# On Solving the Capacitated Open Vehicle Routing Problems with Uncertainty Demand

**Hotman Simbolon**[*]
Department of Mathematics,
HKBP Nommensen University, Siantar Indonesia

**Herman Mawengkang**
Departments of Mahtematics,
University of Sumatera Utara, Medan Indonesia

*Abstract— In open vehicle routing problems, the vehicles are not required to return to the depot after completing service. In this paper, we extend the problem to be more realistic by including the uncertainty of customer demands. Each customer has a demand and each customer must be serviced by a single vehicle and no vehicle may serve a set of customers whose total demand exceeds its capacity. Each vehicle route must start at the depot and end at the last customer it serves. The objective is to define the set of vehicle routes that minimizes the total costs. We solve the stochastic model using a strategy of releasing nonbasic variables from their bounds, combined with the "active constraint" method. This strategy is used to force the appropriate non-integer basic variables to move to their neighborhood integer points. A study of criteria for choosing a nonbasic variable to work with in the integerizing strategy has also been made.*

*Keywords— vehicle routing, separation, stochastic programminng*

## I. INTRODUCTION

The standard version of Vehicle Routing Problems (VRPs), it is defied that the vehicles are required to return to the depot after completing service (see for example [26]). In some instances the vehicles involved are not required to return to the depot. This kind of version is called open VRP. From graph stands of view, we can say that the vehicle routes are not closed paths but open ones, starting at the depot and ending at one of the customers.

At first sight, having open routes instead of closed ones looks like a minor modification. Indeed, if travel costs are asymmetric, there is essentially no difference between the open and closed versions: to transform the open version into the closed one, it suffices to set the cost to zero for traveling from any customer to the depot. However, if travel costs are symmetric, things are more subtle. Indeed, we prove in the next section that, somewhat surprisingly, the open version turns out to be more general than the closed one, in the sense that any closed VRP on n customers can be transformed into an open VRP on *n* customers, but there is no transformation in the reverse direction.

This sort of VRP version naturally can occur, for example, when a company does not own a vehicle fleet and all its deliveries from a central depot are undertaken by hired vehicles that are not obliged to return to the depot. In such situations, the cost of the distribution may be proportional to the distance travelled while loaded. A practical case study of this type is described in [34] and [35]. The same model can also be used for pick-ups, where vehicles start empty at any customer and must pick up the demands of each customer on their route and deliver them to the depot.

There are also applications where the vehicles start at the depot, deliver to a set of customers and then are required to visit the customers in reverse order, picking up items that are required to be backhauled to the depot. If, for each customer, the pick-up demand is no larger than the delivery demand, then an open VRP model can be used. An application of this type for an air express courier is mentioned by [32] in an early article describing features of practical routing problems.

Two further areas of application are described by [11]. The first involves the planning of train services, starting or ending at the Channel Tunnel. The second involves planning a set of school bus routes where in the morning pupils are picked up at various locations and brought to school, and in the afternoon, the routes are reversed to take pupils home. That includes a description of a problem of express airmail distribution in the USA that is essentially an open pick-up and delivery VRP with capacity constraints and time windows.

Open VRPs are easily seen to be strongly NP-hard by reduction from the Hamiltonian path problem. Research on open VRPs has therefore up to now concentrated on devising effective heuristics for solving them. For the version involving only capacity constraints, [31] presented two-phase heuristic involving minimum spanning trees, [34] present a population-based heuristic, and [35] present a heuristic of the threshold-accepting type. For a more general variant involving both capacity and route-length constraints, [4] and [11], [12] describe tabu search heuristics, [14] present a record-to-record travel heuristic, and [24] present an adaptive neighborhood search heuristic. Heuristics have also been devised for open VRPs with other kinds of constraints; see for example [29] and [2].

The classical Vehicle Routing Problem (VRP) determines the optimal set of routes used by a fleet of vehicles to serve a given set of customers on a predefined graph; it aims at minimizing the total travel cost (proportional to the travel times or distances) and operational cost (proportional to the number of vehicles used). The Stochastic VRP (SVRP) arises whenever some parameters of the VRP are random (e.g. demand and travel time).

In this paper we present the *Capacitated Open Vehicle Routing Problem* (COVRP), which is defined as follows. A complete undirected graph $G = (V, E)$ is given, with $V = \{0, \ldots, n\}$. Vertex 0 represents the depot, the other vertices represent customers. The cost of travel from vertex $i$ to vertex $j$ is denoted by $c_{ij}$, and we assume costs are symmetric, so $c_{ij} = c_{ji}$. A fleet of K identical vehicles, each of capacity $Q > 0$, is given. Each customer $i$ has a demand $q_i$, with $0 < q_i \leq Q$. Each customer must be serviced by a single vehicle and no vehicle may serve a set of customers whose total demand exceeds its capacity. Each vehicle route must start at the depot and end at the last customer it serves. The objective is to define the set of vehicle routes that minimizes the total costs.

## II.    MINIMUM UNMET DEMAND ROUTES

We now formulate our problem into a mixed integer programming (MIP) model. We first introduce the notation used and formulate the deterministic version of the problem. We then compare different uncertainty models for this problem.

### A. Notation

We consider a set *K* of vehicles and a set *D* of demand nodes. We identify an additional node, node 0, as the supply node (depot) and let $C = D \cup \{node\ 0\}$ represent the set of all nodes. Indexed on sets *K* and *C*, we define the following deterministic parameters:

$n$ :       initial number of vehicles at the supply node (depot)
$s$ :       amount of supplies at the supply node (depot)
$c$k :       load capacity of vehicle $k$
$dl$i :       service deadline at demand node $i$.

