

A note on asymptotic formulae for one-dimensional network flow problems

Carlos F. Daganzo* and Karen R. Smilowitz*

(to appear in *Annals of Operations Research*)

Abstract

This note develops asymptotic formulae for single-commodity network flow problems with random inputs. The transportation linear programming problem (TLP) where N points lie in a region of \mathbb{R}^1 is one example. It is found that the average distance traveled by an item in the TLP increases with $N^{1/2}$; i.e., the unit cost is unbounded when N and the length of the region are increased in a fixed ratio. Further, the optimum distance does not converge in probability to the average value. These one-dimensional results are a useful stepping stone toward a network theory for two and higher dimensions.

Key Words: Transportation problem; distance approximations

1. Introduction

Asymptotic formulae exist for the traveling salesman problem, “TSP”, (Eilon et al., 1971, Karp, 1977, Daganzo, 1984a), and the vehicle routing problem, “VRP”, (Eilon et al., 1971, Daganzo, 1984b, Haimovich et al., 1985, Newell and Daganzo, 1986 and 1986a). The results apply 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. This limiting result holds for problems in K -dimensional regions. This note presents related but qualitatively different results for single-commodity network flow problems with random inputs on undirected paths. The transportation linear programming problem, “TLP”, with points in a region of \mathbb{R}^1 is an example of this class of problems. One dimensional TLP’s arise in connection with the earthwork minimization problem of highway design.

2. Background

* Institute of Transportation Studies and Department of Civil and Environmental Engineering
University of California, Berkeley CA 94720; daganzo@ce.berkeley.edu.

* Department of Industrial Engineering and Management Sciences, Northwestern University.

Here the TLP is defined as follows. Given are N points, inter-point distances, $\{d_{ij}, \forall i, j=1..N; i \neq j\}$, satisfying the triangle inequality, and 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 shipments, $\{v_{ij}, \forall i, j=1..N; i \neq j\}$, that minimize the total distance traveled while satisfying flow balance constraints.

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

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

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

Equations (1b) specify flow-conservation at each point, ensuring that the net flow from a point i never exceeds the net supply at i . Because distances satisfy the triangle inequality, in an optimum solution origin nodes only emit flow and destination nodes only receive flow. A TLP is feasible only if $\sum_{i \in N} v_i \geq 0$, as can

be seen by summing (1b) across i . For balanced problems constraints (1b) are individually satisfied as pure equalities; for unbalanced problems, they are not. We define an always-feasible problem, ATLP, that includes a fictitious source, $i = 0$, with the least positive net supply to make 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 impose a penalty for items not shipped. Since M is large, the distances satisfy the triangle inequality, and the ATLP is a TLP. If the original TLP is infeasible then the ATLP is a balanced TLP, and the outflow from the fictitious source is v_0 . The penalty component of an

optimum solution is $v_0 M$, and the distance component is $d^* = z^* - v_0 M$, with $v_0 = \left(-\sum_{i=1}^N v_i\right)^+$. If the TLP

is feasible, then $d^* = z^*$; otherwise we take the distance component of the ATLP as its solution.

In the TLP/ATLP it is assumed that excess supplies are left at the origins. We consider a variant, called ‘‘depot-TLP’’, or DTLP, where excess supplies are carried to the extra point, or ‘‘depot’’. In the DTLP the depot distances do not have to be fixed or large but must be non-negative, $(d_{0j}, d_{i0}) \geq 0, \forall i, j$, and satisfy the triangle inequality. The minimum of the DTLP objective function, z_D , is denoted d_D^* . Due to the penalty depot distances, it should be clear that the following is true (a formal proof can be found in Daganzo and Smilowitz, 2002).

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 $d^* = d_D^*$. ■

3. Formulae

This section develops formulae for the averages of d^* and d_D^* over a set of solutions when conditions vary. Points are embedded in the normed linear space R^1 and identified by a single coordinate, x_i . Distance is $d_{ij} = |x_i - x_j|$. We examine balanced and unbalanced versions of the TLP and DTLP. We also consider grid problems where the net supplies change as random variables, but points are fixed on a 1-D grid, and random problems where points are located randomly. The modifiers “G” or “R” for “grid” or “random”, and “U” or “B” for “unbalanced” or “balanced”, identify problem characteristics.

