Probabilistic Analysis of a Combined Aggregation and Math Programming Heuristic for a General Class of Vehicle Routing and Scheduling Problems

Awi Federgruen • Garrett van Ryzin
Graduate School of Business, Columbia University, New York, New York 10027

We propose and analyze a heuristic that uses region partitioning and an aggregation scheme for customer attributes (load size, time windows, etc.) to create a finite number of customer types. A math program is solved based on these aggregated customer types to generate a feasible solution to the original problem. The problem class we address is quite general and defined by a number of general consistency properties. Problems in this class include VRPs with general distance norms, capacitated problems, time window VRPs, pick-up and delivery problems, combined inventory control and routing problems and arc routing.

We provide a probabilistic analysis of this heuristic under very general probabilistic assumptions. In particular, we do not require independence between customer locations and their various attributes. The heuristic is (a.s.) ϵ-optimal as the number of customers n tends to infinity. Further, it runs in $O(n \log n)$ time for a fixed relative error, and can be designed to be asymptotically optimal while still running in polynomial time. We characterize the asymptotic average value of the heuristic and the optimal solution as the limit of a sequence of linear program values. We also provide bounds on the rate of convergence to the asymptotic value and bounds on tail probabilities. Finally, we discuss numerical issues involved in implementing our heuristic.
(*Vehicle Routing; Linear Programming; Probabilistic Analysis; Polynomial Time Algorithms; Heuristics*)

Introduction and Overview

Region partitioning schemes, in which solutions for a large service region are generated by combining solutions formed on smaller subregions, have played a prominent role in the probabilistic analysis and design of effective algorithms for vehicle routing problems (VRPs). The seminal papers of Beardwood et al. (1959) and Karp (1977) for the TSP and Steele's (1981) general theory of subadditive Euclidean functionals all employ variants of region partitioning schemes. For the capacitated VRP, such schemes were used in the pioneering work of Haimovich and Rinnooy Kan (1985), and they form the foundation for much of the subsequent work on VRPs (see Federgruen and Simchi-Levi (1992) for a review). For example, the location-based heuristic of Bramel and Simchi-Levi (1991) is analyzed by bounding its cost by a region partitioning scheme. In Federgruen and van Ryzin (to appear in *Oper. Res.*), we used the partitioning analysis in Bramel et al. (1993) together with a math-programming-based heuristic for general bin packing problems to solve a capacitated VRP with time window constraints.

The appeal of region partitioning schemes lies in their ability to approximate the solution structure of certain classes of VRPs. Only customers within the same subregion are assigned to a given route. The resulting tour structure produces tightly clustered collections of customers connected to the depot by radial arcs, allowing

0025-1909/97/4308/1060$05.00

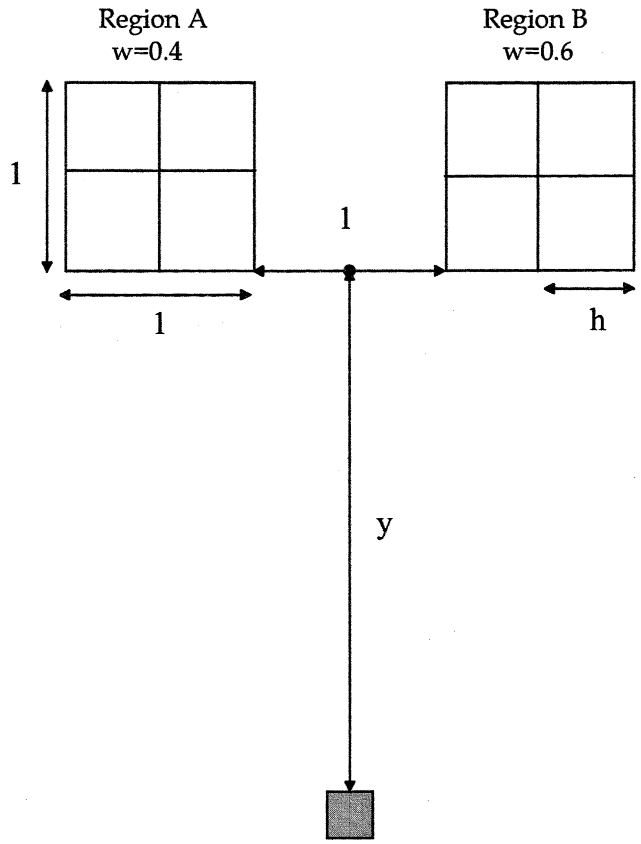
for accurate bounds on the travel cost. The problem is then analyzed by letting the grid size tend to zero as the number of customers grows. Under certain probabilistic assumptions, one can show that with high probability the resulting limiting tour structure is asymptotically optimal (see Federgruen and Simchi-Levi (1992)).

However, while such a solution structure is intuitively appealing, its asymptotic optimality only arises under the (typical) assumption (see Federgruen and Simchi-Levi (1992)) that a customer's location is independent of its other attributes, such as delivery requirements, time windows, etc. As a result, uniformly throughout the service region one finds roughly the same range of customer types in the same proportions, and—at least asymptotically—it makes sense to only consider tours that visit customers within a small neighborhood.

Nevertheless, there are many classes of problems for which region partitioning schemes can perform badly. An example illustrates the point: Consider the simple, capacitated VRP shown in Figure 1, with vehicles of unit capacity. The customers' locations are confined to two squares, A and B, with sides of unit length, and separated from each other by a distance of one unit. The vertical distance between the depot and the squares is given by y. The customers in square A (B) have delivery sizes of 0.4 (0.6). Let n customers be independently assigned to sets A and B with equal probability, and then to some location in A or B according to a given distribution. Let N_A (N_B) denote the number of customers assigned to A (B). ($N_A + N_B = n$). For a sufficiently small grid size, a region partitioning scheme takes only customers within some subregion of A or B. The resulting cost of the partitioning solution is thus $y(N_A/2 + N_B) + O(n)$ as $y \to \infty$. On the other hand, if we maximally match customers in A with those in B and assign each pair a tour, the cost is $y[\min(N_A, N_B) + (n - 2\min(N_A, N_B))] + O(n)]$. Dividing by n and letting $n \to \infty$ we obtain (a.s.) costs of $\frac{3}{4}y + O(1)$ and $\frac{1}{2}y + O(1)$ for the two schemes, implying that for y large the cost of the region partitioning scheme exceeds that of the pairing heuristic by about 50%.

As we show in §5, the reason for the poor performance of the partitioning heuristic in this example is that the asymptotic optimal solution is not a collection of tightly clustered sets of customers connected to the depot by radial arcs. There are many other VRP applications where location dependent customer attributes cause region partitioning schemes to fail; for example, suppose in Figure 1 that the A (B) customers require pick-up (delivery) of a full load. Clearly, it is better to combine a delivery to B with a pick-up from A rather than to make deliveries and pick-ups using separate tours.

Figure 1 **A VRP with Load Size Dependent on Location**



In this paper we propose an alternative class of heuristics based on aggregation schemes in which customers with similar combinations of locations and attribute values (e.g. delivery requirements, scheduling constraints, etc.) are aggregated into a single type. These aggregation schemes enable us to generate provably good heuristics for a wide class of vehicle routing and scheduling problems under very general distributions of problem instances, including problems with location-dependent customer attributes.

Like region partitioning schemes, our heuristics may be viewed as an approximation method for the classical set covering formulation to VRPs, originally due to Charnes and Miller (1956) and Balinski and Quandt (1964), and subsequently employed in several effective algorithms, see e.g. Cullen et al. (1981), Desrochers et al. (1981), Ceria et al. (1995) and Cabrara et al. (1995). Region partitioning schemes amount to a restriction method in which the set of columns (routes) is restricted to those visiting customers in the same subregion. Because customers are partitioned into subregions, the restricted problem is decomposable by subregion. Our aggregation schemes may be viewed as row aggregation methods, converting the set covering problem into a multi-set covering problem. (See Rogers et al. (1991) for a recent survey of aggregation methods in mathematical programming.) By avoiding the sometimes rather severe restrictions imposed by region partitioning schemes, we safeguard the quality of the heuristic solution. Moreover, the recent effective algorithms based on the set covering formulation (Desrochers, Desrosiers, and Solomon (1992); Ceria, Paolo, and Sassano (1995)) all show that the computational effort is primarily determined by its number of rows; row aggregation schemes thus result in a most effective approach to reduce computational complexity while maintaining the quality of the resulting solution.

The approach suggests a powerful class of heuristics which can be designed to be both asymptotically optimal and polynomial (under mild probabilistic assumptions); it also provides a quite complete analysis of the solution cost of a wide class of VRPs. As a corollary, we extend Bramel and Simchi-Levi (1993) to show that the linear programming relaxations of set covering formulations for VRPs are asymptotically tight.

In §1 we specify the class of VRPs that can be addressed with our approach and demonstrate its generality. In §2 we analyze a version of the problem in which the distribution of locations and attribute values is discrete. We characterize the asymptotic optimal solution value as the value of an underlying linear program which depends on the joint probability mass function of the locations and attribute values; we also derive bounds for the tail of the minimum cost value and specify the complete limiting distribution of the minimum cost value. In §3 we address continuous location and attribute distributions. Our approach here is to approximate the continuous distribution by a sequence of progressively finer discretizations of both the service region and the attribute space. This allows us to employ the integer programming analysis of §2. More specifically, we develop a lower and upper bound VRP with discrete sets of customer types. As an effective heuristic, we propose finding a feasible solution for the upper bound problem by solving its LP relaxation and rounding the solution up to integer values; the LP relaxation of the lower bound problem can be used as a lower bound to gauge the heuristic's optimality gap. We show that this math programming heuristic is asymptotically optimal as long as the number of customer types in the aggregation scheme grows sublinearly with n, the number of customers; moreover, its complexity is polynomially bounded, albeit with a high degree polynomial bound depending on M, the maximum number of customers per route, which is uniformly bounded by the properties of the class of VRPs considered. The fact that the rounded up solution of the LP relaxation is close to optimal was identified by Bartholdi (1981) for set covering problems arising in cyclical staff scheduling problems. He as well as Hochbaum (1982, 1983) develop worst-case bounds for the optimality gap of this rounding heuristic.

In §4, we continue the theoretical characterization of the heuristic's complexity by showing that it can be bounded by a polynomial of fixed degree—independent of M—when solving the linear program by a column generation technique, provided the effort to generate columns is polynomial in n itself. We illustrate this complexity characterization for the case of the classical VRP. We then discuss how the proposed heuristic could be solved in practice making use of recent computational results for set covering formulation based heuristics. Finally in §5, we provide computational results for several examples, exhibiting the quality of the heuristic as a function of the number of customers, the aggregation scheme applied and the model parameters.

1. A General Class of Vehicle Routing Problems

1.1. Problem Definition and Examples

We define a generalized vehicle routing problem as follows: Customers are described by a location $x = (x_1, x_2)$ in $[0, 1]^2$ and a d-dimensional vector $w = (w^1, \ldots, w^d)$

of attributes. The restriction of x to the unit square is without loss of generality, provided the service region is bounded. More generally, locations may be in $[0, 1]^r$ for $r \geq 2$, e.g. see Examples 5 and 6 below, but to simplify the exposition we restrict ourselves to the more basic planar case. We assume that attribute values are bounded, and hence can be shifted and scaled to be in [0, 1]. The distance between two locations x, y is given by a pseudo-metric $d(x, y)$ which is homogeneous of degree one, i.e. i) $d(x, y) \geq 0$, ii) $d(x, y) = d(y, x)$, iii) $d(\alpha x, \alpha y) = \alpha d(x, y)$ for all $\alpha > 0$ and iv) $d(x, y) \leq d(x, z) + d(z, y)$. Examples include the Euclidean and Manhattan (l_1) distances or any other pseudo-metric which approximates travel distance/time for a given service region.

A tour is an ordered set of customers $\tau = \{(x_k, w_k); 1 \leq k \leq |\tau|\}$, which starts and ends at a given depot location $x_0 \in [0, 1]^2$. We let $l_k = d(x_k - 1, x_k)$, $k = 1, \cdots |\tau|$, $l_{|\tau|+1} = d(x_{|\tau|}, x_0)$. A tour may visit the same customer type multiple times. Considering all possible locations in $[0, 1]^2$ and attribute vectors in $[0, 1]^d$, let Ω denote the collection of all ordered sets of customers τ that can be visited using a single tour (the collection of all *feasible tours*) with the following consistency properties:

P1. *Every finite set of customers can be serviced by some finite collection of tours in* Ω.

