

**STOCHASTIC OPTIMIZATION FOR FEASIBILITY DETERMINATION:
AN APPLICATION TO WATER PUMP OPERATION IN WATER DISTRIBUTION NETWORK**

Yi-An Tsai
Giulia Pedrielli
Logan Mathesen

School of Computing-IDSE
Arizona State University
699 S Mill Ave
Tempe, AZ 85258, USA

Zelda B. Zabinsky

Department of Industrial and Systems Engineering
University of Washington
3900 E Stevens Way NE
Seattle, WA 98195, USA

Hao Huang

Department of Industrial Engineering &
Management
Yuan Ze University
Taoyuan City, 320, TAIWAN

Antonio Candelieri
Riccardo Perego

Department of Computer Science,
Systems and Communication
University of Milano Bicocca
Milan, 20125, ITALY

ABSTRACT

Water Distribution Networks are a particularly critical infrastructure for the high energy costs and frequent failures. Variable Speed Pumps have been introduced to improve the regulation of water pumps, a key for the overall infrastructure performance. This paper addresses the problem of analyzing the effect of the VSPs regulation on the pressure distribution of a WDN, which is highly correlated to leakages and energy costs. Due to the fact that water network behavior can only be simulated, we formulate the problem as a *black box feasibility determination*, which we solve with a novel stochastic partitioning algorithm, the Feasibility Set Approximation Probabilistic Branch and Bound, that extends the algorithm previously proposed by two of the authors. We use, as black box, EPANet, a widely adopted hydraulic simulator. The preliminary results, over theoretical functions as well as a water distribution network benchmark case, show the viability and advantages of the proposed approach.

1 INTRODUCTION

Water Distribution Networks (WDNs) are large and complex cyber-physical infrastructures, crucial for social, industrial and environmental aspects of our life. Water utilities that manage WDNs must ensure the satisfaction of service supply in terms of quality and quantity of water, while simultaneously achieving customer performance goals. A recent report by Copeland and Carter (2017), reveals that energy consumption by public drinking water and wastewater utilities, which are primarily owned and operated by local governments, can represent 30%-40% of a municipality's energy bill. Also, water infrastructures are known to be characterized by frequent leakages that oftentimes lead to failures of the network which are difficult to identify and, therefore, recover (Candelieri et al. 2017; Candelieri et al. 2015a; Candelieri et al. 2015b; Candelieri et al. 2014). It is of uttermost important to operate these infrastructures in a way that prevents failures, which clearly has impact on the energy costs. In this

direction, one of the key components responsible for safe operations are the pumps, whose functioning is responsible for 80% of the overall energy cost. Pumps are also responsible for the pressure distribution across the infrastructure, which is a key to understanding location and diffusion of potential leakages. The engineering community has carefully looked into this problem proposing the design of a new technical solution to control water pumps, the Variable Speed Pump (VSP). Any VSP can be programmed to run at many different rotational speed settings, each with a different related pump curve. Specifically, it is possible to regulate the pump speed by controlling the relative speed parameter, e.g., when running the pump at half speed, the relative setting is 0.5 (Rossman 2000). Potentially, VSPs can save up to 90% of annual energy cost when compared to traditional binary pumps.

Regulating VSPs to achieve cost savings and efficient pressure distribution is a challenge, due, in part, to the difficulty of analyzing the system. Indeed, computationally costly and extensive computer simulations are required to analyze the behavior of a WDN. Available simulation packages include: EPANet, Finesse, H2Onet, and WaterCAD (Rossman 2000; Balut and Urbaniak 2011; Paluszczyszyn et al. 2015; Schmid 2002). In light of this, the goal of this research is to provide a novel methodology that makes an intelligent use of simulation to enable water utilities to better understand appropriate working conditions for a WDN in terms of pump regulation focusing on the pressure distribution profile.

1.1 Solution Approach

In this paper we explore the problem of determining safe working conditions for a large water network as a stochastic feasibility determination problem. In particular, we formalize the problem of safe working condition characterization as the problem of determining a set of configurations of pump speeds such that the profile pressure across the network is appropriately bounded within values defined by the engineering experts. Our approach can then be coupled with an optimizer/control algorithm according to the process summarized in Figure 1.

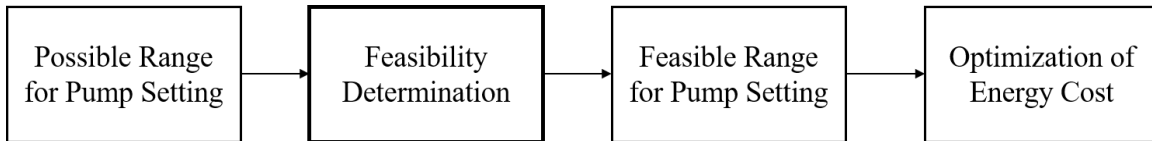


Figure 1: Process of pump scheduling optimization.

2 BACKGROUND

The background of this research is vast and we focused on two main areas: (1) simulation of WDNs, and (2) approaches for the control and analysis of WDNs. In Section 2.1, we introduce in detail an open-source hydraulic simulation software, EPANet. Then, the algorithms proposed in the literature to solve the Pump Scheduling Optimization, are presented in Section 2.2.

2.1 EPANet

In this work, we adopted a state-of-the-art hydraulic simulator, EPANet 2.0 (Rossman 2000), an open-source freeware that is widely adopted in the literature, that describes the detailed workings of a complex hydraulic system. The main network components considered within EPANet are pipes, nodes (junctions), pumps, valves and storage tanks or reservoirs. With these components, EPANet performs an extended period simulation of the hydraulic and water quality behavior. The user can construct the network by setting object properties (i.e., pipe diameter, pump efficiency curve). As output, EPANet provides information about the flow of water in each pipe, the pressure at each node, the height of water in each tank and the concentration of a chemical throughout the network during a user-defined time period characterized by a user-defined demand profile. In addition to running the simulation of the various

complex hydraulic systems configuration, EPANet can also compute the energy cost for keeping the network pumps active during the simulated horizon.