3.1 Balanced TLP problems

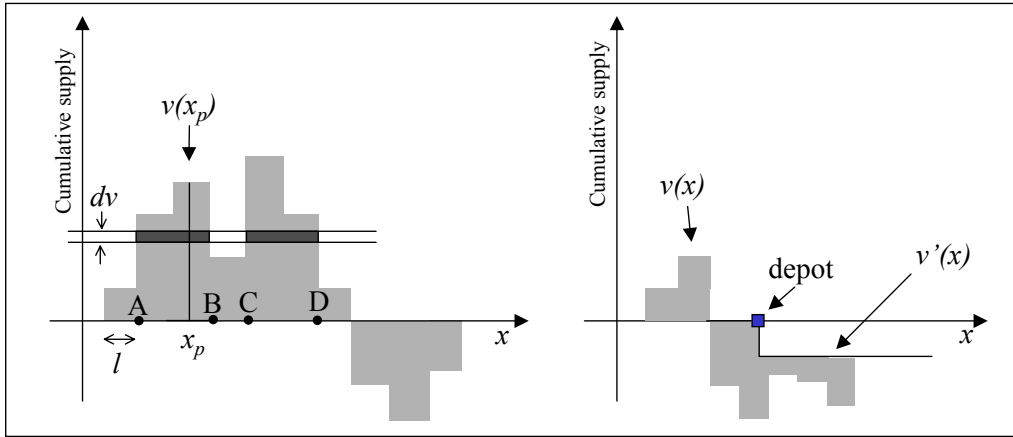


Fig. 1a presents cumulative supply vs. position, $v(x) = \sum_{x_i \leq x} v_i$ for a TLP(B) with random point locations.

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

Result 1. (Formula for TLP(B)). The minimum cost for TLP(B) is the absolute area between $v(x)$ and the x-axis; i.e.,

$$d^* = \int_{-\infty}^{+\infty} |v(x)| dx = \sum_{i=1}^{N-1} |v(x_i)| |x_{i+1} - x_i|. \quad \blacksquare \quad (2)$$

Proof: It suffices to show that (2) is both an upper bound and a lower bound for d^* . For any point, x_p , such as the one in Fig. 1a, the net flow across x_p in any feasible solution is $v(x_p)$ because the aggregate supply and demand on both sides of x_p must be 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. Therefore, the sum on the right

side of (2) is a lower bound for d^* . Conversely, a feasible solution can be constructed by considering horizontal slices of dv items and transporting the items 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 the slices for small dv (still given by (2)) is the distance of a feasible solution and an upper bound to d^* . ■

3.1.1 Grid problems

If points are on a grid with lattice spacing l , then (2) reduces to $d^* = l \sum_{i=1}^{N-1} |v(x_i)|$. We examine the expected value of d^* for TLP(B,G) with random net supplies. 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 must be $\langle v_i v_j \rangle = -\sigma^2 / (N-1)$ since this covariance ensures that the variance of the sum of all 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:

$$\langle v(x_i) \rangle = 0, \quad \langle |v(x_i)| \rangle = i\mu, \quad \text{and} \quad \langle v(x_i)^2 \rangle = i\sigma^2 \left[1 - \frac{i-1}{N-1} \right] \quad (3)$$

Assume now that the v_i have a joint multinormal distribution, and recall that if X is a normal random variable with zero-mean, $\langle |X| \rangle = [2\langle X^2 \rangle / \pi]^{1/2}$. Thus, in the multinormal case, we have:

$$\langle |v(x_i)| \rangle = \left(\frac{2}{\pi} \right)^{1/2} \sigma \left(i \left[1 - \frac{i-1}{N-1} \right] \right)^{1/2}. \quad (4)$$

Result 2. (Expected optimal cost of TLP(B,G)). For the homogeneous, zero-mean TLP(B,G) with normal demand in R^1 ,

$$\langle d_{BG}^* \rangle = \left(\frac{2}{\pi} \right)^{1/2} \sigma l \sum_{i=1}^{N-1} \left(i \left[1 - \frac{i-1}{N-1} \right] \right)^{1/2}. \quad (5a)$$

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

$$\langle d_{BG}^* \rangle \rightarrow \left(\frac{\pi}{32} \right)^{1/2} \sigma l N^{3/2}. \quad \blacksquare \quad (5b)$$