We use *M* as a large constant used to express nonlinear relationships through linear constraints. We also consider the following two parameters to represent the uncertain travel time and demand, respectively

$\tau_{i,j,k}$ :       time required to traverse arc$(i, j)$ for vehicle $k$
$\zeta_i$ :       amount of demand needed at node $i$.

Finally, we define the binary and non-negative decision variables as follows, indexed on sets *K*, *C*:

Binary:
$X_{i,j,k}$ :       flow variables, equal to 1 if $(i, j)$ is traversed by vehicle k and 0 otherwise
$S_{i,k}$ :       service variables, equals to 1 if node $i$ can be serviced by vehicle $k$
Non-negative:
$Y_{i,j,k}$ :       amount of commodity traversing arc$(i, j)$ using vehicle $k$
$U_i$ :       amount of unsatisfied demand of commodity at node $i$
$T_{i,k}$ :       visit time at node $i$ of vehicle $k$
$\delta_{i,k}$ :       delay incurred by vehicle k in servicing $i$.

### B. Deterministic Model

The deterministic, minimize unmet demand problem can be expressed as follows

DP :
$$\text{minimize } \sum_{i \in D} U_i + \kappa \sum_{i \in D, k \in K} T_{i,k}$$
$$\text{subject to constraints } (1) - (17),$$

where the constraints are explained in detail below.

The objective of model DP is to minimize the weighted sum of the total unmet demands over all demand nodes and the total visit time at demand nodes of all vehicles. The $\kappa$ value usually is set to be very small to make the total travel time a secondary objective compared with the unmet demand quantity. However, the travel time is a necessary term in the objective function to guide the route generation after the deadline. Since we model the routing problem in response to a large-scale emergency, the service start times (arrival times) directly associate with when the supply will be shipped and used at the dispensing sites. We would like to serve the dispensing sites as early as possible for life-saving purposes, so the arrival time is a much more important indicator of the service quality than the conventional objectives such as travel times or operational time.

We group the constraints into four parts: route feasibility constraints, time constraints, demand flow constraints and node service constraints. The following constraints (1)-(6) characterize the vehicle flows on the path and enforce the route feasibility.

$$\sum_{i \in D} \sum_{k \in K} X_{0,i,k} \leq n \tag{1}$$

$$\sum_{i \in D} \sum_{k \in K} X_{i,0,k} \leq n \tag{2}$$

$$\sum_{j \in D} X_{0,j,k} = \sum_{k \in K} X_{j,0,k} = 1 \qquad (\forall k \in K) \tag{3}$$

$$\sum_{j \in C} \sum_{k \in K} X_{i,j,k} = 1 \qquad (\forall i \in D) \tag{4}$$

$$\sum_{j \in C} \sum_{k \in K} X_{j,i,k} = 1 \qquad (\forall i \in D) \tag{5}$$

$$\sum_{j \in C} X_{i,j,k} = \sum_{j \in C} X_{j,i,k} \qquad (\forall i \in D \quad k \in K) \qquad (6)$$

Constraints (1) and (2) specify that the number of vehicles to service must not exceed the available quantity ready at the supply node at the beginning of the planning horizon. The number of vehicles to service is stated by the total number of vehicles flowing from and back to the depot. Constraint (3) represents each vehicle flow from and back to the depot only once. Constraints (4) and (5) state that each demand node must be visited only once. Constraint (6) requires that all vehicles who flow into a demand point must flow out of it.

Constraints (7)-(10) guarantee schedule feasibility with respect to time considerations.

$$T_{0,k} = 0 \qquad (\forall k \in K) \qquad (7)$$

$$(T_{i,k} + \tau_{i,j,k} - T_{j,k}) \leq (1 - X_{i,j,k})M \qquad (\forall i, j \in C \quad k \in K) \qquad (8)$$

$$0 \leq T_{i,k} - \delta_{i,k} \leq \sum_{j \in C} X_{i,j,k} \qquad (\forall i \in D \quad k \in K) \qquad (9)$$

$$0 \leq T_{i,k} - \delta_{i,k} \leq dl_i \sum_{j \in C} X_{i,j,k} \qquad (\forall i \in D \quad k \in K) \qquad (10)$$

The fact that all vehicles leave the depot at time 0 is specified by constraint (7). Constraint (8) enforces the time continuity based on the node visiting sequence of a route. Constraint (9) sets the visit time to be zero if the vehicle does not pass a node. The variable $\delta_{i,k}$ represents the delay of the visit time if a vehicle reaches the node after the deadline and is set to zero if it arrives before the deadline in constraint (10).

This model primarily accommodates the emergency situation where late deliveries could lead to fatalities. To maximize the likelihood of saving lives, medication should be received by the affected population within the specified hours of the onset of symptoms to impact the patient survival. This is the rationale behind the preference of using a hard deadline constraint instead of the soft deadline. However, for problems where late deliveries are possible we can translate the proposed model to soft deadlines, having the penalty on the violation represent the worsening in patient condition due to late arrival.

