ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Exact algorithms and heuristics for the Quadratic Traveling Salesman Problem with an application in bioinformatics☆

A. Fischer[a], F. Fischer[b], G. Jäger[c,*], J. Keilwagen[d,e], P. Molitor[f], I. Grosse[e,f,g]

[a] *Department of Mathematics, TU Dortmund, D-44227 Dortmund, Germany*

[b] *Department of Mathematics, Chemnitz University of Technology, D-09107 Chemnitz, Germany*

[c] *Department of Mathematics and Mathematical Statistics, University of Umeå, S-90187 Umeå, Sweden*

[d] *Institute for Biosafety in Plant Biotechnology, Julius-Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, D-06484 Quedlinburg, Germany*

[e] *Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Seeland OT Gatersleben, Germany*

[f] *Institute of Computer Science and Universitätszentrum Informatik, Martin Luther University Halle-Wittenberg, D-06120 Halle, Germany*

[g] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, D-04103 Leipzig, Germany*

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form 1 July 2013
Accepted 24 September 2013
Available online xxxx

Keywords:
Traveling Salesman Problem
Branch-and-bound
Branch-and-cut
Heuristical methods
Exact methods
Bioinformatics

ABSTRACT

In this paper we introduce an extension of the Traveling Salesman Problem (TSP), which is motivated by an important application in bioinformatics. In contrast to the TSP the costs do not only depend on each pair of two nodes traversed in succession in a cycle but on each triple of nodes traversed in succession. This problem can be formulated as optimizing a quadratic objective function over the traveling salesman polytope, so we call the combinatorial optimization problem *quadratic TSP* (QTSP). Besides its application in bioinformatics, the QTSP is a generalization of the Angular-Metric TSP and the TSP with reload costs. Apart from the TSP with quadratic cost structure we also consider the related Cycle Cover Problem with quadratic objective function (QCCP). In this work we present three exact solution approaches and several heuristics for the QTSP. The first exact approach is based on a polynomial transformation to a TSP, which is then solved by standard software. The second one is a branch-and-bound algorithm that relies on combinatorial bounds. The best exact algorithm is a branch-and-cut approach based on an integer programming formulation with problem-specific cutting planes. All heuristical approaches are extensions of classic heuristics for the TSP. Finally, we compare all algorithms on real-world instances from bioinformatics and on randomly generated instances. In these tests, the branch-and-cut approach turned out to be superior for solving the real-world instances from bioinformatics. Instances with up to 100 nodes could be solved to optimality in about ten minutes.

1. Introduction

Given a weighted graph the *Traveling Salesman Problem* (TSP) is the problem of finding a tour with minimal costs where the costs are associated to each pair of nodes that are traversed in succession. The TSP is well-known to be an $\mathcal{NP}$-hard

☆ A preliminary version of this paper appeared in the proceedings of the Second International Conference on Combinatorial Optimization and Applications (COCOA) 2008, Lecture Notes in Comput. Sci. 5165, pp. 211–224.

* Corresponding author. Tel.: +46 90 786 6141; fax: +46 90 786 5222.
E-mail address: gerold.jaeger@math.umu.se (G. Jäger).

0166-218X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dam.2013.09.011

problem. In this paper we consider an extension of the TSP. The *Quadratic Traveling Salesman Problem* (QTSP) is the problem of finding a cost minimal tour where the costs are associated to each *triple* of nodes that is traversed in succession. Because a path of three nodes is contained in a tour if and only if the two corresponding arcs are present, the QTSP can be modeled as optimizing a quadratic objective function over the traveling salesman polytope [8].

The QTSP is inspired by the problem of finding the optimal Permuted Markov (PM) model [10] or the optimal Permuted Variable Length Markov (PVLM) model [40] for a given set of DNA sequences. This problem is important in bioinformatics and genome research for the recognition of *transcription factor binding sites* and *splice sites*. Given a set of short DNA sequences of equal length n such as a set of *splice sites*, the task is to learn a pattern from this set that allows the recognition of further *splice sites* from a set of unknown sequences. Such patterns are typically learned by statistical models, and PM models and PVLM models are two of the most powerful models for the recognition of *splice sites*. Learning a pattern from data by a statistical model is often accomplished by finding the maximum likelihood estimates (MLEs) of the model parameters, and the problem of finding the MLEs of the parameters of PM models or PVLM models results in the QTSP. A more detailed description of the biological background and the connection to the QTSP is given in Section 2.

