

CONTINUOUS REFORMULATIONS AND HEURISTICS FOR THE EUCLIDEAN TRAVELLING SALESPERSON PROBLEM

TUOMO VALKONEN¹ AND TOMMI KÄRKKÄINEN¹

Abstract. We consider continuous reformulations of the Euclidean travelling salesperson problem (TSP), based on certain clustering problem formulations. These reformulations allow us to apply a generalisation with perturbations of the Weiszfeld algorithm in an attempt to find local approximate solutions to the Euclidean TSP.

Mathematics Subject Classification. 90C26, 90C59, 90C27.

Received March 17, 2008.

Published online August 20, 2008.

1. INTRODUCTION

This paper is concerned with the travelling salesperson problem with Euclidean (ℓ^2) distances (undiscretised), *i.e.* the problem of finding the shortest closed path that visits every vertex (or city) in a given finite subset of \mathbb{R}^m exactly once, with the distances given by the Euclidean metric. Whereas various rather efficient algorithms exist for the general and general metric TSP [14], few seem to be able to take advantage of the special features of the variant with Euclidean distances – that still remains NP-hard. The most remarkable of those that do are Arora’s polynomial time (and even “nearly linear time”) approximation schemes (PTAS) [2,3], the good performance of which is, however, only asymptotic. Other methods for Euclidean instances specifically include various heuristics optimised for speed and based on clustering or partitioning of the plane, or spacefilling curves [14].

Here, we make another stab at formulating and finding (local) solutions to the Euclidean TSP. Our approach consists of first reformulating the problem as a continuous diff-convex problem. Instead of attempting to find the optimal path, we attempt to find points that construct the path, constrained to equal one of the input vertices. We then relax this problem, converting the constraint into a mere penalty. Dependent on the formulation of the constraint, the relaxed problem is found to be equivalent to certain clustering problems (including the multisource Weber problem or “ K -spatial medians”) perturbed with the path length penalty. (Perhaps not so coincidentally, Arora’s methods can also be extended to approximate the K -spatial medians [3,4].)

As a continuation of the work in [22,23], in this paper we restrict ourselves to locally solving these penalised reformulations, by applying the so-called “perturbed Weiszfeld method” applicable to finding “semi-critical” points of a sum of Euclidean distances from fixed points, perturbed by a concave function. Although applicable to the multisource Weber problem (providing a sort of dual of the K -means -style algorithm), it is unfortunately not applicable to the problem perturbed with the path length penalty. The algorithm is, however, applicable

Keywords and phrases. Euclidean TSP, clustering, diff-convex, Weiszfeld algorithm.

¹ Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland. tujumava@jyu.fi; tka@mit.jyu.fi

to another clustering formulation presented in [23], perturbed with the path length penalty. It is this latter reformulation we will use in our numerical experiments.

An (approximate) solution of such a continuous reformulation of the Euclidean TSP is not in practice – and not in theory either for big penalty parameters – a permutation of the original vertices. Therefore, along the course of studying these reformulations, we derive a heuristic that we use to “associate” the points of a solution with the original vertices. We also develop some other heuristics to reduce problem sizes, based on this heuristic and the clustering principle.

As for the applicability of our algorithms, we do not have any theoretical proofs of efficiency aside from partial convergence to “semi-critical points” (often local minima), and each step of the basic algorithm being $O(n^2)$ (consisting of n parallel Weiszfeld steps). On the experimental side, our method does seem to provide rather good results in quite few iterations for small problems. For bigger problems the performance however degrades considerably – there are, after all, many more local solutions then. A bigger penalty parameter value might help, but the algorithm we apply has a limit on its magnitude. Clustering heuristics that we develop, however, somewhat remedy the situation. Nevertheless, our numerical results are not remarkable compared to what is achievable with other (non-Euclidean) algorithms [14].

The primary contributions of this work are thus the reformulations that appear new, and perhaps with other methods applied to them, could provide better numerical results. The basic method based on the Weiszfeld algorithm is also new. Our clustering heuristics are related to the classic Karp clustering heuristic, Bentley’s Fast Recursive Partitioning scheme [8], and Litke’s clustering heuristic [15]. The first two of these use a “hard-coded” partitioning approach until the clusters are small enough, after which the sub-problems in the cluster are solved either approximately or exactly. Our approach, by contrast, uses a more dynamic cluster configuration, as defined by a clustering problem objective function. Litke’s method also uses an *ad hoc* dynamic clustering method. None of these methods incorporate TSP path length optimisation in the cluster calculation phase. Finally, our geometric penalisation approach bears some resemblance to various geometric neural net methods for the problem – see [13] and the references therein – as well as the Lazy TSP of [18]. In this latter paper a formulation very similar to the first one of ours, but with squared distances, is analysed along with its convexification. This problem is also considered in [9], in a wider measure-theoretic transport optimisation framework.