The minimum set of data needed to create a model of a WDN consists of: (a) coordinates and altimetry for each node; (b) demand profile (i.e., a “pattern”) over the simulation time horizon of the water request for each consumption node in the network; (c) size and shape of each tank, with an initial level; (d) connectivity of the WDN (links connecting nodes); (e) length, diameter and roughness of each pipe; (f) efficiency curve of each pump (which can be binary or VSPs); and (g) energy tariff over the simulation period. A simple example of a WDN model (named Net1) is presented in Figure 2(a). This network has one variable speed pump, one reservoir, one storage tank, and nine nodes. Examples of the input to EPANet for hourly demand pattern and pump curve are illustrated in Figure 2(b) and Figure 2(c), respectively. Outputs from the EPANet Net1 example are shown in Figure 3. In particular, Figure 3(a) illustrates the pressure at two junctions in the network (red and green curves) over a time period of 24 hours, while Figure 3(b) shows the time series for the tank level.

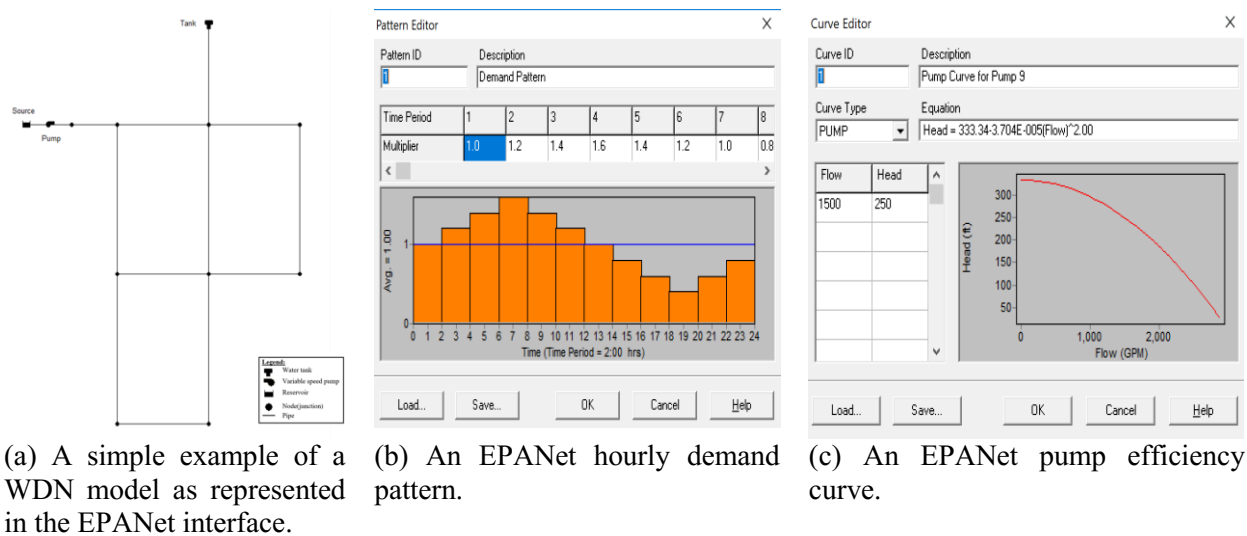
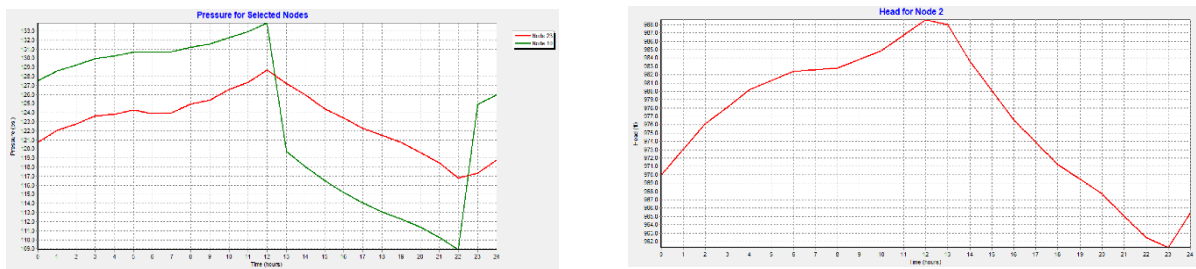


Figure 2: EPANet screen shots of inputs to the Net1 example.



(a) Pressure measured at two nodes of the WDN. (b) Tank level.

Figure 3: EPANet screen shots of outputs from the Net1 example over the simulation time period (24 hours).

2.2 Control and Analysis of WDNs

Most of the operations research literature on operation of WDNs focuses on the Pump Scheduling Optimization (PSO) problem (Mala-Jetmarova et al. 2017). PSO tries to identify the optimal pumps’

operating conditions in terms of hourly pump regulation (e.g., speed ratio in case of Variable Speed Pumps, on/off in case of binary pumps) in order to identify the regulation that guarantees a water supply that meets the demand and minimizes the energy cost. In order to model the network behavior and formulate the energy cost several approximation to the highly non-linear flow dynamics of WDNs have been proposed. These approximations transform the PSO problem to a classic optimization problem for which traditional deterministic algorithms apply. In this direction, (Mala-Jetmarova et al. 2017) reports several classes of existing approaches, including linear programming (Pasha and Lansey 2009), nonlinear programming (Chase and Ormsbee 1989), and dynamic programming (Sterling and Coulbeck 1975). Mathematical programming-based approaches try to formalize the problem by linearizing/convexifying the equations regulating flow, thus greatly simplifying the complex water distribution system. As a result, most of the applications are limited to solve the optimization problem only on simple water distribution networks.