Equation (5b) is true because the right hand side of (5a) is a Riemann sum that becomes a definite integral for $N \rightarrow \infty$, and such integral reduces to (5b). The details are as follows:

$$\left(\frac{2}{\pi}\right)^{1/2} \sigma \int_0^N \left[x \left(1 - \frac{x}{N}\right) \right]^{1/2} dx = \left(\frac{2}{\pi}\right)^{1/2} \sigma N^{3/2} \int_0^1 \left[\left(\frac{x}{N}\right) \left(1 - \frac{x}{N}\right) \right]^{1/2} d\left(\frac{x}{N}\right) = \left(\frac{\pi}{32}\right)^{1/2} \sigma N^{3/2}. \quad (6)$$

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

$$\langle p_{BG}^* \rangle \rightarrow \sqrt{\frac{\pi}{32}} \sigma \delta^{-1} \sqrt{N}. \quad (7)$$

This function increases without limit with the number of points, unlike in the TSP, where the average distance per point tends to a limit. The dependence with \sqrt{N} is caused by the long-range interactions arising from the flow balancing requirements.

3.1.2 Extensions: random problems and non-normal demands

Equations (5b) and (7) are general and hold if the v_i are not normal, but satisfy the regularity conditions of Cramér's version of the central limit theorem for large deviations. Let $F_{v(x)}$ be the c.d.f. of $v(x)$ for a given N and let $N \rightarrow \infty$ with a fixed interval length, L . Then, $v(x)$ is the sum of $xN/L \rightarrow \infty$ net supplies, and the

central limit theorem applies. Recall $\langle d^* \rangle = \int_0^L \langle |v(x)| \rangle dx$. Let Φ be the c.d.f. of the normal approximation to

$v(x)$. Note that $\langle |v(x)| \rangle = \int_{-\infty}^0 F_{v(x)}(z) dz + \int_0^{\infty} [1 - F_{v(x)}(z)] dz \approx \int_{-\infty}^0 \Phi(z) dz + \int_0^{\infty} [1 - \Phi(z)] dz$. The central limit theorem

for large deviations (see corollary on p.553 of Feller, 1971) guarantees that the error in $\langle |v(x)| \rangle$ can be made smaller than any $\varepsilon > 0$ for all $v(x)$. Thus, the error in $\langle d^* \rangle$ is bounded by εL . QED

Equations (5b) and (7) also hold if point locations vary in a segment of length L as a homogeneous Poisson process with rate δ and an average of N points ($N = \delta L$). Since $v(x)$ is a compound Poisson random variable, the central limit theorem applies and $\langle |v(x)| \rangle \rightarrow [2\langle v(x) \rangle / \pi]^{1/2}$. It suffices to show

$\langle v(x) \rangle \rightarrow \sigma^2 N \left[\frac{x}{L} \left(1 - \frac{x}{L}\right) \right]$ because then the expectation of (2), $\langle d^* \rangle = \int_{-\infty}^{+\infty} \langle |v(x)| \rangle dx$, is (6). Let $i(x)$ be

the number of points in $[0, x]$, and note the conditional random variable $(v(x) | i(x))$ has zero mean and

variance $\langle v(x) | i(x) \rangle = i(x) \sigma^2 \left(1 - \frac{(i(x)-1)}{(N-1)}\right)$. For zero mean, the unconditional variance is the expectation of

the conditional variance; i.e.

$$\langle v(x) \rangle = \sigma^2 \left[\frac{\langle i(x) \rangle N}{N-1} - \frac{\langle i(x)^2 \rangle}{N-1} \right] = \sigma^2 \left[\frac{N^2 x}{L(N-1)} - \frac{1}{N-1} \left[\left(\frac{Nx}{L} \right)^2 + N \left(\frac{x}{L} \right) \left(1 - \frac{x}{L} \right) \right] \right] \rightarrow \sigma^2 N \left[\frac{x}{L} \left(1 - \frac{x}{L} \right) \right]. \text{ QED}$$

3.2 Unbalanced TLP problems (grid and random point locations)

An exact expression for TLP(U) is more difficult to obtain, yet it is shown in Daganzo and Smilowitz (2001) 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. Equation (7) 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. 1b; 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. Since the DTLP(U) is a balanced TLP with cumulative demand $v(x) - v'(x)$, Result 1 applies with $v(x) - v'(x)$ substituted for $v(x)$, and we have:

Result 3. (Deterministic DTLP). For both grid and random problems,

$$d_D^* = \int_{-\infty}^{+\infty} |v(x) - v'(x)| dx. \quad \blacksquare \quad (8)$$

For the DTLP(U), the expectation of the integrand of (8) is symmetric with respect to the location of the depot. Thus, for an integration region $[0, L]$ it can be simplified as follows:

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

Result 4. (Expected optimal cost of DTLP(U)).

