

**BOUNDS AND APPROXIMATIONS FOR THE TRANSPORTATION PROBLEM OF LINEAR  
PROGRAMMING AND OTHER SCALABLE NETWORK PROBLEMS**

Carlos F. Daganzo and Karen R. Smilowitz

Institute of Transportation Studies and Department of Civil and Environmental Engineering

University of California, Berkeley CA 94720

(*Transportation Science*: submitted October 2001; revised October 30, 2002;

second revision February 7, 2003; accepted for publication February 2003)

**ABSTRACT**

Bounds and approximate formulae are developed for the average optimum distance of the transportation linear programming problem with homogeneously but randomly distributed points and demands in a region of arbitrary shape. It is shown that if the region size grows with a fixed density of points then the cost per item is bounded from above in  $3^+$  dimensions ( $3^+$ -D), but not in 1-D and 2-D. Lower bounds are also developed, based on a mild monotonicity conjecture. Computer simulations confirm the conjecture and yield approximate formulae. These formulae turn out to have the same functional form as the upper bounds. Curiously, the monotonicity conjecture implies that the cost per item does not depend on zone shape asymptotically, as problem size increases, for  $2^+$ -D problems but it does in 1-D. Therefore, the 2-D case can be viewed as a transition case that shares some of the properties of 1-D (unbounded cost) and some of the properties of 3-D (shape-independence).

The results are then extended to more general network models with subadditive link costs. It is found that if the cost functions have economies of scale, then the cost per item is bounded in 2-D. This explains the prevalence of the “last mile” effect in many logistics applications. The paper also discusses how the results were used to estimate costs under uncertainty for a vehicle-repositioning problem.

## INTRODUCTION

This paper develops bounds and approximations for the average optimum distance of the transportation linear programming problem (TLP) with random supplies and demands. The goal of the paper is not to develop better algorithms, since efficient algorithms for the TLP exist, but to develop simple formulae that can be used to predict system performance. Such formulae are helpful in the planning stages of complex logistics systems involving back-hauls and empty-vehicle repositioning because planning decisions are lasting and must be optimized over time periods when conditions vary in an uncertain way. The approximations developed in this paper have been used to find optimal designs for package delivery systems, recognizing the effects of demand variability (Smilowitz and Daganzo, 2002).

Similar formulae have been developed for the traveling salesman problem, or “TSP” (Beardwood et al, 1959, Eilon et al., 1971, Karp, 1977, Daganzo, 1984a), and for the vehicle routing problem, or “VRP” (Eilon et al., 1971, Daganzo, 1984b, Haimovich et al., 1985, Newell and Daganzo, 1986a and 1986b, and Newell, 1986). The results apply in particular to problems where  $N$  points are randomly and uniformly distributed on a region of a metric plane with area  $A$ , and density  $\delta = N/A$ . In all cases the distance traveled per point for the TSP, or the “detour” distance per point for the VRP, tends to a fixed multiple of  $\delta^{-1/2}$  as  $N$  and  $A$  are increased in a fixed ratio; i.e., the average detour distance per point is bounded. The detour distance of the VRP is the distance traveled in excess of the lower bound formed by the product of the average round trip distance between the depot and a point, and the fraction of a vehicle’s capacity consumed by each point. Systematic design methodologies for complex “many-to-one” and “one-to-many” logistics problems based on these formulae have been developed; see for example Daganzo (1991) and the references therein. Other work has focused more on the performance of heuristics for such logistics problems; see for example Bramel and Simchi-Levi (1997).

Similar asymptotic results have been obtained for other minimization problems--less central to the field of logistics, but still based on the distribution of random points in a region of space. Examples are matching (Papadimitriou, 1978), triangulation (Steele, 1982) and spanning tree problems (Steele, 1988).

Additional information on this subject can be found in Steele (1997) and Yukich (1998). It is known in this field (probability theory of geometric optimization problems) that the distance per point in  $K$ -dimensional space tends to  $\delta^{-1/K}$  for any problem satisfying certain “geometric subadditivity” properties--shared by all the problems mentioned so far. Thus, if for any of these problems one holds  $\delta$  constant while increasing  $N$  (i.e., one enlarges the region) the distance per point tends to a constant; i.e., it is bounded, independent of  $K$ . Unfortunately, this result is not universal. Logistics problems involving random point distributions *and* random flows, such as the TLP, turn out to be quite different.

This paper examines TLP’s where points lie on a region of a linear normed space. It is found that the average optimum distance per item (or distance per point) is not always bounded because TLP solutions must include some long trips to balance interregional flows. Interestingly, as occurs in some applied probability problems (e.g., the Ising model of statistical mechanics and the first passage of time of a random walk), the nature of the solution depends on the dimensionality of the space. We find that if one holds  $\delta$  constant while increasing  $N$  and the size of the region in a constant ratio, then the average distance traveled per item is bounded in three or more dimensions ( $3^+$ -D), but it is of order  $\log(N)$  in 2-D and of order  $N^{1/2}$  in 1-D. One-dimensional problems arise in connection with highway construction projects and the minimization of earthwork haul. Asymptotic formulae with  $N \rightarrow \infty$  are provided for problems with homogeneous demand. Unlike in 1-D, the upper bounds developed for two or more dimensions are independent of region-shape. Approximate TLP formulae that can be used for large-scale network design problems are also given. The paper also examines a more general family of network problems with subadditive link costs. We find that 2+-D problems of this type are bounded if the costs exhibit economies of scale.

The paper provides theoretical results in terms of the abovementioned formulae and related insights, as well as practical results in terms of cost approximations for large-scale network design problems. Section 1 provides background on the TLP and summarizes earlier 1-D results. Section 2 develops upper and lower bounds for 2+-D cubes. Section 3 generalizes the bounds to regions of arbitrary shape, develops an approximate formula, and shows how it can be applied to a vehicle-repositioning problem. Section 4 extends results to non-linear network problems. Section 5 summarizes the key findings and discusses future work.

## 1. BACKGROUND

### 1.1 Definitions

In this paper the TLP is defined as follows. Given are  $N$  points, a set of inter-point distances,  $\{d_{ij}, \forall i, j = 1, \dots, N \text{ with } i \neq j\}$ , satisfying the triangle inequality, and a set of net supplies,  $v_i$ , at each point. Positive  $v_i$  are interpreted as supplies and negative  $v_i$  as demands. The goal is to find a set of shipments,  $\{v_{ij}, \forall i, j = 1, \dots, N \text{ with } i \neq j\}$ , that minimizes the total distance traveled while satisfying flow balance constraints.

$$(TLP) \quad \min \quad z = \sum_{i, j \in N} d_{ij} v_{ij} \quad (1a)$$

$$\text{s. t.} \quad \sum_{j \in N} (v_{ij} - v_{ji}) \leq v_i ; \quad \forall i \in N \quad (1b)$$

$$v_{ij} \geq 0 \quad ; \quad \forall i, j \in N . \quad (1c)$$

Equations (1b) specify flow-conservation at each point, ensuring that the net flow emanating from a point  $i$  never exceeds the net supply at  $i$ . Because distances satisfy the triangle inequality, in an optimum solution origin nodes can only emit flow and destination nodes only receive flow. Otherwise it would be possible to re-route some flow and reduce cost.

Recall that TLP's are feasible only if  $\sum_{i \in N} v_i \geq 0$ , as can be seen by summing (1b) across  $i$ , and that for

balanced problems (where this sum is null) constraints (1b) are individually satisfied as pure equalities. To address infeasible problem instances, we define an always-feasible, auxiliary problem, ATLP, that includes a fictitious source,  $i = 0$ , with the least positive net supply that makes the problem feasible, and distances,

$d_{i0}, d_{0j} = M \gg \sup(d_{ij})$ . This net supply is  $v_0 = -\sum_{i=1}^N v_i$ , if the problem is infeasible, and  $v_0 = 0$ ,

otherwise. The fictitious distances represent a penalty for failing to ship an item. Because  $M$  is large, the ATLP distances satisfy the triangle inequality, and the ATLP is still a TLP. If the original TLP is infeasible then the ATLP is a balanced TLP, and the outflow from the fictitious source must be  $v_0$ . The penalty component of an optimum solution is  $v_0 M$ , and the distance component,  $d^*$ , can be expressed in all cases as

$$d^* = z^* - v_0 M, \quad \text{with} \quad v_0 = \left( - \sum_{i=1}^N v_i \right)^+. \quad (2)$$

Recall that if the TLP is feasible, then  $d^* = z^*$ . When it is infeasible we will take as its solution the distance component of the ATLP.

In the TLP/ATLP it is assumed that if supply exceeds demand, the excess supply is left at the origins. We also consider a variant of the problem, where any excess supplies are carried to the extra point, or “depot”. This version of the problem will be called “depot-TLP”, or DTLP. The DTLP is an ordinary TLP, where the net supply at the depot precisely balances the problem. In the DTLP the depot distances do not have to be fixed or large but they must be non-negative,  $(d_{0j}, d_{i0}) \geq 0, \forall i, j$ , and must satisfy the triangle inequality. The minimum of the DTLP objective function,  $z_D$ , is denoted  $d_D^*$ .

**Proposition 1: (DTLP as an upper bound to TLP).** For any TLP and any associated DTLP,  $d^* \leq d_D^*$ .

Furthermore, if the TLP is balanced then  $d^* = d_D^*$ . ■

**Proof:** Recall that if a TLP is infeasible, a fictitious source is introduced to obtain  $z^*$  and  $d^*$ . The ATLP has the same constraints as the DTLP. Therefore, a set of optimal shipments for the DTLP, denoted  $\mathbf{v}_D^*$ , is a feasible solution of the ATLP. The associated ATLP distance, including penalties, is denoted  $z(\mathbf{v}_D^*)$ . If the DTLP is now adjusted by adding  $M$  to all the depot distances,  $d'_{0j} = d_{0j} + M$  and  $d'_{i0} = d_{i0} + M$ , the objective function value corresponding to  $\mathbf{v}_D^*$  would become  $d_D^* + v_0 M$ , since  $v_0$  is the total flow to/from the depot in the optimum solution of DTLP. All the link distances in the ATLP are less than or equal to those of the adjusted DTLP, and  $\mathbf{v}_D^*$  is feasible in both cases. Therefore,  $z(\mathbf{v}_D^*) \leq d_D^* + v_0 M$ . Since  $z^* \leq z(\mathbf{v}_D^*)$ , it is also true that  $z^* \leq d_D^* + v_0 M$ , and hence that  $d_D^* \geq z^* - v_0 M = d^*$  for the infeasible case. In the feasible case, the non-depot flows ( $i \neq 0, j \neq 0$ ) in  $\mathbf{v}_D^*$  incur a cost  $d_0 \leq d_D^*$  for the DTLP. Furthermore, since

the depot outflows in  $\mathbf{v}_D^*$  are zero, the non-depot flows are a feasible solution of the original TLP. To see this note that (1b) can be rewritten in terms of the non-depot flows as  $\sum_{j \neq 0} (v_{ij} - v_{ji}) \leq v_i + v_{0i} - v_{i0} \quad \forall i \neq 0$ , which implies  $\sum_{j \neq 0} (v_{ij} - v_{ji}) \leq v_i \quad \forall i \neq 0$  when  $v_{0i} = 0$ , as occurs with  $\mathbf{v}_D^*$ . Since the non-depot flows are feasible,  $d_0 \geq z^* = d^*$ , and it follows that  $d^* \leq d_D^*$ . This completes the first part of the proof. The second part is true because if the problem is balanced, then  $v_0 = 0$  and the DTLP and TLP problems coincide. ■

## 1.2 Different Versions of the TLP and DTLP

In what follows, we look for the average of  $d^*$ , denoted  $\langle d^* \rangle$ , over a set of solutions (e.g., over an infinite number of days) when conditions vary. [The brackets  $\langle Y \rangle$  and  $\rangle Y \langle$  will be used for the mean and variance of a variable  $Y$  across problem instances.] It is assumed that points are embedded in a  $K$ -dimensional normed linear space where each point  $i$  is identified by a set of Cartesian coordinates,  $\mathbf{x}_i$ , and that distances,  $d_{ij}$ , are given by any norm of the Cartesian separation vector,  $\|\mathbf{x}_i - \mathbf{x}_j\|$ . This includes the  $L_p$  norms, as well as norms derived from them by coordinate rotations and anisotropic scalings.