Meta-heuristic algorithms, such as Genetic Algorithms (Nicklow et al. 2009; Savic et al. 1997), and Harmony-Search Algorithm (De Paola et al. 2016) have also been proposed. Most of the approaches do not consider the presence of Variable Speed Pumps (VSPs). As a result, the problem is reduced to decision variables of binary pump statuses (0 = pump off, 1 = pump on) during a short time interval (Mala-Jetmarova et al. 2017). More recently, a Sequential Model Based Optimization (SMBO), more precisely Bayesian Optimization (BO) approach has been proposed to solve the PSO problem in both the VSPS and ON/OFF pumps settings (Candelieri et al. 2018). Also, as previously mentioned, most of the literature focuses on energy cost minimization, while fewer contributions look at the problem from a feasibility perspective. Most of the contributions consider the feasibility of pump operation settings by means of a penalty function added to the objective function (Biles et al. 2007). The main idea of penalization is to try to minimize the true objective while also driving the penalty to zero, thus leading to the identification of a feasible, optimal, solution. Lagrangian relaxation is among the most popular techniques used in this area (Bertsekas 2014). Although Lagrangian relaxation (Biles et al. 2007) can guarantee the identification of the optimal solution, it fails to provide insights into feasibility. This motivates us in the direction of studying multiple measures that can lead to characterizing the quality working conditions of a WDN. To do so, we propose an adaptation of a stochastic optimization method designed for level set approximation (Huang and Zabinsky 2013; Zabinsky and Huang 2018) in order to apply it for feasibility determination. Feasibility has been investigated in Szechtman and Yucesan (2008), but for Ranking and Selection problem. According to our approach, the output from the feasibility determination algorithm is a set of sub-regions with corresponding feasibility measures providing an approximation of the true, unknown feasible region. Such an approximation helps practitioners understand constraints and possible stressing working conditions/vulnerable nodes within the system, and improve the efficiency of the optimization process.

3 METHODOLOGY

We formulate the problem of identifying pump configuration with non-dangerous pressure profiles as a feasibility determination problem. We propose to adapt a stochastic optimization algorithm, Probabilistic Branch and Bound (PBnB) (Zabinsky et al. 2011; Huang and Zabinsky 2013; Zabinsky and Huang 2018), which approximates a level set with probabilities of being correct and incorrect (within α error). PBnB can be applied to mixed continuous-integer black box problems. While it has good performance in optimization settings, it has not been formally applied to feasibility determination, and this will be an outcome of this paper.

3.1 Introduction of Probabilistic Branch and Bound

The Probabilistic Branch and Bound (PBnB) algorithm is a partitioning-based random search simulation optimization approach, which is designed for optimizing noisy and deterministic functions over mixed continuous-integer domains (Zabinsky et al. 2011; Huang and Zabinsky 2013; Zabinsky and Huang

2018). Aiming to approximate a user-defined target level set, PBnB iteratively maintains, prunes and branches sub-regions to approximate the unknown target level set with a probabilistic guarantee of the accuracy. When the algorithm terminates, it provides sub-regions with a relatively high concentration of high quality solutions, and returns an estimation of the global optima. Many simulation optimization algorithms have been proposed but most of the approaches focus on finding location estimates for the local/global solution, rather than a *set* of high quality solutions. PBnB provides a set of solutions that capture the target level set, allowing decision makers to make analyze a broader set of good solutions. A summary of PBnB is presented in Figure 4.

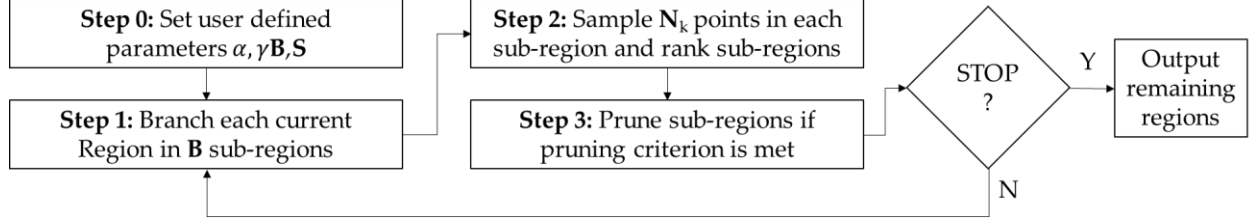


Figure 4: Algorithm Flow.

3.2 PBnB for Feasibility Determination: FSA-PBnB

The PBnB algorithm assumes the feasible region is defined by box constraints (upper and lower bounds on the decision variables) and does not allow general constraints. However, we are interested in solving the feasibility determination problem, identifying the feasibility set $\mathbb{X}_F$ inside the box constrained set $\mathbb{X}_B$:

$$\begin{aligned} \mathbb{X}_B &= \{\mathbf{x} \in \mathbb{R}^n, l_i \leq x_i \leq u_i, i = 1, \dots, n\} \\ \mathbb{X}_F &= \{\mathbf{x} \in \mathbb{X}_B, \vartheta_c^l \leq f_c(\mathbf{x}) \leq \vartheta_c^u, c = 1, \dots, C\}. \end{aligned} \quad (1)$$