P2. *If* $\tau \in \Omega$, *then for every* $(x, w) \in \tau$, $\tau - \{(x, w)\} \in \Omega$.

P3. *If the possible locations x and attributes w are both restricted to a finite set, then the collection of feasible tours in* Ω *is finite.*

Property P1 assures the existence of a feasible solution for any finite instance. Property P2 says that a feasible tour remains feasible when a customer is removed. In view of P1 and P2, any single customer can be served using a dedicated tour. P3 is a boundedness condition. One technical difficulty concerning P3 is the possible existence of *null* customers, i.e. those that can be added to any feasible tour without violating its feasibility (e.g. a customer with zero demand in the classical VRP). P3 ensures that no such customers exist, since by adding k null customers to a feasible tour τ for $k = 1, 2, \ldots$, an infinite sequence of additional feasible tours arises. That is, even though some (or all) attribute values of a customer are zero, there is still an upper bound M on the number of such customers that can be served in any single tour. This assumption is quite reasonable and nonrestrictive in practice. If some attributes take on a continuum of values, we need two additional properties:

P4. *A tour* $\tau \in \Omega$ *remains feasible when an attribute value of any of its customers is reduced.*

P5. *Suppose the pseudo-metric* $d(x, y)$ *is replaced by a function (not necessarily pseudo-metric or homogeneous)* $\underline{d}(x, y)$ *satisfying* $\underline{d}(x, y) \leq d(x, y)\ \forall\ x, y \in [0, 1]^2$. *Then any tour* τ *that is feasible under d remains feasible under* $\underline{d}$.

Whether or not P4 holds for a given problem can depend on which variables are used to represent attributes (see Example 3 below). It implies that M is a uniform upper bound for the number of customers in any feasible tour. P5 says that a similar property holds for distances; namely, if they are reduced, the tours remain feasible.

For some classes of problems, such as the time window VRP, feasibility of tours depends on the times traveled on the various legs of the tour, and a further property, P6 below, is needed. Informally, we require that customers have an on-site service time and also that feasibility depends only on the *times* at which customers are visited (possibly including the return time to the depot to include tour length constraints) and not on distances traveled. It is convenient to emphasize the service time attribute notationally and describe customers by a triple (x, s, w) where s is the on-site service time requirement and w is a vector of remaining attributes. As before, we assume $x \in [0, 1]^2$ and $(s, w) \in [0, 1]^d$. Moreover, vehicles travel at a fixed velocity $\nu > 0$. Formally, we require:

P6. *Let s denote an on-site service attribute and* $\tau = \{(x_k, s_k, w_k); 1 \leq k \leq |\tau|\}$ *be a feasible tour in* Ω. *Let* τ_h *denote the tour formed by replacing* s_k *by* $s_k - h/\nu$ *for each customer* $k = 1, \ldots, |\tau|$ *in* τ. *Let* d_h *denote a distance metric satisfying* $d_h(x, y) \leq d(x, y) + h$, $\forall\ x, y \in [0, 1]^2$. *Then for all* $h > 0$, τ_h *is feasible under the distance metric* d_h.

The cost of every tour $\tau \in \Omega$ is given by a function $c : \Omega - R$ satisfying:

P7. *$c(\tau)$ is uniformly bounded ($c(\tau) \le C \ \forall \ \tau \in \Omega$); nondecreasing (component-wise) in w_k, $k = 1, \cdots |\tau|$; and nondecreasing (component-wise) and first order Lipschitz in l_k and s_k (when applicable), $k = 1, \ldots, |\tau| + 1$.*

The objective of the generalized VRP is to find a least cost collection of feasible tours visiting a given set of customers S. Let $c^*(S)$ denote the optimal cost.

For probabilistic analyses, we consider a *joint* probability measure μ on the pair of vectors (x, w) (or triple (x, s, w) if applicable), permitting dependence between locations and attributes. Without loss of generality μ has support on $[0, 1]^{d+2}$. We shall look at instances that are formed by drawing an independent sequence of customers from μ. For our primary results in Theorems 1 and 4, even independence is not required and we only assume that the sequence is stationary.

When x and w are restricted to finite sets, any set of customers S can be represented by a vector $z = (z_1, \ldots, z_I) \in \mathbb{Z}_+^I$ where $i = 1, \ldots, I$ indexes the *type* of object (i.e. a particular combination of a discrete location x and a discrete attribute vector w) and z_i denotes the number of customers of type-i contained in the set S. In this case, μ is a discrete probability mass function, denoted by $\pi = (\pi_1, \ldots, \pi_I)$.

Examples of VRPs that fall under this general framework include:

EXAMPLE 1. CLASSICAL CAPACITATED VRP. Each customer $k = 1, \ldots, n$ has a delivery size v_k, and Ω is the set of all tours for which $\Sigma_{k\in\tau} v_k \le 1$, i.e. tours are constrained only by the vehicle capacity, which is identical for all vehicles and normalized to one. We assume $v \in [\delta, 1]$ for some $\delta > 0$. Since feasibility does not depend on travel time, no on-site service time is required. A customer is thus characterized by its location $x \in [0, 1]$ and a single attribute $w = v - \delta \in [0, 1]$. The cost of a tour is $c(\tau) = \Sigma_{k\in\tau} l_k$, the total distance, where the pseudo-metric $d(x, y)$ is the Euclidean or any other metric. Since $v \ge \delta$, the number of customers per tour is bounded by δ^{-1}, and its cost by $C = \delta^{-1} d((0, 0), (1, 1))$. It is easy to verify that P1–P5 and P7 are satisfied, while P6 is not applicable.

EXAMPLE 2. TIME WINDOW VRP. Each customer has a location, scalar load size v, a service time s for loading/unloading the vehicle, an earliest delivery time e and a latest delivery time f. A unit time planning period is available during which all on-site service must start and end. (Vehicle travel is not necessarily restricted to this window.) We define $r = 1 - f$ and the vector of attributes $w = (v, e, r)$. A tour is feasible if it visits all customers within their specified time windows and the unit capacity of the vehicle is not exceeded. That is, there exists start times (a schedule) $t_0, t_1, \ldots, t_k$ such that for $k = 1, \ldots, |\tau|$: i) $t_k \ge t_{k-1} + s_{k-1} + d(x_{k-1}, x_k)/\nu$, ii) $t_k \ge e_k$ and iii) $t_k \le 1 - s_k - r_k$, and $\Sigma_{k\in\tau} v_k \le 1$. $c(\tau)$ can be chosen to minimize the total distance, the number of vehicles or some more general cost structure satisfying Property P5. P4 is satisfied by the definition of r_k. In this case tour feasibility depends on time; however, one easily verifies that P6 holds. The remaining properties hold under the conditions of Example 1. We note that several versions of this problem have been analyzed (see Bramel et al. (1991); Bramel and Simchi-Levi (1992, 1993) and Federgruen and van Ryzin (1992)) under the assumption that locations x_k are independent of the values (v_k, s_k, e_k, r_k). A special case of this problem is the VRP with a total load size and time constraint.

EXAMPLE 3. A PICK-UP AND DELIVERY PROBLEM. In this class of problems one may have people to transport (dial-a-ride systems) or backhaul opportunities; customers may need a delivery, a pick-up or both. See Bodin et al. (1983) and Casco et al. (1988) for surveys of this problem class. To date, no probabilistic analysis of this problem has appeared in the literature.

As in Example 1, a tour is feasible if its load never exceeds the vehicle capacity (of one). A customer is described by a triplet of values $(x, v, \bar{v})$, with x its location, v the size of the delivery and $\bar{v}$ the size of the pick-up. A tour $\tau = \{(x_1, v_1, \bar{v}_1), \ldots, (x_{|\tau|}, v_{|\tau|}, \bar{v}_{|\tau|})\}$ is feasible if $\Sigma_{k=1}^{l} \bar{v}_k + \Sigma_{k=l+1}^{|\tau|} v_k \le 1$, $l = 0, \ldots, |\tau|$, i.e., the sum of the loads that have been picked up and those yet to be delivered must not exceed the capacity of the vehicle. The cost function can be chosen as in Examples 1 or 2. Again, P1, P2, P4 and P5 are satisfied. P3 and P7 are satisfied provided delivery and pick-up load sizes are uniformly bounded away from zero. P6 does not apply since feasibility is not affected by the distances traveled.

EXAMPLE 4. AN INVENTORY ROUTING PROBLEM. Here the customers represent retailers, which are replenished from a central warehouse with unlimited supply of a specific item and with vehicles of unit capacity. Retailer k faces continuous demands for the item at a constant rate of d_k units, and incurs holding costs at a constant rate h per unit of product and time. Demands at each retailer must be met over an infinite horizon without shortages or backlogging. The frequency with which a given retailer can be visited is bounded from above by f. Various additional constraints may apply to the routes used to supply retailers, e.g. a time constraint as in Example 2 or an upper bound on the number of stops (i.e., retailers visited) per route. The objective is to find a replenishment strategy which minimizes infinite horizon average costs.

Most existing approaches to the above class of inventory-routing problems (see Anily and Federgruen 1990, and Bramel and Simchi-Levi 1991) restrict themselves to the class of fixed-partition strategies in which the retailers are partitioned into sets such that all retailers in a set are always replenished together and independently of any retailer outside the set. Chan et al. (1993) recently showed that when routes are constrained by the vehicle capacity and frequency constraints only, and when locations and demand rates are independent attributes, fixed-partition strategies are in fact asymptotically close-to-optimal (see Chan et al., to appear in *Oper. Res.*, for details).

Under a fixed-partition strategy, it is clearly optimal to serve every region with constant replenishment intervals and such that retailers are replenished only when their inventories are depleted. Consider a region and a corresponding tour τ through its retailers and the depot. Let $D = \sum_{k=1}^{|\tau|} d_k$ and $L = \sum_{k=1}^{|\tau|} l_k + l_{|\tau|+1}$ denote the total demand rate of the region and the length of the tour, respectively. If the region is replenished with intervals of length T, then $DT \le 1$ (to satisfy the capacity constraint) and $T \ge f^{-1}$ (to satisfy the frequency constraint). The cost of the tour τ is given by $LT^{-1} + hDT/2$. Minimizing this function over the feasible interval $[f^{-1}, D^{-1}]$ we obtain

$$c(\tau) = \begin{cases} Lf + \frac{1}{2}hDf^{-1} & \text{if } \sqrt{\frac{2L}{hD}} < f^{-1}, \\ \sqrt{2hLD} & \text{if } f^{-1} \le \sqrt{\frac{2L}{hD}} \le D^{-1}, \\ LD + \frac{1}{2}h & \text{otherwise.} \end{cases}$$

Retailers are thus characterized by a location x_k and a single attribute value d_k. A tour is feasible if and only if $D = \sum_{k=1}^{|\tau|} d_k \le f$. The cost $c(\tau)$ is bounded and is first order Lipschitz in the l_k quantities provided the demand rates d_k are uniformly bounded away from zero. The set of feasible tours Ω is the same as in Example 1 with the load size w_k given by d_k/f. Alternatively, tours may be constrained by a total time constraint in addition to the volume constraint, in which case Ω is defined as in Example 2.

Other generalizations include settings where the holding cost rate and maximum frequency parameter are retailer specific, i.e. each retailer is characterized by a location x_k and a triple of attributes $w_k = (d_k, h_k, v_k)$ where $v_k = f_k^{-1}$, and h_k, f_k denote, respectively, the holding cost rate and maximum delivery frequency at retailer k. The maximum frequency for a given tour τ is then $\min\{f_k : 1 \le k \le |\tau|\}$ or equivalently $1/v_\tau^*$ where $v_\tau^* \doteq \max\{v_k : 1 \le k \le |\tau|\}$. One can verify that in this case

$$c(\tau) = \begin{cases} L/v_\tau^* + \frac{1}{2}v_\tau^* \sum_{k=1}^{|\tau|} h_k d_k & \text{if } \sqrt{\frac{2L}{\sum_{k=1}^{|\tau|} h_k d_k}} < v_\tau^*, \\ \sqrt{2L \sum_{k=1}^{|\tau|} h_k d_k} & \text{if } v_\tau^* \le \sqrt{\frac{2L}{\sum_{k=1}^{|\tau|} h_k d_k}} \le D^{-1}, \\ LD + \frac{1}{2} \frac{\sum_{k=1}^{|\tau|} h_k d_k}{\sum_{k=1}^{|\tau|} w_k} & \text{otherwise.} \end{cases}$$

Assuming as before that all the attributes d_k, h_k, v_k are bounded away from zero, one easily verifies that Properties P1–P5 and P7 hold in this case, while P6 does not apply.