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

Note that in all cases, $\langle d^* \rangle = O(N^{3/2})$ when one holds l constant as N is increased. However, $\langle d^* \rangle = O(N^{1/2})$ if one holds the total region size, $L = lN$, constant.

3.2.1 Convergence for unbalanced TLP cost approximations

The form of convergence of d^* is examined next. Unlike the TSP d_{UD}^* does not converge in probability to (9) because its coefficient of variation tends to a positive constant as $N \rightarrow \infty$.

Result 5. (Asymptotic coefficient of variation of the optimal distance for DTLP(U)) The asymptotic coefficient of variation tends to a positive constant as N increases; i.e., $d_{UD}^* \langle^{1/2} / \langle d_{UD}^* \rangle \rightarrow \text{constant} > 0$. ■

Proof: Recall d_{UD}^* is a Riemann sum of the distance in small intervals of $[0, L]$, with a second moment,

$$\langle d_{UD}^{*2} \rangle = \langle \int_0^L \int_0^L |v(x) - v'(x)| |v(y) - v'(y)| dx dy \rangle = \int_0^L \int_0^L \langle |v(x) - v'(x)| |v(y) - v'(y)| \rangle dx dy. \quad \text{By symmetry,}$$

$$\langle d_{UD}^{*2} \rangle = 2 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| |v(y)| \rangle dx dy + 2 \int_0^{L/2} \left(\int_{L/2}^L \langle |v(x)| [|v(y) - v'(y)]| \rangle dy \right) dx. \quad \text{For large } N, \text{ curve } v(x) \text{ has}$$

independent increments. Thus, the integrand of the second integral involves the product of independent quantities since the absolute net supplies being multiplied correspond to non-overlapping regions of $[0, L]$. The expression can be rewritten using the product of the expectations as:

$$\begin{aligned} \langle d_{UD}^{*2} \rangle &= 2 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| |v(y)| \rangle dx dy + 2 \int_0^{L/2} \left(\int_{L/2}^L \langle |v(x)| \rangle \langle [|v(y) - v'(y)]| \rangle dy \right) dx \\ &= 2 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| |v(y)| \rangle dx dy + 2 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| \rangle \langle |v(y)| \rangle dx dy. \end{aligned}$$

The second inequality follows from symmetry. Note now that the first integrand satisfies:

$$\langle |v(x)| |v(y)| \rangle = \langle |v(x)| \rangle \langle |v(y)| \rangle + \text{cov}\{|v(x)|, |v(y)|\}. \quad \text{Thus,}$$

$$\langle d_{UD}^{*2} \rangle = 4 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| \rangle \langle |v(y)| \rangle dx dy + 2 \int_0^{L/2} \int_0^{L/2} \text{cov}(|v(x)|, |v(y)|) dx dy$$

The covariance integrand is strictly positive for all x and y not equal to 0, because $v(x)$ and $v(y)$ share the net supplies from 0 to $\min(x, y)$. An expression for $\text{cov}(|v(x)|, |v(y)|)$ is of the form $\sigma^2 N \rho(x/L, y/L)$, and ρ is positive if both its arguments are positive. The second moment is then:

$$\begin{aligned} \langle d_{UD}^{*2} \rangle &= 4 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| \rangle \langle |v(y)| \rangle dx dy + 2 \sigma^2 N \int_0^{L/2} \int_0^{L/2} \rho(x/L, y/L) dx dy \\ &= 4 \int_0^{L/2} \int_0^{L/2} \langle |v(x)| \rangle \langle |v(y)| \rangle dx dy + 2 \sigma^2 N L^2 \int_0^{1/2} \int_0^{1/2} \rho(x', y') dx' dy' \end{aligned}$$