Four versions of the problem are considered depending on which data are allowed to vary. Points are either fixed on a  $K$ -dimensional cubic lattice (grid) with Cartesian spacing  $l$ , or randomly and uniformly distributed in space. The net supplies,  $v_i$ , are assumed to be identically distributed normal random variables with mean 0 and variance  $\sigma^2$ , with a covariance  $\langle v_i v_j \rangle = -\sigma^2 / (N-1)$  for balanced problems and independent for unbalanced problems. The given covariance ensures that both the mean and variance of  $\sum_i v_i$  are zero for balanced problems. The modifiers ‘‘G’’ or ‘‘R’’, referring to ‘‘grid’’ or ‘‘random’’ point locations, and ‘‘B’’ or ‘‘U’’ referring to ‘‘balanced’’ or ‘‘unbalanced’’ demand, are used to identify problem versions; e.g., DTLP(B,R) and DTLP(U,R) designate the unbalanced and balanced versions of the DTLP with randomly distributed points.

Dimensional analysis yields the general functional form of  $\langle d^* \rangle$  for any TLP version in terms of its input constants. For additional information on dimensional analysis see Bridgman (1963) or Barenblatt (1996); the well-known result on which our assertion is based is customarily called the “ $\pi$ -theorem”, Buckingham (1914). For problems with random point locations in a region of specific shape (e.g., a cube) these constants could be  $\sigma$ ,  $\delta$  (the expected spatial density of points), and  $A$  (the region volume). The same parameters can be used to define grid problems, with the convention:  $\delta = l^{-K}$ . In general we look for  $\langle d^* \rangle$ , or alternatively for the average distance per point, defined as  $\langle p^* \rangle = \langle d^* \rangle / N = \langle d^* \rangle / \delta A$ .

For these problems, only two independent dimensionless parameters can be formed with the input constants and the solution value:  $\delta A$  and  $\langle p^* \rangle \delta^{1/K} / \sigma$ . According to the  $\pi$ -theorem, the exact solution for  $\langle p^* \rangle$  must then be of the form:

$$\langle p^* \rangle = \sigma \delta^{-1/K} f(\delta A), \quad (3)$$

where “ $f$ ” is the only unknown left to be determined. This function generally depends on the type of problem, the norm, region shape and the dimensionality of the space. The subscripts “T” and “D” will be used with  $f$  when it refers specifically to a TLP or DTLP. The subscripts “TU”, “DU” and “B” are used to specify the unbalanced and balanced versions of these problems. Only one subscript is used for balanced problems because  $\text{TLP(B)} \equiv \text{DTLP(B)}$ . Used alone, the subscript “U” refers generically to both  $\text{TLP(U)}$  and  $\text{DTLP(U)}$ . If no subscript is used, then  $f$  denotes a generic problem. Since  $f$  is dimensionless it is called the “dimensionless distance per point”. Up to a multiplicative constant, it is the average number of lattice spacings traveled by an item.

### 1.3 One-dimensional results

The 1-D case is special in that points can be ordered along a unique shortest path, which allows for a different solution approach. Results are presented below, and derived in the appendix. The  $\text{TLP(B,G)}$  with

zero-mean, multinormal demand, defined in an interval of  $\mathbb{R}^1$  ( $d_{ij} = |x_i - x_j|$ ) satisfies:

$$f_B(N) / \sqrt{N} \rightarrow \sqrt{\frac{\pi}{32}}, \quad \text{as } N \rightarrow \infty. \quad (4)$$

The  $\sqrt{N}$ -dependence of  $f_B$  is caused by the long-range interactions arising from the flow balancing requirements. Equation (4) also holds if the  $v_i$  are not normal, but satisfy the conditions of the central limit theorem, and also if the point locations vary across days as a homogeneous Poisson process with rate  $\delta$  (where  $N = \delta A$ ).

The asymptotic expression for DTLP(U) in intervals of  $\mathbb{R}^1$  with grid or random point locations and centrally located depots is:

$$f_{DU}(N) / \sqrt{N} \rightarrow \sqrt{\frac{4}{9\pi}}, \quad \text{as } N \rightarrow \infty. \quad (5)$$

It is easy to see from these formulae that the asymptotic distance depends on region shape. Simply, consider two intervals of equal length ( $L/2$ ) separated by a much greater distance ( $W$ ), with the depot in the middle. Then, the total distance traveled by the depot flows outside the intervals must be proportional to  $WN^{1/2}$ , since the depot flow is itself proportional to  $N^{1/2}$ . This external distance is always comparable with the distance predicted by (4) or (5) for the case with  $W = 0$ . Hence,  $\langle p_{DU}^* \rangle$  depends on  $W$  asymptotically.

The following two sections show that the dependence of  $f$  with  $N$  is weaker in higher dimensions. It is shown that these functions are  $f(N) = O(\log(N))$  in 2-D, and  $f(N) = O(N^0)$  in 3<sup>+</sup>-D. Section 2 examines cubic zones and Sec.3 extends the results to regions of arbitrary shape. It is found that the asymptotic formulae do not depend on shape in 2<sup>+</sup>-D. This is fortuitous because a single asymptotic formula can then be used for all zone shapes in the important 2-D case.

## 2. BOUNDS FOR K-DIMENSIONAL CUBES

This section develops lower bounds for  $f$  in Section 2.1 and upper bounds in Section 2.2. We assume that points are homogeneously distributed in a cube with density  $\delta$ , either as a Poisson process or on a grid, and



that net supplies are zero-mean, normal random variables with standard deviation  $\sigma$ . Note that for any such random variable,  $\langle |v_i| \rangle \leq \sigma$ . The random variables are assumed to be independent for unbalanced problems and correlated for balanced ones. The service region diameter is  $\alpha A^{1/K}$ , where  $\alpha$  is a norm-specific and dimension-specific constant ( $\alpha = K^{1/p}$  for the  $L_p$  metric).

## 2.1 A lower bound for all versions

Every item transported must travel to a nearest neighbor, or farther. Since every item is associated with one origin and one destination, the expected distance per point for balanced problems cannot be less than one half of the product of the expected distance between nearest neighbors  $\eta \delta^{-1/K}$  and the expected absolute net supply per point  $\langle |v_i| \rangle \leq \sigma$ . (Note that  $\eta = 1$  for grid problems with an  $L_p$  metric.) Thus,

$$\langle p_B^* \rangle \geq \frac{1}{2} \eta \delta^{-1/K} \sigma, \quad (6)$$

for balanced problems. This bound also applies to the DTLP(U) since this problem is a balanced TLP. The

inequality holds asymptotically for the TLP(U), because the supply not shipped,  $\left( \sum_{i \in N} v_i \right)^+$ , becomes a

negligible fraction of the total as  $\delta \rightarrow \infty$ . This is true because the expected total absolute net supply

$\left( \sum_{i \in N} v_i^+ \right)$  is of order  $\delta$ , while  $\left( \sum_{i \in N} v_i \right)^+$  is of order  $\delta^{d/2}$ .

## 2.2 Upper bounds

The formulae of Sec. 1.3 showed that  $f(N) = O(\sqrt{N})$  for several versions of the 1-D problem. We now show that the following theorem is true for all versions.

**Theorem 1: (Upper bounds).** For DTLPs and TLPs defined on cubes  $f(N)$  is  $O(\sqrt{N})$  in 1-D,  $O(\log(N))$  in 2-D and  $O(N^{\theta})$  in 3<sup>+</sup>-D. ■

Figure 1 provides a schematic of the proof, with arrows indicating logical connections between the six cases that can arise. The two gray arrows indicate that the theorem holds for the second row if it holds for the first row—since Proposition 1 guarantees that an upper bound for DTLP(U) is also an upper bound for TLP(U). The black arrows imply the same for the third row—this will be proven in Sec. 2.2.4. Thus, it suffices to prove the first row. We begin by describing a bilevel decomposition method used in the proofs.

### 2.2.1 Bilevel decomposition

Assume that the service region has been partitioned into a finite number of subregions,  $C_I$ , each with its own sub-depot,  $I$ , and that specific net supplies,  $v_{i(I)}$ , from each point  $i$  have been allocated to each subdepot,  $I$ . [In what follows, the subscripts  $i$  and  $j$  are reserved for the original points, including the main depot in the case of the DTLP, and capital letters,  $I, J$ , are used for subdepots.] The allocated flows satisfy:

$$v_i = \sum_I v_{i(I)}, \forall i, \text{ and } v_I = \sum_i v_{i(I)}, \forall I, \text{ where } v_I \text{ is the cumulative net supply associated with sub-depot } I.$$

The following algorithm can now be defined for the DTLP.

#### ***Bilevel algorithm.***

Step 1 (lower level): For all  $I$ , solve a DTLP with the  $\{v_{i(I)}\}$  as data and sub-depot  $I$  as the depot. This transfers a flow  $v_I$  to each sub-depot.

Step 2 (upper level): Route the  $v_I$  net supplies from/to each subdepot as per a feasible DTLP solution with the main depot as the depot. ■

Obviously, the net flow of every subdepot after both steps is zero; i.e., items just pass through these points. It is therefore possible to express the result of the algorithm in path form, by simply specifying the number of items  $v_{ijk}$  that share the  $k^{\text{th}}$  path from origin  $i$  to destination  $j$ , including any intermediate subdepots. The net flows at all points satisfy the conservation equations of the original DTLP. The set of origin-destination

flows obtained by summing the path flows for every origin-destination pair,  $\mathbf{v} = \left\{ v_{ij} = \sum_k v_{ijk}, \forall i, j \right\}$ ,

including the main depot as an origin or a destination, is a feasible solution of the original DTLP with distance  $d_D(\mathbf{v}) = \sum_{ij} v_{ij} d_{ij}$ . Thus,  $d_D(\mathbf{v}) \geq d_D^*$ . If we now let  $d^{(b)}$  be the combined distance for both steps of

the bilevel algorithm, it is possible to show the following.

