Notes on the Travelling Salesman Problem with forced edges

Giovanni Mottola
2025-03-24

1) Definitions and counterexamples

We have N segments in a plane; for each segment, there are two waypoints W_i and W_j at its extremes. The goal is to find the shortest path that passes over all waypoints and over each segment. This is a special case of the *Travelling Salesman Problem* (TSP): here, the connection graph is *complete*, meaning that the path can go from any point to any other point. Some graph edges (W_i, W_j) are *forced*, one for each segment; S is the set of such edges. We assume that no two W_i's coincide and that no two segments cross; the path starts and ends at a base point I, distinct from the W_i's. Every W_i is connected to one and only one edge in S.

For simplicity, we assume that all W_i's can be visited in one path; in ongoing developments, we want to expand the results presented here to the case where the total path cost cannot be higher than a given value T_{max}. Before reaching T_{max}, the path must go back to I.

For each pair of points, the Euclidean distance is $d_{ij} = \|W_i - W_j\|$; we then define a matrix of distances $\mathbf{D} = (d_{i,j})$. We now introduce *modified* distances $\tilde{d}_{i,j}$, such that, when the TSP is solved using the $\tilde{d}_{i,j}$ as weights in the graph, the optimal path passes through the edges in S. Seven possible methods (denoted as A, B, ..., G) of defining the $\tilde{d}_{i,j}$'s are detailed in Tab. 1.

One could naively expect that the simplest method A, where the forced edges have zero length (and thus zero cost), guarantees that all of them are part of the optimal path, but this is not the case, and infinitely many counterexamples exist; experimentally, it is found that in most cases with realistic designs the shortest path skips some segments. Similarly, method B (where the non-forced edges are penalized by adding to their cost the maximum point-to-point distance Δ) can fail, although counterexamples are harder to find; in all experiments on realistic designs, method B actually converges to

(W_i, W_j)	A	B	C	D	E	F	G
$\in S$	0	$d_{i,j}$	0	$d_{i,j} - \Delta$	δ	0	0
$\notin S$	$d_{i,j}$	$d_{i,j} + \Delta$	$d_{i,j} + \Delta$	$d_{i,j} + \Delta$	$d_{i,j} + \Delta$	$d_{i,j} + \Delta - \delta$	$d_{i,j} + d_{f_i,i} + d_{f_j,j}$

Table 1: modified distances $\tilde{d}_{ij}$, depending on whether an edge is forced (first row) or not, with the methods we considered (A to G); here, $\Delta = \max_{i,j}(d_{ij})$ and $\delta = \min_{i,j}(d_{ij})$. For method G, if $(W_i, W_j) \notin S$, let W_{f_i} be the waypoint such that $(W_i, W_{f_i}) \in S$ (see method G); similarly, W_{f_j} is such that $(W_j, W_{f_j}) \in S$. By the assumptions in Sec. 1, W_{f_i} and W_{f_j} exist and are unique; also, $(I, W_j) \notin S$ for all waypoints, in which case we set $W_{f_I} = I$ and thus $\tilde{d}_{I,j} = d_{I,j} + d_{f_j,j}$.