Where l_i and u_i are lower and upper bounds on $x_i, i = 1, \dots, n$, $f_c(\mathbf{x})$ is the c -th constraint value (deterministic) of sample point $\mathbf{x}$, ϑ_c^l and ϑ_c^u in (1) are the lower bound and upper bound reference values for the c^{th} constraint, and C is the number of constraints. In order to identify the feasible set in (1), we propose the Feasibility Set Approximation-Probabilistic Branch and Bound (FSA-PBnB) method which uses the following metric as the main driver to indicate how far a solution is from feasibility:

$$\mathcal{D}(\mathbf{x}) = \sqrt{\sum_{c=1}^C \left[\max \left(0, f_c(\mathbf{x}) - \vartheta_c^u, \vartheta_c^l - f_c(\mathbf{x}) \right) \right]^2}. \quad (2)$$

The metric introduced in (2) is the Euclidean distance between the point $\mathbf{x}$ and the feasible frontier. If the sample point $\mathbf{x}$ is feasible, the metric $\mathcal{D}(\mathbf{x}) = 0$.

For each sampled solution $\mathbf{x}$, we measure the amount of infeasibility, and instead of just penalizing the value (as is traditional in Lagrangian-type approaches), we seek to approximate the *set* of feasible solutions. Approximating the feasible region, with a metric for infeasibility, provides more insight into the characterization of a given solution $\mathbf{x}$.

To manage the feasibility determination as a standard optimization problem, we can see the feasibility measure $\mathcal{D}(\mathbf{x})$ as an objective function. Therefore, the problem becomes a single-objective optimization problem, the goal of which is to identify the feasible set of solutions that have $\mathcal{D}(\mathbf{x}) = 0$. Before providing the details of the algorithm, Table 1 contains the main inputs to the FSA-PBnB algorithm.

Table 1: Notation of the main inputs to FSA-PBnB.

Notation	Description
α	$0 < \alpha < 1$, control the quality of the feasible set approximation
δ	Define the δ -quantile for the $\mathcal{D}(\mathbf{x})$ lower bound
B	Number of disjoint partitions to generate when branching any sub-region
$\mathbb{X}_B$	Potential solution set considering only box constraints for the control parameters

The main algorithm steps of FSA-PBnB are presented below.

Step 0. Initialization Input the user defined parameters $\alpha, \delta, B, \mathbb{X}_B$. Initialize the feasible M , infeasible P , and current U sub-region sets, and the iterate counters: $\tilde{\Sigma}_k^M = \emptyset$, $\tilde{\Sigma}_k^P = \emptyset$, $\tilde{\Sigma}_k^U = \mathbb{X}$, $k = 1$. Set $\alpha_1 = \frac{\alpha}{2}$.

Step 1. Partition. Branch each current region into B disjoint sub-regions. Denote $\tilde{\Sigma}_k^{U'}$ as the union of all newly generated sub-regions after partitioning.

Step 2. Sample and Rank. Let $N_i^k = \left\lceil \frac{\ln(\alpha_k)}{\ln(1-\delta)} \right\rceil$, and uniformly sample N_i^k points in each current sub-regions $\sigma_{i,k} \in \tilde{\Sigma}_k^{U'}$, where i is the index of the sub-region and k is the index of the algorithm iteration. Evaluate the sampled points by the distance function $\mathcal{D}(\mathbf{x}_{i,j,k})$ using (2), where $\mathbf{x}_{i,j,k} \in \sigma_{i,k}$, $i = 1, \dots, |\tilde{\Sigma}_k^{U'}|$ and for $j = 1, \dots, N_i^k$, and calculate the average distance value for $\sigma_{i,k}$ $\bar{\mathcal{D}}(\sigma_{i,k}) = \frac{\sum_{j=1}^{N_i^k} \mathcal{D}(\mathbf{x}_{i,j,k})}{N_i^k}$. Rank the current sub-regions according to $\bar{\mathcal{D}}(\sigma_{i,k})$ with $\sigma_{(i)}$ denoting the i^{th} best sub-region, that is, translate regions into feasibility rankings, $\sigma_{(1)}$ being the most feasible sub-region. $\bar{\mathcal{D}}(\sigma_{(1),k}) \leq \bar{\mathcal{D}}(\sigma_{(2),k}) \leq \dots \leq \bar{\mathcal{D}}(\sigma_{(|\tilde{\Sigma}_k^{U'}|),k})$. Rank all the sample points $\mathbf{x}_{(i),j,k} \in \sigma_{(i),k}$ according to $\mathcal{D}(\mathbf{x}_{(i),j,k})$ with $\sigma_{(i),(j),k}$ denoting the j^{th} best point in the i^{th} best sub-region, $\mathcal{D}(\mathbf{x}_{(i),(1),k}) \leq \mathcal{D}(\mathbf{x}_{(i),(2),k}) \leq \dots \leq \mathcal{D}(\mathbf{x}_{(i),(N_i^k),k})$. From the ranking, define $\sigma_{(1),k}$ as the most feasible sub-region and $\mathbf{x}_{(1),(N_i^k),k}$ as the most infeasible point in $\sigma_{(1),k}$. We arbitrarily break the ties in case we have multiple sub-regions that return the same value, e.g., $\bar{\mathcal{D}}(\sigma_{(\cdot),k}) = 0$.

Step 3. Identify Feasible Sub-Region. Define the indicator functions $\mathcal{I}_{m,k}$ for $m = 1, \dots, |\tilde{\Sigma}_k^{U'}|$, $\mathcal{I}_{m,k} = \begin{cases} 1, & \text{if } \bar{\mathcal{D}}(\sigma_{(m),k}) = 0 \\ 0, & \text{otherwise} \end{cases}$, where $\mathcal{I}_{m,k} = 1$ indicates $\sigma_{(m),k}$ can be categorized into feasible set $\tilde{\Sigma}_k^M$.