In summary, we have shown that a large class of inventory-routing problems of considerably greater generality than those discussed in the literature can be treated as special cases of our general routing problem, when restricting oneself to the class of fixed partition strategies.

EXAMPLE 5. ARC ROUTING PROBLEMS. In some routing problems, there is a need to cover a specific collection of arcs in a network. Examples include urban services such as snow removal, garbage collection or postal carrier routes. A list of origin-destination pairs $\{x_k = (y_k, z_k) : k = 1, \dots, n\}$ is given with $y_k, z_k \in [0, 1]^2$; each of these links is to be covered by at least one tour. Each link $x_k = (y_k, z_k)$ can be viewed as a customer residing in $[0, 1]^4$. A tour τ is described by an ordered set of (not necessarily distinct) customers $\{x_k : k = 1, \ldots, n\}$. In some versions of the problem, tours are constrained only by a total time constraint, as in Example 2; more complex versions would include time windows for some or all customers.

The distance between a pair of customers $x_1 = (y_1, z_1)$ and $x_2 = (y_2, z_2)$ is given by $d(x_1, x_2) = \hat{d}(z_1, y_2)$ with $\hat{d}$ a pseudo-metric in R^2, i.e. when $y_2 \neq z_1$, $d(x_1, x_2)$ denotes the cost of dead-heading between z_1, the tail of x_1, and y_2, the head of x_2; $d(x_1, x_2) = 0$ if and only if $y_2 = z_1$. Observe that the distance metric d is a pseudo-metric, which is homogeneous of degree one. The objective may be to minimize the total dead-head distance or minimize the total number of vehicles used, etc.

Each customer has an on-site service time requirement $s = d(y, z)$, where $d(y, z)$ is the time to traverse and service the link. Note that in this case customer locations and attribute values are highly dependent, as the on-site service time is a function of the location value x. One can verify that Properties P1–P7 are satisfied provided that for all x, y in an instance, $d(x, y) \geq \delta > 0$ for some $\delta > 0$, i.e. there is a uniform lower bound on the arc lengths.

EXAMPLE 6. MANY-TO-MANY ROUTING PROBLEMS. In many routing problems, deliveries must be made from a variety of sources to a set of corresponding destinations, as opposed to the routing problems in Examples 1–4 in which all deliveries are made from or to a single depot. Examples include delivery problems for common carriers and railroad freight.

A list of origin-destination pairs $\{x_k = (y_k, z_k) : k = 1, \ldots, n\}$ is given with $y_k, z_k \in [0, 1]^2$, and with an associated delivery quantity $v_k \leq 1$. In the simplest version of the problem, tours are only constrained by the vehicles' unit capacity. As in Example 5, we consider each origin-destination pair x_k as a customer with $x_k \in [0, 1]^4$, $k = 1, \ldots, n$; however, in contrast to Example 5, we describe a tour τ as a sequence of the form $\{(\xi_1, \nu_1), \ldots, (\xi_{|\tau|}, \nu_{|\tau|})\}$ with $\xi_i \in [0, 1]^2$ and $\nu_i \in \{-1, 1\}$, such that if $\nu_i = 1$, $\xi_i = y_{k(i)}$ and if $\nu_i = -1$, $\xi_i = z_{k(i)}$ for some $k(i) \in \{1, \ldots, n\}$. (Though this definition of a tour is not exactly in the same form as specified above, it does satisfy Property P2.) A tour is feasible if for all $i = 1, \ldots, |\tau|$, (1) $\xi_i = y_{k(i)}$ implies $\xi_j = z_{k(i)}$ for some $j > i$, (2) $\xi_i = z_{k(i)}$ implies $\xi_j = y_{k(i)}$ for some $j < i$, and (3) $\sum_{l=1}^{i} \nu_l w_{k(l)} \leq 1$ for all $i = 1, \ldots, |\tau|$.

The distance between any consecutive pair (ξ_i, ν_i) and (ξ_{i+1}, ν_{i+1}) is given by $\hat{d}(\xi_i, \xi_{i+1})$ with $\hat{d}(\cdot, \cdot)$ any of the standard metrics in R^2. Again, one could have an objective of minimizing total distance traveled or the number of vehicles used. Under the assumptions of Examples 1 and 5, Properties P1–P5 and P7 hold, while P6 does not apply.

We conclude that in all six example models, all required properties are satisfied under the minor assumptions made there.

2. The General VRP with Discrete Customer Types

In the discrete VRP, each location x and attribute vector w can take on only a finite number of values. Thus, there are a finite number of possible customer *types*, which we index by i, $i = 1, \ldots, I < +\infty$. For the discrete case, we do not require Properties P4, P5, and P6, and only the boundedness in P7 is needed.

We represent any set of customers S by a vector $z \in \mathcal{Z}_+^I$ where z_i denotes the number of type i customers in S. By Property P3, if there are only finitely many possible vectors w, there are finitely many feasible tours τ. Let j, $j = 1, \ldots, J < +\infty$ index these tours and a_{ij} and c_j denote, respectively, the number of type i customers served in and the cost of a type j tour. Define the matrix $A = [a_{ij}]$, let A^j denote its jth column and define $c = (c_1, \ldots, c_J)$. Then the general VRP is equivalent to the following integer program:

$$c^*(z) = \min\{c^T y : Ay \geq z, y \geq 0, y \text{ integer}\}. \quad (1)$$

Here y_j represents the number of type j tours to be driven, $j = 1, \ldots, J$.

This formulation is closely related to the set covering formulation due originally to Charnes and Miller (1956) and Balinski and Quandt (1964). Here each feasible tour $\tau = 1, \ldots, m$ through a subset of customers is enumerated. Let $\gamma_{kr} = 1$ if customer k belongs to tour r and $\gamma_{kr} = 0$ otherwise. Let c_r denote the cost of tour r.

$$c^{SC}(S) = \min \sum_{r=1}^{m} c_r x_r \quad (2)$$

s.t.

$$\sum_{r=1}^{m} \gamma_{kr} x_r \geq 1, \quad k = 1, \ldots, n, \quad (3)$$

$$x_r \in \{0, 1\}, \quad r = 1, \ldots, m. \quad (4)$$

Let c^{SCLP} denote the LP-relaxation of c^{SC}. Note that problem (1) arises from c^{SC} by aggregating rows corresponding to customers of the same type (and variables corresponding with identical columns after this row aggregation step). Note that the set covering formulation has a very large number of constraints and variables. In contrast, program (1) can be solved in $O(Jn^I)$ time using dynamic programming as shown in Federgruen and van Ryzin (1992). Though polynomial in n, this approach is clearly not practical. An alternative approach and one that is more amenable to analysis, is to consider the linear programming relaxation of (1),

$$c^{LP}(z) = \min\{c^T y : Ay \geq z, y \geq 0\}. \quad (5)$$

The next lemma summarizes some useful properties about c^* and c^{LP}. In this lemma and throughout the paper, $|x|$ denotes the l_1-norm of the vector x. The proof is straightforward, see Federgruen and van Ryzin (1992).

LEMMA 1. a) *$c^*(z)$ and $c^{LP}(z)$ are, component-wise, nondecreasing in z.*

b) *$c^{LP}(z)$ is continuous and convex in z.*

c) *$c^{LP}(\alpha z) = \alpha c^{LP}(z)$ for all $z \geq 0$ and $\alpha \geq 0$.*

d) *$c^{LP}(z) - \bar{c}|z_0| \leq c^{LP}(z + z_0) \leq c^{LP}(z) + \bar{c}|z_0|$ and $c^*(z) - \bar{c}|z_0| \leq c^*(z + z_0) \leq c^*(z) + \bar{c}|z_0|$ (equivalently $c^{LP}(z_0) - \bar{c}|z - z_0| \leq c^{LP}(z) \leq c^{LP}(z_0) + \bar{c}|z - z_0|$ and $c^*(z_0) - \bar{c}|z - z_0| \leq c^*(z) \leq c^*(z_0) + \bar{c}|z - z_0|$) for all nonnegative z, z_0.*

e) *$c^{LP}(z) \leq c^*(z) \leq c^{LP}(z) + CI$, with C defined in P7.*

2.1. Almost Sure Convergence

Lemma 1 allows us to characterize the asymptotic behavior of c^* when instances are drawn according to a stationary sequence $\{(x_k, w_k) : k \geq 1\}$. Specifically, let $\pi = (\pi_1, \ldots, \pi_I)$ be the probability measure on the set of possible customer types and let $S^{(n)} = \{(x_k, w_k); 1 \leq k \leq n\}$ denote a set formed from the first n objects in the stationary sequence $\{(x_k, w_k) : k \geq 1\}$; let $z^{(n)}$ be its vector representation. Kingman's subadditive ergodic theory (1976) establishes the (a.s.) convergence of $c^*(z^{(n)})/n$ to a constant $\gamma(\mu)$. The following theorem provides a simple proof of this fact and characterizes the constant $\gamma(\mu)$:

THEOREM 1. $\lim_{n\to\infty} c^*(z^{(n)})/n = c^{LP}(\pi)$ *(a.s.).*

PROOF. First note that for the stochastic sequence $z^{(n)}$, we have by the strong law of large numbers for stationary sequences (see Revesz 1968), that $z^{(n)}/n \to \pi$ (a.s.). Thus, by the continuity of $c^{LP}(\cdot)$, $c^{LP}(z^{(n)}/n) \to c^{LP}(\pi)$ (a.s.). By Lemma 1 a) and c), we have $c^{LP}(z^{(n)}/n) \leq (1/n)c^*(z^{(n)}) \leq c^{LP}(z^{(n)}/n) + CI/n$. Taking limits proves the result. □

This proof is constructive, since the upper bound of Lemma 1 e) represents the cost of the heuristic obtained by solving the linear program (5) for $z^{(n)}$ and rounding up each variable in the resulting solution to the closest integer. Note that solving this linear program takes $O \log n$ time on a machine with finite word length, since its dimensions (i.e. numbers of rows and columns) are fixed. The complexity of this heuristic is therefore $O(n \log n)$ due to the cost of constructing the vector $z^{(n)}$. We discuss computational issues in more detail in §4.

2.2. Bounds on Tail Probabilities and Asymptotic Distribution

A tail probability result can also be obtained by direct analysis of the math program (1). It shows rapid (exponential) convergence of $c^*(z^{(n)})$ to its asymptotic value. The proof is a minor modification of that given in Federgruen and van Ryzin (1992) and is therefore omitted.

THEOREM 2. *For every $\epsilon > 0$, $\alpha > 0$, there exist a constant $n_0 = n_0(\epsilon, \alpha)$ such that,*

$$P\left\{\left|\frac{c^*(z^{(n)})}{n} - c^{LP}(\pi)\right| > \epsilon\right\} \leq 3e^{-[n\epsilon^2/25(1+\alpha)]}$$

for all $n \geq n_0$.

In the discrete case, it is also possible to characterize the full asymptotic distribution of $c^*(z^{(n)})$. Let V denote the $I \times I$ matrix with $V_{ii} = \pi_i(1 - \pi_i)$ and $V_{ij} = -\pi_i\pi_j$ for $i \neq j$. The following theorem may be viewed as a central limit theorem for the value of the VRP. It also provides bounds for its mean and variance. See Federgruen and van Ryzin (1992) for a proof.

THEOREM 3. *Assume the linear program $c^{LP}(\pi)$ has a unique optimal dual vector λ. Then*

$$\frac{c^*(z^{(n)}) - n\lambda^T\pi}{\sqrt{n\lambda^T V\lambda}}$$

converges in distribution to a standard normal random variable, i.e.

$$\lim_{n\to\infty} P\left\{\frac{c^*(z^{(n)}) - n\lambda^T\pi}{\sqrt{n\lambda^T V\lambda}} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\xi^2/2} d\xi, \quad (6)$$

$$n\lambda^T\pi \le E[c^*(z^{(n)})] \le n\lambda^T\pi + O(1) \quad (7)$$

and

$$|\mathrm{var}[c^*(z^{(n)})] - n\lambda^T V\lambda| = O(1). \quad (8)$$

3. The General VRP with Continuous Distributions

Assume now that the attribute vector $w \in [0, 1]^d$ and the location vector $x = (x_1, x_2)$ take on a continuum of values. In this case, we assume the conditional distributions of w given a specific location x are continuous. The case where these conditional distributions are discrete or where they are given by a mixture of discrete and continuous distributions can be handled with minor adjustments.