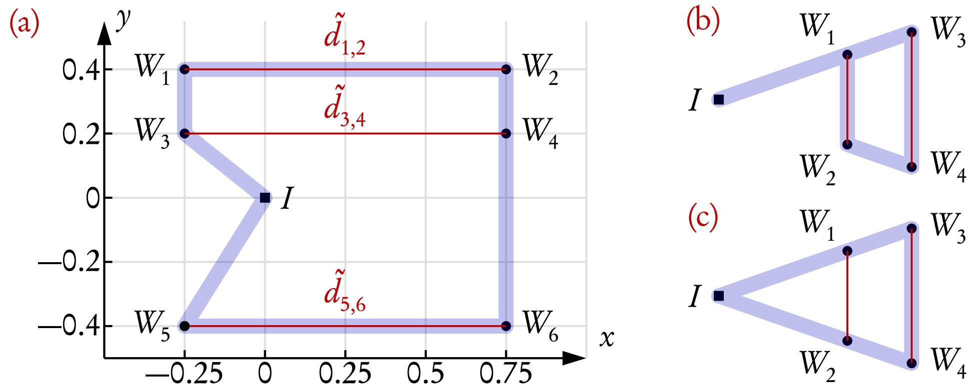

Figure 1: (a): a case where methods A and B fail. The shortest path is in blue, while the segments are in red. The $W_i's$ are marked by black circles, while $I = (0, 0)$ is marked by a square. The optimal path does not pass through (W_3, W_4). (b) and (c): with method A, paths I-1-2-4-3-I and I-1-3-4-2-I have the same cost, as $d_{2,4} + d_{3,I} = d_{1,3} + d_{4,2} + d_{2,I}$, so an optimizer may converge to the latter.

a solution that respects the constraints. A counterexample valid for both methods A and B is reported in Fig. 1.

We are interested in defining methods that always lead to solutions that respect the constraints. Indeed, we can prove that methods C to G always work. We prove this by contradiction, assuming that the optimal path (Fig. 2) misses at least one required segment (W_i, W_j), and then showing that a different path can be found with lower total cost T.

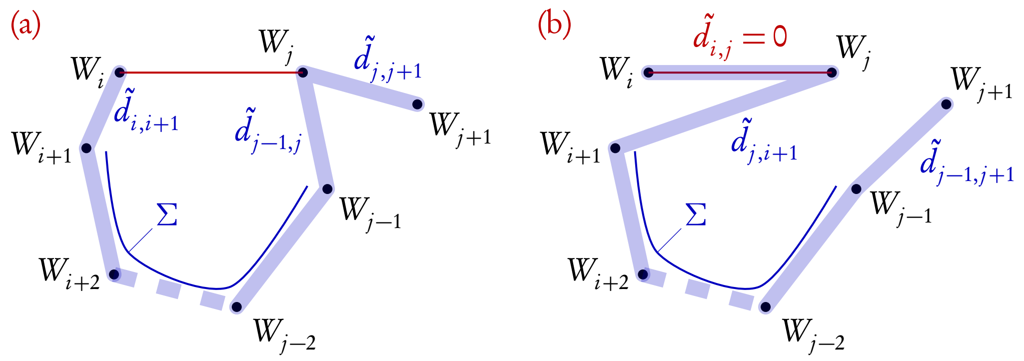

Figure 2: a subset of the W_i's and two possible optimal paths. ■ is the cost of the path (represented only in part, as a dashed line) from W_{i+1} to W_{j-1}, given by the sum of the $\tilde{d}_{i,j}$'s; ■ is the same both in (a) and (b). Since $(W_i, W_j) \in S$, then (W_i, W_{i+1}), (W_{j-1}, W_j), (W_j, W_{i+1}) and (W_j, W_{j+1}) are not in S, as no two segments share a point; (W_{j-1}, W_{j+1}) may or may not be in S.

2) Proofs for methods C to G

Consider the part of the path that starts at W_i, ends at W_{j+1} (the point visited immediately after W_j), and does not pass by I between W_i and W_j (remembering that the total path is closed); thus, either $W_{j+1} \equiv I$ or I is not in this part of the path. We show that the alternative sub-path in Fig. 2(b), which starts and ends at the same points as the one in Fig. 2(a) and visits the same points, but passes through the segment, has lower cost (for all distance modification methods C to G); thus, the first path cannot be optimal, contradicting our start assumption. Consider first method C. The total cost T of the paths in the two cases (a) and (b) is respectively

$$T_a = \tilde{d}_{i,i+1} + \blacksquare + \tilde{d}_{j-1,j} + \tilde{d}_{j,j+1} = (d_{i,i+1} + \Delta) + \blacksquare + (d_{j-1,j} + \Delta) + (d_{j,j+1} + \Delta) \quad (1)$$

and

$$T_b = 0 + \tilde{d}_{j,i+1} + \blacksquare + \tilde{d}_{j-1,j+1} = (d_{j,i+1} + \Delta) + \blacksquare + (d_{j-1,j+1} + \Delta) \quad (2)$$

The first term $\tilde{d}_{i,j}$ in T_b is zero, since the first segment (W_i, W_j) is forced.

