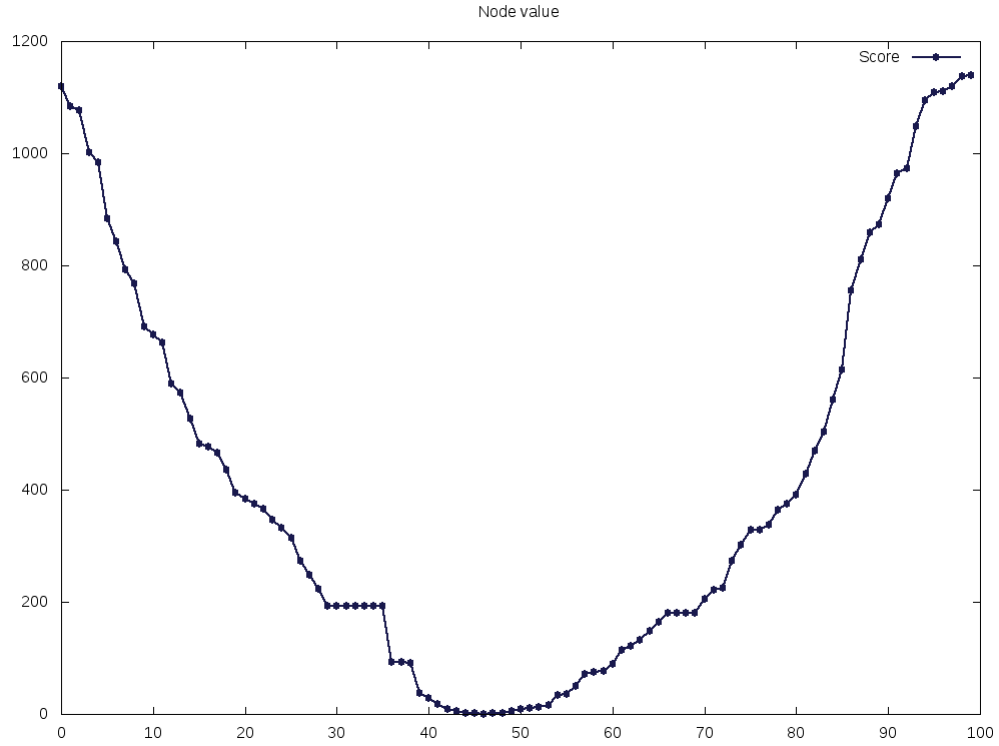


Figure 2: Example of a curve generated with  $n = 100$ ,  $center = 0.5$  and  $\sigma = 5$