**Proposition 2: (Bilevel upper bound to DTLP).**  $d^{(b)} \geq d_D^*$ . ■

**Proof:** We have seen that  $d_D(\mathbf{v}) \geq d_D^*$ . Since the triangle inequality ensures that  $d_{ij}$  is a lower bound to the length of every path from  $i$  to  $j$ , we have:  $d^{(b)} = \sum_{ijk} v_{ijk} d_{ijk} \geq \sum_{ijk} v_{ijk} d_{ij} = \sum_{ij} v_{ij} d_{ij} = d_D(\mathbf{v}) \geq d_D^*$ . ■

### 2.2.2. Proof of Theorem 1 for DTLP(U,R)

Proposition 2 is now used to establish the following preliminary result.

**Lemma 1: (Recursive relation for the dimensionless distance of DTLP(U,R)).**

$$f_{DUR}(n^K) + \frac{1}{2}\alpha(m-1)n^{1-K/2} \geq f_{DUR}(n^K m^K), \quad \text{for } n \geq 0, m = 1, 2, 3, \dots \quad \blacksquare \quad (7)$$

[We use  $n^K$  instead of  $\delta A$  or  $N$  for the expected number of points because this simplifies future derivations.]

**Proof:** Partition the region into  $m^K$  identical cubes with sub-depots at their centers as in Figure 2(a), and apply the bilevel algorithm to the problem assuming that origins and destinations are exclusively associated with their nearest centroid. Then, the expected total cost for the lower level problem is an aggregation of the costs of  $m^K$  scaled-down, random-location DTLP(U,R) problems of the same type as the original. Since (3) applies to each sub-problem with  $A$  replaced by  $A/m^k$ , the expected total lower level cost is

$$\langle d_{UR}^{(b)} \rangle_L = \delta A \sigma \delta^{-1/K} f_{DUR} \left( \delta \frac{A}{m^K} \right), \quad (8a)$$

since the expected total number of points at the lower level is  $\delta A$ .

The high level problem is an unbalanced grid problem with  $m^K$  points and lattice spacing  $A^{1/K} / m$ . The variance of the net supply at each point is that of the excess supply in one cube. Conditional on the number of points in the cube,  $P$ , the mean and variance of the excess supply are 0 and  $P\sigma^2$ . Therefore, the unconditional variance is the mean of the conditional variance; i.e.,  $\langle |v_I| \rangle = \delta A \sigma^2 / m^K$ . Thus,

$\langle |v_I| \rangle \leq \sigma \sqrt{\delta A / m^K}$ . A feasible solution to the high level problem is to send items directly to/from the depot.

Since the distance traveled by any such item cannot exceed the radius of the convex hull for all subdepots, which is  $\frac{1}{2} \alpha A^{1/K} \left(1 - \frac{1}{m}\right)$ , it follows that:

$$\langle d_{UR}^{(b)} \rangle_H \leq m^K \left[ \frac{1}{2} \alpha A^{1/K} \left(1 - \frac{1}{m}\right) \right] \sigma \sqrt{\delta A / m^K}. \quad (8b)$$

Recall now that the original DTLP problem satisfies:

$$\langle d_{DUR}^* \rangle = (\delta A) \sigma \delta^{-1/K} f_{DUR}(\delta A). \quad (8c)$$

Since Proposition 2 holds for each instance of the problem, it also holds for the expectation. Thus,

$\langle d_{UR}^{(b)} \rangle_L + \langle d_{UR}^{(b)} \rangle_H \geq \langle d_{DUR}^* \rangle$ . Substituting (8a– 8c) for these terms, and dividing both sides of the resulting

inequality by  $\delta A \sigma \delta^{-1/K}$ , we find:

$$f_{DUR} \left( \frac{\delta A}{m^K} \right) + \frac{\frac{1}{2} \alpha (1 - 1/m) m^{K/2}}{(\delta A)^{1/2 - 1/K}} \geq f_{DUR}(\delta A), \quad \text{for } \delta A \geq 0, m = 1, 2, 3, \dots \quad (8d)$$

This system of inequalities can be simplified with the change of variable  $\delta A = n^K m^K$  (where  $n$  is real and non-negative). The result is (7). ■

We are now ready to prove the theorem.

**Proposition 3: (Upper bounds for DTLP(U,R)).** The DTLP(U,R) satisfies Theorem 1. ■

**Proof:** For a fixed  $n_o$ , consider the subset of (7), corresponding to  $m = 2$  and  $n = 2n_o, 4n_o, 8n_o \dots$

$$f_{DUR}(n^K) + \frac{1}{2}\alpha n^{(1-K/2)} \geq f_{DUR}((2n)^K), \quad n = n_o 2^j \quad (j = 1, 2, \dots), \quad (9a)$$

and the related set of equalities,

$$f_{DUR}(n^K) + \frac{1}{2}\alpha n^{(1-K/2)} = f_{DUR}((2n)^K), \quad n = n_o 2^j \quad (j = 1, 2, \dots). \quad (9b)$$

We look for the highest possible function,  $\tilde{f}_{DUR}$ , with domain  $D(n_o) = \{n^K : n = n_o 2^j, j = 1, 2, \dots\}$  that satisfies (9a) and matches  $f_{DUR}$  when  $n = n_o$ . Since ((9a) and (9b)) have a recursive structure, such a function can be constructed by iterating (9b) starting with the given initial value,  $n^K = n_o^K$ . The unique result, given below, is an upper bound for  $f_{DUR}$  in  $D(n_o)$ .

$$\tilde{f}_{DUR}(n^K) = f_{DUR}(n_o^K) + \frac{1}{2}\alpha \left[ \frac{n^{(1-K/2)} - n_o^{(1-K/2)}}{2^{(1-K/2)} - 1} \right] \quad \text{if } K \neq 2, \quad (10a)$$

$$= f_{DUR}(n_o^K) + \frac{1}{2}\alpha \log_2(n/n_o) \quad \text{if } K = 2. \quad (10b)$$

A bound for all  $n \geq \varepsilon$ , where  $\varepsilon$  is an arbitrary positive real number, is now derived. Since every real  $n \geq \varepsilon$  belongs to a  $D(n_o)$  with  $n_o \in [\varepsilon/2, \varepsilon]$ , it suffices to find a common upper bound to all the instances of (10a) and (10b) with  $n_o \in [\varepsilon/2, \varepsilon]$ . It is first shown that, for  $n_o \in [\varepsilon/2, \varepsilon]$ , the first terms of (10) are bounded by  $\frac{1}{2}\alpha\varepsilon$ . Recall that the optimum total distance of a DTLP(U,R) instance, given by (8c), is bounded from above by the product of the total flow, which is itself bounded by  $(\delta A)\sigma$  and  $\frac{1}{2}$  the diameter of the region,  $\frac{1}{2}\alpha A^{1/K}$ . Therefore,  $(\delta A)\sigma \delta^{-1/K} f_{DUR}(\delta A) \leq \frac{1}{2}(\delta A)\sigma \alpha A^{1/K}$ , or equivalently,  $f_{DUR}(\delta A) \leq \frac{1}{2}\alpha(\delta A)^{1/K}$ . Thus,  $f_{DUR}(n_o^K) \leq \frac{1}{2}\alpha(n_o^K)^{1/K} = \frac{1}{2}\alpha n_o \leq \frac{1}{2}\alpha\varepsilon$ , as stated. Note now that the second terms of (10) decrease with  $n_o$ . Therefore, they are bounded from above by their values at  $\varepsilon/2$ . With these bounds, we find:

$$f_{DUR}(n^K) \leq \frac{1}{2}\alpha\varepsilon + \frac{1}{2}\alpha \left[ \frac{n^{(1-K/2)} - (\varepsilon/2)^{(1-K/2)}}{2^{(1-K/2)} - 1} \right] \quad \text{if } K \neq 2 \text{ and } n \geq \varepsilon \quad (11a)$$

$$\leq \frac{1}{2}\alpha\varepsilon + \frac{1}{2}\alpha \log_2(2n/\varepsilon) \quad \text{if } K = 2 \text{ and } n \geq \varepsilon. \quad (11b)$$

Upon changing  $N$  for  $n^K$  in (11) it becomes apparent that  $f_{DUR}(N)$  is  $O(\sqrt{N})$  if  $K=1$ ,  $O(\log(N))$  if  $K=2$ , and  $O(N^0)$  if  $K \geq 3$ . This completes the proof. ■

### 2.2.3 Proof of Theorem 1 for DTLP(U,G)

We now turn our attention to the top right cell of Fig. 1. It is assumed for this case that  $N = n^K$ , where  $n$  is a natural number. The recursive arguments of Sec 2.2.2 are modified because the cells  $C_I$  will in general include different numbers of points. To deal with this issue, it is useful to imagine (without losing generality from the original assumptions) that supplies are continuously generated throughout the cube by a  $K$ -dimensional, zero-mean, homogeneous Brownian process,  $V$ , and that each net supply,  $v_i$ , is the amount generated by  $V$  in the cubic cell,  $C_i$ , of point  $i$ . The variance per unit volume of process  $V$  is denoted  $\gamma$  (items<sup>2</sup>/distance <sup>$K$</sup> ). For a cubic lattice of subdepots,  $\mathbf{L}_m$ , with  $m$  points to a side ( $M = m^K$  subdepots), the lattice spacing is  $l_m = A^{1/K}/m$  and the variance of the net supplies in one of its cells,  $C_I^m$ , is  $\sigma_m^2 = \gamma A / M$ .

Consider the lattices  $\mathbf{L}_n$  and  $\mathbf{L}_m$ , where  $n > m$ , and apply the bilevel algorithm to problem  $\mathbf{L}_n$ , using  $\mathbf{L}_m$  as the set of sub-depots and allocating to  $I$  from  $i$  the cumulative net supply generated in the overlap region of the two cells:  $v_{i(I)} = V(C_i^n \cap C_I^m)$ ; see Figure 2(b). Because the overlap regions are a partition of the service region the allocated flows are both feasible and independent. Furthermore,  $v_I = V(C_I^m)$ . This guarantees that the upper level problem is in the family of problems under consideration. This result will be used in conjunction with the following lemma, which bounds the distance of the lower level problem.

[Note as an aside that a proper set of  $v_{i(I)}$  can also be generated directly from the conditional distribution of  $\{v_{i(I)} | v_i\}$  without using the auxiliary process  $V$ . This conditional distribution can be easily obtained from the joint distribution of  $\{v_{i(I)}, v_i\}$ , which is known from the overlap rules; the relevant formulae can be found for example in Johnson and Kotz (1972).]