Step 4. Prune Infeasible Sub-Region

Denote the most feasible sub-region $\sigma_{(1),k}$ and the most feasible point $\mathbf{x}_{(1),(1),k}$ as σ_k^* and $\mathbf{x}_{(1),k}^*$, respectively. Define the indicator functions $\mathcal{I}_{p,k}$ for $p = 2, \dots, |\tilde{\Sigma}_k^{U'}|$, $\mathcal{I}_{p,k} =$

$\begin{cases} 1, & \text{if } \mathcal{D}(\mathbf{x}_{(p),(1),k}) > \mathcal{D}(\mathbf{x}_{(1),k}^*) \\ 0, & \text{otherwise} \end{cases}$, where $\mathcal{I}_{p,k} = 1$ indicates $\sigma_{(p),k}$ can be categorized into infeasible set $\tilde{\Sigma}_k^P$.

Step 5. Update Current Region. Update the current sub-regions which are not classified as feasible or infeasible at the k -th iteration: $\tilde{\Sigma}_{k+1}^U = \tilde{\Sigma}_k^{U'} \setminus \{\sigma_{(m),k} : \mathcal{I}_{m,k} = 1\} \cup \{\sigma_{(p),k} : \mathcal{I}_{p,k} = 1\}$. Update the set of the feasible sub-regions: $\tilde{\Sigma}_{k+1}^M = \tilde{\Sigma}_k^M \cup \{\sigma_{(m),k} : \mathcal{I}_{m,k} = 1\}$.

Terminate FSA-PBnB. If the criteria to terminate PBnB is met, output the sub-regions $\tilde{\Sigma}_{k+1}^U$ and $\tilde{\Sigma}_{k+1}^M$. Otherwise, let $\alpha_{k+1} = \frac{\alpha_k}{2}$, $k \leftarrow k + 1$, and go back to Step 1.

4 PRELIMINARY RESULTS

In the following, we present the results of FSA-PBnB implementation on one test function in different dimensions (Section 4.1) and on one simple Water Distribution Network (Section 4.2).

4.1 Test Functions

To showcase the performance of FSA-PBnB over generally constrained optimization problems, we apply FSA-PBnB to a test function with varying dimensionality. In particular, we use a sinusoidal function in different dimensions. This is a well-known single-objective optimization problem function frequently used in the global optimization literature (Ali et al. 2005), namely:

$$f_x(\mathbf{x}) = -2.5 \prod_{i=1}^n \sin\left(\frac{\pi x_i}{180}\right) - \prod_{i=1}^n \sin\left(\frac{\pi x_i}{36}\right).$$

We added the following constraints:

$$f_x(\mathbf{x}) \leq -2.3, 0 \leq x_i \leq 180, i = 1, \dots, n.$$

The global optimum is located in $\mathbf{x}^* = (90, \dots, 90)$ with a function value $f(\mathbf{x}^*) = -3.5$. FSA-PBnB algorithm parameters were: $\delta = 0.1$, $\alpha = 0.25$, $B = 3$. FSA-PBnB terminates when the number of iterations reaches the maximum number K set by the user, the values adopted are displayed in Table 2 as a function of the problem dimension. For purposes of the experiment, we also estimated R(TV), the ratio of the true volume of the feasible region $V(\mathbb{X}_F)$ to the initial volume $V(\mathbb{X}_B)$.

Table 2: Criteria to terminate the algorithm.

Dimension	2	3	4
Iteration (K)	10	13	15

For each experiment, we collected the metrics in Table 3. Figure 5 shows the approximated feasible region resulting from FSA-PBnB for the 2-dimensional sinusoidal function. The undecided regions (in orange) and the maintained regions (in blue) are the result from 10 iterations of FSA-PBnB. The inner space of the black dashed line is the true feasible set for the constraint $f_x(\mathbf{x}) \leq -2.3$. We can observe that the more inside the feasible boundary, the larger the blue rectangles are, that is, the regions are “maintained” earlier. Also, it is possible to observe how the undecided regions are mostly located at the boundaries of the feasible set, hinting that the regions around the boundaries require more iterations as more partitioning and sampling are needed to classify the regions.

Table 3: Definition of metrics for experiment results.

Metric	Definition
$\hat{\gamma}$	Estimated probability of the optimal solution being in the remaining region (remaining region: $\tilde{\Sigma}_k^U \cup \tilde{\Sigma}_k^M$)
T Pts	Total number of points sampled at termination, iteration K
R(PV)	Ratio of the pruned volume to the initial volume $V(\mathbb{X}_B)$
R(UV)	Ratio of the undecided volume to the initial volume $V(\mathbb{X}_B)$
R(MV)	Ratio of the maintained volume to the initial volume $V(\mathbb{X}_B)$
R(TV)	Ratio of the true volume to the initial volume $V(\mathbb{X}_B)$

We performed 100 macro replications for each experiment. The detailed performance is presented in Table 4. For each experiment, the top result for each metric represents the average performance obtained across the 100 macro-replications, while the values in parentheses are the coefficients of variation associated with the average performance. As an exception, $\hat{\gamma}$ does not report the variability, since this metric is the Monte Carlo estimate of the probability γ obtained by averaging the result (1 if optimum is remained 0 otherwise) across the macro-replications.

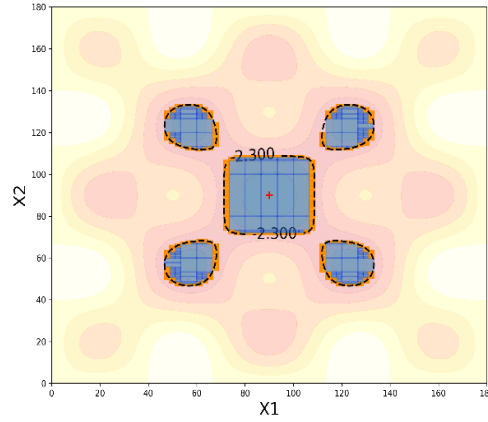


Figure 5: Result of FSA-PBnB for 2-dimensional test function.