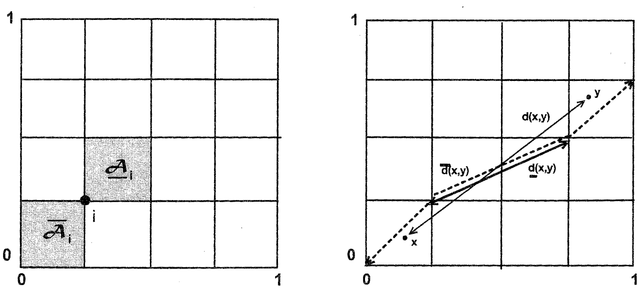
3.1. Discretization Schemes

The idea we use is to discretize the location space $[0, 1]^2$ and the attribute space $[0, 1]^d$ and use this grid to approximate locations and attributes. This produces a problem with a finite set of customer types which can then be analyzed using the results of the previous section. We then analyze a sequence of discretized problems formed by letting the grid size tend to zero.

To discretize attributes we divide, as in Federgruen and van Ryzin (1992), each coordinate of the vector w into I intervals, each of length $h = 1/I$, for I a nonnegative integer. The aggregation scheme produces $(1 + I)^d$ different attribute vectors, corresponding with the grid points $\{0, 1, \ldots, I\}^d$ on the attribute space $[0, 1]^d$.

Let $\bar{\mathcal{A}}_i$ ($\underline{\mathcal{A}}_i$) be the subcube of all points which are transformed into grid point i when their components are rounded up (down) to the nearest multiple of h. That is, $\bar{\mathcal{A}}_i = \{w = (w^1, \ldots, w^d) : h(\lceil w^1/h\rceil, \ldots, \lceil w^d/h\rceil) \leftrightarrow i\}$ and $\underline{\mathcal{A}}_i = \{w = (w^1, \ldots, w^d) : h(\lfloor w^1/h\rfloor, \ldots, \lfloor w^d/h\rfloor) \leftrightarrow i\}$. We extend this definition to include subcubes that are not subsets of $[0, 1]^d$. Examples of these sets for the case $d = 2$ are shown in Figure 2.

Figure 2 **Examples of the Sets $\bar{\mathcal{A}}_i$ and $\underline{\mathcal{A}}_i$ and the Distance Metrics $\underline{d}(x, y)$, $d(x, y)$ and $\bar{d}(x, y)$**



We discretize the locations x using a slightly different scheme. Partition each side of the unit square into $I = 1/h$ equal sized intervals forming I^2 subsquares of size $h \times h$. We associate locations with *subsquares* rather than grid points. Let $\{S_i, i = 1, \ldots, (1/h)^2\}$ denote the subsquares and consider a customer to be at location i if $x \in S_i$. We let i_x denote the index of the subsquare i containing x. Define the distance between sets S_i and S_j by $d(S_i, S_j) = \min\{d(x, y) : x \in S_i, y \in S_j\}$, and let $\|(h, h)\|$ denote the diameter of a subsquare. For all $i, j \in \{1, \ldots, (1/h)^2\}$ we define two distance functions on this finite set of locations:

$$\underline{d}(i, j) = d(S_i, S_j) \quad \text{and}$$

$$\bar{d}(i, j) = \underline{d}(i, j) + 2\|(h, h)\| = \underline{d}(i, j) + 2h\|(1, 1)\|, \quad (9)$$

where the last equality follows from $d(\cdot, \cdot)$ being homogeneous of degree one. We also use the notation $\bar{d}(x, y) \doteq \bar{d}(i_x, i_y)$ and $\underline{d}(x, y) \doteq \underline{d}(i_x, i_y)$ when referring to distances as a function of the original locations x, y. Clearly, $\underline{d}(x, y) \le d(x, y) \le \bar{d}(x, y)\ \forall\ x, y \in [0, 1]^2$. The geometrical relationship between these distance functions is illustrated in Figure 2.

This combined discretization of attributes and locations produces a finite number of customer types, each defined by a particular attribute grid point and a particular subregion location, and indexed by $i, i = 1, \ldots, I'$ where $I' = I^2(I + 1)^d = O(1/h^{d+2})$. Abusing our notation somewhat, we use $\bar{\mathcal{A}}_i$, ($\underline{\mathcal{A}}_i$) to denote the set of vectors w that when rounded up (down) produce the attribute grid point of a type i customer, and S_i to denote the subset of locations corresponding to a type i customer. With this notational convention S_i and S_j may refer to the same subregion if type i and j customers share a common subregion but have different attribute values. Likewise $\bar{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_j$ may refer to the same region of the attribute space if i and j have the same attribute grid point but different locations.

3.2. Construction of the Upper and Lower Bound Problems

We define two problems on this discrete set of customer types. The first uses the distance metric $\underline{d}(i, j)$ and the second the metric $\bar{d}(i, j)$. The discretization scheme together with a distance function defines a discrete version of the VRP with a finite number of types which, by Property P4, has a finite set of feasible tours in Ω. Thus we obtain two problems of the form

$$\begin{array}{ll} \underline{c}^*(z) = \min \underline{c}^T y & \bar{c}^*(z) = \min \bar{c}^T y \\ \text{s.t.} & \text{s.t.} \\ \underline{A} y \geq z, & \bar{A} y \geq z, \\ y \geq 0, & y \geq 0, \\ y \text{ integer}, & y \text{ integer}, \end{array} \tag{10}$$

where $\underline{A}$ ($\bar{A}$) denotes the matrix whose columns correspond to feasible tours with distance functions $\underline{d}$ ($\bar{d}$) and $\underline{c}$ ($\bar{c}$) the corresponding vector of tour costs with distance function $\underline{d}$ ($\bar{d}$). $\underline{c}^{LP}(z)$ and $\bar{c}^{LP}(z)$ denote the optimal values of the linear programming relaxations of these two problems.

Note by Property P5 that the columns of $\bar{A}$ are a subset of the columns of $\underline{A}$; thus, we can partition $\underline{A}$ as $\underline{A} = [\bar{A} \,|\, B]$, with B the set of columns which are feasible under $\underline{d}$ but not under $\bar{d}$. If tour feasibility does not depend on distances, as in the classical VRP, $\bar{A} = \underline{A}$ (i.e. $B = \emptyset$). By the monotonicity of the cost function in the distance traveled, it follows that for each column j in $\bar{A}$ we have $\underline{c}_j \leq \bar{c}_j$. Thus, for all nonnegative integer z, we have $\underline{c}^*(z) \leq \bar{c}^*(z)$.

To relate these problems to the original continuous problem some more work is needed. Consider a set of customers $S^{(n)} = \{(x_1, w_1), \ldots, (x_n, w_n)\}$ and define for $i = 1, \ldots, I'$, $\bar{z}_i^{(n)} = |\{(x_k, w_k) \in S^{(n)} : x_k \in S_i, w_k \in \bar{\mathcal{A}}_i\}|$, and $\underline{z}_i^{(n)} = |\{(x_k, w_k) \in S^{(n)} : x_k \in S_i, w_k \in \underline{\mathcal{A}}_i\}|$, as the number of type i customers produced by rounding up (down) the attribute vector and discretizing the locations.

LEMMA 2. *If Ω satisfies P1–P5 and the cost function $c(\tau)$ satisfies P7, then*

$$\underline{c}^*(\underline{z}^{(n)}) \leq c^*(S^{(n)}) \leq \bar{c}^*(\bar{z}^{(n)})$$

PROOF. By P4 and P5 any tour which is feasible for the rounded up attributes and the distance function $\bar{d}$ is also feasible for the original attribute values and original distance function d. Furthermore, by P7 the cost is not increased when reducing sizes and distances. Thus $c^*(S^{(n)}) \leq \bar{c}^*(\bar{z}^{(n)})$ and $\underline{c} \leq c^*(S^{(n)})$ follows by a similar argument. □

For our probabilistic analysis, consider two distinct probability measures, $\bar{\pi}$ and $\underline{\pi}$, on the set of I' customer types by defining for $i = 1, \ldots, I'$, $\bar{\pi}_i = \mu(\bar{\mathcal{A}}_i \times S_i) = P\{w \in \bar{\mathcal{A}}_i$ and $x \in S_i\}$, and $\underline{\pi}_i = \mu(\underline{\mathcal{A}}_i \times S_i) = P\{w \in \underline{\mathcal{A}}_i$ and $x \in S_i\}$. We write $\underline{\pi}(h)$ and $\bar{\pi}(h)$ when we want to make the dependence on h explicit. Let $S^{(n)}$ denote a set formed by taking the first n objects of an i.i.d. sequence $\{(x_k, w_k) : k \geq 1\}$ with distribution μ. Note $\underline{z}^{(n)}/n \to \underline{\pi}$ and $\bar{z}^{(n)}/n \to \bar{\pi}$ (a.s.) by the strong law of large numbers; see Revesz (1968) and

$$\gamma(\mu) = \lim_{n \to \infty} \frac{c^*(S^{(n)})}{n} \quad \text{(a.s.)} \tag{11}$$

by Kingman's theory of subadditive ergodic processes (Kingman 1976). It follows from Lemma 2 and Theorem 1 that:

LEMMA 3. $\underline{c}^{LP}(\underline{\pi}(h)) \leq \gamma(\mu) \leq \bar{c}^{LP}(\bar{\pi}(h))$.

We must therefore compare the two linear program values in Lemma 3. We will show that they converge under appropriate conditions on the distribution μ and for h tending to zero. Note first that the upper bound in Lemma 3 is constructive and corresponds to the asymptotic value of the following heuristic:

Math Programming (MP) Heuristic

1. Fix $h > 0$.
2. Round attributes up to the nearest multiple of h and discretize locations based on h as described above to form $I = h^{-2}(1 + 1/h)^d$ customer types.
3. Count the number of customers of each type and form the vector $\bar{z}^{(n)}$.
4. Replace distances by $\bar{d}$ and generate the set of feasible columns $\bar{A}$.
5. Solve the corresponding linear program. Round the basic variables up to the nearest integer to produce a feasible solution.

For fixed h, the dimension of the linear program in Step 5 is fixed, and thus the running time of the heuristic is $O(n \log n)$ as in the discrete case. Of course, one could also solve the integer program directly in the same time since its size too is fixed. However, the linear-programming-based upper bound is more efficient to compute and more convenient to analyze.

3.3. Analysis of the Discretized VRP

As mentioned above, we assume that the conditional distributions of the attribute vector w given x are continuous, with density functions $f(w \mid x)$. Let $\mu_x(\cdot)$ denote the marginal probability measure of the locations x. By definition, we take $f(w \mid x) = 0$ if $(x, w) \notin [0, 1]^{d+2}$. We begin by relating the probability measures $\underline{\pi}$ and $\overline{\pi}$ in Lemma 4. (See the Appendix for a proof.) To do so, we need:

DEFINITION 1. A density function $f: [0, 1]^d \to [0, \infty)$ is *Lipschitz continuous of order* s on a set $\mathcal{A} \subseteq R^d$ if

$$|f(x) - f(y)| \le K|x - y|^s \quad \forall\, x, y \in \mathcal{A},$$

where we define $f(w) = 0$ for $w \notin [0, 1]^d$.

For example, triangular densities on [0, 1] are first order Lipschitz continuous everywhere while the uniform density is not Lipschitz continuous on R^d since it is discontinuous at the boundary of the unit cube; however, the uniform density is Lipschitz continuous of all orders on the interior of the unit cube.

LEMMA 4. (a) *Let* $D = \{(x, w) \in R^{d+2} : f(w \mid x)$ *is discontinuous in* $w\}$. *If* $\mu(D) = 0$, *then* $|\underline{\pi}(h) - \overline{\pi}(h)| \to 0$ *as* $h \to 0$.

(b) *If the densities* $f(w \mid x)$ *are uniformly bounded on* R^{d+2} *and first order Lipschitz continuous in* w *on the interior of* $[0, 1]^d$ *for every* x, *then* $|\underline{\pi}(h) - \overline{\pi}(h)| = O(h)$.

(c) *If the densities* $f(w \mid x)$ *are uniformly bounded on* R^{d+2} *and Lipschitz continuous of order* s *in* w *everywhere on* R^d *for every* x *then* $|\underline{\pi}(h) - \overline{\pi}(h)| = O(h^s)$.