**Lemma 2: (Upper bound for lower level distance of DTLP(U,G)).** If  $n/m \leq 2$  then the lower level distance for the bilevel algorithm with the overlap rule satisfies,

$$\langle d_L^{(b)} \rangle \leq (2^{1+K/2}) \alpha \gamma^{1/2} (A^{1/2+1/K}) (n^{K/2-1}). \quad (12)$$

**Proof:** Since  $n/m \leq 2$ , the lower level distance traveled by any item cannot exceed the diameter of the largest cell,  $C_I^m$ , or twice the diameter of the smallest cell  $C_i^n$ ; see Figure 2(b). Hence,

$$\langle d_L^{(b)} \rangle \leq (2\alpha A^{1/K} / n) \left\langle \sum_{il} |v_{i(l)}| \right\rangle = (2\alpha A^{1/K} / n) \left\langle \sum_{il} \langle |v_{i(l)}| \rangle \right\rangle \leq (2\alpha A^{1/K} / n) \left( \sum_{il} (\gamma A_{il})^{1/2} \right),$$

where  $A_{il}$  denotes the volume of the overlap region  $C_i^n \cap C_I^m$ . The last factor in this expression can be expressed as  $\sum_i (\sum_{l'} (\gamma A_{il'})^{1/2})$ . Since  $\sum_{l'} A_{il'}$  is a constant,  $A/N$ , and the inner sum is a concave and symmetric function of the  $A_{il'}$ , it is maximized by  $A_{il} = (A/N) / (\#)$ , where  $\#$  is the number of non-zero terms in the sum. Thus, the inner sum cannot exceed  $(\# \gamma A/N)^{1/2}$ . Since this sum can have at most  $2^K$  non-zero terms (see Figure 2(b)) it is bounded by  $(2^K \gamma A/N)^{1/2}$ . It follows that  $\langle d_L^{(b)} \rangle \leq (2\alpha A^{1/K} / n) (N \sqrt{2^K \gamma A / N})$ , which reduces to (12). ■

It is now possible to establish the desired result.

**Proposition 4: (Upper bounds for DTLP(U,G)).** The DTLP(U,G) satisfies Theorem 1. ■

**Proof:** Recall that  $\langle d_{DUG}^* \rangle = (\delta A) \sigma \delta^{-1/K} f_{DUG}(\delta A)$ ; see (8c). If we substitute  $n^K/A$  for  $\delta$  and  $(\gamma A n^{-K})^{1/2}$  for  $\sigma$  in this expression we obtain:  $\langle d_{DUG}^* \rangle = \gamma^{1/2} A^{1/2+1/K} n^{K/2-1} f_{DUG}(n^K)$ . Since the upper level problem on  $\mathbf{L}_m$  is in the same family as the original problem (with  $m$  instead of  $n$ ), and the lower level problem is bounded by Lemma 2, Proposition 2 now yields the following recursive inequality for  $f_{DUG}$ :

$$\gamma^{1/2} (A^{1/2+1/K}) (n^{K/2-1}) f_{DUG}(n^K) \leq \gamma^{1/2} (A^{1/2+1/K}) (m^{K/2-1}) f_{DUG}(m^K) + (2^{1+K/2}) \alpha \gamma^{1/2} (A^{1/2+1/K}) (n^{K/2-1}).$$

This inequality can be rewritten, after dividing by  $\gamma^{1/2} (A^{1/2+1/K}) (n^{K/2-1})$ , as:

$$f_{DUG}(n^K) \leq \left(\frac{n}{m}\right)^{1-K/2} f_{DUG}(m^K) + (2^{1+K/2})\alpha, \quad \text{if } n/m \leq 2, \quad (13a)$$

or alternatively, as:

$$f_{DUG}(n^K) - f_{DUG}(m^K) \leq \left[\left(\frac{n}{m}\right)^{1-K/2} - 1\right] f_{DUG}(m^K) + (2^{1+K/2})\alpha, \quad \text{if } n/m \leq 2. \quad (13b)$$

If we now choose  $n/m = 2$  in (13a) we obtain:

$$f_{DUG}((2m)^K) \leq 2^{1-K/2} f_{DUG}(m^K) + 2^{1+K/2} \alpha. \quad (14)$$

This recursion can be treated in the same way as (9a) (considering exponential sequences of the form  $m_j = 2^j m_0$ ) with the same result. It is found that, for any such sequence (with  $N_j = m_j^K$ ),  $f_{DUG}(N_j)$  is  $O(\sqrt{N_j})$  if  $K=1$ ,  $O(\log(N_j))$  if  $K=2$ , and  $O(N_j^0)$  if  $K \geq 3$ . Note from (13b), however, that the difference between the value of  $f_{DUG}$  for any  $m$  and the value for the closest  $m_j$  from below ( $m/m_j < 2$ ) is of the same order as  $f_{DUG}(m_j^K)$ . Hence, the bounds hold for all natural  $m$ ; i.e., for all  $N$  that can be expressed as the  $K^{\text{th}}$  power of a natural number. ■

We have now established the validity of Theorem 1 for the top row of Fig. 1, and therefore of the second row too. The third row (balanced problems) is examined next.

#### 2.2.4 Proof of Theorem 1 for DTLP(B,G) and DTLP(B,R)

It is shown below the difference between the optimal costs per point for balanced and unbalanced problems is  $O(N^{1/K-1/2})$ ; i.e., of a lesser order than the upper bound. Therefore, Theorem 1 holds.

The main complication for balanced problems is that the net supplies are not independent. This is remedied by defining three sets of zero-mean, homogeneous net supplies for any fixed set of  $N$  points:  $u = \{u_i\}$ ,  $v = \{v_i\}$  and  $w = \{w_i\}$ , with the following properties:  $\langle u_i \rangle = \sigma^2$  and  $\langle u_i u_j \rangle = 0$  if  $i \neq j$ ;  $w_i = -\sum_j (u_j/N)$ ; and  $v_i = u_i + w_i$ . The four sets of net supplies are related by,

$$v = u + w \quad (15a)$$

$$u = v + (-w). \quad (15b)$$



Note that  $\{w_i\}$  always defines a balanced problem, since  $\sum_i w_i = 0$ . We imagine that (15a) is used to generate one balanced problem instance for every unbalanced instance; i.e., the set of balanced instances so generated are in a 1:1 correspondence with the set of unbalanced instances. Note that the set of  $u$ 's defines an unbalanced problem with variance  $\sigma^2$  and the set of generated balanced instances,  $\{v\}$ , defines a balanced problem with variance  $\sigma^2(1-1/N)$ . The set of  $w$ 's in the correspondence define a problem with identical net supplies at all points and variance  $\sigma^2/N$ . Obviously, a balanced problem with any desired variance,  $\sigma_B^2$ , can be constructed from an unbalanced (independent) problem with larger variance,  $\sigma_B^2/(1-1/N)$ . The difference between the two optimal solutions depends on  $\{w\}$  and this is exploited below.

Let the  $N$  points be arbitrarily located with a maximum distance from the depot,  $\phi_D$ , and consider the optimum distance for the problems defined by the  $w$  alone. Clearly,

$$\langle p^* \rangle_{\text{identical}} = 0 \quad \text{for TLP, and} \quad (16a)$$

$$\leq \phi_D \sigma N^{-1/2} \quad \text{for DTLP.} \quad (16b)$$

Equation (16b) is true because all flows must go to/from the depot, so that the average distance to the depot is bounded by  $\phi_D$ , and because the expected net supply per point is bounded in absolute value by its standard deviation,  $\sigma N^{-1/2}$ . If all the points (including the depot) are contained within a cube of volume  $A$ , then  $\phi_D \leq \alpha A^{1/K}$ , and (16b) becomes:

$$\langle p^* \rangle_{\text{identical}} \leq \alpha \sigma N^{1/K-1/2} \left( \frac{A}{N} \right)^{1/K} \quad \text{for DTLP.} \quad (16c)$$

The results about to be derived use (16c) and the following well known property of problems with linear constraints, such as (1b).

**Superposition property of DTLP.** If  $\mathbf{u} = \{u_{ij}\}$  is a feasible solution of a DTLP with net supplies,  $u = \{u_i\}$ , and  $\mathbf{w} = \{w_{ij}\}$  is a feasible solution of the same problem when the net supplies are  $w = \{w_i\}$ , then  $\mathbf{v} = \mathbf{u} + \mathbf{w}$  is a feasible solution of the same problem with net supplies  $v = u + w$ . ■

The central result of this subsection can now be stated.

**Lemma 3: (Asymptotic equivalence of balanced and unbalanced DTLP's in  $2^+-D$ ).** For the DTLP (grid or random),  $[f_B(\delta A) - f_{DU}(\delta A)] = O((\delta A)^{1/K-1/2})$ . ■

**Proof:** We continue to assume that the number of points,  $N$ , and their locations are fixed (either on a grid or randomly). Equation (15a) and the superposition property imply that the sum of two optimal solutions with  $u$  and  $w$  as data is a feasible solution of the balanced problem with  $v$  as data. Thus, the sum of these two optimal costs bounds from above the optimal cost of the balanced problem. Since this is true for every realization of the set of net supplies and point locations ( $v, u, w, \mathbf{x}$ ) it must be true of their expected values; i.e., the average cost per point must satisfy (for fixed  $N$ ):

$$\sigma \sqrt{1 - 1/N} \left(\frac{A}{N}\right)^{1/K} f_B(N) \leq \sigma \left(\frac{A}{N}\right)^{1/K} f_{DU}(N) + \alpha \sigma N^{1/K-1/2} \left(\frac{A}{N}\right)^{1/K}.$$

[For the most part of this proof, the random versions of  $f$  stand for the dimensionless distance per point with random locations but  $N$  fixed. These functions,  $f_R^{(fixed\ number)}$  still obey (3). They are related to the conventional  $f$ 's with random  $N$  by  $\langle f_R^{(fixed\ number)}(N) \rangle = f_R^{(conventional)}(\delta A)$ . The superscript “(fixed number)” is omitted in this proof.] The first two terms in the above inequality are definitions and the third term is a result of (16c).

The same logic applied to (15b) (with  $w$  replaced by  $-w$ ) yields

$$\sigma \left(\frac{A}{N}\right)^{1/K} f_{DU}(N) \leq \sigma \sqrt{1 - 1/N} \left(\frac{A}{N}\right)^{1/K} f_B(N) + \alpha \sigma N^{1/K-1/2} \left(\frac{A}{N}\right)^{1/K};$$

i.e.,

$$\sigma \sqrt{1 - 1/N} \left(\frac{A}{N}\right)^{1/K} f_B(N) \geq \sigma \left(\frac{A}{N}\right)^{1/K} f_{DU}(N) - \alpha \sigma N^{1/K-1/2} \left(\frac{A}{N}\right)^{1/K}.$$

After dividing these two inequalities by  $\sigma \sqrt{1 - 1/N} \left(\frac{A}{N}\right)^{1/K}$ , they can be jointly written as:

$$[f_B(N) - f_{DU}(N)] = f_{DU}(N) \left[ \left( \frac{N}{N-1} \right)^{1/2} - 1 \right] \pm \alpha N^{1/K-1/2} \left( \frac{N}{N-1} \right)^{1/2}, \text{ or}$$

$$[f_B(N) - f_{DU}(N)] = f_{DU}(N)O(N^{-1}) \pm \alpha N^{1/K-1/2}O(N^0). \quad (17)$$

Since  $f_{DU}(N) = O(N^{1/K+1/2})$  by virtue of Propositions 3 and 4, we see from (17) that the lemma holds for grid problems. When  $N$  is a Poisson random variable with mean  $\delta A$ , as occurs in the random location case, the difference  $f_{BR}^{(conventional)}(\delta A) - f_{DUR}^{(conventional)}(\delta A)$  is the average of the difference for fixed numbers (17), which is also  $O((\delta A)^{1/K-1/2})$ . ■

Note that Lemma 3 is consistent with (4) and (5) since it predicts that the difference between these two expressions is bounded by a constant. Since Lemma 3 shows that the absolute difference between the asymptotic formula of an unbalanced problem and its balanced version is of an order equal or less than the order of the upper bound for unbalanced problems, a simple corollary is

**Proposition 5: (Upper bound for balanced problems).** Theorem 1 also applies to balanced DTLP's. ■

### 3. APPLICATIONS TO LOGISTICS PROBLEMS: SHAPE INDEPENDENCE

The bounds developed in Section 2 are converted into practical approximate formulae. The work was motivated by an application—a vehicle-repositioning problem for a large package delivery carrier (Smilowitz and Daganzo, 2002)—that would have been impossible without a closed-form formula. The application and the formula are described in Section 3.2. First, Section 3.1 shows that any such formula should apply to regions of arbitrary shape. Thus, formulae of this type should be quite general.