Constraints (11)-(13) state node service constraints.

$$\delta_{i,k} \leq (1 - S_{i,k})M \qquad (\forall i \in D \quad k \in K) \qquad (11)$$

$$S_{i,k} \leq \sum_{j \in C} X_{i,j,k} \qquad (\forall i \in D, \quad k \in K) \qquad (12)$$

$$S_{i,k}M \geq \left( \sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k} \right) \qquad (\forall i \in D, \quad k \in K) \qquad (13)$$

Binary decision variables $S_{i,k}$ are used to indicate whether a node *i* can be serviced by vehicle *k* (when it equals to 1). That is, if the vehicle *k* visits node *i* before the deadline, then the vehicle can drop off some commodities at this node. However, the vehicle does not necessarily do it when $S_{i,k}$ equals to 1 since there might not be enough supply at the depot so the vehicle may not carry any commodities when it visits a later node in the route. We use these binary variables to keep the feasible region of this problem non-empty all the time. Constraints (4) and (5) will still enforce each node to be visited once and only once no matter before or after the deadline; however, those visits after the deadline cannot service the node any more. Constraint (11) states the deadline constraint and it can only be violated when $S_{i,k}$ equals to zero. Constraint (12) illustrates the relationship between the binary flow variables and the binary service variables. It implies the service variable can only be true when a vehicle physically passes a node. Constraint (13) requires that no commodity flows in a node after the deadline. On the other hand, there is no compulsory dropping-off commodities at nodes visited before the deadline since there may not be enough supplies to meet the demand. It establishes the connection between the commodity flow and the vehicle flow.

Constraints (14)-(16) state the construction on the demand flows.

$$s - \sum_{k \in K} \left[ \sum_{j \in C} Y_{0,j,k} - \sum_{j \in C} Y_{j,o,k} \right] \geq 0 \qquad (14)$$

$$\sum_{k \in K} \left[ \sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k} \right] + U_i - \zeta_i \geq 0 \qquad (\forall i \in D) \qquad (15)$$

$$X_{i,j,k} c_k \geq Y_{i,j,k} \qquad (\forall i, j \in C, \quad k \in K) \qquad (16)$$

Constraint (14) requires the total shipment of commodity from the depot not exceeding its current supply inventory level. Constraint (15) enforces the balanced material flow requirement for the demand nodes.
Constraint (16) allows the flow of commodities as long as there is sufficient vehicle capacity. It also connects the commodity flow and the vehicle flow.

$$X_{i,j,k}, S_{i,j} \text{ binary}; \quad Y_{i,j,k} \geq 0; \quad U_i \geq 0; \quad T_{i,k} \geq 0; \quad \delta_{i,k} \geq 0 \qquad (17)$$

Constraint (17) states the binary and non-negativity properties of the decision variables.

### C. Stochastic Model

The parameters $\tau_{i,j,k}$ in constraint (8) and $\zeta_i$ in constraint (15) represent the uncertain travel time and demand parameters of our problem, respectively. If we ignore the uncertainty and replace these random quantities by representative values, such as their mean $\mu\tau_{i,j,k}$ and $\mu\zeta_i$ or mode values, we can solve a deterministic problem DP to obtain a simple solution for this problem. This deterministic solution will be helpful as a benchmark to compare the quality of routes and demonstrate the merits of other more sophisticated methods we discuss next. There are two other ways to handle uncertainty that for this problem lead to the solution of a single deterministic problem DP: chance constrained programming and robust optimization. The solution of this routing problem through other methods of representing uncertainty, such as stochastic programming and markov-decision processes require more involved solution procedures and will not be explored in this paper.

In chance constrained programming (CCP) we assume that the parameters $\tau_{i,j,k}$ and $\zeta_i$ are unknown at the time of planning but follow some known probability distributions. We assume they are uniformly and independently distributed. We let $\alpha_D$ and $\alpha_T$ represent the confidence level of the chance constraints defining the unmet demand at each node and the arrival time of each vehicle at each node respectively. Thus, the constraints with stochastic parameters must hold with these given probabilities. For a given distribution on $\tau_{i,j,k}$ and $\zeta_i$, we can rewrite constraint (8) and constraint (15) in the chance constrained fashion with levels $\alpha_T$ and $\alpha_D$ as follows:

$$P\{\tau \big| (T_{i,k} + \tau_{i,j,k} - T_{j,k}) \le (1 - X_{i,j,k})M\} \ge 1 - \alpha_T \quad (\forall i, j \in C \quad k \in K) \tag{18}$$

$$P\left\{\zeta \big| \sum_{k \in K}\left[\sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k}\right] + U_i - \zeta_i \ge 0\right\} \ge 1 - \alpha_D \quad (\forall i \in D) \tag{19}$$

We call this chance constrained model (CCP model), which is modified based on the DP model in section 3.2, by replacing constraints (8) and (15) with constraints (18) and (19). Under some assumption of their distribution, constraint (18) and constraint (19) can be transformed to their deterministic counterpart. From this point onward in this paragraph, we use short notation $\tau$ and $\zeta$ to substitute $\tau_{i,j,k}$ and $\zeta_i$ for simplicity. For example, we assume $\tau$ and $\zeta$ follow a lognormal distribution with mean $\mu_\tau$ and standard deviation $\sigma_\tau$ and mean $\mu_\zeta$ and standard deviation $\sigma_\zeta$ respectively. The logarithm $\log(\tau)$, $\log(\zeta)$ are normally distributed as normal($\mu_\tau', \sigma_\tau'$) and normal($\mu_\zeta', \sigma_\zeta'$). The relationship between the parameters of lognormal distribution and normal distribution is stated as: $\mu' = \log\mu - \frac{1}{2}\sigma'^2$, $\sigma'^2 = \log(\frac{\mu^2 + \sigma^2}{\mu^2})$. We let $\kappa_T$ and $\kappa_D$ represent the $Z$ value for the normal distribution corresponding to the confidence level $\alpha_T$ and $\alpha_D$ and we call them "safety factors" in the later experimental results section. Therefore, the deterministic counterpart of constraint (18) and constraint (19) can be expressed as:

$$(T_{i,k} + e^{\mu'_{\tau_{i,j,k}} + \kappa_T \sigma'_{\tau_{i,j,k}}} - T_{j,k}) \le (1 - X_{i,j,k})M \quad (\forall i, j \in C \quad k \in K) \tag{20}$$

$$\sum_{k \in K}\left[\sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k}\right] + U_i \ge e^{\mu'_{\zeta_i} + \kappa_D \sigma'_{\zeta_i}} \quad (\forall i \in D) \tag{21}$$