Part (b) above is satisfied whenever the marginal densities $f(w \mid x)$ are continuously differentiable on the interior of the unit d-cube, and the result can be generalized to the case where the densities have discontinuities on the interior provided these occur only along a finite number of sufficiently smooth surfaces. Part (c) shows that the rate of convergence of $|\underline{\pi}(h) - \overline{\pi}(h)|$ is a function of the smoothness of the underlying density.

The above lemma allows us to characterize the asymptotic value of the continuous version of the generalized VRP. Namely,

THEOREM 4. *Suppose* μ *satisfies the conditions of Lemma 4 a). Then* $\lim_{n\to\infty} c^*(S^{(n)})/n = \gamma(\mu)$ *(a.s.) where* $\gamma(\mu) \equiv \lim_{h\to 0} \underline{c}^{LP}(\underline{\pi}(h))$, *and the MP heuristic can be designed to be (a.s.)* ϵ*-optimal as* $n \to \infty$ *by an appropriate choice of* h. *In addition, if* μ *satisfies the conditions of Lemma 4 b), then for any* $h > 0$,

$$\underline{c}^{LP}(\underline{\pi}(h)) \le \gamma(\mu) \le \underline{c}^{LP}(\underline{\pi}(h)) + O(h).$$

PROOF. *Case* 1. $B = \emptyset$: We have for every tour τ,

$$\bar{c}(\tau) \le \underline{c}(\tau) + Kh|\tau|, \tag{12}$$

for some constant K since by (9) each distance l_k, $k = 1, \ldots, |\tau|$ is increased by at most $2h\|(1, 1)\|$ when replacing the distance function $\underline{d}$ by $\bar{d}$, and by Property P7 the cost $c(\tau)$ is first order Lipschitz in these distances. Consider now $\underline{c}^{LP}(\overline{\pi}(h))$ and let y^* be an optimal solution of the associated LP. Note y^* is feasible for $\bar{c}^{LP}(\overline{\pi}(h))$ since $B = \emptyset$. In view of (12) and Lemma 1 d), we have

$$\begin{aligned}\bar{c}^{LP}(\overline{\pi}(h)) &= \bar{c}^T y^* \le \underline{c}^T y^* + Kh \sum_{i=1}^{I'} \overline{\pi}_i(h) \\ &= \underline{c}^{LP}(\overline{\pi}(h)) + Kh \le \underline{c}^{LP}(\underline{\pi}(h)) \\ &\quad + C|\underline{\pi}(h) - \overline{\pi}(h)| + Kh.\end{aligned}$$

In view of Lemma 3, we have $\gamma(\mu) - C|\underline{\pi}(h) - \overline{\pi}(h)| - Kh \le \underline{c}^{LP}(\underline{\pi}(h)) \le \gamma(\mu)$, and the result follows from Lemma 4.

Case 2: $B \ne \emptyset$. As mentioned, in this case we describe customers by a triple (x, s, w). To construct an upper bound on $\bar{c}^*(\bar{z}^{(n)})$, consider a tour $\tau = \{(x_k, s_k, w_k) : 1, \ldots, |\tau|\}$ that is feasible under the distance metric $\underline{d}(x, y)$. Let $a = \|(1, 1)\|$ and choose h small enough that $s_k \ge \lceil 2a/\nu \rceil h$, $k = 1, \ldots, |\tau|$. Then by P6 and (9), the tour $\tau' = \{(x_k, s_k - \lceil 2a/\nu \rceil h, w_k) : 1, \ldots, |\tau|\}$ is feasible under the distance metric $\bar{d}(x, y)$. For each such tour τ, define a new tour cost $\bar{\bar{c}}(\tau) = \underline{c}(\tau) + K(2ah + \lceil 2a/\nu \rceil h)|\tau|$, where K is the Lipschitz constant from property P7. Note also that by P6, P7 and the definition of the cost $\bar{\bar{c}}(\tau)$ that for all τ and τ' defined above,

$$\begin{aligned}\bar{c}(\tau') &\le \underline{c}(\tau') + K(2ah)|\tau| \\ &\le \underline{c}(\tau) + K\lceil 2a/v \rceil h|\tau| + K(2ah)|\tau| = \bar{\bar{c}}(\tau).\end{aligned} \tag{13}$$

Now for a given $h > 0$, consider the following construction:

1. Remove all customers in the set $\Delta_h(S^{(n)}) = \{(x_k, s_k, w_k) : s_k + h\lceil 2a/\nu \rceil > 1\}$. Service each customer in $\Delta_h(S^{(n)})$ with a separate vehicle.

2. Set $s_k \leftarrow s_k + \lceil 2a/\nu \rceil h$ for all $k \notin \Delta_h(S^{(n)})$. Using these modified values of s_k, apply the MP discretization scheme to these remaining customers, forming a vector $\bar{\bar{z}}^{(n)}$ of discrete types. More formally, let $\mathcal{A}_i = \bar{T}_i \times \bar{U}_i$, where $\bar{T}_i$ ($\bar{U}_i$) denotes the set of values s (w) that, when rounded up, correspond to those of a type i object. Then for $i = 1, \ldots, I'$, $\bar{\bar{z}}_i^{(n)} = |\{(x_k, s_k, w_k) \in S^{(n)}\Delta_k(S^{(n)}) : x_k \in S_i, w_k \in \bar{U}_i, s_k + \lceil 2a/\nu \rceil h \in \bar{T}_i\}|$.

3. Solve the problem

$$\bar{\bar{c}}^*(\bar{\bar{z}}^{(n)}) = \min\{\bar{\bar{c}}^T y : \underline{A} y \geq \bar{\bar{z}}^{(n)}, y \geq 0, y \text{ integer}\} \quad (14)$$

The problem solved in Step 3 uses the columns of the lower bound matrix $\underline{A}$ along with the cost vector $\bar{\bar{c}}$. By P6, any tour generated by this problem will be feasible for the original values (x_k, s_k, w_k). Considering the tours that are optimal in the program $\underline{c}^*(\bar{\bar{z}}^{(n)})$, we get

$$\bar{\bar{c}}^*(\bar{\bar{z}}^{(n)}) \leq \underline{c}^*(\bar{\bar{z}}^{(n)}) + K(2a + \lceil 2a/\nu \rceil)hn. \quad (15)$$

Considering those that are optimal in (14) we get by (13) and the definition of $\Delta_h(S^{(n)})$ that

$$\bar{c}^*(\bar{z}^{(n)}) \leq \bar{\bar{c}}^*(\bar{\bar{z}}^{(n)}) + C|\Delta_h(S^{(n)})|. \quad (16)$$

Let $\bar{\bar{T}}_i = \{s : s + \lceil 2a/\nu \rceil h \in \bar{T}_i\}$ denote the set of s values that, when shifted and rounded up, produce the same s value as a type i customer. Define a new measure on the set of discrete types, $\bar{\bar{\pi}}_i = \mu(S_i \times \bar{\bar{T}}_i \times \bar{U}_i)$. Combining (15) and (16), dividing both sides above by n and taking limits we obtain (a.s.) by Theorem 1,

$$\begin{aligned} \bar{c}^{LP}(\bar{\pi}) &\leq \underline{c}^{LP}(\bar{\bar{\pi}}) + CP\{s > 1 - \lceil 2a/\nu \rceil h\} + C'h \\ &\leq \underline{c}^{LP}(\underline{\pi}) + C|\bar{\pi} - \underline{\pi}| + C|\bar{\bar{\pi}} - \bar{\pi}| \\ &\quad + CP\{s > 1 - \lceil 2a/\nu \rceil h\} + C'h \end{aligned} \quad (17)$$

where $C' \doteq C(2a + \lceil 2a/\nu \rceil)$. Clearly, $C'h < \epsilon/5$ for $h < \epsilon/5C'$.

Also, by the continuity of $f(w|x)$, we can select $h > 0$ so that $C|\bar{\pi} - \underline{\pi}| < \epsilon/5$, by Lemma 4 a), and $C|\bar{\bar{\pi}} - \bar{\pi}| < \epsilon/5$ by recognizing that for all $i = 1, \ldots, I'$, $\bar{\bar{\pi}}_i$ and $\bar{\pi}_i$ are obtained by integrating the same probability measure over regions that are shifted by only an amount $\lceil 2a/\nu \rceil h$ along the s-axis. Again, by the continuity of f, $P\{s > 1 - \lceil 2a/\nu \rceil h\} = O(h)$ so that the last term in (17) can also be made less that $\epsilon/5$ for an appropriate choice of $h > 0$. The rest of the proof follows as in Case 1. □

So far we have characterized the performance of the MP heuristic for a fixed discretization level $h > 0$. In particular, we have shown for any $\epsilon > 0$, that the heuristic is asymptotically ϵ-optimal (a.s.) when choosing a fixed h sufficiently small, and that its running time is $O(n \log n)$. We now show that the heuristic can be made asymptotically (fully) optimal while maintaining polynomial complexity by choosing h as a function of n:

THEOREM 5. *Suppose μ satisfies the conditions of Lemma 4 a). Choose h as function of n such that $nh^{d+2} \to \infty$. Then the MP heuristic is both polynomial in n and asymptotically optimal (a.s.).*

PROOF. Recall that every feasible tour τ has $|\tau| \leq M$ and $I' = O(h^{-(d+2)})$. The number of columns in the linear program solved in the MP heuristic is therefore bounded by the number of ways in which M objects can be spread over I' types, i.e., I'^M. Then, using any of several interior point methods (see Goldfarb and Todd 1989), the dual of the linear program can be solved in $O(I'^4 L)$ where $L = O(I'^M)$ is the length of the input string. The complexity is thus $O(I'^{M+4}) = O(h^{-(d+2)(M+4)}) = o(n^{M+4})$ since $nh^{d+2} \to \infty$. □

The above analysis can be extended to obtain upper bounds on the rate of convergence of $E[c^*(S^{(n)})]$ to its asymptotic value. These rates of convergence also apply to the MP heuristic. The results require that the density be Lipschitz of order one or higher.

THEOREM 6. a) *Suppose μ has conditional densities $f(w|x)$ satisfying the conditions of Lemma 4(b). Then*

$$n\gamma(\mu) \leq E[c^*(S^{(n)})] \leq n\gamma(\mu) + O(n^{(d+3)/(d+4)}).$$

b) *If the conditional densities $f(w|x)$ are Lipschitz of order s in w everywhere and the cost function $c(\tau)$ is also Lipschitz of order s in the distances l_k, $k = 1, \ldots, |\tau|$, then*

$$n\gamma(\mu) \leq E[c^*(S^{(n)})] \leq n\gamma(\mu) + O(n^{(s+d+2)/(2s+d+2)}).$$

PROOF. We only show part b), since the proof of part a) is analogous. First, the lower bound follows from the fact that $E[c^*(S^{(n)})] \geq E[\underline{c}^{LP}(\underline{z}^{(n)})] \geq \underline{c}^{LP}(n\underline{\pi}(h)) = n\underline{c}^{LP}(\underline{\pi}(h))$ by Jensen's inequality (e.g., Royden (1968),

p. 110) and Lemma 1. Letting $h \to 0$, we have by Theorem 4 that $E[c^*(S^{(n)})] \geq n \lim_{h\to 0} \underline{c}^{LP}(\underline{\pi}(h)) = n\gamma(\mu)$.

For the second inequality, consider first the case where $\underline{A} = \bar{A}$. Proceeding as in Theorem 4 and using Lemma 4 and the fact that the conditional densities and cost function are Lipschitz order s, we obtain, by Lemma 1, that

$$\begin{aligned} E[c^*(S^{(n)})] &\leq E[\bar{c}^{LP}(\bar{z}^{(n)})] + I' \leq n\bar{c}^{LP}(\bar{\pi}(h)) \\ &\quad + CE[|\bar{z}^{(n)} - n\bar{\pi}(h)|] + I' \\ &\leq n\underline{c}^{LP}(\underline{\pi}(h)) + nC|\underline{\pi}(h) - \bar{\pi}(h)| \\ &\quad + nKh^s + CE[|\bar{z}^{(n)} - n\bar{\pi}(h)|] + I' \\ &\leq n\gamma(\mu) + O(nh^s) + nKh^s \\ &\quad + O((n/h^{d+2})^{1/2}) + O(1/h^{d+2}), \end{aligned}$$

since

$$\begin{aligned} E[|\bar{z}^{(n)} - n\bar{\pi}(h)|] &= \sum_{i=1}^{I'} E[|\bar{z}_i^{(n)} - n\bar{\pi}_i(h)|] \\ &\leq \sum_{i=1}^{I'} \sqrt{E[(\bar{z}_i^{(n)} - n\bar{\pi}_i(h))^2]} \\ &= \sum_{i=1}^{I'} \sqrt{n\bar{\pi}_i(h)(1 - \bar{\pi}_i(h))} \\ &\leq \sum_{i=1}^{I'} \sqrt{n\bar{\pi}_i(h)} \leq \sqrt{nI'}, \end{aligned}$$

where the last inequality follows from $\Sigma_i \sqrt{\pi_i}$ being a concave separable function of the π_i probabilities with $\Sigma_i \pi_i = 1$. The convergence rate in part b) above is then obtained by letting $h = n^{\alpha}$ and selecting α to minimize the largest exponent in the error terms. Doing this yields $h = n^{-1/(2s+d+2)}$, from which the convergence rate follows.