#### 3.1 Shape effects

This subsection extends the ideas in Section 2 to any region  $S$  formed by a non-overlapping assembly of  $Q$  cubes of volume  $A$ . Both an upper bound and a lower bound are developed—the latter, based on a monotonicity conjecture. If  $\phi$  is the diameter of such a region, then the following upper bound holds:

**Theorem 2: (Shape-based upper bound for DTLP and TLP(B)).**

$$f_S(\delta A Q) \leq f_D(\delta A) + \phi A^{-1/K} (\delta A)^{1/K-1/2} \quad \blacksquare \quad (18a)$$

**Proof:** We begin first with DTLP(U) and move to DTLP(B)≡TLP(B). Assume that the bilevel algorithm is applied to S with sub-depots at the centroid of each individual cube, and that the lower level flows are optimum DTLP(U) (grid or random) solutions within each cube. Clearly, the lower level distance is then  $Q(\sigma\delta^{-1/K}(\delta A)f_{DU}(\delta A))$ . The upper level distance is bounded above by the product of the maximum distance in S, the expected absolute value of the net supplies at a centroid, and the number of centroids. This product is bounded above by  $\phi(\sigma(\delta A)^{1/2})Q$ . From Proposition 2:

$$\sigma\delta^{-1/K}(\delta A Q)f_{DUS}(\delta A Q) \leq Q[\sigma\delta^{-1/K}(\delta A)f_{DU}(\delta A)] + \phi[\sigma(\delta A)^{1/2}]Q.$$

Hence, the inequality of the theorem applies to  $f_{DUS}$  and  $f_{DU}$ :

$$f_{DUS}(\delta A Q) \leq f_{DU}(\delta A) + \phi A^{-1/K} (\delta A)^{1/K-1/2} \quad \blacksquare \quad (18b)$$

Since the last term of (18b) is of an equal or lesser order than  $f_{DU}(\delta A)$  (see Theorem 1),  $f_{DUS}$  also satisfies Theorem 1. The steps of Lemma 3 can be repeated to show that  $[f_{BS}(\delta A) - f_{DUS}(\delta A)] = O((\delta A)^{1/K-1/2})$ . Again, since this is of an equal or lesser order than the right side of (18b) the theorem also holds for balanced problems.  $\blacksquare$

For a lower bound, we conjecture the following:

**Conjecture 1: (Monotonicity).** Since  $f(\delta A)$  is monotonic in 1-D, it is reasonable to conjecture that it is also monotonic in 2+-D. The 2-D simulation in Section 3.2 lends further support to the conjecture.  $\blacksquare$

Let B be a cube with volume  $B$  and diameter  $\phi_B$  that contains S and the depot. Let S' be the complement of S in B,  $S'=B-S$ , and  $Q'$  the number of elementary cubes forming S'. If we now superpose an optimal DTLP solution in S with a bi-level solution in S', where a DTLP is solved for each elementary cube and the elementary overflows are routed to the depot, we obtain a feasible DTLP solution for B. Therefore,

the distance of the superposed solutions is an upper bound to the optimum DTLP distance in B, and we can write:

$$\sigma\delta^{-1/K}\delta B f_D(\delta B) \leq \sigma\delta^{-1/K}\delta A Q f_S(\delta A Q) + \sigma\delta^{-1/K}\delta A Q' f_D(\delta A) + \phi_B \sigma Q'(\delta A)^{1/2}.$$

The first term on the right side is the optimum DTLP distance in S, the second term is the distance of the (lower level) DTLP's in S', and the third term an upper bound to the (upper level) distance linking the elementary cubes and the depot—since  $\sigma(\delta A)^{1/2}$  is an upper bound to the expected flow to/from an elementary cube,  $\phi_B$  is an upper bound to the distance between a cube and the depot, and there are  $Q'$  cubes. If we now substitute  $A(Q+Q')$  for  $B$  in the left hand side of the above and divide both sides by  $\sigma\delta^{-1/K}\delta A Q$  we find that  $f_S(\delta A Q) \geq [(Q+Q')/Q] f_D(\delta B) - [Q'/Q] f_D(\delta A) - [Q'/Q]\phi_B A^{-1/K}(\delta A)^{1/K-1/2}$ . The conjecture allows us to substitute  $f_D(\delta A)$  for  $f_D(\delta B)$  on the right side of this inequality, and this yields:

$$f_S(\delta A Q) \geq f_D(\delta A) - [Q'/Q]\phi_B A^{-1/K}(\delta A)^{1/K-1/2}. \quad (19)$$

Equations (18a) and (19), combined, imply that  $f_S(\delta A Q)/f_D(\delta A) \rightarrow (\text{constant})$  as  $\delta \rightarrow \infty$  for  $K \geq 2$ ; i.e., as the density of points tends to infinity in a given region, the boundaries of the region do not influence the number of lattice spacings traveled per point for problems defined in  $2^+-D$ . It was shown in Sec. 1.3 that this is not true in 1-D. Therefore, the 2-D case can be viewed as a transition case that shares some of the properties of 1-D (unboundedness) and some of the properties of 3-D (shape-independence).

### 3.2 Simulation of empty vehicle repositioning: The TLP(B,R)

This subsection develops a closed-form approximate formula for the Euclidean TLP(B,R) in 2-D. Such a formula has been used to estimate the yearly cost of repositioning empty vehicles among the terminals of a package delivery carrier as a function of its number of terminals, while recognizing that demand varies from day to day randomly (Smilowitz and Daganzo, 2002). Although in an ideal deterministic world the number of long-haul vehicles arriving at each long-haul terminal would perfectly match the requests at the terminal for empty vehicles, obviating the need for repositioning, in reality fluctuations from the averages create demands and oversupplies that need to be managed. Smilowitz and Daganzo (2002) assumed that this was

done by solving a TLP with zero-mean net supplies each day, using buffers of empty trucks at each terminal to absorb the trip time effects. In this way the TLP(B,R) expression completed a module for repositioning costs that itself was part of a larger analytical expression for total logistic cost. This expression was eventually used to obtain efficient system designs. This application, of course, is only one of many where an expression for the average cost of a TLP(B,R) could be of use. The formula was developed by simulation.

For simplicity, and in view of Lemma 3 and the shape-independence property, we simulated a battery of TLP(B,R) problems with fixed  $N$  in Euclidean 2-D squares. The formula is generic, however. Results are presented for twenty-five test cases, with various levels of  $A$ ,  $N$  and  $\sigma$ . In total, 769 simulations were completed with  $A$  ranging from 4,000 to 90,000 units;  $N$  from 25 to 5,000; and  $\sigma$  from 4.9 to 12.6 items. A subset of runs was used for calibration and its complement for validation. Figure 3 displays the calibration results, plotting  $\langle p_{BR}^* \rangle \delta^{1/2} / \sigma$  vs.  $\log_2(N)$ . Each point in the figure is the average of fifteen problem instances with identical properties. The equation of the line is:

$$f_{BR}(N) = 0.43 + 0.03 \log_2(N), \quad \text{for } N \in [25, 5000]. \quad (20)$$

The deviations from the line are consistent with the standard errors estimated from the simulation. The figure clearly shows that the solution indeed has the functional form of the upper bound.

Expression (20) was used with the validation test group to verify the errors. Figure 4 shows the percentage error between costs approximated with (20) and the average costs for test cases with identical parameters as a function of  $\log_2(N)$ . The errors for lower bounds using expression (6) are also presented.

### 3.3 Observations

It is interesting to compare (20) with the results for the TSP. Since the average number of items supplied per point is  $\langle (v_i)^+ \rangle = \sigma / \sqrt{2\pi}$  in the case of normal demands, we see that the average TLP distance traveled per item in the Euclidean case is  $f_B(N) \sqrt{2\pi} \delta^{-1/2}$ . That is,

$$\langle \text{distance per item} \rangle \approx \delta^{-1/2} (1 + 0.078 \log_2(N)).$$

As a point of reference, this distance is about twice as long as for the Euclidean TSP, for the values of  $N$  one is likely to encounter in actual logistics problems ( $N \approx 2^5$  to  $2^{10}$ ).

#### 4. SCALABLE NETWORKS

Many transportation network problems are complicated by the existence of non-linear edge costs. For example, if vehicles incur fixed costs, independent of their loads, and they make single-edge trips, the edge costs should be modeled as a step function of the number of items flowing on the edge. More generally, one can assume that edge costs are subadditive functions of flow; see Daganzo (1999) for background. We extend our results here to these kinds of networks. It is also possible to extend the results to multi-commodity networks in which subadditive costs are associated with multi-link vehicle routes, but length restrictions prevent the development of this idea.

Assume a directed graph characterized by sets,  $I(i)$  and  $O(i)$ , that identify the edges pointing in and out of a node  $i$ . We introduce a subadditive function  $\psi(\cdot) \geq 0$  and two constants,  $v_0 > 0$  and  $p \geq 0$ , to define edge costs. This information, combined with the sets of edges and nodes, and a distance norm, is called a “network”. For a given network and set of net supplies, we seek the edge flows  $v_{(e)}$  that solve the following network problem (NP),

$$(NP) \quad \min \quad z = \sum_e d_e v_{(e)}^p \psi(v_{(e)} / v_0) \quad (21a)$$

$$\text{s. t.} \quad \sum_{e \in O(i)} v_{(e)} - \sum_{e \in I(i)} v_{(e)} \leq v_i, \quad \forall i \quad (21b)$$

$$v_{(e)} \geq 0, \quad \forall e. \quad (21c)$$

Note that (21a) is subadditive in  $v_{(e)}$  if  $p \leq 1$ . Note too the similarity to (1). Again, one can define feasible, auxiliary (ANP) versions and DNP versions where one of the nodes is designated as a “depot”.

The results in prior sections for fixed-point locations extend to problem NP if the network is “scalable”; i.e., the network belongs to a family whose members are fully characterized by a scale parameter

$l$  and a size parameter  $N = m^K n^K$ , with the following properties. First, the graph can be partitioned into  $m^K$  identical subgraphs that define sub-networks in the family with scale parameter  $l$  and size parameter  $n^K$ . Second, paths can be extracted from the original graph to connect the sub-depots and form a graph that belongs to the original family with scale parameter  $nl$  and size parameter  $m^K$ . Depots are not required to be centrally located; see the 1-D example in Figure 5. The networks in this figure belong to a family in which members have equally spaced nodes, connected from the left to the nearest neighbor and to the right from another neighbor. Note that the sub-networks are tiles that can be joined to fill the space and make larger networks. Scalable networks in 2 and 3 dimensions can also be formed by joining tiles.

