Constructing a Small Compact Binary Model for the Travelling Salesman Problem

J. Fabian Meier [a]

[a]*Institute of Transport Logistics, TU Dortmund, Leonhard-Euler-Str. 2, 44227 Dortmund, Germany*

Abstract

A variety of formulations for the Travelling Salesman Problem as Mixed Integer Program have been proposed. They contain either non-binary variables or the number of constraints and variables is large. We want to give a new formulation that consists solely of binary variables; the number of variables is of order $O(n^2 \ln(n)^2)$ and the number of constraints is of order $O(n^2 \ln(n))$.

Keywords: Integer Program, Landau function, Traveling Salesman Problem

1. Introduction

The Travelling Salesman Problem is *the* classical discrete optimization problem. The main difficulty in formulating the Travelling Salesman Problem as (Mixed) Integer Program is the avoidance of subtours. The most prominent
5 approaches to avoid subtours are the subtour elimination constraints of Dantzig [DFJ54] (which are exponentially many) and the Miller-Tucker-Zemlin formulation [MTZ60] which has few variables, but weak constraints with real-valued variables. The first model is numerically superior. In the following section we will define a small model formulation and use the subsequent sections to give
10 an estimate of its size.

2. The Model Formulation

As usual, we consider as set V of n given points with distances c_{ij}, $i, j \in V$. For the construction of the model, we need a virtual split of one of the vertices (we call it 1) into a source and a sink, which we call $\underline{1}$ and $\overline{1}$. This is necessary
15 because the following constraints for subtour elimination forbid any kind of circle. To avoid excessive writing, we define $\underline{V}$ to be all vertices except $\overline{1}$ and vice versa for $\overline{V}$. Let E be the resulting set of arcs, i.e. all pairs (i, j) with $i \in \underline{V}$, $j \in \overline{V}$, $i \neq j$ except for $(\underline{1}, \overline{1})$.

---

*Corresponding author
Email address:* `meier@itl.tu-dortmund.de` (J. Fabian Meier)

Following the usual model definition, we define the binary variables x_{ij} to be one if the arc $i \to j$ is used and zero if not. We use the standard objective function and the two well-known constraints:

$$\text{Min} \sum_{(i,j) \in E} c_{ij} x_{ij}$$

$$\sum_{i:(i,j) \in E} x_{ij} = 1 \qquad\qquad \forall j \in \overline{V} \qquad\qquad (1)$$

$$\sum_{j:(i,j) \in E} x_{ij} = 1 \qquad\qquad \forall i \in \underline{V} \qquad\qquad (2)$$

$$x_{ij} \in \{0,1\} \qquad\qquad \forall (i,j) \in E \qquad\qquad (3)$$

To construct our model formulation, we use some simple number theory. Let $p_1, \ldots, p_\alpha$ be pairwise relative prime numbers with $\prod_k p_k \geq n$. Then we define, for every $(i,j) \in E$, the following variables:

$$x_{ij}^{1,0}, x_{ij}^{1,1}, \ldots, x_{ij}^{1,p_1-1}$$

$$x_{ij}^{2,0}, x_{ij}^{2,1}, \ldots, x_{ij}^{2,p_2-1}$$

$$\vdots$$

$$x_{ij}^{\alpha,0}, x_{ij}^{\alpha,1}, \ldots, x_{ij}^{\alpha,p_\alpha-1}$$

Because we have $|E| = n(n-1)$, we get $n(n-1)\big(\sum_k p_k + 1\big)$ variables in total. The additional constraints are the following:

$$\sum_a x_{ij}^{k,a} = x_{ij} \qquad\qquad \forall (i,j) \in E, k \leq \alpha \quad (4)$$

$$\sum_{i:(i,j) \in E} x_{ij}^{k,a} = \sum_{i:(j,i) \in E} x_{ji}^{k,(a+1 \mod p_k)} \quad \forall j \in \underline{V} \cap \overline{V}, k \leq \alpha, a \leq p_k - 1 \quad (5)$$

The number of constraints is even smaller than the number of variables, in total $n(n-1)\alpha + (n-1)\sum_k p_k$.

Why does this formulation work? Assume a solution forms a circle. Let us consider the resulting $x_{ij}^{k,a}$ variables for a fixed value k and a given circle $i_0, i_1, \ldots, i_m, i_0$. Constraint (4) implies that for every i_β, $i_{\beta+1}$ and k, there is a unique a with $x_{i_\beta i_{\beta+1}}^{k,a} = 1$. Assume that for $\beta = 0$ we have $a = 0$, so that $x_{i_0 i_1}^{k,0} = 1$. Constraint (5) now implies that $x_{i_1 i_2}^{k,1} = 1$. This implication can be furthered until we come back to the beginning. In total, this shows that a circle can only exist if its length is divisible by p_k.

As we have shown this for every k, and the values p_k are relative prime, we know that a circle has at least the length $p_1 p_2 \cdots p_k \geq n$ which is impossible.

2

3. Estimating the Size of the Model

In the last section we constructed a model for which the number of variables is bounded by $n(n-1)\left(\sum_k p_k + 1\right)$ for values p_k with $\prod_k p_k \geq n$. Of course, one would pick the p_k so that the sum $\sum_k p_k$ is minimized. But how small is this sum, compared to n? For that, we make the following definition

$$P_n = \left\{ P \subset \mathbb{N} \,\middle|\, \mathrm{lcm}_{p \in P}(p) \geq n \right\},$$

where lcm denotes the least common multiple. Then we define

$$f(n) = \min_{P \in P_n} \sum_{p \in P} p \tag{6}$$

Obviously, the optimal $P \in P_n$ consists of pairwise relative prime numbers. Furthermore, these numbers are prime powers because $k \cdot m \geq k + m$ for every $k, m \geq 2$. For practical purposes, it is easy to give an estimate by explicitly giving values for p_k.

n	UB $f(n)$
10	$3 + 4$
100	$9 + 11$
1000	$9 + 11 + 13$
10000	$9 + 11 + 13 + 17$
100000	$7 + 9 + 11 + 13 + 17$

We see that the size of $f(n)$ is rather small and grows roughly on a logarithmic scale. Can we give a general estimate for high values of n?