Note the second double integral is a positive constant; the second term is a positive multiple of $\sigma^2 N L^2$ and the first term is $\langle d_{UD}^* \rangle^2$. The second moment becomes $\langle d_{UD}^{*2} \rangle = \langle d_{UD}^* \rangle^2 + \rho_o \sigma^2 N L^2$, for some $\rho_o > 0$, and the variance is $\langle d_{UD}^{*2} \rangle - \langle d_{UD}^* \rangle^2 = \rho_o \sigma^2 N L^2 = \rho_o \sigma^2 l^2 N^3$. Since $\langle d_{UD}^* \rangle \rightarrow \sqrt{\frac{4}{9\pi}} \sigma l N^{3/2}$, the coefficient of variation is independent of σ , l and N , and it tends to a positive constant as stated. ■

4. Conclusions and comparison with other TLP problems

It is interesting to compare these results with those for other TLP versions (see Smilowitz and Daganzo, 2002). The result, $\langle p^* \rangle = \langle d^* \rangle / N = O(N^{1/2})$ when one holds δ constant, does not extend to planar problems or problems in higher dimensions. The dependence on N is damped in higher dimensions. In 2-D, the dependence is only of order $\log(N)$. In 3-D, $\langle p^* \rangle$ is bounded and the standard deviation of the optimum distance per point declines with N . This implies that in 2-D the optimum distance converges in probability to the expected value.

It is also possible to show that the asymptotic 1-D formulae depend on region shape, even for a constant region size. If one considers a region consisting of two intervals of equal length, $L/2$, separated by a greater distance, D , with the depot in the middle, then $\langle p_{UD}^* \rangle$ depends on D asymptotically. To see that this is true, note that in the optimal solution the total distance traveled by the depot flow increases linearly with $DN^{1/2}$, since this flow is proportional to $N^{1/2}$. This quantity cannot be neglected because it is shown in Section 3 that the internal distances within each sub-zone are $O(N^{1/2})$ when region size is held constant. Fortunately for practical applications, shape does not have an asymptotic effect in 2-D. Thus, the 2-D TLP is quite similar to the TSP in that the optimum distance per point is shape-independent and appears to converge in probability to the asymptotic mean. Although the 2-D TLP distance per point is unbounded when one holds density constant (unlike in the case of the TSP), said distance increases with N so slowly that it may be treated as a constant for problems where N only varies by a factor of 10.

5. Acknowledgements

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

6. References

- Daganzo, C.F. (1984a) "The length of tours in zones of different shapes", *Trans. Res. B* 18B, 135-146.
Daganzo, C.F. (1984b) "The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application", *Transportation Science* 18 (4), 331-350.

- Daganzo, C.F. and Smilowitz, K, (2000) "Asymptotic approximations for the transportation LP and other scalable network problems" ITS Working Paper, University of California, Berkeley;
www.ce.berkeley.edu/Programs/Transportation/Daganzo/tlp.pdf; revised as:
- Daganzo, C.F. and Smilowitz, K, (2002) "Bounds and approximations for the Transportation LP" (under revision, *Transportation Science*.)
www.ce.berkeley.edu/Programs/Transportation/Daganzo/TLP_revised.pdf.
- Eilon, S., Watson-Gandy, C.D.T. and Christofides, N. (1971) *Distribution Management: Mathematical Modelling and Practical Analysis*, Hafner, NY, NY.
- Feller, W. (1971) *An Introduction to Probability Theory and Applications, Vol.2* Wiley, NY,NY.
- Haimovich, M. A., A. H. G. Rinnooy-Kan and L. Stougie (1985) "Bounds and heuristics for capacitated routing problems", *Math. Opns. Res.* 10, 527-542.
- Karp, R.M. (1977) "Probabilistic analysis of partitioning algorithms for the traveling salesman problem", *Math. Opns. Res.* 2, 209-224.
- Newell, G.F. and Daganzo, C.F. (1986) "Design of multiple vehicle delivery tours--I: A ring-radial network", *Trans. Res. B* 20B(5), 345-364.
- Newell, G.F. and Daganzo, C.F. (1986a) "Design of multiple vehicle delivery tours--II: Other metrics", *Trans. Res. B* 20B(5), 365-376.