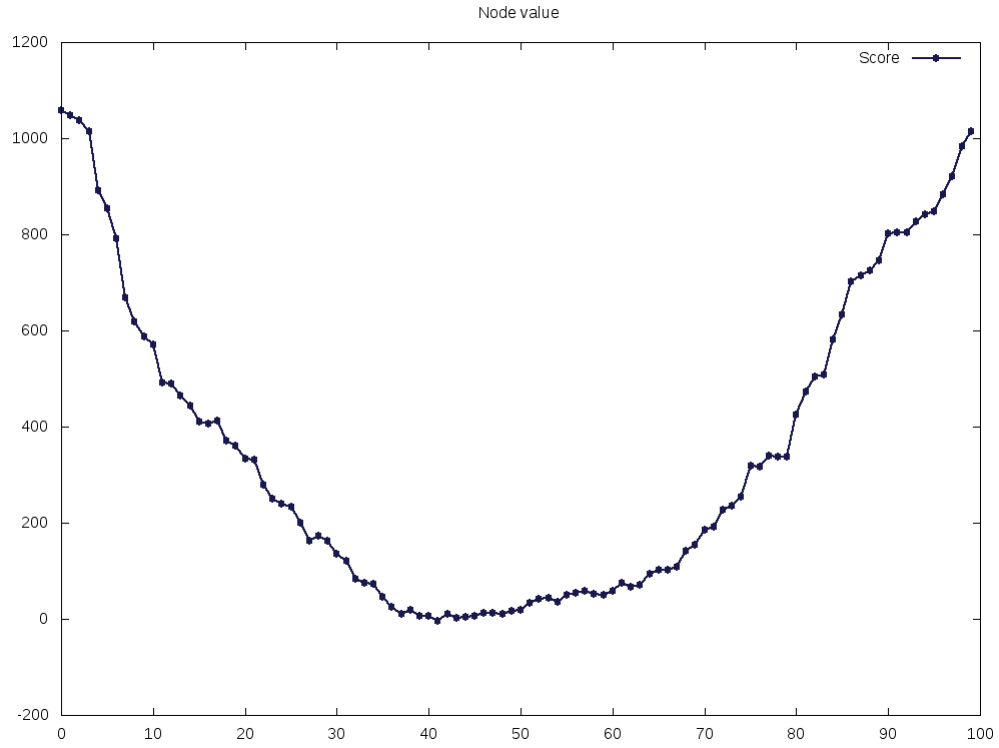
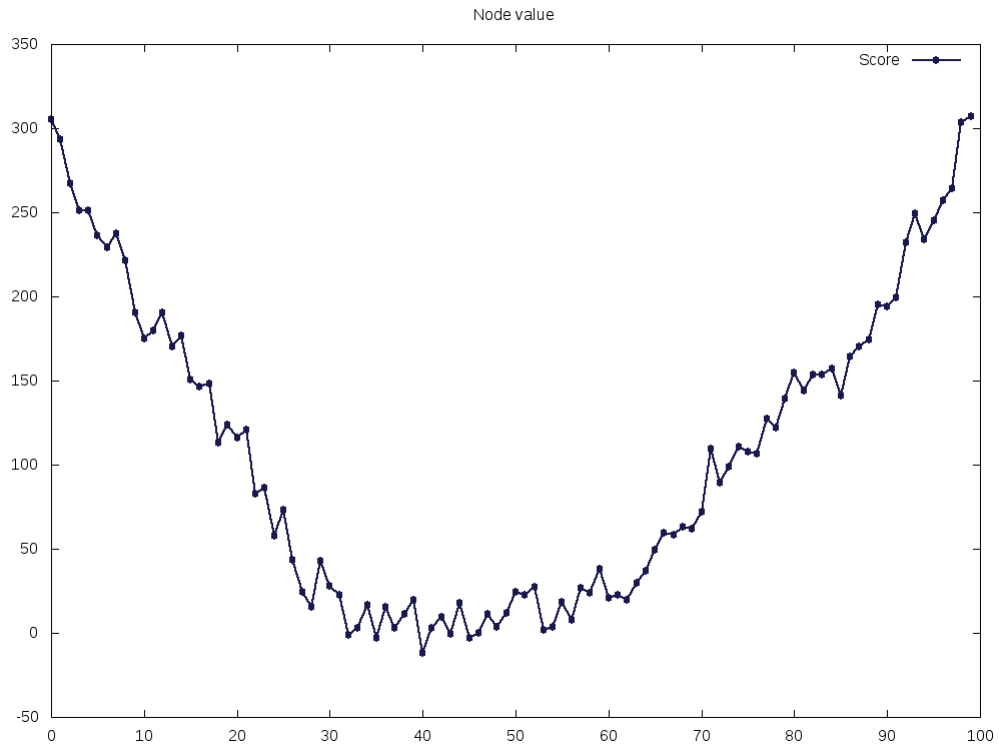


Figure 3: Example of a curve generated with  $n = 100$ ,  $center = 0.5$  and  $\sigma = 10$



## 2 Bisection Algorithms

### 2.1 Traditional Bisection

This is the simplest bisection algorithm we implemented. The basic idea of this algorithm is to divide the original problem in two halves and, by analysing the neighbours of the node in the middle, determine which half has the minimum and then solve this half recursively. To guarantee the optimality of the algorithm (when the input respects the U-Curve assumption) we solve both halves when the neighbours can't determine if the minimum lies to the left or right of the middle point.

---

**Algorithm 2**

---

```
1: procedure BISECTION( $v$ )
2:    $n \leftarrow v.length$ 
3:    $i \leftarrow n/2$ 
4:   if ( $valley(v, i)$ ) then
5:     return  $v[i]$ 
6:   else
7:      $direction \leftarrow SelectSide(v, i)$ 
8:     if  $direction = Left$  then
9:       return  $Bisection([v_i, \dots, v_n])$ 
10:    else if  $direction = Right$  then
11:      return  $Bisection([v_0, \dots, v_{i-1}])$ 
12:    else ▷ Unknown direction
13:      return  $min(Bisection([v_0, \dots, v_{i-1}]), Bisection([v_i, \dots, v_n]))$ 
14:    end if
15:  end if
16: end procedure
```

---

---

**Algorithm 3**

---

```
1: procedure SELECTSIDE( $v, i$ )
2:    $d = v[i + 1] - v[i - 1]$ 
3:   if  $|d| < \epsilon$  then
4:     return Unknown
5:   else if  $d > 0$  then
6:     return Left
7:   else
8:     return Right
9:   end if
10: end procedure
```

---

## 2.2 Mid-neighbour Bisection

As an attempt to minimize the effects of noise in the traditional bisection algorithm we developed the Mid-neighbour Bisection, which brings two new ideas to the traditional bisection, that changes the evaluated points and also how we divide the problem in smaller problems.