From the numerical results it is possible to observe how, independently from the dimension of the problem, the true optimal solution is always included in the remaining region (union of the undecided and maintained sub-regions). Observing the metric R(PV), it is clear that the algorithm succeeded in pruning a large portion of the region in all the dimensional cases. Further observations can be made by observing the ratio R(Remaining)/R(TV). It is important to highlight that R(TV) was approximated by partitioning $\mathbb{X}_B$ using a grid with controlled granularity. For each hypercube generated by the grid, we evaluate the hypercube volume as feasible as long as the center point of the volume is feasible. We then compute the ratio of the volume of feasible hyper-cubes against the volume of all the hyper-cubes which is $V(\mathbb{X}_B)$.

Table 4: Results of FSA-PBnB on test function.

Dimension	T Pts	$\hat{\gamma}$	R(PV)	R(UV)	R(MV)	R(Remaining)/R(TV)
2	71.475×10^3 (14.38%)	1	90.54% (1.13%)	1.88% (3.43%)	7.58% (1.13%)	9.46% / 8.77%
3	12.591×10^5 (27.48%)	1	97.43% (0.15%)	1.33% (16.83%)	1.25% (6.92%)	2.58% / 1.92%

Dimension	T Pts	$\hat{y}$	R(PV)	R(UV)	R(MV)	R(Remaining)/R(TV)
4	47.607×10 ⁵	1	99.39	0.5%	0.1%	0.6% / 0.3%
	(20.63%)		(0.08%)	(19.57%)	(24.17%)	

From Table 4, we see that the ratio of the remaining volume against the true volume increases as the dimension of the problem grows, i.e., the algorithm accuracy is affected by the problem dimension. We also recall that R(remaining) consists of R(UV) and R(MV). By looking into the portion of R(UV) and R(MV) in R(remaining), the 2-dimensional case shows that R(MV) makes up 80% and R(UV) makes up 20% of R(remaining) at iteration 10. In the 3-dimensional case, at iteration 13, R(MV) makes up 52% and R(UV) makes up 48% of R(remaining). These ratios suggest that the majority of the feasible regions are captured in the maintained regions. However, in the 4-dimensional case, R(MV) makes up only 17% of R(remaining), that is, R(UV) may contain a large portion of infeasible solutions. This reflects the increasing complexity of the function which translates into the need for a larger effort to lead to identification.

4.2 Application to a Water Distribution Network: the Net1 Example

In this section, we present preliminary results from the application of FSA-PBnB to a Water Distribution Network (WDN) example. In order to run FSA-PBnB, we exposed the simulator as a web service and we created the connection between the simulator and algorithm, which was programmed in Python 3.6.

Preliminary results were obtained for the simple network case known as Net1 (refer to Figure 1). Results for Net1 were gathered for a 2 hour simulated time horizon, with two time slots of 1 hour each, where the variable pump speed can take on a value between zero and one for each hour, i.e., $x_i \in [0,1]$ for $i = 1,2$. The electricity tariff is set to 0.0244 [\$/kWh]. For the Net1 Example, feasibility is entirely determined by the pressure at each node of the network. EPANet tracks the pressure at each node over a time period and returns a unique pressure value for each node after every hour of simulated time. As an example, a simulation of two hours on Net1 results in a 9×2 pressure matrix, each column corresponding to an hour and each row corresponding to a node in the network. It is important, as specified within the manual of EPANet for this network, to guarantee that the minimum pressure across all nodes of the network is larger than the lower bound $\vartheta_c^l = 0$ [Pa], resulting in two constraints (one for each time slot).

To construct a pressure benchmark for the Net1 as a function of the VSPs speed settings, we ran over 10,000 regulations of the variable speed pump (uniformly sampled in the interval (0,1) for the two time-slots referring to the regulations during the first and second hour, respectively).

Formally, the constraint values of each sample point $\mathbf{x} \in [0,1] \times [0,1]$ can be defined as:

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}. \quad (3)$$

Where $f(\mathbf{x}) = \min_{n=1,\dots,N} p_{t,n}$, where $p_{t,n}$ is the pressure at the n -th node during time slot $t = 1,2$ and $N = 9$ is the number of nodes in the network. From (3), we derive the Euclidean Distance from the lower

bound $\vartheta_c^l = 0$ [Pa] as: $\mathcal{D}(\mathbf{x}) = \sqrt{\sum_{c=1}^2 ((\vartheta_c^l - f_c(\mathbf{x}))^+)^2}$.

When this aggregated pressure measure satisfies $\mathcal{D}(\mathbf{x}) = 0$, the corresponding pump speed configuration is infeasible. The $\mathcal{D}(\mathbf{x})$ projected plot resulting from the simulation study is shown in Figure 6, where “x_time_slot1” and “x_time_slot2” represent each control value set for the pump in the two hour long time-slots of $\mathbf{x} \in [0,1] \times [0,1]$, respectively. The “Pressure_distance” refers to the aggregated pressure measure. It is apparent that most of the pump configurations end up being critical ($\mathcal{D}(\mathbf{x}) > 0$) and that feasible configurations can quickly turn into bad solutions with small variations in the control decisions.