## III. CHANCE-CONSTRAINED PROGRAMMING

A large class of optimization problems arising from important planning and design applications in uncertain environments involve service level or reliability constraints. Consider, for example, the problem of locating service centers for responding to medical emergencies. Requiring 100% coverage over all possible emergency scenarios is physically and economically impractical and so typically emergency preparedness plans calls for some minimum response reliability [1, 4]. Service level agreements in telecommunication contracts require network providers to guarantee, with high probability, that packet losses will not exceed a certain percentage [19, 37]. In financial portfolio planning, investors often require that, with high probability, portfolio losses do not exceed some threshold (value-at-risk) while maximizing expected returns [13, 23]. Mathematical models for planning/designing reliability constrained systems such as these lead to optimization with chance constraints or probabilistic constraints.

A generic chance-constrained optimization problem can be formulated as

$$\min_{x \in X} f(x) \text{ subject to } \Pr\{G(x, \xi) \le 0\} \ge 1 - \varepsilon, \tag{22}$$

where $X \subset \square^n$ represents a deterministic feasible region, $f : \square^n \to \square$ represents the objective to be minimized, $\xi$ is a random vector whose probability distribution is supported on set $\Xi \subset \square^n$, $G : \square^n \times \square^d \to \square^m$ is a constraint mapping, 0 is an $m$-dimensional vector of zeroes, and $\varepsilon \in (0, 1)$ is a given risk parameter (significance level). Formulation (22) seeks a decision vector $x$ from the feasible set $X$ that minimizes the function $f(x)$ while satisfying the chance constraint $G(x, \xi) \le 0$ with probability at least $1 - \varepsilon$. It is assumed that the probability distribution of $\xi$ is known.

By way of illustration, consider the following simple facility sizing example. We need to decide capacities of $n$ facilities servicing uncertain customer demand. The cost-per-unit capacity installed for each facility is given, as is the joint demand distribution. The goal is to determine the cheapest capacity configuration so as to guarantee that the installed capacity exceeds demand with probability $1 - \varepsilon$. This chance-constrained problem can be formulated as follows.

$$\min_{x \geq 0} \sum_{i=1}^{n} c_i x_i \quad \text{subject to} \quad \Pr\{\xi i - xi \leq 0, \ i = 1,...,n\} \geq 1 - \varepsilon. \tag{23}$$

Here $x_i$, $c_i$ and $\xi_i$ denote the capacity, cost, and random demand for facility $i$, respectively. It is assumed that the (joint) probability distribution of the random vector $\xi = (\xi_1, ..., \xi_n)$ is known (otherwise the probabilistic constraint in (23) is not defined). Note that the probabilistic (chance) constraint of (23) can be considerably weaker than trying to satisfy the demand for all possible realizations of $\xi$. Note also that (23) is an example of (22) with $G(x, \xi) = \xi - x$.

In this example, we require that the reliability requirement be applied to all facilities jointly. One could also consider the individual chance constraints $\Pr\{\xi_i \leq x_i\} \geq 1 - \varepsilon_i$, $i = 1, ..., n$, applied to each facility separately. This leads to a much simpler problem, since $\Pr\{\xi_i \leq x_i\} \geq 1 - \varepsilon_i$ is equivalent to $F_i^{-1}(x_i) \geq 1 - \varepsilon_i$, where $F_i$ is the cumulative distribution function (cdf) of $\xi_i$. Note, however, that in order to ensure the joint chance constraint by enforcing the individual chance constraints, the corresponding risk parameters $\varepsilon_i$ should be considerably smaller than especially when $n$ is large.

Beginning with the work of Charnes, Cooper and Symonds [8], chance-constrained stochastic programs have been extensively studied. In addition to the facility location, telecommunication and finance examples cited earlier, chance constrained models have been used in numerous other applications including production planning [20, 17], chemical processing [15, 16] and water resources management [28, 33]. See [27] for background and an extensive list of references. Despite important theoretical progress and practical importance, chance-constrained stochastic problems of the form (22) are still largely intractable except for some special cases. There are two primary reasons for this difficulty.

1. In general, for a given $x \in X$, computing $\Pr\{G(x, \xi)\} \leq 0$ accurately, i.e., checking whether x is feasible to (1), can be hard. In multidimensional situations this involves calculation of a multivariate integral which typically cannot be computed with a high accuracy.

2. The feasible region defined by a chance constraint generally is not convex even if $G(x, \xi)$ is convex in $x$ for every possible realization of $\xi$. This implies that even if checking feasibility is easy, optimization of the problem remains difficult. For example, the facility sizing example (23) with $n$ facilities and $m$ equiprobable realizations of the demand vector is equivalent to a maximum clique problem on a graph with n nodes and m edges, and is therefore strongly NP-hard, [18].