The first idea consists in changing the points that are evaluated to decide which side to go. Instead of looking to the neighbours of the middle point, we now evaluate the points that are in the middle of the left half and right half. After evaluating these points, we calculate the difference between the left-middle point and middle point and also the difference between the right-middle point and the middle point to get two abstract slopes that can guarantee us which fraction of the input has the minimum value.

The second idea considers the "reliability" of the abstract slopes mentioned before, for instance, if we say that the left slope (middle point - left-middle point) is negative, it's more likely that the function is going to increase from the left-middle point and before if the distance between these two points is greater. That happens because if the function decreases from the left-middle point and before, then the left-middle point suffered some noise that caused a peak (an inverted u-shape) and, considering that the points should describe a u-shape that would imply in a bigger noise for bigger distances of middle point and left-middle point. We modeled the "reliability" of the abstract slope as a linear function that decreases from one to zero in  $\lg n$  steps.

The abstract slopes are reduced to three different cases: 1 for growing, -1 for decreasing and 0 for plain areas or unreliable difference. Since we have three different slopes, there are nine different cases that need to be handled by the algorithm. Suppose that we are using the mid-neighbour bisection on a vector  $v = \{v_1, \dots, v_n\}$  and the right-middle, middle and left-middle point are, respectively,  $lm$ ,  $m$ , and  $rm$ . Then, the following pictures represent how the algorithm finds the minimum recursively.

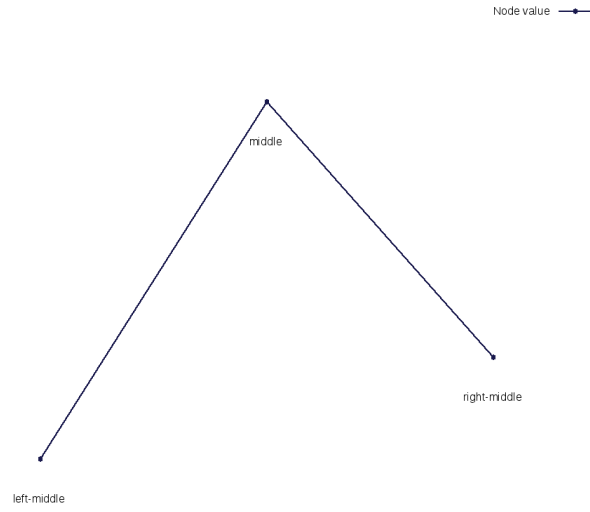


Figure 4: In this case we split the original problem into the two problem of finding the minimum of  $\{v_1, \dots, v_{m-1}\}$  and  $\{v_m, \dots, v_n\}$ .

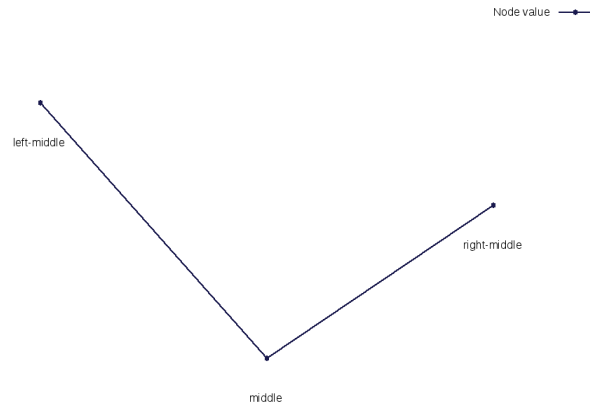


Figure 5: In this case we reduce the problem to finding the minimum of  $\{v_{lm}, \dots, v_{rm}\}$ .