The TSP is closely related to the *Cycle Cover Problem* (CCP) that asks for a cost minimal set of disjoint cycles over all nodes. The CCP corresponds to a linear assignment problem, which can be solved by the Hungarian method in polynomial time [28]. However, we will show that its quadratic counterpart is $\mathcal{NP}$-hard. A special case of the QTSP is the *Angular-Metric TSP* [1], which for given n points in the Euclidean space aims to determine a tour with minimal total angle change. It has applications in robotics, *e.g.*, it allows to determine optimal robot paths w.r.t. energetic aspects. In particular, robots often tend to have a higher energy demand if the path bends sharply, so one is interested in tours without high curvature. An extension of Angle-TSP is the so called *TSP for Dubins vehicle* [38,33,32], where the task is to determine a shortest trajectory of a nonholonomic vehicle w.r.t. given curvature constraints. In their solution approach, the authors of [33] determine an optimal tour where the changes in direction are weighted against the length of the tour, which is a special case of QTSP, too. The TSP with reload costs is a further special case of QTSP. Here, given an arc-colored graph, the task is to determine a tour with minimal weighted sum of the color changes along the tour. Such cost structures arise, *e.g.*, in transport networks if the costs for loading processes are high in comparison to the transportation costs [2].

In this work, we present several exact and heuristical algorithms for the solution of the QTSP. We consider three exact algorithms. The first algorithm transforms the QTSP to an STSP to be solved by the state-of-the-art solver Concorde [7]. The second algorithm is a Branch-and-Bound algorithm based on combinatorial lower bounds. Here, we use a lower bound on the optimal value of the $\mathcal{NP}$-hard QCCP as a lower bound for QTSP. The third algorithm employs a Branch-and-Cut (BnC) algorithm based on a linear relaxation of a linearized integer programming formulation of the QTSP, which allows us to handle the well-known subtour elimination constraints [8]. Apart from these constraints we present further valid inequalities of the integer program that can be added during the BnC algorithm. Using the new cutting planes we could reduce the running times of the instances from bioinformatics significantly in comparison to BnC without the new cutting planes.

However, our experiments show that exact methods may lead to large running times or to a large number of nodes in the BnB/BnC-tree, especially for instances with costs taken uniformly at random from a given set. In particular the BnB algorithm benefits from good start solutions. This motivates the investigation of heuristical approaches for the QTSP. We present several heuristics that are extensions of classical heuristics for the TSP.

Finally, we experimentally compare the algorithms for several real-world instances from bioinformatics and for some randomly generated instances, partially motivated by other applications described above. Most heuristics lead to almost-optimal solutions for the real-world instances, and the Branch-and-Cut algorithm is the fastest exact algorithm for both random and real-world instances. This algorithm is capable of solving large *real-world* instances to proven optimality in reasonable time.

Note that some conclusions from our experiments are rather interesting not only for this specific problem, but possibly also for other combinatorial optimization problems. For example, in our experiments we present example instances, where local search algorithms benefit from relatively poor starting solutions, and other example instances, where the currently leading TSP solver Concorde behaves poorly.

The paper is organized as follows. In Section 2 we describe the motivating problem from bioinformatics that leads to QTSP. In Section 3 we formally introduce the QTSP, QCCP and related problems and study their computational complexity. Furthermore, we develop a polynomial reduction from QTSP to TSP, which is the basis for our first exact algorithm. Our exact and heuristical algorithms for the QTSP are presented in Sections 5 and 6, respectively. We compare the exact methods and the heuristics on several random and real-world instances in Section 7. Finally, we summarize this paper and give suggestions for future research in Section 8.

2. Motivation from Bioinformatics

Gene regulation in higher organisms is accomplished by several cellular processes, two of which are transcription and RNA splicing. In order to better understand these processes, it is desirable to have a good understanding of how transcription factors bind to their DNA binding sites and how the spliceosome binds to RNA splice sites. Many approaches for the computational recognition of transcription factor binding sites or splice sites rely on statistical models, and two of the most promising models for this task are Permuted Markov (PM) models [10] and Permuted Variable Length Markov (PVLM)

Please cite this article in press as: A. Fischer, et al., Exact algorithms and heuristics for the Quadratic Traveling Salesman Problem with an application in bioinformatics, Discrete Applied Mathematics (2013), http://dx.doi.org/10.1016/j.dam.2013.09.011