For these types of problems, dimensional analysis yields the following general solution:

$$\langle d^* \rangle = m^K n^K l \sigma^p f_N(N), \quad (22)$$

where the subscript “ $N$ ” indicates that the dimensionless distance per point pertains to a network problem with a specific geometry and set of cost parameters  $\psi(\cdot)$ ,  $p$ , and  $v_o$ . The bilevel algorithm for the DNP yields flows  $\mathbf{v}_L$  and  $\mathbf{v}_H$  that satisfy (21) with the net supply data relevant for each level. Scalability implies that the resulting average total costs can be expressed as:

$$\langle d^{(b)} \rangle_L = \langle z(\mathbf{v}_L) \rangle = m^K n^K l \sigma^p f_N(n^K), \quad (23a)$$

$$\langle d^{(b)} \rangle_H = \langle z(\mathbf{v}_H) \rangle = m^K \left( \sigma n^{K/2} \right)^p (nl) f_N(m^K). \quad (23b)$$

If  $p \leq 1$ , then (21a) is subadditive. Subadditivity guarantees that  $\langle z(\mathbf{v}_L) \rangle + \langle z(\mathbf{v}_H) \rangle \geq \langle z(\mathbf{v}_L + \mathbf{v}_H) \rangle$ , and since  $(\mathbf{v}_L + \mathbf{v}_H)$  is a feasible solution of the original problem,  $\langle z(\mathbf{v}_L + \mathbf{v}_H) \rangle \geq \langle z^* \rangle$ . Thus, if  $p \leq 1$ , then the sum of (23a) and (23b) is an upper bound for (22), and the following inequality results:

$$f_N(n^K) + n^{[1-K(1-p/2)]} f_N(m^K) \geq f_N(m^K n^K); \quad m, n = 1, 2, 3 \dots \quad (24)$$

If we put  $m = 2$  in this expression the result is analogous to (9a) when NP is linear ( $p = 1$ ). Thus, the logic in the proof of Proposition 3 also applies to linear NPs when  $m = 2$ , and the bounds of the proposition (and Theorem 1) continue to hold.



An even stronger result holds in the non-linear case with diseconomies of scale,  $p < 1$ . Recall from the proof of Proposition 3 that the solution is bounded from above by a constant when the second term on the left side of (9a) declined with  $n$ . This is also true now. Thus, in the nonlinear case the solution is bounded by a constant if  $1 - K(1 - p/2) < 0$ . This implies that 2-D network problems with diseconomies are bounded.

## 5. EXTENSIONS AND CONCLUSIONS

This paper developed bounds and formulae for the optimum cost of spatially homogeneous transportation problems. This was done with dimensional analysis, using a bilevel algorithm. We found that the 2-D bounds share some properties with those for 1-D problems (unboundedness) and other properties (shape-independence) with those in 3-D. Simulations showed that the 2-D solution has the functional form of the upper bounds. The results have been used elsewhere to estimate costs for a vehicle-repositioning problem arising in the package delivery industry.

Results were also extended to more general single-commodity network problems with non-linear edge costs. It was found that the costs per item for these types of networks are bounded with respect to network size if link costs have economies of scale. For such networks, the negligible contribution of long distance travel to total cost explains the prevalence of the “last-mile” effect. Future work should examine the asymptotic properties of multi-commodity networks with route-based costs.

The results of this paper can also be used for inhomogeneous transportation problems if the mean and variance of the net supplies at each location are known. A feasible solution to these problems can always be formed by the superposition a deterministic solution to the TLP (with the mean net supplies as data) and a stochastic component (with the deviations as data). It should be intuitive that the average TLP cost is bounded from below by the deterministic cost component and from above by the sum of the deterministic and stochastic costs—a proof of this statement is given in Daganzo and Smilowitz (2000). The formulae in this paper can be used to estimate the stochastic component. When this component is inhomogeneous (different variances) it can itself be bounded from above by a homogeneous problem with larger variance. More discussion of these issues can be found in Daganzo and Smilowitz (2000).

## ACKNOWLEDGEMENTS

Research supported in part by the University of California Transportation Center.

## REFERENCES

- G. I. Barenblatt, *Scaling: Dimensional Analysis and Intermediate Asymptotics*, Cambridge University Press, Cambridge, U. K. (1996).
- J. Beardwood, J. H. Halton and J. M. Hammersley, "The Shortest Path Through Many Points," *Proceedings Cambridge Philosophical Society* 55, 299-327 (1959).
- J. Bramel and D. Simchi-Levi, *The Logic of Logistics*, Springer, Heidelberg, Germany (1997).
- P. W. Bridgman, *Dimensional Analysis*, Yale University Press, New Haven, CT (1963).
- E. Buckingham, "On Physically Similar Systems: Illustrations of the Use of Dimensional Equations," *Physics Review* 4, 345-376 (1914).
- C. F. Daganzo, "The Length of Tours in Zones of Different Shapes," *Transportation Research* 18B, 135-146 (1984a).
- C. F. Daganzo, "The distance traveled to Visit  $N$  Points with a Maximum of  $C$  Stops per Vehicle: An Analytic Model and an Application", *Transportation Science* 18, 331-350 (1984b).
- C. F. Daganzo, *Logistics Systems Analysis*, Springer, Heidelberg, Germany (1991) (and 3rd edition, 1999)
- C. F. Daganzo and K. R. Smilowitz, "Asymptotic Approximations for the Transportation LP and Other Scalable Network Problems", Working Paper UCB-ITS-WP-2000-2, Institute of Transportation Studies, University of California, Berkeley, 2000.
- S. Eilon, C. D. T. Watson-Gandy, and N. Christofides, *Distribution Management: Mathematical Modelling and Practical Analysis*, Hafner, New York, N.Y. (1971).
- M. A. Haimovich, A. H. G. Rinnooy-Kan and L. Stougie, "Bounds and Heuristics for Capacitated Routing Problems," *Mathematics of Operations Research* 10, 527-542 (1985).
- E. Hillie and R. S. Phillips, "*Functional Analysis and Semi-groups*," Waverly Press, Baltimore, MD (1957).

- N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York, N.Y. (1972).
- R. M. Karp, "Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem," *Mathematics of Operations Research* 2, 209-224 (1977).
- G. F. Newell, "Design of Multiple Vehicle Delivery Tours--III: Valuable Goods", *Transportation Research* 20B, 377-390 (1986).
- G. F. Newell and C. F. Daganzo, "Design of Multiple Vehicle Delivery Tours--I: A Ring-Radial Network", *Transportation Research* 20B, 345-364 (1986a).
- G. F. Newell and C. F. Daganzo, "Design of Multiple Vehicle Delivery Tours--II: Other Metrics", *Transportation Research* 20B, 365-376 (1986b).
- C. H. Papadimitriou, "The Probabilistic Analysis of Matching Heuristics," Proceedings 15<sup>th</sup> Allerton Conference on Communication, Control and Computing, University of Illinois, Urbana, IL, 363-378 (1978).
- K. R. Smilowitz, "Design and Operation of Multi-Mode, Multi-Service Logistics Systems" Ph.D. thesis, University of California, Berkeley (2001).
- K. R. Smilowitz and C. F. Daganzo, "Cost Modeling and Solution Techniques for Complex Transportation Systems", *Transportation Science*, submitted (2002).
- J. M. Steele, "Optimal Triangulation of Random Samples in the Plane," *Annals of Probability* 10, 548-553 (1982).
- J. M. Steele, "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges," *Annals of Probability* 16, 1767-1787 (1988).
- J. M. Steele, *Probability Theory and Combinatorial Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA (1997).
- J. E. Yukich, *Probability Theory of Classical Euclidean Optimization Problems*, Lecture Notes in Mathematics #1675, Springer, Berlin, Germany (1998).

## APPENDIX: ASYMPTOTIC FORMULAE FOR ONE-DIMENSIONAL TLPs

### A.1. Balanced TLP problems and derivation of Eq. (4)

We define a curve of cumulative supply vs. position,  $v(x) = \sum_{x_i \leq x} v_i$ , as in Fig. A1a for a balanced TLP problem with random point locations.

**Result A1: (Formula for TLP(B)).** The minimum cost for TLP(B) is the absolute area between  $v(x)$  and the x-axis; i.e.,

$$d^* = \int_{-\infty}^{+\infty} |v(x)| dx = \sum_{i=1}^{N-1} |v(x_i)| |x_{i+1} - x_i|. \quad (A1)$$

**Proof:** It suffices to show that (A1) is an upper bound and a lower bound for  $d^*$ . For any point,  $x_p$ , (see Fig.A1a) the net flow across  $x_p$  in a feasible solution is  $v(x_p)$  since the aggregate supply and demand on both sides of  $x_p$  are satisfied. Thus,  $|v(x_p)| dx$  is a lower bound to the optimal distance traveled in any small interval,  $(x_p, x_p + dx)$  where  $v(x)$  is constant, and the sum on the right side of (A1) is a lower bound for  $d^*$ .

Conversely, a feasible solution can be constructed by considering horizontal slices of  $dv$  items (as in the figure) and transporting these quantities from points where the slice intersects a rising portion of curve  $v(x)$  to adjoining points where it intersects a falling portion. In the figure,  $dv$  items would be carried from A to B and from C to D. Thus, the summation of all slices for small  $dv$  (still given by (A1)) is the distance of a feasible solution and an upper bound to  $d^*$ . ■

For the TLP(B,G) with lattice spacing  $l$ , (A1) reduces to  $d^* = l \sum_{i=1}^{N-1} |v(x_i)|$ . For any homogeneous, balanced problem the net supplies must have zero mean,  $\langle v_i \rangle = 0$ , the same absolute mean  $\langle |v_i| \rangle = \mu$ , and the same variance,  $\langle v_i^2 \rangle = \sigma^2$ . The (equal) covariances are  $\langle v_i v_j \rangle = -\sigma^2 / (N-1)$  since this covariance ensures that the variance of the sum of the net supplies is zero. The general formulae for the mean, absolute mean, and variance of the sum of the first  $i$  net supplies,  $v(x_i)$ , are then:

$$\langle v(x_i) \rangle = 0, \quad \langle |v(x_i)| \rangle = i\mu, \quad \text{and} \quad \langle v(x_i) \rangle = i\sigma^2 \left[ 1 - \frac{i-1}{N-1} \right] \quad (\text{A2})$$

If  $X$  is a normal random variable with zero-mean,

$$\langle |X| \rangle = [2\langle X^2 \rangle / \pi]^{1/2}. \quad (\text{A3})$$

Thus, if the  $v_i$  have a joint multinormal distribution, we have:

$$\langle |v(x_i)| \rangle = \left( \frac{2}{\pi} \right)^{1/2} \sigma \left( i \left[ 1 - \frac{i-1}{N-1} \right] \right)^{1/2},$$

and the following is true.