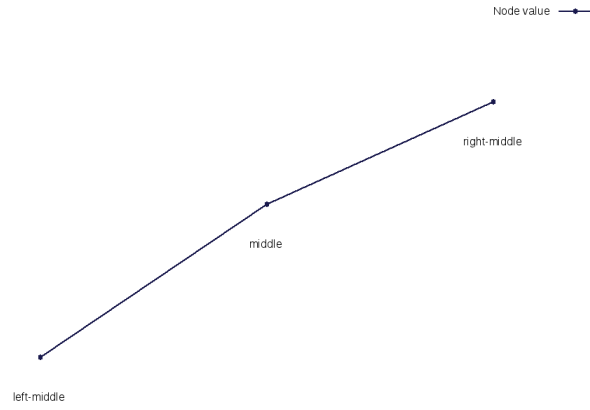


Figure 6: For this case we reduce the problem to finding the minimum of  $\{v_1, \dots, v_m\}$ .

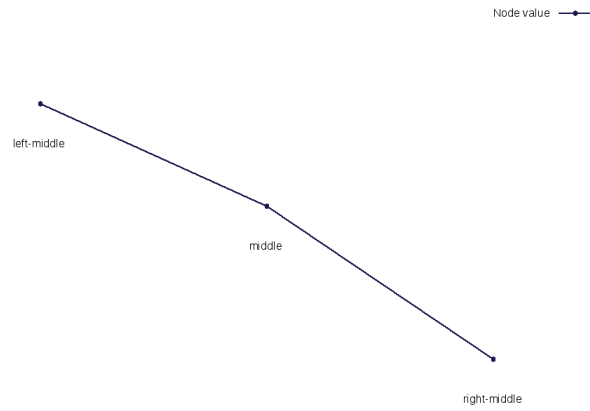


Figure 7: In this case we reduce the problem to finding the minimum of  $\{v_m, \dots, v_n\}$ .

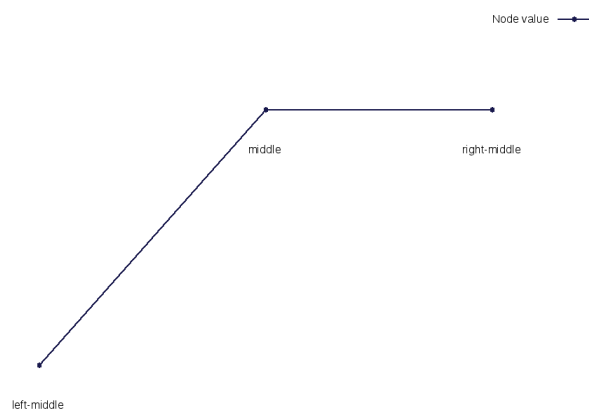


Figure 8: For this case we reduce the problem to finding the minimum of  $\{v_1, \dots, v_m\}$ .

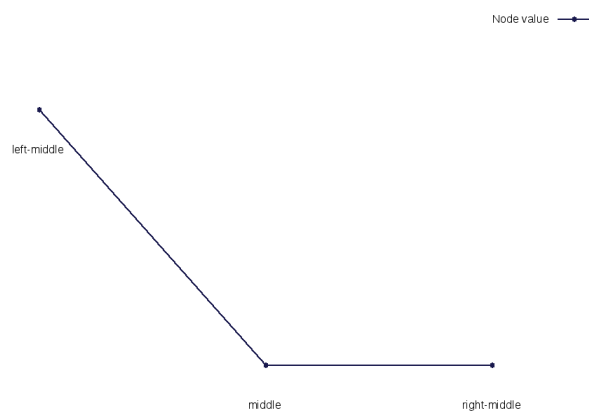


Figure 9: For this case we reduce the problem to finding the minimum of  $\{v_{lm}, \dots, v_m\}$ .



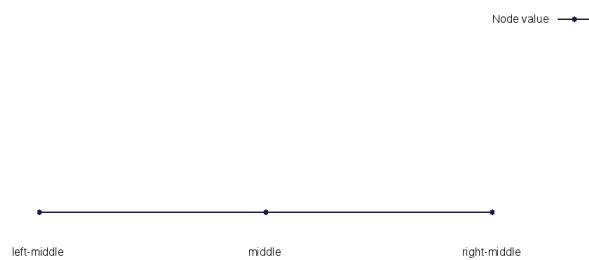


Figure 10: In this case we split the original problem into the two problem of finding the minimum of  $\{v_1, \dots, v_{m-1}\}$  and  $\{v_m, \dots, v_n\}$ .