In light of the above difficulties, existing approaches for chance-constrained stochastic programs can be classified as follows. First are the approaches for problems where both difficulties are absent, i.e., the distribution of $\xi$ is such that checking feasibility is easy, and the resulting feasible region is convex. A classical example of this case is when $G(x, \xi) = v - \xi^T x$ and $\xi$ has a multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma$. Then for $\varepsilon \in (0, 0.5)$,

$$\left\{ x \in \square^n : \Pr\{\xi^T x \geq v\} \geq 1 - \varepsilon \right\} = \left\{ x \in \square^n : v - \mu^T x + z_\varepsilon \sqrt{x^T \Sigma x} \leq 0 \right\},$$

where $z_\varepsilon = \Phi^{-1}(1 - \varepsilon)$ is the $(1 - \varepsilon)$-quantile of the standard normal distribution. In this case, under convexity of $X$, the chance-constrained problem reduces to a deterministic convex optimization problem. The second class of approaches are for problems where only the second difficulty is absent, i.e., the feasible region of the chance constraint is guaranteed to be convex. The best known example of this case is when $G(x, \xi) = \xi - Ax$, where A is a deterministic matrix and $\xi$ has a log-concave distribution. In this case the chance constraint feasible set is convex [25]. However it may still be difficult to compute $\Pr\{G(x, \xi)\} \leq 0$ exactly. Solution methods in this class are primarily based on classical nonlinear programming techniques adapted with suitable approximations of the chance constraint function and its gradients (see [26]). The third class of approaches are for problem where the first difficulty is absent, i.e., computing $\Pr\{G(x, \xi)\} \leq 0$ is easy, e.g., when $\xi$ has a finite distribution with a modest number of realizations (in this case the feasible region is typically non-convex). A number of approaches based on integer programming and global optimization have been developed for this class of problems [9, 10, 30]. Finally, more recently, a number of approaches have been proposed to deal with both difficulties [6, 7, 21, 22, 3]. The common theme in these approaches is that they all propose convex approximations of the non-convex chance constraint that yield solutions which are feasible, or at least highly likely to be feasible, to the original problem. Thus the difficulty of checking feasibility as well as non-convexity is avoided. Unfortunately, often, the solutions produced by these approaches are quite conservative.

In this paper we consider an approximation of the chance constraint problem (22) where the true distribution of $\xi$ is replaced by an empirical distribution corresponding to a Monte Carlo sample. The resulting sample average approximation problem can be used to provide good candidate solutions along with optimality gap estimates. The sampled approximation problem is a chance-constrained problem with a discrete distribution and can be quite difficult. We discuss integer programming based approaches for solving it.

## IV.    SAMPLE AVERAGE APPROXIMATION

In order to simplify the presentation we assume, without loss of generality, that the constraint function $G : \square^n \times \square^d \to \square$ in (22) is scalar valued. Of course, a number of constraints $Gi(x, \xi) \leq 0$, $i = 1, ..., m$, can be equivalently replaced by one constraint $G(x, \xi) := \max_{1 \leq i \leq m} G_i(x, \xi) \leq 0$. The chance-constrained stochastic program (22) can be rewritten as

$$\min_{x \in X} f(x) \text{ subject to } q(x) \leq \varepsilon \tag{24}$$

where $q(x) := \Pr\{G(x, \xi) > 0\}$

Now let $\xi^1, \dots \xi^N$ be an independent identicaly distributed (iid) sample of $N$ realizations of random vector $\xi$. Given $x \in X$ let

$$\hat{q}_N(x) := N^{-1} \sum_{j=1}^{N} 1_{(0,\infty)}(G(x, \xi^j)),$$

where $1_{(0,\infty)} : \square \to \square$ is the indicator function of $(0, \infty)$. That is, $\hat{q}_N(x)$ is equal to the proportion of realizations with $G(x, \xi^j) > 0$ in the sample. For some given $\gamma \in (0, 1)$ consider the following optimization problem associated with a sample $\xi^1, \dots, \xi^N$,

$$\min_{x \in X} f(x) \text{ subject to } \hat{q}_N(x) \leq \gamma \tag{25}$$

We refer to problems (24) and (25) as the true and sampled average approximate (SAA) problems, respectively, at the respective risk levels $\varepsilon$ and $\gamma$.

The SAA problem is a chance-constrained stochastic problem with a different (discrete) distribution and a different risk level than (24). Unless $N$ is prohibitively large, the chance-constrained problem SAA does not suffer from the first difficulty (computing $\hat{q}_N(x)$ mentioned in Section 1, however it may still be difficult to solve. Assuming we have a scheme for solving SAA, what can we say about an optimal solution and the optimal value of SAA in relation to that of the true problem (24)? Intuitively, assuming $N$ is large enough, if $\gamma \leq \varepsilon$, then a feasible solution of SAA is likely to be feasible to the true problem, and $\gamma \geq \varepsilon$ if then the optimal value of SAA is likely to be a lower bound to that of the true problem. Thus the SAA problem can be used to obtain both candidate feasible solutions to the true problem as well as optimality gap estimates. Next we discuss these concepts slightly more rigorously.

We assume that X is compact, $f(.)$ is continuous, $G(x, .)$ is measurable for every $x \in \square^n$, and $G(\cdot, \xi)$ is continuous for almost every . Then the functions $q(x)$ and $\hat{q}_N(x)$ are lower-semicontinuous, and the true problem (24) and the SAA problem (25) are guaranteed to have optimal solutions if they are feasible. Let $X^*(\varepsilon)$ and $\hat{X}_N(\gamma)$ denote the set of optimal solutions of the true and SAA problems, respectively, $v(\varepsilon)$ and $\hat{v}_N(\gamma)$ denote the optimal value of the true and SAA problems, respectively.