**Result A2: (Expected optimal cost of TLP(B,G)).** For the homogeneous, zero-mean TLP(B,G) with normal demand in  $R^1$ ,

$$\langle d_{BG}^* \rangle = l \sum_{i=1}^{N-1} |v(x_i)| = \left( \frac{2}{\pi} \right)^{1/2} \sigma l \sum_{i=1}^{N-1} \left( i \left[ 1 - \frac{i-1}{N-1} \right] \right)^{1/2}. \quad (\text{A4a})$$

Furthermore, the limit of this expression for  $N \rightarrow \infty$  is

$$\langle d_{BG}^* \rangle \rightarrow \left( \frac{\pi}{32} \right)^{1/2} \sigma l N^{3/2}. \quad (\text{A4b})$$

Equation (A4b) is true because the right hand side of (A4a) is a Riemann sum that becomes a definite integral for  $N \rightarrow \infty$ , and such integral reduces to (A4b). The details are as follows:

$$\left( \frac{2}{\pi} \right)^{1/2} \sigma l \int_0^N \left[ x \left( 1 - \frac{x}{N} \right) \right]^{1/2} dx = \left( \frac{2}{\pi} \right)^{1/2} \sigma l N^{3/2} \int_0^1 \left[ \left( \frac{x}{N} \right) \left( 1 - \frac{x}{N} \right) \right]^{1/2} d \left( \frac{x}{N} \right) = \left( \frac{\pi}{32} \right)^{1/2} \sigma l N^{3/2}. \quad (\text{A5})$$

In terms of the point density,  $\delta = 1/l$ , the average distance per point,  $\langle p_{BG}^* \rangle = \langle d_{BG}^* \rangle / N$ , is:

$$\langle p_{BG}^* \rangle \rightarrow \sqrt{\frac{\pi}{32}} \sigma \delta^{-1} \sqrt{N}, \quad (\text{A6})$$

which reduces to (4).

## A.2. Unbalanced TLP problems and derivation of Eq.(5)

Equation (A6) holds for DTLP(B) since  $\text{DTLP(B)} \equiv \text{TLP(B)}$ . The expression for DTLP(U) is different, but qualitatively similar. To derive it, first define the cumulative demand for the depot  $v'(x)$  as shown in Fig. A1b; i.e.,  $v'(x) = -v_0 H(x - x_0)$ , where  $H$  is the Heaviside unit step function and  $x_0$  is the depot location.

Then, the same arguments used with Result A1 establish the following:

**Result A3: (Deterministic DTLP).** For both grid and random problems,

$$d_D^* = \int_{-\infty}^{+\infty} |v(x) - v'(x)| dx . \blacksquare \quad (\text{A7})$$

For the DTLP(U), the expectation of the integrand of (A7) is symmetric with respect to the location of the depot. Thus, for an integration region  $[0, L]$  it can be simplified as follows:

$$\begin{aligned} \langle d_{UD}^* \rangle &= \int_0^L \langle |v(x) - v'(x)| \rangle dx = 2 \int_0^{L/2} \langle |v(x) - v'(x)| \rangle dx = 2 \int_0^{L/2} \langle |v(x)| \rangle dx, \text{ which from (A.3) is} \\ \langle d_{UD}^* \rangle &= 2 \int_0^{L/2} \langle |v(x)| \rangle dx \rightarrow 2 \int_0^{L/2} [(2/\pi)\sigma^2(Nx/L)]^{1/2} dx = \sqrt{\frac{4}{9\pi}} \sigma LN^{1/2}, \text{ i.e., we have} \end{aligned}$$

**Result A4: (Expected optimal cost of DTLP(U)).**

$$\langle d_{UD}^* \rangle \rightarrow \sqrt{\frac{4}{9\pi}} \sigma LN^{\frac{3}{2}}, \text{ and } \langle p_{UD}^* \rangle \rightarrow \sqrt{\frac{4}{9\pi}} \sigma LN^{\frac{1}{2}}, \text{ and } f_{DU}(N) / \sqrt{N} \rightarrow \sqrt{\frac{4}{9\pi}} . \blacksquare \quad (\text{A8})$$

This is Eq. (5) of the text.

An exact expression for TLP(U) is more difficult to obtain. Yet, it is shown in Daganzo and Smilowitz (2000) that  $|\langle p_{BG}^* \rangle - \langle p_{UG}^* \rangle| = O(N^{1/2})$ , implying that  $\langle p_{UG}^* \rangle$  is also  $O(N^{1/2})$  for unbalanced problems.

**LIST OF FIGURES**

Figure 1: Variants of the TLP

Figure 2: Application of the bilevel algorithm to (a) DTLP(U,R) and (b) DTLP(U,G)

Figure 3: Calibration results:  $\langle p_{BR}^* \rangle \delta^{1/2} / \sigma$  vs.  $\log_2(N)$

Figure 4: Error comparison

Figure 5: Scalable networks

Figure A1: Graphical solutions of 1-D problems (a) TLP(B); (b) DTLP(U)

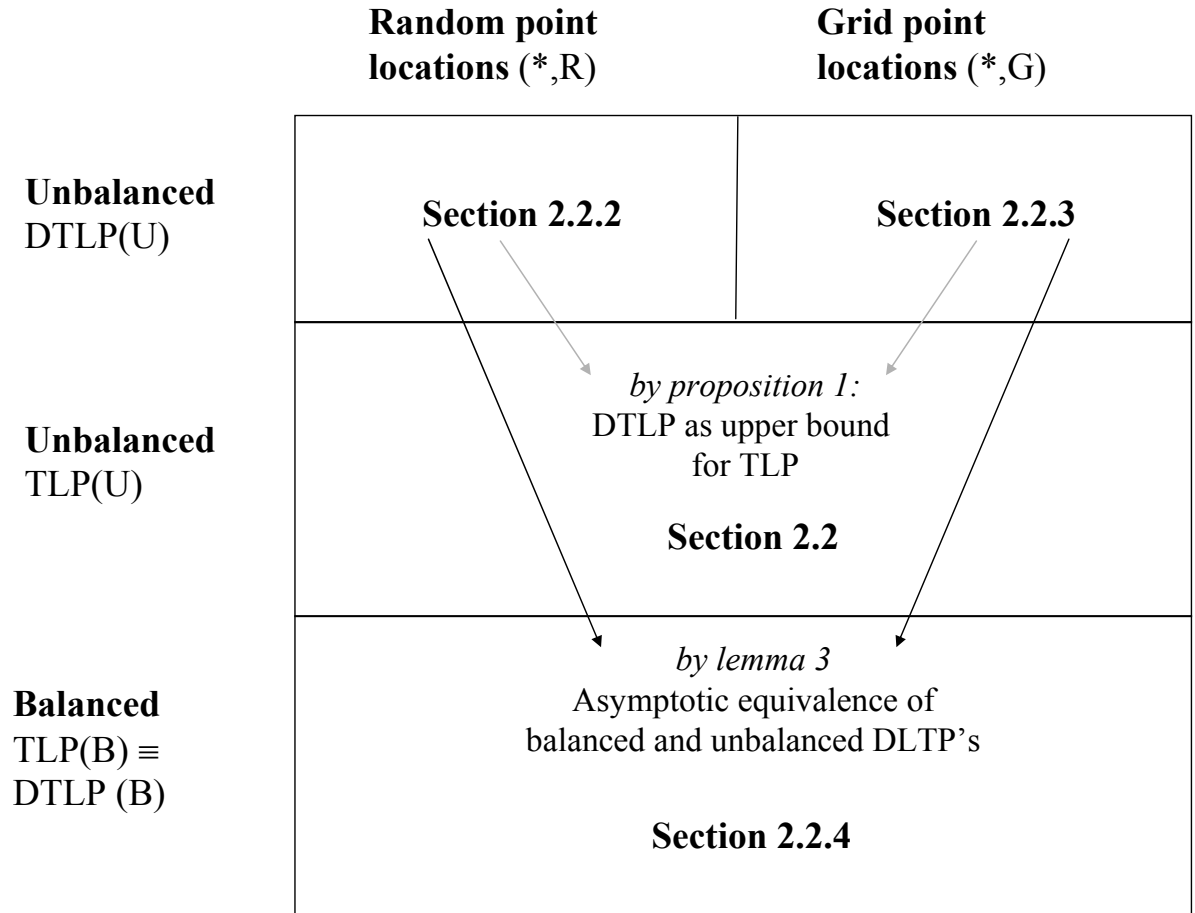


Figure 1: Variants of the TLP



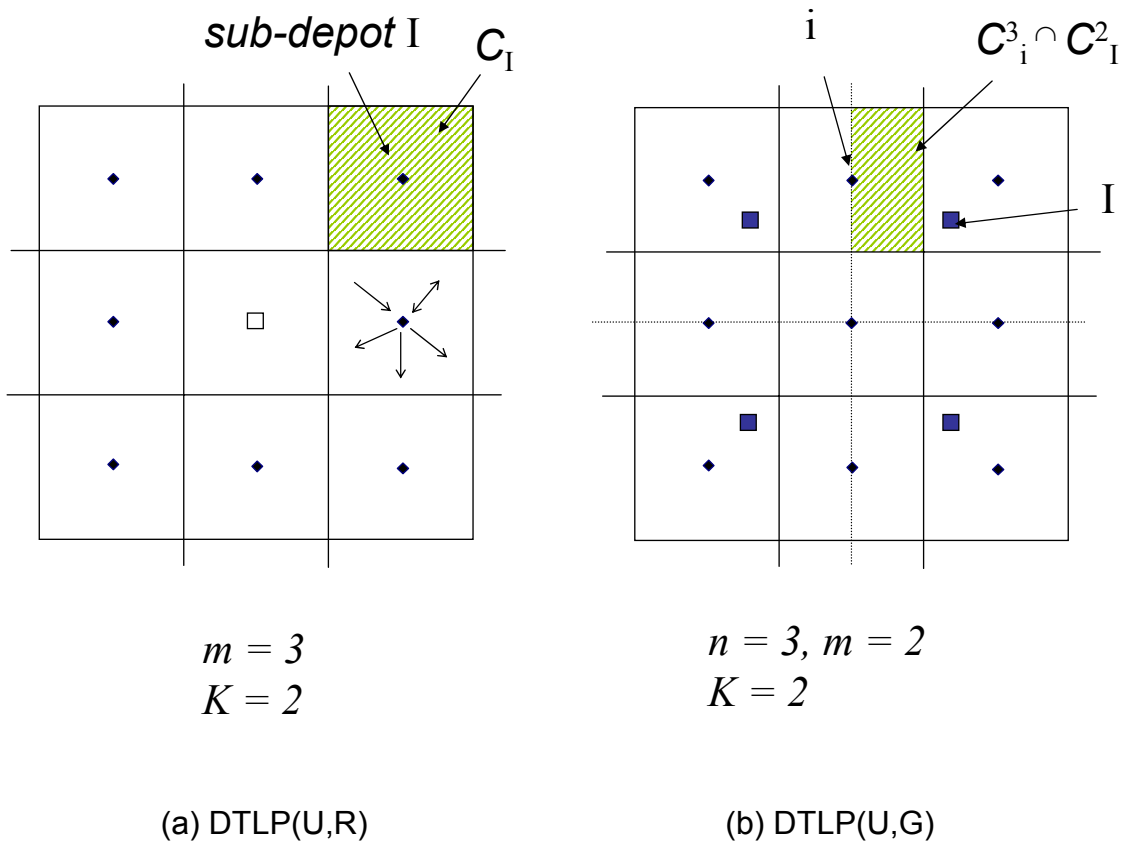


Figure 2: Application of the bilevel algorithm to (a) DTLP(U,R) and (b) DTLP(U,G) .