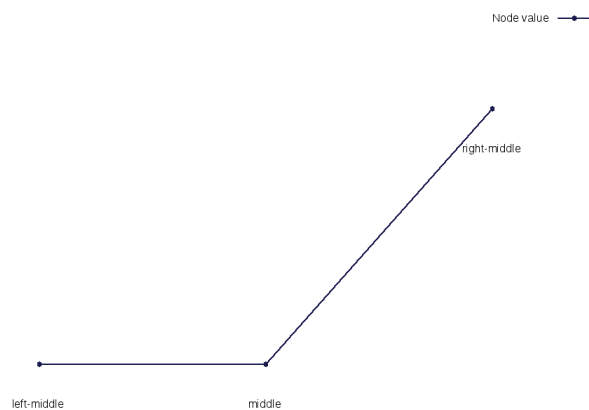


Figure 11: In this case we reduce the problem to finding the minimum of  $\{v_1, \dots, v_m\}$ .

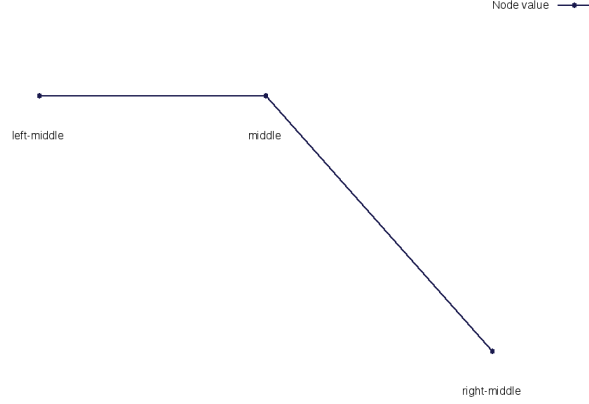


Figure 12: In this case we reduce the original problem to finding the minimum of  $\{v_m, \dots, v_n\}$

### 2.3 U-Curve Probabilistic Bisection

The U-Curve Probabilistic Bisection (UPB) is a stochastic version of the algorithm we called "Traditional Bisection". The UPB algorithm keeps a probability mass function  $f_i(x)$ , for every iteration  $i$  and every node  $x$  of the search space, that can represent the algorithm belief, on iteration  $i$ , that the node  $x$  is the answer to the problem. This algorithm has two key points: selecting the node to be evaluated and updating the probability mass function.

Different from the Traditional Bisection, we don't select the element that is the middle of the input to evaluate, instead we select the  $m$ -th element of the input vector  $v$  such that:

$$m = \min_{1 \leq i \leq n} \{i \mid f(v_i) \geq \frac{1}{2}\}.$$

, i.e the median of the possible solutions. The algorithm then compares the adjacent nodes of  $v_m$  and decides which side the solution should be the same way it was done in the traditional bisection.

We call  $p_c$  the probability that we chose the correct side when evaluating  $v_m$ 's neighbours and  $\alpha = \sum_{j=1}^m f_i(v_j)$ . In the case where the actual solution lies in the right side of  $v_m$ , the

probability mass function update is calculated by:

$$f_{i+1}(y) = \begin{cases} (\frac{1}{1-\alpha})p_c f_i(y) & \text{if } y \geq v_m \\ \alpha(1 - p_c)f_i(y) & \text{if } y < v_m \end{cases}$$

. This update is similar for the case when the solution lies in the left side of  $v_m$ . If there is no defined value for  $f_0$ , we simply assume that the probability of being a solution is distributed uniformly between all nodes of the chain.

The algorithm terminates when the first quantile or when the third quantile, of the possible solutions, are the same node as the second quantile (median), which is the solution given by the algorithm. This termination condition avoids an early termination, that would be caused by a local minimal element if we used the same termination condition as we did in the traditional bisection.

---

#### Algorithm 4

---

```

1: procedure UPB( $v, pmf$ )
2:    $n \leftarrow v.length$ 
3:    $Quantiles \leftarrow FindQuantiles(pmf)$ 
4:   while  $Quantiles_1 \neq Quantiles_2$  and  $Quantiles_2 \neq Quantiles_3$  do
5:      $direction \leftarrow SelectSide(v, i)$ 
6:     if direction is Unknown then ▷ Unknown direction
7:       return  $SplitUPB(v, pmf, Quantile_2)$ 
8:     else
9:        $UpdatePMF(pmf, Quantile_2, F(Quantile_2), direction)$ 
10:    end if
11:     $Quantiles \leftarrow FindQuantiles(pmf)$ 
12:  end while
13:  return  $Quantile_2$ 
14: end procedure