Figure 6 shows the projection of the distance measure onto $x1_time_slot1$ and $x2_time_slot2$, the blue dots refer to the configurations satisfying $\mathcal{D}(\mathbf{x}) = \mathbf{0}$ and the red dots those such that $\mathcal{D}(\mathbf{x}) > \mathbf{0}$. Only 3,876 out of 10,000 points are feasible from the simulation study. Hence, the Monte Carlo estimate of the percentage of the true feasible region to the initial box is 39%. Figure 7 and Figure 8 present the results obtained by FSA-PBnB at the 6th and 7th iteration, respectively, of FSA-PBnB. The feasible sub-regions are in blue, infeasible in red, and undecided in orange. From Figure 7 and Figure 8, it can be observed that the leftmost red rectangle is pruned at the 1st iteration, which indicates the search space is reduced by 1/3, saving substantial sampling effort over the following iterations. Also, we can see that the approximation of the feasible region, (in blue) in Figure 7, is close to the one in Figure 6. Figure 7 shows that, at iteration 6, after 3,707 simulation runs have been executed, the algorithm identifies the profile of the feasibility. The feasible region is small compared to the overall search space and the undecided regions form an orange belt which can be represented as the feasibility boundary in terms of the pressure. After the 7th iteration, the number of simulation runs grows up to 8025 but only a small part of undecided sub-regions shifts to the maintained/pruned sub-regions.

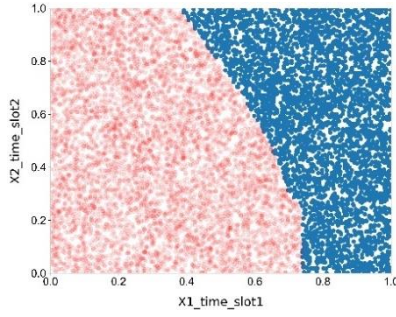


Figure 6: Projection of pressure distance.

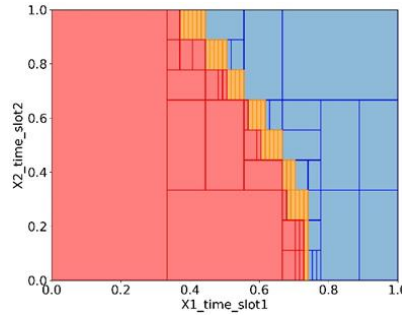


Figure 7: Approximation of feasible region (iteration=6).

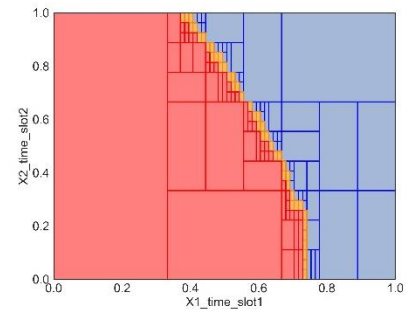


Figure 8: Approximation of feasible region (iteration=7).

Table 5 shows the quantitative results related to Figure 8. Through the comparison of R(Remaining) and R(TV), we can see that the approximated feasible set by FSA-PBnB is very close to the true feasible region.

Table 5: Experiment results on Net1.

Iteration	T Pts	R(PV)	R(MV)	R(UV)	R(Remaining)/R(TV)
6	3,707	59%	36%	5%	41% / 39%
7	8,025	60%	38%	2%	40% / 39%

5 CONCLUSIONS

We propose FSA-PBnB extending Probabilistic Branch and Bound (PBnB) for the solution of a black box feasibility determination problem. This algorithm allows us to tackle a pressing problem in the analysis of complex water distribution networks, i.e., the analysis of feasible working conditions in water networks with Variable Speed Pumps. From the results on a test function in different dimensions as well as the results from a benchmark example, the largest portion of the infeasible region is identified very early in the search thus making the approach very promising. Nevertheless, the results suggest that a better management of the budget later in the search could help in dramatically reducing the amount of simulations required. This research only considers the pressure measure to evaluate the feasibility of pump operation. However, to achieve long-term planning as well as short-term (hourly) control of networks, it is necessary to explore more feasibility measures, such as, load and supply. Also, more

effective partitioning schemes different from hyperplane partitioning are being investigated. In this direction, advanced machine learning techniques can be applied to create non linear separation planes thus possibly improving the performance of the overall algorithm. Applying the algorithm to different setups is also interesting such as investigating risk scenarios for critical infrastructures where the control dimensions are not only regulation but also external uncontrollable factors. Such an application will make the work impactful in important areas such as disaster response.