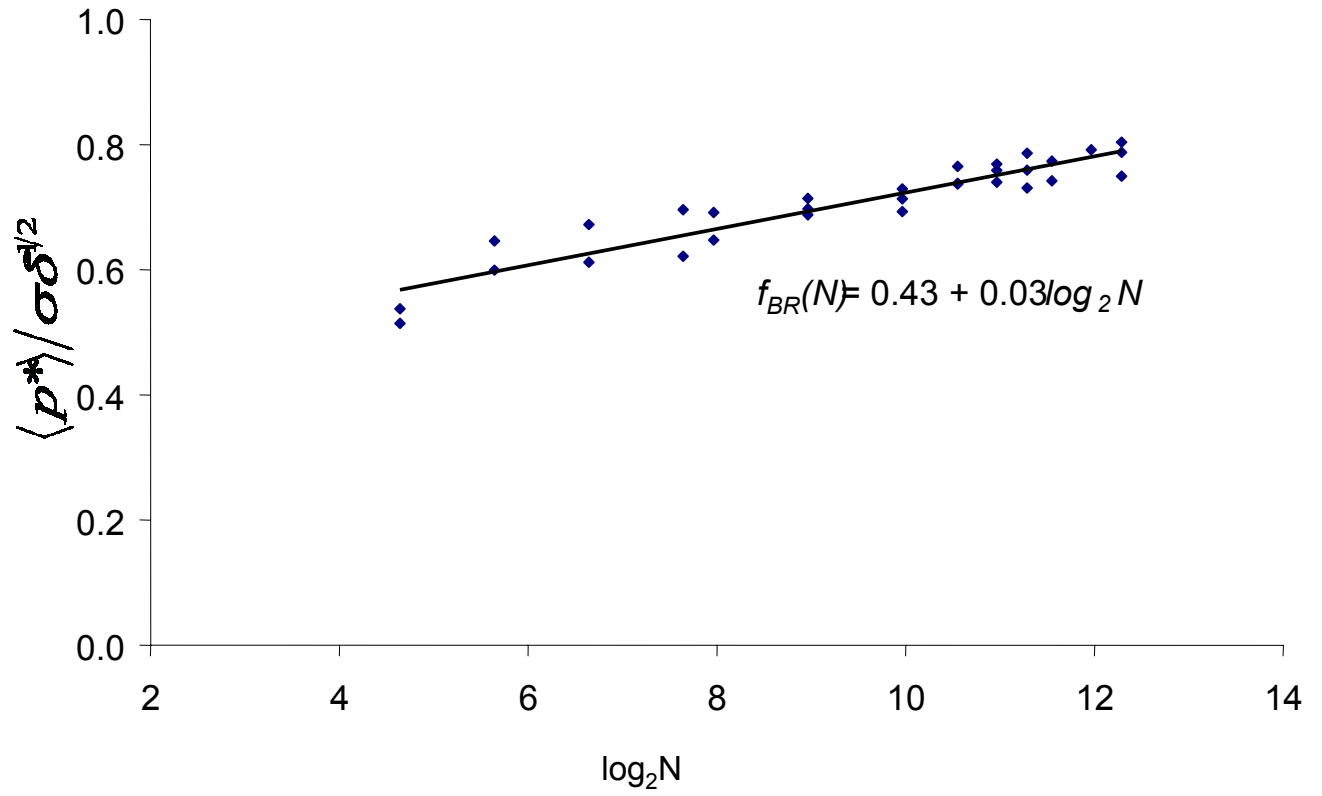


Figure 3: Calibration results:  $\langle p_{BR}^* \rangle \delta^{1/2} / \sigma$  vs.  $\log_2(N)$

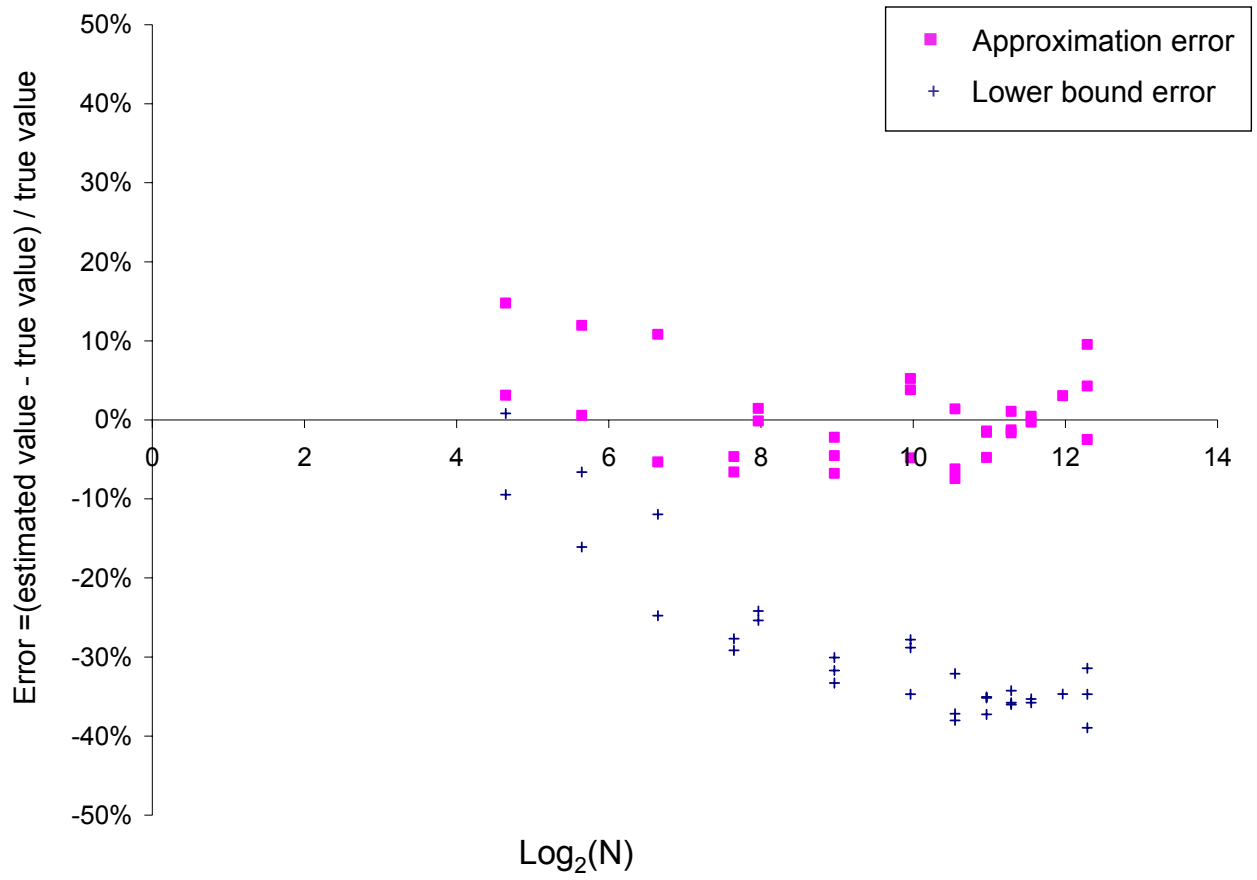
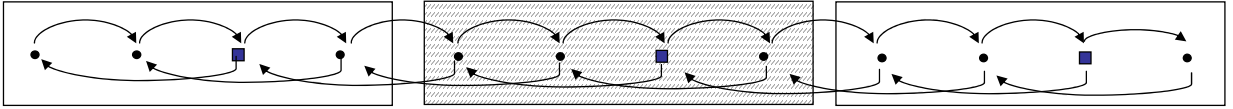
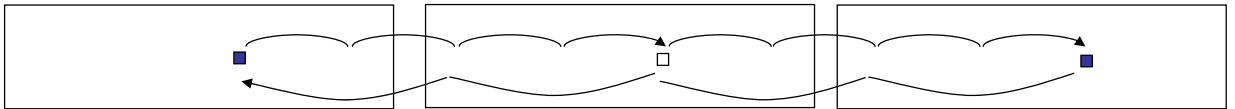


Figure 4: Error comparison

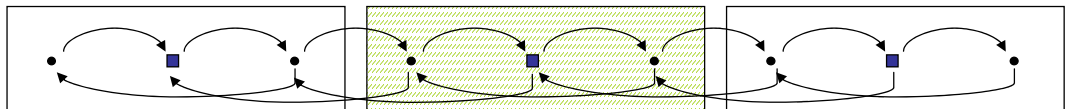
Lower level problem:  $m=3; n=4$



Upper level problem:  $m=3; n=4$



Lower level problem:  $m=3; n=3$



Upper level problem:  $m=3; n=3$

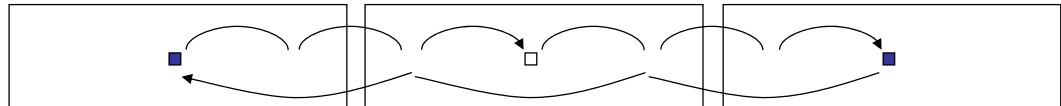


Figure 5: Scalable networks

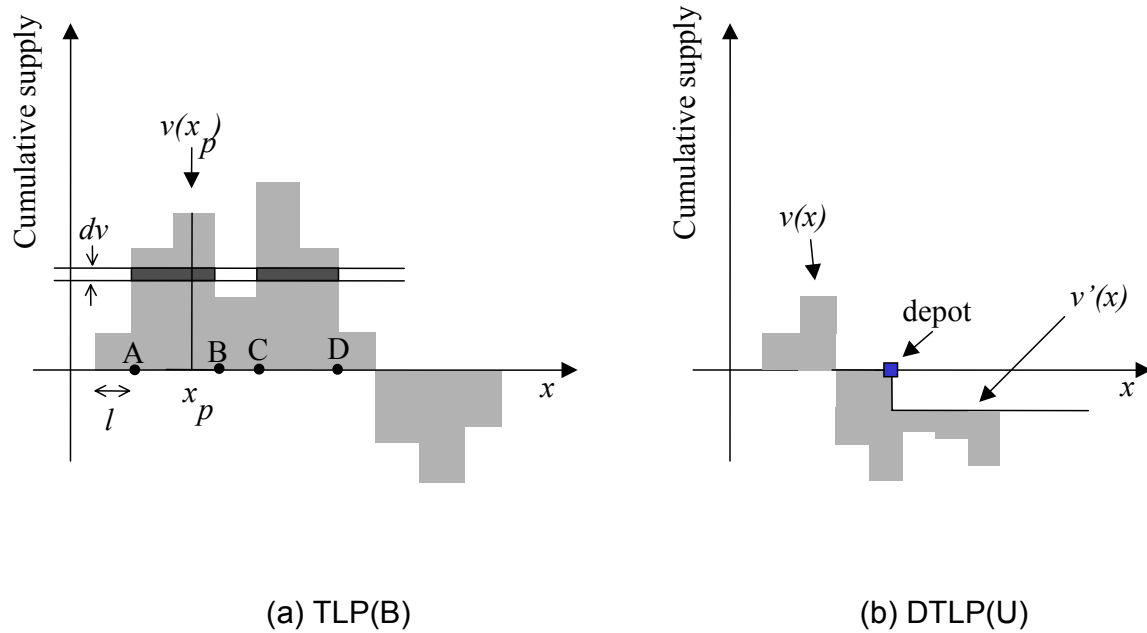


Figure A1: Graphical solutions of 1-D problems (a) TLP(B); (b) DTLP(U)