```

---

## 2.4 Mid-neighbour Probabilistic Bisection

The Mid-neighbour Probabilistic Bisection (MPB) is a probabilistic algorithm that partitionates the search space into four different parts and, according to the value of the points between the subsets, decides in which portion of the chain the minimum should be. The partition on the chain is made according to the quantiles of the  $pmf$  and the points we evaluate

are  $v_{lm}$ ,  $v_m$  and  $v_{rm}$ , which we define as:

- $lm = \min_{1 \leq i \leq n} \{i \mid f(v_i) \geq \frac{1}{4}\}.$
- $m = \min_{1 \leq i \leq n} \{i \mid f(v_i) \geq \frac{2}{4}\}.$
- $rm = \min_{1 \leq i \leq n} \{i \mid f(v_i) \geq \frac{3}{4}\}.$

After finding the points  $v_{lm}$ ,  $v_m$  and  $v_{rm}$  we calculate the cost  $c$  of these points and update the  $pmf$  according to the following rules:

- If  $c(v_{rm}) - c(v_{lm}) > \epsilon$  : then we guess that the minimum should be to the left of  $v_{rm}$ .
- If  $c(v_{rm}) - c(v_{lm}) < -\epsilon$  : then we guess that the minimum should be to the right of  $v_{lm}$ .
- If  $|c(v_{rm}) - c(v_{lm})| < \epsilon$  and  $v_m < \min(v_{rm}, v_{lm})$  : then we guess that the minimum should be between  $v_{lm}$  and  $v_{rm}$
- If  $|c(v_{rm}) - c(v_{lm})| < \epsilon$  and  $v_m \geq \min(v_{rm}, v_{lm})$  : then we can't guess where the minimum is. To overcome this, we divide the problem in the node  $v_m$  so we can solve, recursively, two problems and return the minimum of the two solutions.

---

**Algorithm 5**

---

```
1: procedure MPB( $v, pmf$ )
2:    $n \leftarrow v.length$ 
3:    $Quantiles \leftarrow FindQuantiles(pmf)$ 
4:    $v_{lm} \leftarrow Quatile_1$ 
5:    $v_m \leftarrow Quatile_2$ 
6:    $v_{rm} \leftarrow Quatile_3$ 
7:   while  $v_{lm} \neq v_m$  and  $v_{rm} \neq v_m$  do
8:      $d \leftarrow c(v_{rm}) - c(v_{lm})$ 
9:     if  $abs(d) < \epsilon$  then
10:      if  $v_m < \min(v_{rm}, v_{lm})$  then
11:         $UpdatePMF(pmf, v_{lm}, F(v_{lm}), Right)$ 
12:         $UpdatePMF(pmf, v_{rm}, F(v_{lm}), Left)$ 
13:      else ▷ Unknown direction
14:        return  $SplitMPB(v, pmf, v_m)$ 
15:      end if
16:    else if  $d > \epsilon$  then
17:       $UpdatePMF(pmf, v_{rm}, F(v_{rm}), Left)$ 
18:    else
19:       $UpdatePMF(pmf, v_{lm}, F(v_{lm}), Right)$ 
20:    end if
21:     $Quantiles \leftarrow FindQuantiles(pmf)$ 
22:     $v_{lm} \leftarrow Quatile_1$ 
23:     $v_m \leftarrow Quatile_2$ 
24:     $v_{rm} \leftarrow Quatile_3$ 
25:  end while
26:  return  $v_m$ 
27: end procedure
```

---

## 2.5 Mid-neighbour Belief Bisection

The Mid-neighbour Belief Bisection (MBB) is another algorithm based on the Mid-neighbour algorithm and it relies on the knowledge of distribution of the noise prior to the bisection. This algorithm assumes that the noise present on the chain has a Gaussian distribution with mean zero and standard deviation  $\sigma$  and it's also i.i.d. in the nodes of the chain, as we described in 1.2

The MBB algorithm starts by locating 3 points that divide the chain into 4 equally sized portions, identically to what we did in the Mid-neighbour algorithm; these three points are  $v_x$ ,  $v_y$  and  $v_z$ , where  $x < y < z$ . Then, we calculate the observable cost of the points  $x$  and  $z$ ,  $c'(v_x)$  and  $c'(v_z)$  respectively, and, using the knowledge of the standard deviation of the noise distribution, we can make assumptions on the real cost function  $c$  and decide in which portion of the chain the solution is located, with a high probability.