Comparing Eqs. (1) and (2) and deleting common terms, we find $T_a > T_b$ if and only if

$$d_{i,i+1} + d_{j-1,j} + d_{j,j+1} + \Delta > d_{j-1,j+1} + d_{j,i+1} \quad (3)$$

which can be obtained by adding elementwise the following inequalities:

$$d_{i,i+1} > 0 \quad (4.1)$$

$$d_{j-1,j} + d_{j,j+1} \geq d_{j-1,j+1} \quad (4.2)$$

$$\Delta \geq d_{j,i+1} \quad (4.3)$$

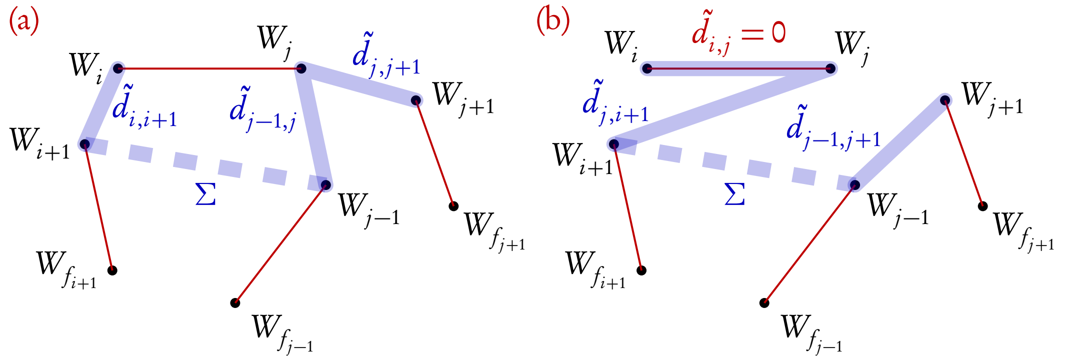

Figure 3: schematic for method G. Here, points $W_{f_{i+1}}$, $W_{f_{j-1}}$ and $W_{f_{j+1}}$ may or may not be in the subpath.

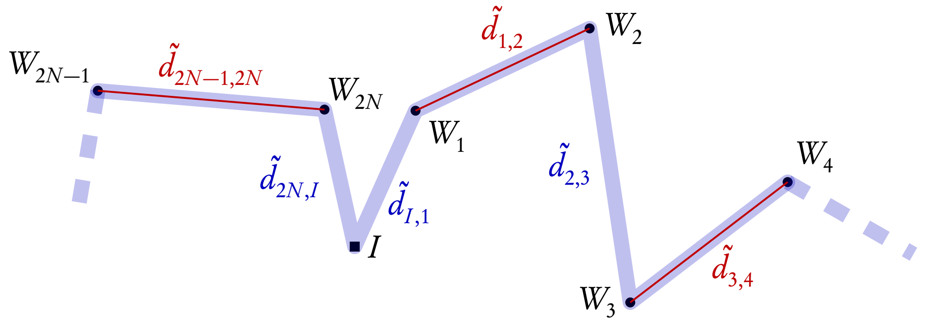

Figure 4: start and end of the optimal path. We renumber the W_i's in the order in which they appear along the path, thus after W_{2N} (where N is the number of segments) the path goes back to I.