For the case where feasibility of the tours depends on the distances, we can proceed the same way by using the bound (17) in the proof of Theorem 4, noting that the last term in (17) can be improved to $C'h^s$ due to the Lipschitz condition on the cost function. Similarly, the fourth term in (17) is $O(h^s)$ by the Lipschitz condition on the densities. The rest of the proof then follows as in the case where $\underline{A} = \bar{A}$. □

A tail probability result can be obtained using an argument similar to that in Theorem 2 (cf. Federgruen and van Ryzin 1992). Specifically,

THEOREM 7. *Under the conditions of Theorem 4, for every $\epsilon > 0$, $\alpha > 0$ there exists a $n_0 = n_0(\epsilon, \alpha)$ such that*

$$P\left\{\left|\frac{c^*(S^{(n)})}{n} - \gamma(\mu)\right| > \epsilon\right\} \leq 6e^{-n\epsilon^2/25(1+\alpha)}$$

for all $n \geq n_0$, where $\gamma(\mu) = \lim_{h\to 0} \underline{c}^{LP}(\underline{\pi}(h))$.

REMARK. The above convergence rates and tail probability bounds also apply to the LP relaxation of the set covering formulation by our observations in §3.4.

3.4. Relationship to the LP Relaxation of the Classical Set Covering Formulation

Bramel and Simchi-Levi (1993) recently showed for a time-window VRP that the LP relaxation, c^{SCLP}, of the set covering formulation (2)–(4) is asymptotically optimal. Here we show, as a simple corollary of Theorem 4, that the same result applies to our general class of VRPs.

COROLLARY 1. *Under the conditions stated in Theorem 4, the value of the linear programming relaxation of the classical set covering formulation, c^{SCLP}, satisfies*

$$\lim_{n\to\infty} \frac{c^{SCLP}(S^{(n)})}{n} = \gamma(\mu) \quad (a.s.).$$

In addition, Theorems 6 and 7 apply to $c^{SCLP}(S^{(n)})$, under the conditions stated there.

PROOF. The result follows by verifying that $\underline{c}^{LP}(\underline{z}^{(n)}) \leq c^{SCLP}(S^{(n)}) \leq c^*(S^{(n)})$. The second inequality follows since $c^{SCLP}(\cdot)$ is a LP-relaxation of $c^*(\cdot)$ and the first is obtained by showing that $\underline{c}^{LP}(\underline{z}^{(n)})$ is obtained from $c^{SCLP}(S^{(n)})$ by a series of relaxation steps. First, aggregate for all $i = 1, \ldots, I'$ all constraints in (2)–(4) that correspond to customers with $(x, w) \in S_i \times \underline{\mathcal{A}}_i$. The τ-th column in the resulting aggregated constraint matrix is identical to a specific column in $\underline{A}$, say column $j(\tau)$, and $\underline{c}_{j(\tau)} \leq c_\tau$ by P7. Replacing for $\tau = 1, \ldots, m$ the objective function coefficients c_τ by $c_{j(\tau)}$ results in a second relaxation. Finally, $\underline{c}^{LP}(\underline{z}^{(n)})$ is obtained by adding all columns of $\underline{A}$ which do not correspond to a column in the aggregated constraint matrix of c^{SCLP}, a third relaxation. (A specific column $\underline{A}_j$ and corresponding coefficient $\underline{c}_j$

may appear multiple times in the resulting linear program.) □

4. Alternative Implementation of the MP Heuristic

The complexity of the MP heuristic is determined by the effort required to solve the linear program in Step 5. We have shown that the complexity can be limited to $O(n \log n)$ (to be polynomial in n) all while guaranteeing asymptotic ϵ-optimality (full optimality). These bounds apply when the LP is solved by a standard (interior point) method. Unfortunately, the size of the LP can grow very quickly in h^{-1}, unless M, the number of customers in a single tour, is explicitly or implicitly restricted to be small, for example 2–4 stops per tour. (This indeed often arises, see e.g. Bell et al. (1983) or implicitly Cullen et al. (1981).) In this section, we discuss alternative implementations for settings where M is large.

First, the LP may be solved by column generation techniques. Bell et al. (1983) and Cullen et al. (1981) describe how large scale instances of c^{SCLP} can be solved effectively via such methods. From a theoretical, worst case point of view, we showed in Federgruen and van Ryzin (1992) that if generating columns is polynomial in $1/h$, then by using the Grötschel-Lovasz-Schrijver (GLS) (1981, 1988) ellipsoid algorithm, so is the entire LP. Karmarkar and Karp (1982), used the same approach for the worst-case analysis of classical bin packing. To solve an LP with I' variables within a fixed tolerance ϵ, the GLS algorithm requires at most $O(I'^2 \log(1/\epsilon))$ calls to a column generating routine and $O(I'^4 \log(1/\epsilon))$ time per call (excluding the running time of the column generating routine itself). Several examples of polynomial time column generation procedures for some packing problems are given in Federgruen and van Ryzin (1992). See Karmarkar and Karp (1982) for more details on the GLS algorithm. This method allows us to achieve better complexity bounds on the MP heuristic in special cases.

Consider first the classical VRP. In a discretization of this problem, each type i corresponds to a load size, $w_i = i/h$, $i = 1, \ldots, I'$, in a particular subregion S_i. Each feasible tour τ is described by a column j of $\bar{A}$, with $\sum_{i=1}^{I'} a_{ij} w_i \leq 1$. All nonnegative, integer vectors $(a_{1j}, \ldots, a_{I'j})$ that satisfy the above inequality are contained in $\bar{A}$. The cost of the column, $\bar{c}_j$, is the length, under distance metric $\bar{d}(x, y)$, of the shortest tour through the set of customers corresponding to column j.

A column generation routine verifies for an initial feasible basis and associated vector of dual prices y whether the basis is optimal. This is the case iff $y^T \bar{A}_j \leq \bar{c}_j$ for all feasible tours $j = 1, \ldots, J$, which can be checked by solving the following dynamic program:

$$V(w, i) = \max\{y_{(k,j)} - \bar{d}(i, j) + V(w - kh, j) : \\ k = 0, 1, \ldots, 1/h, j = 0, 1, \ldots, (1/h)^2, kh \leq w\}$$

with the boundary condition $V(0, i) = d(i, 0)$, $i = 0, 1, \ldots, (1/h)^2$, where i indexes the sets of the partitioned service region, $i = 0$ indicates the depot subregion, k indexes the discretized load sizes, $y_{(k,i)}$ is the dual value corresponding to a customer with load size kh in subregion i and $\bar{d}(i, j)$ is the distance from subregion i to subregion j using the metric $\bar{d}$. $V(w, i)$ is the optimal value given a remaining capacity of w and a current location i; hence, $V(1, 0)$ is the maximum value of $y^T \bar{A}_j - \bar{c}_j$. If $V(1, 0) > 0$, then the dynamic programming solution can be traced back to identify a column to be added to the LP. If $V(1, 0) \leq 0$, the current LP solution is optimal. Note the running time of the dynamic program is polynomial in $1/h(O((1/h)^6))$.

One can use the GLS algorithm with the above column generating DP to solve the dual of the linear program in the MP heuristic. If we let h decrease at the rate $n^{-1/5}$, then the error term in Theorem 6 a) of $O(n^{4/5})$ is obtained, provided one chooses $\epsilon = O(n^{-1/5})$. This gives a polynomial running time of $O(n^{3.6} \log^2(n))$. One could improve on this running time by letting h and ϵ decrease more slowly than $n^{-1/5}$ at the expense of a slower convergence rate. In practice, column generation may be more effectively implemented in conjunction with simplex-based or Karmarkar-type interior point methods. (See Cullen et al. 1981, Goffin and Vial 1990 and Ye 1992.)

Similar dynamic programs can be formulated for problems with time windows, pick-up and delivery constraints and inventory routing costs and constraints. However, the dynamic programs become more complex. Nevertheless, using column generation with the MP heuristic provides a theoretically powerful approach to constructing provably asymptotically optimal, polynomial time heuristics for a wide class of VRPs. In addition, it corresponds closely to the math programming approaches that one frequently encounters in practice.

5. Numerical Examples

We have used the MP heuristic to generate solutions to several variants of the example in Figure 1. Though stylized, this example is easy to understand intuitively; hence, it is a useful aid in developing insights. It also illustrates some fundamental problem characteristics that impact on the heuristic's numerical performance.

5.1. Description of Experiments and Observations

In the example, we used three values for h ($h = 1, \frac{1}{2}$ and $\frac{1}{4}$), three values of y (5, 10 and 100) and three values of n (32, 320 and 3,200). For each combination of parameters, we generated a random instance and computed the values of the MP heuristic solution (c_{MP}), the value of the partition heuristic (c_{PART}), the linear program relaxation of both the upper and lower bound problems, ($\bar{c}^{LP}(\bar{z}^{(n)}(h))$ and $\underline{c}^{LP}(\underline{z}^{(n)}(h))$, and the asymptotic estimates $n\underline{c}^{LP}(\underline{\pi}(h))$. An example solution for the case $h = \frac{1}{2}$ is shown in Figure 1.

We constructed the complete set of feasible columns off-line. CPU times on a 486 66 Mhz PC (using the LINDO solver) were quite modest: well under one second for $h = 1$, under 1.5 seconds for $h = \frac{1}{2}$ and approximately 15 seconds for $h = \frac{1}{4}$. These times were independent of the number of customers n. Tables 1–3 show the numerical results.

Several observations can be made based on these data. First, the performance of the MP heuristic relative to the lower bound $\underline{c}^{LP}(\underline{z}^{(n)}(h))$ improves as y increases. This is expected, since the tour lengths under the two distance metrics $\bar{d}$ and $\underline{d}$—and hence the values of the upper and lower bound problems—are closer when y is large.

As the grid size h is reduced, the solution value of the MP heuristic decreases while the lower bound, $\underline{c}^{LP}(\underline{z}^{(n)}(h))$, increases, improving the optimality gap. This gap ranges from 50% for the case $y = 5$ and $h = 1$ to about 0.7% when $y = 100$ and $h = 1/4$. The optimality gap is less sensitive to the value of h when y is large for the reasons discussed above.

By comparing the values of c_{MP} and $\bar{c}^{LP}(\bar{z}^{(n)}(h))$, one gauges the error introduced by rounding in the MP heuristic. For these examples, the error is minimal. Indeed, in most cases the LP solution was integral and no rounding was required. Only for the case $y = 5$ and $n = 320$ was some rounding error introduced, and even this is quite small (less than 0.3% for all values of h). As a result, the optimality gap is not strongly affected by the number of customers n, especially for h small (see e.g. $h = \frac{1}{4}$). However, these characteristics are dependent on the problem, as we discuss in more detail below.