The rest of this paper is organised as follows: in Sections 2 and 3 we present our continuous reformulations. Then in Section 4 we consider the sensitivity of the solutions of the penalised reformulations with respect to the solutions of the original problem, as the penalty parameter is varied. Section 5 considers heuristic approaches that could be used to improve or speed up results. Finally in Section 6 we present and discuss the results of our numerical experiments, and conclude the paper in Section 7. Appendix 7 presents some auxiliary results.

2. FIRST REFORMULATION

Consider the Euclidean travelling salesperson problem:

$$\min_{\sigma} \sum_{i=1}^n \|a_{\sigma i} - a_{\sigma(i+1)}\|, \quad (2.1)$$

where $\bar{a} \triangleq (a_1, \dots, a_n) \in \mathbb{R}^{mn}$ are distinct vertices, also called cities, and σ is a permutation of the numbers $\{1, \dots, n\}$, with $\sigma(n+1) \triangleq \sigma 1$. We shall henceforth use this identification without explicit mention. We denote by $\hat{\sigma}$ any of the optimal permutations that minimise (2.1). There are always at least n of these, every “shift” of a solution being one.

Let us now reformulate the problem as finding $\bar{p} \triangleq (p_1, \dots, p_n)$ that solves

$$\min f_{\text{TSP}}(\bar{p}) \triangleq \sum_{i=1}^n \|p_i - p_{i+1}\| \quad \text{subject to } p_i = a_{\sigma i} \text{ for some permutation } \sigma.$$

Here again we identify $p_{n+1} \triangleq p_1$. The qualification condition may be written as

$$f_{\text{KM}}(\bar{p}; \bar{a}) \triangleq \sum_{i=1}^n \min_{j=1, \dots, n} \|a_i - p_j\| = 0.$$

The function f_{KM} is precisely the multisource Weber problem (or “ n -spatial medians”) objective function, when the number of data points and cluster prototypes are equal [23]. This function is diff-convex, as may be seen by rewriting $f_{\text{KM}}(\bar{p}; \bar{a}) = f(\bar{p}; \bar{a}) - \nu_{\text{KM}}(\bar{p}; \bar{a})$ with

$$f(\bar{p}; \bar{a}) \triangleq \sum_{i=1}^n \sum_{j=1}^n \|a_i - p_j\| \quad \text{and} \quad \nu_{\text{KM}}(\bar{p}; \bar{a}) \triangleq \sum_{i=1}^n \max_{j=1, \dots, n, k \neq j} (\sum \|a_i - p_k\|). \tag{2.2}$$

These considerations suggest relaxing problem (2.1) to the problem

$$\min_{\bar{p}} f_{\text{KM}}(\bar{p}; \bar{a}) + \lambda f_{\text{TSP}}(\bar{p}), \quad \lambda > 0, \tag{2.3}$$

or

$$\min_{\bar{p}} \left(\sum_{i=1}^n \min_{j=1, \dots, n} \|a_i - p_j\| + \lambda \sum_{i=1}^n \|p_i - p_{i+1}\| \right).$$

Notice that for permutations σ of the vertices, $\bar{p} = \bar{a}_\sigma \triangleq (a_{\sigma 1}, \dots, a_{\sigma n})$ are precisely all the global minimisers of (2.3) for $\lambda = 0$. The function f_{TSP} therefore acts as a perturbation to the multisource Weber problem, penalising such permutations that result in long paths. For small enough perturbation parameter λ , a minimiser \hat{p} of (2.3) actually equals $\bar{a}_{\hat{\sigma}}$ for one of the optimal permutations $\hat{\sigma}$, as Theorem 2.5 below shows. First we need some preliminary results and definitions, however.

Definition 2.1. The vertices a_k ($k = 1, \dots, n$) are *collinear* (on the line L) if there are vectors $z, v \in \mathbb{R}^m$ such that for the line $L \triangleq \mathbb{R}z + v$, $\{a_1, \dots, a_n\} \subset L$. Otherwise the points are *non-collinear*.