## V. SOLVING SAMPLE APPROXIMATIONS

We have seen that we can generate as well as validate candidate solutions to the chance constrained problem (24) by solving (several) sampled approximations (4). In this section we explore approaches for solving these problems.

If $\gamma = 0$ then the SAA problem reduces to

$$\min_{x \in X} f(x) \text{ subject to } G(x, \xi^j) \leq 0, \quad j = 1, \dots, N \tag{26}$$

When the functions $f(\cdot)$ and $G(\cdot, \xi^j)$ for $j = 1, \dots, N$ are convex (linear) and the set $X$ is convex (polyhedral) then (26) is a convex (linear) program, and can usually be solved efficiently using off-the-shelf software. We can then consider increasing the risk level $\gamma$ in the SAA problem. However with $\gamma > 0$ the SAA problem is a chance constrained optimization problem (with a finite distribution) and is NP-hard even in very simple settings [18]. A wide variety of approaches have been proposed to solve different classes of chance-constrained optimization problems under finite distributions (cf. [9, 10, 27] and references therein). In this tutorial we consider an integer programming based approach.

The SAA problem (25) can be formulated as the following mixed-integer problem (MIP)

$$\min \quad f(x)$$

subhect to

$$G(x, \xi^j) \leq M_j z_j \quad j = 1, \dots, N$$

$$\sum_{j=1}^{N} z_j \leq \gamma N \tag{27}$$

$$z_j \in \{0, 1\}$$

$$x \in X$$

where $z_j$ is a binary variables and $M_j$ is a large positive number such that $M_j \geq \max_{x \in X} G(x, \xi^j)$ for all $j = 1, \dots, N$. Note that if $z_j$ is 0 then the constraint $G(x, \xi^j) \leq 0$ corresponding to the realization $j$ in the sample is enforced. On the other hand $z_j = 1$ does not pose any restriction on $G(x, \xi^j)$. The cardinality constraint $\sum_{j=1}^{N} z_j \leq \gamma N$ requires that at least N of the N constraints $G(x, \xi^j) \leq 0$ for $j = 1, \dots, N$ are enforced.

Even in a linear setting (i.e., the functions *f* and *G* are linear in *x* and the set *X* is polyhedral) moderate sized instances of the MIP (10) are typically very difficult to solve as-is by state-of-the-art MIP solvers. The difficulty is due to the fact that the continuous relaxation of (27) (obtained by dropping the integrality restriction on the *z* variables) provides a weak relaxation, and hence slows down the branch-and-bound algorithm that is the work-horse of MIP solvers. This difficulty can be alleviated by strengthening the formulation (27) by addition of valid inequalities or reformulation. Such improved formulations have tighter continuous relaxation gaps and can serve to significantly cut down solve times.

A variety of approaches for strengthening special classes of the MIP (27) have been proposed. Here we discuss an approach for the case of joint probabilistic constraints where the uncertain parameters only appear on the right-hand side, i.e.,

$$G(x, \xi) = \max_{i=1,\ldots,m} \{\xi_i - G_i(x)\}$$

Note that the facility sizing example (23) is of this form. By appropriately translating, we assume that $\xi_j^i \geq 0$ for all *i* and *j*. The MIP (27) can then be written as

$$\min \quad f(x)$$

$$\text{subject to}$$

$$G_i(x) \geq v_i \qquad i = 1,\ldots,m$$

$$v_i + \xi_i^j z_j \geq \xi_i^j \quad i = 1,\ldots,m, \quad j = 1,\ldots,N \tag{28}$$

$$\Sigma_{j=1}^N z_j \leq \gamma N$$

$$z_j \in \{0,1\} \qquad j = 1,\ldots,N$$

$$x \in X, \quad v_i \geq 0 \quad i = 1,\ldots,m$$

Note that we have introduced the auxiliary variables $v_i$ for *i* = 1, …, *m* to conveniently represent $G_i(x)$. As before, if $z_j$ is 0 then the constraints $G_i(x) \geq \xi_i^j$ for *i* = 1, …, *m* corresponding to the realization *j* in the sample is enforced.

## VI. THE ALGORITHM

Let $x = [x] + f, \ 0 \leq f \leq 1$ be the (continuous) solution of the relaxed problem, $[x]$ is the integer component of non-integer variable $x$ and $f$ is the fractional component.

Step 1.  Get row $i^*$ the smallest integer infeasibility, such that
$\delta_{i^*} = min\{f_i, 1 - f_i\}$

Step 2.  Calculate
$v_{i^*}^T = e_{i^*}^T B^{-1}$
This is a pricing operation.