Note finally, the partitioning heuristic cost is significantly higher than the MP cost in all cases, approaching 50% when $y = 100$, as expected. This performance is due to the MP heuristic forming routes that match one customer in Region A with one customer in Region B (see Figure 3); such routes are not considered in a pure partitioning approach. The performance of the partitioning heuristic actually deteriorates in several cases when the grid size h is reduced (see e.g. $y = 100$, $n = 32$). This occurs because when n is small, many subregions in

Table 1 **Numerical Results:** $y = 5$

	n = 32			n = 320			n = 3200		
	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4
c_{PART}	337.7	308.0	316.9	3,253.6	2,824.1	2,746.4	33,717	29,401	28,141
% Over LB	91.0%	58.5%	56.1%	86.5%	49.2%	39.5%	84.3%	46.7%	34.5%
c_{MP}	267.3	239.6	225.6	2,627.2	2,336.9	2,193.5	27,520	24,647	23,220
% Over LB	51.2%	23.3%	11.1%	50.6%	23.5%	11.4%	50.4%	23.0%	11.0%
$\bar{c}^{LP}(\bar{z}^{(n)}(h))$	267.3	239.6	225.6	2,620.1	2,330.7	2,187.1	27,520	24,647	23,220
$\underline{c}^{LP}(\underline{z}^{(n)}(h))$	176.8	194.3	203.0	1,743.9	1,892.7	1,968.1	18,293	20,035	20,914
$n\underline{c}^{LP}(\underline{\pi}(h))$	176.8	193.9	202.5	1,768.0	1,938.9	2,025.0	17,680	19,389	20,249

Table 2 **Numerical Results:** $y = 10$

	n = 32			n = 320			n = 3200		
	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4
c_{PART}	590.5	575.7	624.5	5,796.1	5,460.0	5,385.4	56,412	52,928	51,692
% Over LB	57.7%	47.5%	57.3%	63.1%	47.1%	42.0%	26.8%	50.2%	43.3%
c_{MP}	465.0	435.5	419.6	4,459.3	4,163.5	4,019.1	42,589	39,733	38,308
% Over LB	24.2%	11.6%	5.7%	25.5%	12.2%	6.0%	26.8%	12.8%	6.2%
$\bar{c}^{LP}(\bar{z}^{(n)}(h))$	465.0	435.5	419.6	4,459.3	4,163.5	4,019.1	42,589	39,733	38,308
$\underline{c}^{LP}(\underline{z}^{(n)}(h))$	374.4	390.2	397.0	3,554.2	3,710.9	3,792.8	33,598	35,237	36,060
$n\underline{c}^{LP}(\underline{\pi}(h))$	336.4	353.0	361.3	3,363.8	3,529.6	3,612.8	33,638	35,298	36,128

Region A contain only one point, making it worthwhile to form tours that combine points from two different subregions. Such tours are not considered in the partitioning heuristic. When $h = 1$, all points are in the same subregion and maximal pairing can take place. When y is large, the additional cost introduced by the coarser grid size is more than offset by the reduction in the number of tours. The MP heuristic, on the other hand, allows pairing across subregions, and hence does not suffer from this effect; its performance consistently improves as the grid size h is reduced.

5.2. Rounding Error

Another series of examples shows that the error from rounding the LP solution, which was small in the previous cases, can be highly dependent on the problem characteristics. Consider Figure 3 with 15 randomly distributed customers and the solution produced by the MP heuristic. The thin lines represent tours with one stop in Region A and one stop in Region B (*mixed tours*); the thick lines represent tours that visit one region only (*pure-A* or *pure-B* tours). Note in this case nearly all customers are served using mixed tours; only one customer in Region B is served with a pure-B tour. The cost of this collection of tours is 203.8. (See Table 4.)

Consider the same set of customer locations, but with customers in Region A having a weight of 0.1 and customers in Region B having weight 0.9. Note the same solution in Figure 3 is still feasible for these new weights; however, the MP heuristic generates a more costly solution (19% higher; see Table 4) with more pure tours, as shown in Figure 4. Figure 5 shows the MP solution when the weights in Region A (B) are 0.01 (0.99). It uses only pure-A and pure-B tours, and the cost is 48% higher than the Figure 3 solution. (See Table 4.)

Table 3 **Numerical Results:** $y = 100$

	n = 32			n = 320			n = 3200		
	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4	h = 1	h = 1/2	h = 1/4
c_{PART}	5,081.6	5,254.2	5,842.8	49,003	49,121	49,184	486,605	483,925	483,349
% Over LB	40.6%	44.8%	60.7%	50.5%	50.1%	49.9%	51.3%	49.7%	49.2%
c_{MP}	3,704.5	3,674.4	3,658.5	33,463	33,169	33,026	330,621	327,655	326,200
% Over LB	2.5%	1.2%	0.6%	2.8%	1.4%	0.7%	2.8%	1.4%	0.7%
$\bar{c}^{LP}(\bar{z}^{(n)}(h))$	3,704.5	3,674.4	3,658.5	33,463	33,169	33,026	330,621	327,655	326,200
$\underline{c}^{LP}(\underline{z}^{(n)}(h))$	3,614.0	3,629.1	3,635.9	32,558	32,717	32,800	321,590	323,140	323,943
$n\underline{c}^{LP}(\underline{\pi}(h))$	3,216.0	3,232.1	3,240.1	32,160	32,321	32,401	321,603	323,210	324,013

Figure 3 **MP Solution: Weight** A = 0.4, **Weight** B = 0.6

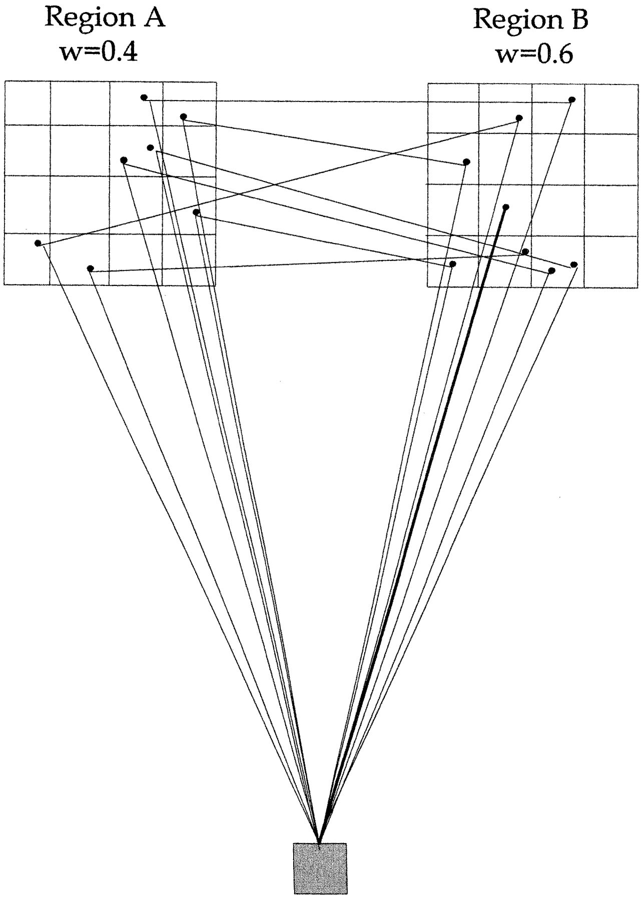
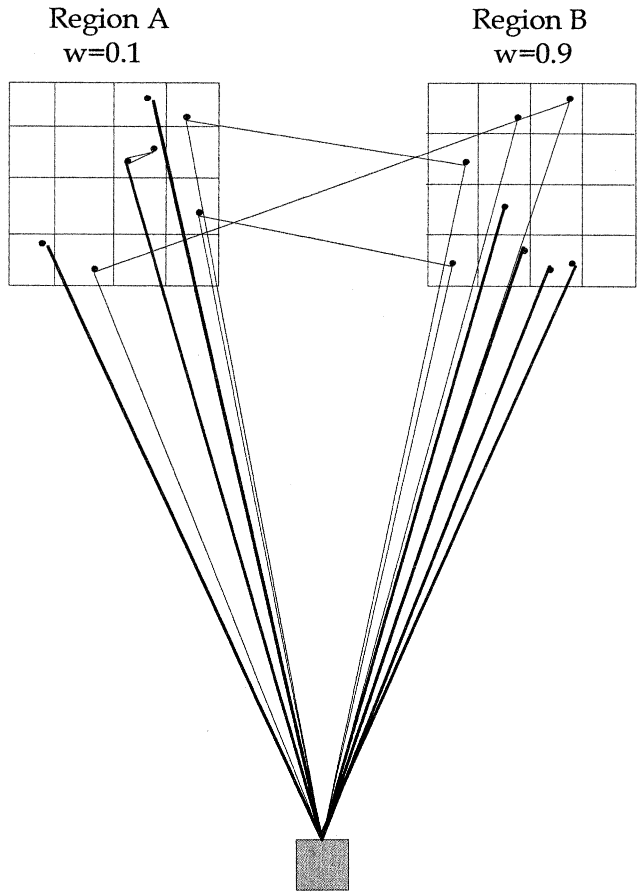


Figure 4 **MP Solution: Weight** A = 0.1, **Weight** B = 0.9



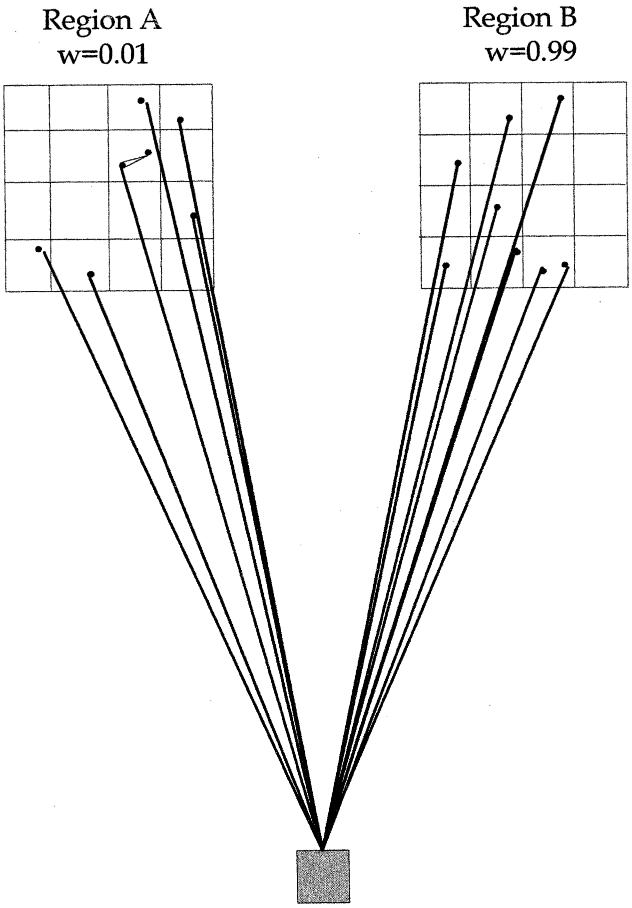
To understand this behavior, note that in varying the weights, the set of feasible mixed and pure-B tours remains the same; however, new pure-A tours are introduced; e.g. with weights in Region A (B) of 0.01 (0, 99), tours that visit 100 customers in a single subregion of A become feasible. Thus, the LP solution can cover each A customer with 1/100-th of a column of this type—and in fact does so for all but two customers. Rounding these small fractional values up to the nearest integer (1 in this case) significantly increases the solution cost (see Table 4). However, the rounding error goes down as the number of customers increases. For an example with n = 32,000 customers and A (B) weights of 0.01 (0.99), the MP solution also consists of only pure tours; however, the rounding error is only 0.4% and the optimality gap is only 3.8%.

Table 4 **Performance of Heuristics with Varying Region Weights**

Weight A	Weight B	c_{MP}	$\bar{c}^{LP}(\bar{z}^{(n)}(h))$	$\underline{c}^{LP}(\underline{z}^{(n)}(h))$	% Rounding
0.40	0.60	203.8	203.8	193.6	0.0%
0.10	0.90	242.8	185.3	176.7	30.9%
0.01	0.99	301.9	172.2	166.46	75.3%

Finally, a potential rounding problem arises when object sizes are discretized. Consider discretizing continuous weights on [0, 1] using a small grid size h. This produces tours that can cover a large number of small objects (e.g. a tour that visits $1/h$ points of weight h in a single subregion). For a given n—no matter how large—a sufficiently small h will introduce significant rounding errors. This is precisely the reason for Prop-

Figure 5 **MP Solution: Weight A = 0.01, Weight B = 0.99**



erty P3, which guarantees that the maximum number of customers per tour remains bounded as $h \to 0$. This behavior, along with the proper scaling of h as a function of n (e.g. $h = o(n^{-1/(d+2)})$) so that the number of customers of each type increases with n (a.s.), ensures that the number of customers of each type becomes significantly larger than any component in a feasible column (a.s.). Hence, the rounding error in the MP heuristic becomes negligible as $n \to \infty$. For example, if the weights in our example were uniformly distributed on the interval $[\delta, 1]$ for $\delta > 0$, then a tour would contain at most $M = 1/\delta$ customers, regardless of the grid size h.