The algorithm has three different branches that determine how we divide and prune the chain to get the minimum cost, which are:

- when  $c(x) < c(z)$ , with a high probability, we have that the cost of the elements  $\{v_x, \dots, v_{n-1}\}$  have a higher cost than the elements  $\{v_0, \dots, v_{x-1}\}$ , therefore we can solve the problem by finding the minimum of the later chain.
- when  $c(x) > c(z)$ , with a high probability, we have that the cost of the elements  $\{v_0, \dots, v_x\}$  have a higher cost than the elements  $\{v_{x+1}, \dots, v_{n-1}\}$ , therefore we can solve the problem by finding the minimum of the later chain.
- when there's not enough information about the difference  $c(x) - c(z)$  we have to split the problem into two different problems on the chains  $\{v_0, \dots, v_y\}$  and  $\{v_{y+1}, \dots, v_{n-1}\}$  and give as a solution the best solution of both sub-problems.

Suppose that  $c'(x) < c'(z)$ , we can determine whether  $P(c(x) < c(z))$  is high or not by deriving the following calculations:

$$P(c(v_x) < c(v_z)) = P(c(v_x) - c(v_z) < 0)$$

Now we use the fact that  $c'(v_i) = c(v_i) + \epsilon_i$  for every  $v_i$  of the input chain, and

$$\begin{aligned} P(c(v_x) - c(v_z) < 0) &= P(c'(v_x) - \epsilon_x - c'(v_z) + \epsilon_z < 0) \\ &= P(\epsilon_x - \epsilon_z > c'(v_x) - c'(v_z)) \end{aligned}$$

Assuming that  $\epsilon_x$  and  $\epsilon_z$  are Gaussian  $(0, \sigma)$  and they are independent, we have that the random variable  $\epsilon_x - \epsilon_z$  also have Gaussian distribution with mean zero and standard deviation  $2\sigma$ . Therefore:

$$P(\epsilon_x - \epsilon_z > -2\sigma) = 0.8413$$

Then, if we know that  $c'(v_x) - c'(v_z) < -2\sigma$ , we can guarantee that  $P(\epsilon_x - \epsilon_z > c'(v_x) - c'(v_z)) = P(c(v_x) < c(v_z)) \geq 0.8413$

It's important to notice that, to determine a high probability, we used  $-2\sigma$  as a boundary that forces  $P(c(v_x) < c(v_z))$  to be at least 0.8413, and also, when the difference  $c'(v_x) - c'(v_z)$  doesn't reach the boundary we split the problem into two different pieces with no prune of the chain. Therefore, there's a trade-off when choosing the boundary for a high probability: as we lower the boundary, the probability  $P(c(v_x) < c(v_z))$  approaches one, but this also implies that the difference  $c(v_x) - c(v_z)$  should be higher otherwise no prune will be made, i.e we have better correctness and worse run time; on the other hand, if we raise the boundary we will make prunes more often, which can imply in the deletion of global minimal, i.e better run time and worse correctness.

---

**Algorithm 6**

---

```
1: procedure MBB( $v, \sigma$ )
2:    $n \leftarrow v.length$ 
3:    $y = n/2$ 
4:    $x = y/2$ 
5:    $z = (n - y)/2$ 
6:   if  $x = z$  then
7:     return  $v_x$ 
8:   end if
9:   if  $c'(v_x) < c'(v_z)$  then
10:     $w = \min(x, y)$ 
11:     $d = v_w - v_z$ 
12:    if  $d < -2\sigma$  then
13:      return  $MBB(\{v_0, \dots, v_z\}, \sigma)$ 
14:    else
15:      return  $\min(MBB(\{v_0, \dots, v_{y-1}\}, \sigma), MBB(\{v_y, \dots, v_{n-1}\}))$ 
16:    end if
17:  else
18:     $w = \min(y, z)$ 
19:     $d = v_w - v_x$ 
20:    if  $d < -2\sigma$  then
21:      return  $MBB(\{v_{x+1}, \dots, v_{n-1}\}, \sigma)$ 
22:    else
23:      return  $\min(MBB(\{v_0, \dots, v_{y-1}\}, \sigma), MBB(\{v_y, \dots, v_{n-1}\}))$ 
24:    end if
25:  end if
26: end procedure
```

---



### 3 Results