Step 3.  Calculate $\sigma_{ij} = v_{i^*}^T \cdot a_j$

With $j$ corresponds to $\min_j \left\{ \left| \frac{d_j}{\sigma_{ij}} \right| \right\}$

I.  *For nonbasic $j$ at lower bound*

If $\sigma_{ij} < 0$ and $\delta_{i^*} = f_i$ calculate $\Delta = \frac{(1 - \delta_{i^*})}{-\sigma_{ij}}$

If $\sigma_{ij} > 0$ and $\delta_{i^*} = 1 - f_i$ calculate $\Delta = \frac{(1 - \delta_{i^*})}{\sigma_{ij}}$

If $\sigma_{ij} < 0$ and $\delta_{i^*} = 1 - f_i$ calculate $\Delta = \frac{\delta_{i^*}}{-\sigma_{ij}}$

If $\sigma_{ij} > 0$ and $\delta_{i^*} = f_i$ calculate $\Delta = \frac{\delta_{i^*}}{\sigma_{ij}}$

II.  *For nonbasic $j$ at upper bound*

If $\sigma_{ij} < 0$ and $\delta_{i^*} = 1 - f_i$ calculate $\Delta = \frac{(1 - \delta_{i^*})}{-\sigma_{ij}}$

If $\sigma_{ij} > 0$ and $\delta_{i^*} = f_i$ calculate $\Delta = \frac{(1 - \delta_{i^*})}{\sigma_{ij}}$

If $\sigma_{ij} > 0$ and $\delta_{i^*} = 1 - f_i$ calculate $\Delta = \frac{\delta_{i^*}}{\sigma_{ij}}$

If $\sigma_{ij} < 0$ and $\delta_{i^*} = f_i$ calculate $\Delta = \frac{\delta_{i^*}}{-\sigma_{ij}}$

Otherwise go to next non-integer nonbasic or superbasic $j$ (if available). Eventually the column $j^*$ is to be increased form LB or decreased from UB. If none go to next $i^*$.

Step 4.  Calculate

$$\alpha_{j^\bullet} = B^{-1}\alpha_{j^\bullet}$$

i.e. solve $B\alpha_{j^\bullet} = \alpha_{j^\bullet}$ for $\alpha_{j^\bullet}$.

step 5.  Ratio test; there would be three possibilities for the basic variables in order to stay feasible due to the releasing of nonbasic $j^*$ from its bounds.

*If $j^*$ lower bound*
Let

$$A = \min_{i' \neq i^* | \alpha_{ij^\bullet} > 0} \left\{ \frac{x_{B_{i'}} - l_{i'}}{\alpha_{ij^\bullet}} \right\}$$

$$B = \min_{i' \neq i^* | \alpha_{ij^\bullet} < 0} \left\{ \frac{u_{i'} - x_{B_{i'}}}{-\alpha_{ij^\bullet}} \right\}$$

$$C = \Delta$$

The maximum movement of $j^*$ depends on:

$$\theta^* = min(A, B, C)$$

*If $j^*$ upper bound*
Let

$$A' = \min_{i' \neq i^* | \alpha_{ij^\bullet} < 0} \left\{ \frac{x_{B_{i'}} - l_{i'}}{\alpha_{ij^\bullet}} \right\}$$

$$B' = \min_{i' \neq i^* | \alpha_{ij^\bullet} > 0} \left\{ \frac{u_{i'} - x_{B_{i'}}}{-\alpha_{ij^\bullet}} \right\}$$

$$C' = \Delta$$

The maximum movement of $j^*$ depends on:

$$\theta^* = min(A', B', C')$$

step 6.  Exchanging basis for the three possibilities

1. If $A$ or $A'$
   - $x_{B_{i'}}$ becomes nonbasic at lower bound $l_{i'}$
   - $x_{j^\bullet}$ becomes basic (replaces $x_{B_{i'}}$)
   - $x_{i^\bullet}$ stays basic (non-integer)

2. If $B$ or $B'$
   - $x_{B_{i'}}$ becomes nonbasic at upper bound $u_{i'}$
   - $x_{j^\bullet}$ becomes basic (replaces $x_{B_{i'}}$)
   - $x_{i^\bullet}$ stays basic (non-integer)

3. If $C$ or $C'$
   - $x_{j^\bullet}$ becomes basic (replaces $x_{i^\bullet}$)
   - $x_{i^\bullet}$ becomes superbasic at integer-valued

Repeat from step 1.

## VII.    CONCLUSION

The Stochastic VRP (SVRP) arises whenever some parameters of the VRP are random (e.g. demand and travel time). In this paper we present the *Capacitated Open Vehicle Routing Problem* (COVRP), in which the demand is uncertain. The model of the problem turns out to be a chance-constrained stochastic program. We use Sample Average Approximation that transforms the model into a mixed integer programming model. Then we solve the integer model using a strategy of releasing nonbasic variables from their bounds, combined with the "active constraint" method.

**REFERENCES**
[1]     Aly and J. White., "Probabilistic formulation of the emergency service location problem". *The Journal of the Operational Research Society*, 29:1167–1179, 1978.

[2]    D.Aksen, Z.Özyurt and N.Aras, "The Open Vehicle Routing Problem with Driver Nodes and Time Deadlines", *J. of the Opl Res. Soc.*, 2006, to appear.

[3]    A. Ben-Tal and A. Nemirovski., "Robust convex optimization". *Mathematics of Operations Research*, 23:769–805, 1998.

[4]    P. Beraldi, M. Bruni, and D. Conforti., "Designing robust emergency medical service via stochastic programming". *European Journal of Operational Research*, 158:183– 193, 2004.

[5]    J. Brandão, "A Tabu Search Algorithm For The Open Vehicle Routing Problem", *Eur. J. Opl Res.*, 157: 552–564, 2004.

[6]    G. C. Calafiore and M. C. Campi., "Uncertain convex programs: randomized solutions and confidence levels". *Mathematical Programming*, 102:25–46, 2005.

[7]    G. C. Calafiore and M. C. Campi., "The scenario approach to robust control design". *IEEE Transactions on Automatic Control*, 5 1:742–753, 2006.