Here, (4.1) holds since no two W_i's coincide, (4.2) is the triangle inequality, and (4.3) is from the definition of Δ. Edge (W_{j-1}, W_{j+1}) may be forced: if so, the last term in Eq. (2) is $\tilde{d}_{j-1,j+1} = 0$, so T_b is even lower. The proof for method D is identical, since the $\tilde{d}_{i,j}$'s are the same for the non-forced edges, and are even lower (as $d_{i,j} - \Delta \leq 0$) for the forced one.

We can also replace (4.1) with $d_{i,i+1} \geq \delta$, with δ being the smallest point-to-point distance. Adding together this with the other inequalities in (4), we find that $T_a \geq T_b$ for both methods E and F, so in the worst-case scenario there may be multiple paths (at least one of which passing through all segments) having the same minimum cost T; notice that the inequalities (4.2) and (4.3) are tight, so that the equal sign may actually hold. To guarantee that the chosen TSP solver converges to a solution that respects all constraints, we can simply set δ as slightly smaller than the minimum distance between the W_i's, so $d_{i,i+1} > \delta$.

To show that method G always works, refer to Fig. 3; the total cost for path (a) is

$$T_a = (d_{i,i+1} + d_{i,j} + d_{i+1,f_{i+1}}) + \blacksquare + (d_{j-1,j} + d_{i,j} + d_{j-1,f_{j-1}}) + (d_{j,j+1} + d_{i,j} + d_{j+1,f_{j+1}}) \quad (5.1)$$

$$T_a = (d_{i,i+1} + d_{i,j} + d_{i+1,f_{i+1}}) + \blacksquare + (d_{j-1,j} + d_{i,j} + d_{j-1,f_{j-1}}) + (d_{j,j+1} + d_{i,j}) \quad (5.2)$$

For path (b) we have instead

$$T_b = 0 + (d_{j,i+1} + d_{i,j} + d_{i+1,f_{i+1}}) + \blacksquare + (d_{j-1,j+1} + d_{j-1,f_{j-1}} + d_{j+1,f_{j+1}}) \quad (6.1)$$

$$T_b = 0 + (d_{j,i+1} + d_{i,j} + d_{i+1,f_{i+1}}) + \blacksquare + (d_{j-1,j+1} + d_{j-1,f_{j-1}}) \quad (6.2)$$

Here, Eqs. (5.1) and (6.1) hold if $W_{j+1} \not\equiv I$ (thus I is not in the subpath), and Eqs. (5.2) and (6.2) otherwise; Eqs. (5) are derived from Eq. (1) and Eqs. (6) from Eq. (2). If $(W_{j-1}, W_{j+1}) \in S$, the last term $\tilde{d}_{j-1,j+1}$ in Eqs. (6) is zero, so again we ignore this case. Then, $T_a > T_b$ if and only if

$$d_{i,i+1} + d_{j-1,j} + d_{i,j} + d_{j,j+1} + d_{i,j} > d_{j-1,j+1} + d_{j,i+1} \quad (7)$$

which is the sum of inequality (4.2) with the following:

$$d_{i,j} > 0 \quad (8.1)$$

$$d_{i,i+1} + d_{i,j} \geq d_{j,i+1} \quad (8.2)$$

Here, (8.1) is trivial as (4.1), while (8.2) is again due to the triangle inequality. Notice that method G can be combined with method E, by setting the cost of segments in S to δ instead of zero; alternatively, it can be combined with method F, by subtracting δ to the cost of non-forced edges. The proof of the validity of these methods is identical to the one above.