Appendix

PROOF OF LEMMA 4. First consider the case where the conditional densities $f(w \mid x)$ are continuous in w everywhere in R^d for every value of $x \in R^2$, where by convention we define $f(w \mid x)$ to be zero when x lies outside $[0, 1]^2$ or w lies outside $[0, 1]^d$. Then the mean value theorem implies $\int_{\underline{\mathcal{A}}_i} f(w \mid x) dw = h^d f(\underline{w}_i \mid x)$ and $\int_{\bar{\mathcal{A}}_i} f(w \mid x) dw = h^d f(\bar{w}_i \mid x)$ for some $\underline{w}_i \in \underline{\mathcal{A}}_i$ and some $\bar{w}_i \in \bar{\mathcal{A}}_i$, which may depend on x. Therefore, unconditioning we have $\underline{\pi}_i = \int_{S_i} h^d f(\underline{w}_i \mid x) \mu_x\{dx\}$, $\bar{\pi}_i = \int_{S_i} h^d f(\bar{w}_i \mid x) \mu_x\{dx\}$ and $|\underline{\pi}_i - \bar{\pi}_i| \le \int_{S_i} h^d |f(\bar{w}_i \mid x) - f(\underline{w}_i \mid x)| \mu\{dx\}$. Since continuity of $f(\cdot \mid x)$ implies uniform continuity over the bounded region $[-h, 1 + h]^d$ and $|\bar{w}_i - \underline{w}_i| \le dh$, it follows that for every $\epsilon > 0$ we can select a $\delta > 0$ such that $|f(\bar{w}_i \mid x) - f(\underline{w}_i \mid x)| \le \epsilon/(1 + h)^d$ for all i, x and $h < \delta$. Thus, for $h < \delta$, $|\underline{\pi} - \bar{\pi}| = \Sigma_{i=1}^{I'} |\underline{\pi}_i - \bar{\pi}_i| \le \epsilon h^d/(1 + h)^d \, \Sigma_{i=1}^{I'} \int_{S_i} \mu_x\{dx\}$. It follows by our construction that there are exactly $(1 + 1/h)^d$ customer types that each have locations x corresponding to the same subsquare (one for each attribute type), so the sets S_i are identical for exactly $(1 + 1/h)^d$ indices i, while the collection of *distinct* sets S_i, $i = 1, \ldots, I'$ partitions $[0, 1]^2$. Thus,

$$\sum_{i=1}^{I'} \int_{S_i} \mu_x\{dx\} = (1 + 1/h)^d \tag{18}$$

and therefore $|\underline{\pi} - \bar{\pi}| \le \epsilon(h^d(1 + 1/h)^d)/(1 + h)^d = \epsilon$. This establishes part (a) where the marginal densities $f(w \mid x)$ are continuous in w everywhere on R^d for every x.

Suppose now that some of the conditional densities $f(w \mid x)$ has discontinuities in w but that $\mu(D) = 0$. Define $\mathcal{C}$ to be the set of indices i for which either $(\underline{\mathcal{A}}_i \times S_i)$ or $(\bar{\mathcal{A}}_i \times S_i)$ intersects $\mathcal{D}$ and $\mathcal{C}_2 = \{1, \ldots, I'\} - \mathcal{C}_1$. Then we can write $|\underline{\pi} - \bar{\pi}| = \Sigma_{i \in \mathcal{C}_1} |\underline{\pi}_i - \bar{\pi}_i| + \Sigma_{i \in \mathcal{C}_2} |\underline{\pi}_i - \bar{\pi}_i|$. Since for $i \in \mathcal{C}_2$, $f(w \mid x)$ is continuous over both $\underline{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_i$ for every x, by the same argument as above we can select an h such that the second sum above is no more than $\epsilon/2$. Further, $\Sigma_{i \in \mathcal{C}_1} |\underline{\pi}_i - \bar{\pi}_i| \le \mu\{\cup_{i \in \mathcal{C}_1} S_i \times (\underline{\mathcal{A}}_i \cup \bar{\mathcal{A}}_i)\}$. But the fact that $\mu(D) = 0$ implies that for sufficiently small h, the term $\mu\{\cup_{i \in \mathcal{C}_1} S_i \times (\underline{\mathcal{A}}_i \cup \bar{\mathcal{A}}_i)\}$ can be made arbitrarily small since $\lim_{h \to 0} \mu\{\cup_{i \in \mathcal{C}_1} S_i \times (\underline{\mathcal{A}}_i \cup \bar{\mathcal{A}}_i)\} = \mu(D)$. Thus, h can be chosen small enough so that the first term is no more that $\epsilon/2$ as well, which combined with the bound on the second term implies $|\underline{\pi} - \bar{\pi}| \le \epsilon$. This completes the proof of part (a).

To prove part (b), note that if the densities $f(w \mid x)$ are Lipschitz continuous of order one on the interior of $[0, 1]^d$ for all x, then for all indices i with both $\underline{\mathcal{A}}_i$ and $\bar{\mathcal{A}}_i$ subsets of $[0, 1]^d$, we have as before $|\underline{\pi}_i - \bar{\pi}_i| \le \int_{S_i} h^d |f(\bar{w}_i \mid x) - f(\underline{w}_i \mid x)| \mu_x\{dx\} \le K h^{d+1} \int_{S_i} \mu_x\{dx\}$. The total contribution from all such indices to $|\underline{\pi} - \bar{\pi}|$ is therefore at most $\Sigma_{i=1}^{I'} K h^{d+1} \int_{S_i} \mu_x\{dx\} = O(h)$ by (18). For an index i with either $\underline{\mathcal{A}}_i$ or $\bar{\mathcal{A}}_i$ outside $[0, 1]^d$, the boundedness of the conditional densities $f(w \mid x)$ implies $|\underline{\pi}_i - \bar{\pi}_i| \le h^d M \int_{S_i} \mu_x\{dx\}$ and since there are $O(1/h^{d+1})$ indices i on the boundary, these contribute at most $O(h)$ as well.

For part (c), note that if the densities $f(w \mid x)$ are Lipschitz continuous of order s everywhere, then $|\underline{\pi}_i - \bar{\pi}_i| \le K h^{d+s} \int_{S_i} \mu_x\{dx\}$ for all i and the result follows as in the previous cases.[1]

[1] The research of both authors was supported by an internal grant from the Columbia Graduate School of Business.

References

Anily, S. and A. Federgruen, "One Warehouse Multiple Retailer Systems with Vehicle Routing Costs," *Management Sci.*, 36 (1990), 92–114.

Balinski, M. and R. Quandt, "On an Integer Program for a Delivery Problem," *Oper. Res.*, 12 (1964), 300–304.

Bartholdi, J. III, "A Guaranteed-Accuracy Round-Off Algorithm for Cyclic Scheduling and Set Covering," *Oper. Res.*, 29 (1981), 501–510.

Beardwood, J., J. H. Halton, and S. M. Hammersley, "The Shortest Path Through Many Points," *Proc. Cambridge Philos. Soc.*, 55 (1959), 299–327.

Bell, W., L. Dalberto, M. Fisher, A. Greenfield, R. Jaikumar, P. Kedia, R. Mack, and P. Prutzman, "Improving the Distribution of Industrial Gases with an On-Line Computerized Routing and Scheduling Optimizer," *Interfaces*, 13 (1983), 4–23.

Bodin, L., B. Golden, A. Assad, and M. Ball, "Routing and Scheduling of Vehicles and Crew—State of the Art," *Comp. and Oper. Res.*, 10 (1983), 63–211.

Bramel, J., E. Coffman, P. Shor, and D. Simchi-Levi, "Probabilistic Analysis of the Capacitated Vehicle Routing Problem with Unsplit Demands," *Oper. Res.*, 40 (1992), 1095–1106.

——, C. L. Li, and D. Simchi-Levi, "Probabilistic Analysis of Heuristics for the Vehicle Routing Problem with Time Windows," *American J. Math. and Management Sci.*, 13(1993), 267–322.

—— and D. Simchi-Levi, "A Location Based Heuristic for General Routing Problems," *Oper. Res.* 43 (1995), 649–660.

—— and ——, "Probabilistic Analysis and Practical Algorithms for the Capacitated Vehicle Routing Problem with Time Window Constraints," *Oper. Res.* 44 (1996), 501–509.

—— and ——, "On the Effectiveness of Set Partitioning Formulations for the Vehicle Routing Problem," *Oper. Res.* 45 (1991), 295–301.

Cabrara, Fischetti and Toth, Working Paper in preparation, University of Belogna, Bologna, Italy, 1995.

Casco, D., B. Golden, and E. Wasil, "Vehicle Routing with Backhauls: Models, Algorithms and Case Studies," in Vehicle Routing: Methods and Studies, B. Golden and A. Assad, Eds., North-Holland, Amsterdam, 1988, pp. 127–148.

Ceria, S., N. Paolo, and A. Sassano, "A Lagrangian-Based Heuristic for Large-Scale Set Covering Problems," Working Paper, Columbia University, New York, 1995.

Chan, L., A. Federgruen, and D. Simchi-Levi, "Probabilistic Analysis and Practical Algorithms for Inventory Routing Models," to appear in *Oper. Res.*

Charnes, A. and M. Miller, "A Model for the Optimal Programming of Railway Freight Train Movements," *Management Sci.*, 3 (1956), 74–92.

Cullen, F., J. Jarvis, and D. Ratliff, "Set Partitioning Based Heuristics for Interactive Routing," *Networks*, 11 (1981), 125–143.

Desrochers, M., J. Desrosiers, and M. Solomon, "A New Optimization Algorithm for the Vehicle Routing Problem with Time Windows," *Oper. Res.*, 40 (1992), 342–354.

Devroye, L. and L. Gyorfi, *Nonparametric Density Estimation, The L_1 View*, John Wiley, New York, 1985.

Federgruen, A. and D. Simchi-Levi, "Analytical Analysis of Vehicle Routing and Inventory Routing Problems," *Handbooks in Operations Research and Management Science*, volume on "Networks and Distribution," M. Ball, T. Magnanti, C. Monma and G. Nemhauser, Eds., 1992.

—— and G. J. van Ryzin, "The Probabilistic Analysis of a Generalized Bin Packing Problem with Applications to Vehicle Routing and Scheduling Problems," to appear in *Oper. Res.*

Goffin, J. and J. Vial, "Cutting Planes and Column Generation Techniques with the Projective Algorithm," *J. Optimization Theory and Applications*, 65 (1990), 409–429.

Goldfarb, D. and M. J. Todd, "Linear Programming," in *Optimization*, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, (Eds.), Handbooks in Operations Research, Elsevier, Amsterdam, 1989.

Grotschel, M., L. Lovasz, and A. Schrijver, "The Ellipsoid Method and its Consequences in Combinatorial Optimization," *Combinatorica*, 1 (1981), 169–197.

——, ——, and ——, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, Berlin, 1988.

Haimovich, M. and A. H. G. Rinnooy Kan, "Bounds and Heuristics for Capacitated Routing Problems," *Math. Oper. Res.*, 10 (1985), 527–541.

Hochbaum, D., "Approximation Algorithms for the Set Covering and Vertex Cover Problems," *SIAM J. Comp.*, 11 (1982), 555–557.

——, "On the Fractional Solution of the Set Covering Problem," *SIAM J. Algebra and Discrete Meth.*, 4 (1983), 221–223.

Karmarkar, N. and R. M. Karp, "An Efficient Approximation Scheme for the One-Dimensional Bin-Packing Problem," Proc. IEEE 23rd Annual Symposium on the Foundations of Computer Science, Chicago, IL, November 3–5, 1982.

Karp, R. M., "Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem," *Math. Oper. Res.*, 2 (1977), 209–224.

Kingman, J. F. C., *Subadditive Processes*, Lecture Notes in Math., 539, Springer-Verlag, Berlin, 1976.

Magnanti, T., "Combinatorial Optimization and Vehicle Fleet Planning: Perspectives and Prospects," *Networks*, 11 (1981), 179–213.

Revesz, P., *The Laws of Large Numbers*, Academic Press, New York, 1968.

Rogers, D. F., R. D. Plate, R. T. Wong, and J. R. Evans, "Aggregation and Disaggregation Techniques and Methodology in Optimization," *Oper. Res.*, 29 (1991), 553–582.

Royden, H., *Real Analysis*, 2nd Edition, MacMillan, Toronto, Canada, 1968.

Steele, M. J., "Subadditive Euclidean Functionals and Non-Linear Growth in Geometric Probability," *Ann. Probability*, 9 (1981), 365–376.

Ye, Y., "A Potential Reduction Algorithm Allowing Column Generation," *SIAM J. Optimization*, 2 (1992), 7–20.

Accepted by Thomas M. Liebling; received December 1993. This paper has been with the authors 10 months for 1 revision.