[8]    A. Charnes, W. W. Cooper, and G. H. Symonds., "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil". *Management Science*, 4:235–263, 1958.

[9]    M.-S. Cheon, S. Ahmed, and F. Al-Khayyal., "A branch-reduce-cut algorithm for the global optimization of probabilistically constrained linear programs". *Mathematical Programming*, 108:617–634, 2006.

[10]   D. Dentcheva, A. Prékopa, and A. Ruszczyński., "Concavity and efficient points of discrete distributions in probabilistic programming". *Mathematical Programming*, 89:55–77, 2000.

[11]   Z.Fu, R.Eglese and L.Y.O.Li, "A New Tabu Search Heuristic For The Open Vehicle Routing Problem". *J. of the Opl Res. Soc.*, 56: 267–274, 2005.

[12]   Z.Fu, R.Eglese and L.Y.O.Li, "Corrigendum To A New tabu Search Heuristic for the open vehicle routing problem*". J. of the Opl Res. Soc.*, 57, p. 1018, 2006.

[13]   A. Gaivoronski and G. Pflug., "Value-at-risk in portfolio optimization: Properties and computational approach". *The Journal of Risk*, 7:1–33, 2005.

[14]   F.Li, B.Golden and E.Wasil, "The Open Vehicle Routing Problem: Algorithms, Largescale Test Problems, And Computational Results". *Comp. & Oper. Res.*, 2006, to appear.

[15]   R. Henrion, P. Li, A. Möller, M. C. Steinbach, M. Wendt, and G. Wozny., "Stochastic optimization for operating chemical processes under uncertainty". In M. Grötschel, S. Krunke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 45 7–478. Springer, 2001.

[16]   R. Henrion and A. Möller., "Optimization of a continuous distillation process under random inflow rate". *Computers and Mathematics with Applications*, 45:247–262, 2003.

[17]   M. A. Lejeune and A. Ruszczyński. "An efficient trajectory method for probabilistic inventory-production-distribution problems". *Operations Research*, 55:378–394, 2007.

[18]   J. Luedtke, S. Ahmed, and G. Nemhauser., "An integer programming approach for linear programs with probabilistic constraints". In M. Fischetti and D. Williamson, editors, *IPCO 2007*, Lecture Notes in Computer Science, pages 410–423, Berlin, 2007. Springer-Verlag.

[19]   V. Marianov and M. Rios., "A probabilistic quality of service constraint for a location model of switches in ATM communications networks". *Annals of Operations Research*, 96:237–243, 2000.

[20]   M. R. Murr and A. Pr´ekopa., "Solution of a product substitution problem using stochastic programming". In S. P. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*, pages 252–271. Kluwer Academic Publishers, 2000.

[21]   A. Nemirovski and A. Shapiro., "Scenario approximation of chance constraints". In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design Under Uncertainty*, pages 3–48. Springer, London, 2005.

[22]   A. Nemirovski and A. Shapiro. "Convex approximations of chance constrained programs". *SIAM Journal on Optimization*, 17:969–996, 2006.

[23]   B. Pagnoncelli, S. Ahmed, and A. Shapiro., "Computational study of a chance constrained portfolio selection problem". Submitted for publication, 2008. Preprint available at www. optimization-online. org.

[24]   D.Pisinger and S.Ropke, A General Heuristic For Vehicle Routing Problems". Comp. & Oper. Res., 2006, to appear.

[25]   Prékopa., "On probabilistic constrained programming". In H. W. Kuhn, editor, *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113– 138, Princeton, N.J., 1970. Princeton University Press.

[26]   A. Prékopa., *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[27]   A. Prékopa., "Probabilistic programming". In A. Ruszczyński and A. Shapiro, editors, Stochastic Programming, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267–351. Elsevier, 2003.

[28]   A. Prékopa and T. Szántai., "Flood control reservoir system design". *Mathematical Programming Study*, 9:138–151, 1978.

[29]   P.P.Repoussis, C.D.Tarantilis and G. Ioannou, "The Open Vehicle Routing Problem With Time Windows". J. of the Opl Res. Soc., 2006, to appear.

[30]   A. Ruszczyński., "Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedral". *Mathematical Programming*, 93:195–215, 2002.

[31]   D.Sariklis and S.Powell, "A heuristic method for the open vehicle routing problem", J. of the Opl Res. Soc., 2000, 51: 564–573.

[32]   L.Schrage, "Formulation And Structure Of More Complex/Realistic Routing And Scheduling Problems", Networks, 11: 229–232, 1981.

[33]   A. K. Takyi and B. J. Lence., "Surface water quality management using a multiple-realization chance constraint method". *Water Resources Research*, 35:1657–1670, 1999.

[34]   C.D.Tarantilis, D.Diakoulaki and C.T.Kiranoudis, "Combination Of Geographical Information System And Efficient Routing Algorithms For Real Life Distribution Operations", *Eur. J. Opl Res.,* 152: 437–453, 2004.

[35]   C.D.Tarantilis, G.Ioannou, C.T.Kiranoudis and G.P. Prastacos, Solving The Open Vehicle Routeing Problem Via A Single Parameter Metaheuristic Algorithm", *J. of the Opl Res. Soc.*, 56: 588–596, 2005.

[36]   P.Toth and D.Vigo, (eds.) *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications, 2002.

[37]   A. N. Ventetsanopoulos and I. Singh., "Topological optimization of communication networks subject to reliability constraints". *Problem of Control and Information Theory*, 15:63–78, 1986.