For that we consider the *Landau function* $g(m)$. It is defined for every natural number m to be the largest order of an element of the symmetric group S_m. Equivalently, $g(m)$ is the largest least common multiple of any partition of m. Let us prove that

$$f(g(m)) \leq m.$$

For that let $m = p_1 + \ldots + p_\alpha$ be a partition of m with largest least common multiple. Then $\{p_1, \ldots, p_\alpha\}$ forms a valid set for $f(n)$ with $n = g(m)$. As $f(x)$ and $g(x)$ are monotone functions, the inequality holds.

Our aim is to use this relation to construct an upper bound for $f(n)$ for large n. For that we exploit some properties of $g(m)$. From [Nic97] we know that

$$\lim_{n \to \infty} \frac{g(m+1)}{g(m)} = 1$$

which implies that for every $\hat{m}$ there is a constant $C_{\hat{m}}$ with

$$g(m+1) \leq C_{\hat{m}} g(m) \qquad\qquad \forall m \geq \hat{m} \tag{7}$$

Furthermore, [MNR89] shows that

$$\ln g(m) \geq \sqrt{m \ln m} \qquad\qquad \forall m \geq 906 \qquad\qquad (8)$$

The estimate (8) clearly implies

$$\big(\ln g(m)\big)^2 \geq m \qquad\qquad \forall m \geq 906 \qquad\qquad (9)$$

Now, for a given n large enough, choose m with

$$g(m) \leq n \leq g(m+1)$$

Note that $g(m)$ is clearly unbounded, see e.g. (8). Then we have, using the monotonicity of f, g and ln:

$$
\begin{aligned}
f(n) &\leq f(g(m+1)) \\
&\leq m+1 \\
&\leq \big(\ln g(m+1)\big)^2 \\
&\leq \big(\ln C_m g(m)\big)^2 \\
&\leq \big(\ln C_m n\big)^2 \\
&= \big(\ln C_m + \ln n\big)^2
\end{aligned}
$$

This shows that $f(n)$ is of size $O(\ln(n)^2)$, so that the total number of variables can be estimated by $O(n^2 \ln(n)^2)$. Note furthermore that

https://oeis.org/A000793/b000793.txt

lists the first 10000 values of $g(m)$. From this is would be possible to derive good upper bounds for $f(n)$ for every reasonable value of n (considering that $g(10000)$ is 837248314781153836222627527908638519798608214039954646068895157070382622912184532863125113007409531931972405106926383184266365066811182400).

For the number of constraints, we see that the term $n(n-1)\alpha$ clearly dominates the other terms. Assuming that $P \in P_n$ is chosen to minimize $\sum_{p \in P} p$ as before, what can be said about the size of $\alpha = |P|$ in relation to n? For that, we establish a small lemma:

Lemma 1. *If $P \in P_n$ is chosen to minimize $\sum_{p \in P} p$, then $\prod_{p \in P} p < 2n$.*

Proof. As noted before, all $p \in P$ are prime powers. Let q_0 be the smallest involved prime number and $p_1 = q_0^t$. Because $\hat{p}_1 = (q_0 - 1)q_0^{t-1}$ is also relative prime to $p_2, \ldots, p_\alpha$ or equal to one, we know that $\hat{p}_1 p_2 \ldots p_\alpha < n$ (otherwise we could choose a better P, contradicting the optimality). Now we have

$$
\begin{aligned}
p_1 p_2 \ldots p_\alpha &= \frac{q_0}{q_0 - 1}\hat{p}_1 p_2 \ldots p_\alpha \\
&< \frac{q_0}{q_0 - 1}n \leq 2n
\end{aligned}
$$

$\square$

4

Now we know:

$$2^{\alpha} \leq p_1 \ldots p_{\alpha} < 2n$$

implying

$$\alpha < C \ln(n)$$

for some constant C. In total, we see that the number of constraints is $O(n^2 \ln(n))$.

4. Conclusion

This paper explains a new modelling approach for the TSP. It consists of $O(n^2 \ln(n)^2)$ binary variables and $O(n^2 \ln(n))$ constraints. In its pure form, it will probably not useful for solving the problem numerically, but it may form the starting point for a different view on subtour elimination constraints.

References

[DFJ54] George Dantzig, Ray Fulkerson, and Selmer Johnson. Solution of a large-scale traveling-salesman problem. *Journal of the operations research society of America*, 2(4):393–410, 1954.

[MNR89] Jean-Pierre Massias, Jean-Louis Nicolas, and Guy Robin. Effective bounds for the maximal order of an element in the symmetric group. *Mathematics of computation*, pages 665–678, 1989.

[MTZ60] Clair E Miller, Albert W Tucker, and Richard A Zemlin. Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, 7(4):326–329, 1960.

[Nic97] Jean-Louis Nicolas. On Landau's function $g(n)$. In *The Mathematics of Paul Erdös I*, pages 228–240. Springer, 1997.

[Sch96] E Robert Schulman. How to write a scientific paper. *Annals of Improbable Research*, 2(5):8, 1996.