REFERENCES

- Ali, M.M., C. Khompatraporn, and Z.B. Zabinsky. 2005. "A Numerical Evaluation of Several Stochastic Algorithms on Selected Continuous Global Optimization Test Functions". *Journal of Global Optimization* 31(4):635–672.
- Balut, A., and A. Urbaniak. 2011. "Management of Water Pipeline Networks supported by Hydraulic Models and Information Systems". In *Carpathian Control Conference (ICCC)*, 2011 12th International, 16–21. IEEE.
- Bertsekas, D. P. 2014. "Constrained Optimization and Lagrange Multiplier Methods". Academic press.
- Biles, W. E., J. P. Kleijnen, W. Van Beers, and I. Van Nieuwenhuyse. 2007. "Kriging Metamodeling in Constrained Simulation Optimization: an Explorative study". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson et al., 355–362. Piscataway, New Jersey, IEEE.
- Candelieri, A., D. Soldi, D. Conti, and F. Archetti. 2014. "Analytical Leakages Localization in Water Distribution Networks through Spectral Clustering and Support Vector Machines: the ICeWater approach". *Procedia Engineering* 89, 1080–1088.
- Candelieri, A., D. Soldi, and F. Archetti. 2015a. "Cost-effective Sensors Placement and Leak Localization – The Neptun pilot of the ICeWater project". *Journal of Water Supply: Research and Technology - AQUA* 64(5):567–582.
- Candelieri, A., D. Soldi, and F. Archetti. 2015b. "Short-term Forecasting of Hourly Water Consumption by using Automatic Metering Readers Data". *Procedia Engineering* 119:844–853.
- Candelieri, A., I. Giordani, and F. Archetti. 2017. "Automatic Configuration of Kernel-based Clustering: an Optimization Approach". *Lecture Notes in Computer Science*, 10556 LNCS, 34–49.
- Candelieri, A., R. Perego, and F. Archetti. 2018. "Bayesian Optimization of Pump Operations in Water Distribution Systems". *Journal of Global Optimization* 71(1): 213–235.
- Chase, D. V., and L.E. Ormsbee. 1989. "Optimal Pump Operation of Water Distribution Systems with Multiple Storage Tanks". In *Water Resources Planning and Management* 733–736.
- Copeland, C., and N. Carter. 2017. "Energy-Water Nexus: The Water Sector's Energy Use". *Congressional Research Service Report R43200*.
- De Paola, F., N. Fontana, M. Giugni, G. Marini, and F. Pugliese. 2016. "An Application of the Harmony-Search Multi-Objective (HSMO) Optimization Algorithm for the solution of Pump Scheduling Problem". *Procedia Engineering* 162:494–502.
- Huang, H., and Z.B. Zabinsky. 2013. "Adaptive Probabilistic Branch and Bound with Confidence Intervals for Level Set Approximation," In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy et al., 980–991. Piscataway, New Jersey: IEEE.
- Mala-Jetmarova, H., N. Sultanova, and D. Savic. 2017. "Lost in Optimisation of Water Distribution Systems? A Literature Review of System Operation". *Environmental Modelling & Software* 93:209–254.
- Nicklow, J., P. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld, A. Singh, and E. Zechman. 2009. "State of the Art for Genetic Algorithms and beyond in Water Resources Planning and Management". *Journal of Water Resources Planning and Management*, 136(4):412–432.
- Pasha, M. F. K., and K. Lansey. 2009. "Optimal Pump Scheduling by Linear Programming". In *World Environmental and Water Resources Congress: Great Rivers* (pp. 1-10).

- Paluszczyzyn, D., P. Skworcow, and B. Ulanicki. 2015. "Modelling and Simulation of Water Distribution Systems with Quantised State System Methods". *Procedia Engineering* 119:554–563.
- Rossman, L. A. 2000. "EPANet 2 Users Manual". Water Supply and Water Resources Division, National Risk Management Research Laboratory. Cincinnati, United States Environmental Protection Agency.
- Savic, D. A., G.A. Walters, and M. Schwab. 1997. "Multiobjective Genetic Algorithms for Pump Scheduling in Water Supply". In *AISB international workshop on evolutionary computing*, 227-235. Springer, Berlin, Heidelberg.
- Schmid, R. 2002. "Review of Modeling Software for Piped Distribution Networks". Skat Foundation. St. Gallen.
- Sterling, M. J. H., and B. Coulbeck. 1975. "A Dynamic Programming solution to Optimization of Pumping Costs". *Proceedings of the Institution of Civil Engineers* 59(4):813-818.
- Szechtman, R., and E. Yucesan. 2008. "A new Perspective on Feasibility Determination". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S.J. Mason et al., 273-280. Piscataway, New Jersey, IEEE.
- Zabinsky, Z. B., W. Wang, Y. Prasetio, A. Ghate, and J.W. Yen. 2011. "Adaptive Probabilistic Branch and Bound for Level Set Approximation". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain et al., 4151-4162. Piscataway, New Jersey, IEEE.
- Zabinsky, Z. B., and H. Huang. 2018. "A Partition-based Optimization Approach for Level Set Approximation: Probabilistic Branch and Bound." An invited book chapter to be published in, *Women in Industrial and Systems Engineering: Key Advances and Perspectives on Emerging Topics*, Alice E. Smith, editor, Springer Nature.

AUTHOR BIOGRAPHIES

YI-AN TSAI got her master degree in Industrial Engineering at Arizona State University on water systems and operations research. Her email address is: ytsai43@asu.edu.

RICCARDO PEREGO is a PhD student at Milano Bicocca pursuing a Degree in Computer Science focused on Machine Learning and Optimization. His email address is: r.perego12@campus.unimib.it

GIULIA PEDRIELLI is an Assistant Professor for the School of Computing, Informatics and Decision Sciences Engineering at Arizona State University, US. Her research is in learning for simulation and simulation optimization. Her email address is: giulia.pedrielli@asu.edu

HAO HUANG is an Assistant Professor in the Department of Industrial Engineering and Management at the Yuan Ze University at Taoyuan, Taiwan. His research interests include simulation optimization and healthcare applications. His e-mail address is: haohaung@saturn.yzu.edu.tw.

ZELDA B. ZABINSKY is a Professor in the Department of Industrial and Systems Engineering at the University of Washington. She is an IIE Fellow, on the editorial board of the Journal of Global Optimization, and a board member of the Pacific Institute of Mathematical Sciences (PIMS). Her research interests are in global optimization under uncertainty for complex systems. Her email address is: zelda@u.washington.edu.

ANTONIO CANDELIERI is an Assistant Professor at University of Milano-Bicocca in the department of Computer Science, Systems and Communication. One of his research activity has focused on Machine Learning and Optimization in Smart Water Distribution Networks. His email address is: antonio.candelieri@unimib.it

LOGAN MATHESSEN is a second year PhD student at Arizona State University pursuing a Degree in Industrial Engineering focused in Operations Research and Industrial Statistics. Two research interests are: efficient global optimization, and design of experiments for statistical modeling. His email address is: lmathese@asu.edu.