3) Equivalence of the TSP solution and generalizations

We also show that, if a path minimizes the sum of the $\tilde{d}_{i,j}$'s, then it also minimizes the total distance T_R, which is the sum of the $d_{i,j}$'s. For methods C to G, we have respectively (Fig. 4)

$$T_R = d_{I,1} + d_{1,2} + d_{2,3} + d_{3,4} + \ldots + d_{2N,I} \quad (9.1)$$

$$T_C = d_{I,1} + \Delta + 0 + d_{2,3} + \Delta + 0 + \ldots + d_{2N,I} + \Delta \quad (9.2)$$

$$T_D = d_{I,1} + \Delta + d_{1,2} - \Delta + d_{2,3} + \Delta + d_{3,4} - \Delta + \ldots + d_{2N,I} + \Delta \quad (9.3)$$

$$T_E = d_{I,1} + \Delta + \delta + d_{2,3} + \Delta + \delta + \ldots + d_{2N,I} + \Delta \quad (9.4)$$

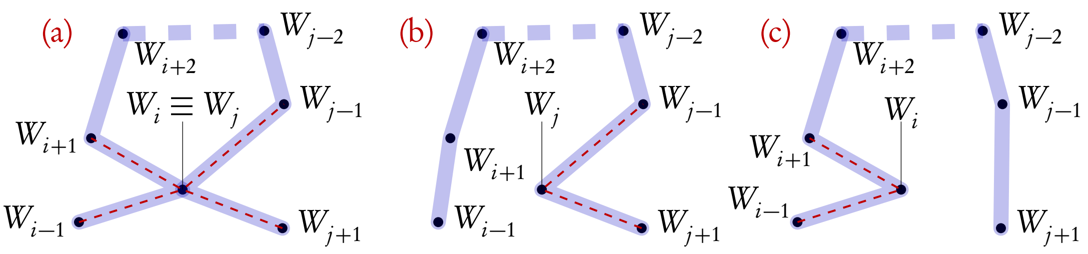

Figure 5: (a): a path that visits $W_i \equiv W_j$ twice. At most one of edges (W_i, W_{i-1}), (W_i, W_{i+1}), (W_i, W_{j-1}), and (W_i, W_{j+1}) is in S. (b) and (c): two shorter paths, with the same start and end points.

$$T_F = d_{I,1} + \Delta - \delta + 0 \qquad + d_{2,3} + \Delta - \delta \qquad + 0 \qquad + \ldots + d_{2N,I} + \Delta - \delta \tag{9.5}$$

$$T_G = d_{I,1} + d_{1,2} \quad + 0 \qquad + d_{1,2} + d_{2,3} + d_{3,4} + 0 \qquad + \ldots + d_{2N,I} + d_{2N-1,2N} \tag{9.6}$$

where each expression contains $2N + 1$ terms. Define $L_S = d_{1,2} + d_{3,4} + \ldots$ as the total length of the N segments in S: comparing Eq. (9.1) with Eqs. (9.2–9.6), we find that

$$\begin{aligned} T_C &= T_R + (N+1)\Delta - L_S & T_D &= T_R - \Delta \\ T_E &= T_R + (N+1)\Delta + \delta N - L_S & T_F &= T_R + (N+1)\Delta - \delta(N+1) - L_S \\ & & T_G = T_R + L_S & \end{aligned} \tag{10}$$

For a given case, Δ, δ, L_S and N are fixed, so the differences in T with the $\tilde{d}_{i,j}$'s and with the $d_{i,j}$'s are constant, and the path minimizing T_R also minimizes T_C and so on (and viceversa).

Finally, one may ask whether the shortest path through all edges in S actually visits each point once; without this constraint, we have the *Rural Postman Problem* (RPP). The path still needs to go back to I, so any point is connected to an even number M of edges ($M = 2$ for the TSP). If a point W_i is visited two or more times (Fig. 5a), then shorter paths exist that visit the same edges in S, start and end at W_{i-1} and W_{j+1}, but pass by W_i once: see Fig. 5(b), if an edge in S is visited after W_{j-1}, and Fig. 5(c) otherwise. If $W_i = I$, both (b) and (c) are valid. Here, the edge costs are the $d_{i,j}$, so the triangle inequality holds for (W_i, W_{i-1}, W_{i+1}) and (W_j, W_{j-1}, W_{j+1}). Thus, generalizing to the RPP would not in fact give a shorter path.