Definition 2.2. Given a path/permutation σ , there is said to be a *degenerate angle* at the point $a_{\sigma k}$, if $(a_{\sigma(k+1)} - a_{\sigma k})^T (a_{\sigma(k-1)} - a_{\sigma k}) = \|a_{\sigma(k+1)} - a_{\sigma k}\| \|a_{\sigma(k-1)} - a_{\sigma k}\|$.

Since the collinear case is trivial, we will only consider the case of

Assumption 2.3. The vertices $a_k \in \mathbb{R}^m$ ($k = 1, \dots, n$) are non-collinear and distinct.

The following result is well-known, but we provide the proof for completeness:

Lemma 2.4. *Suppose that Assumption 2.3 holds on the points $a_k \in \mathbb{R}^m$ ($k = 1, \dots, n$). Then the points of an optimal path $\bar{a}_{\hat{\sigma}}$ form a simple closed curve. In particular, there are no degenerate angles.*

Proof. Assume without loss of generality that $\hat{\sigma}$ is the identity permutation. Suppose two (open) straight line segments of the path (a_k, a_{k+1}) and (a_i, a_{i+1}) with $i \neq k$, cross at a point c . Then replacing the former segments with (a_k, a_i) and (a_{k+1}, a_{i+1}) , and reversing part of the remaining path, produces a valid path with one less crossing. Now

$$\begin{aligned} \|a_k - a_i\| + \|a_{k+1} - a_{i+1}\| &\leq \|a_k - c\| + \|a_i - c\| + \|a_{k+1} - c\| + \|a_{i+1} - c\| \\ &= \|a_k - a_{k+1}\| + \|a_i - a_{i+1}\|, \end{aligned}$$

with the inequality strict if c does not lie on one (and then both) of the segments (a_k, a_i) or (a_{k+1}, a_{i+1}) . Thus the path can in that case be improved by removing the crossing.

If $c \in (a_k, a_i) \cap (a_{k+1}, a_{i+1})$, then these points are collinear, and (a_k, a_{k+1}) or (a_i, a_{i+1}) contains an endpoint of the other; say $a_k \in [a_i, a_{i+1}]$, the other cases being analogous. The path can therefore visit a_k during

this segment, not increasing the cost. Furthermore, if this segment is part of the optimal path, the smaller problem with a_k removed will have equal optimal path length. If removing a_k does not improve the path length by going from a_{k-1} directly to a_{k+1} , it must be that a_{k-1}, a_{k+1}, a_i and a_{i+1} are collinear. Therefore, if recursively applying the argument never improves the path, all the points must be collinear. This is in contradiction to our assumptions. \square

We denote by $B(x, r)$ the closed ball centred at $x \in \mathbb{R}^m$ of radius r . Note that $\partial \|\cdot - a\| (a) = B(0, 1)$, yielding (by local convexity) that $\partial f_{\text{KM}}(\bar{a}_\sigma; \bar{a}) = \prod_{i=1}^n B(0, 1)$ when the points are distinct.

Theorem 2.5.

- (i) For $\lambda \in (0, 1/2]$, every global minimiser \bar{p} of (2.3), is a permutation of \bar{a} .
- (ii) For $\lambda \in (0, 1/2)$, global minimisers of (2.3), coincide with optimal TSP paths $\bar{a}_{\hat{\sigma}}$; the same holds for $\lambda = 1/2$ under Assumption 2.3.
- (iii) However, for every permutation σ , \bar{a}_σ is a strict local minimiser of (2.3) for $\lambda \in [0, 1/2)$ and a (possibly non-strict) local minimiser for $\lambda = 1/2$.

Proof. Let $\bar{p} = (p_1, \dots, p_n) \in \mathbb{R}^{mn}$ be arbitrary. Suppose that for some p_j ($j = 1, \dots, n$) the following property holds: for every a_k ($k = 1, \dots, n$) and some $i(k) \neq j$, $\|a_k - p_{i(k)}\| \leq \|a_k - p_j\|$. The point p_j then does not contribute to f_{KM} , and we may assume that it lies on the straight line segment from p_{j-1} to p_{j+1} , for otherwise the cost could be decreased by making this alteration. We may in fact freely move p_j on the path composed of the remaining points p_i ($i \neq j$). Therefore, we can arrange the points in such a way that whenever p_j minimises $i \mapsto \|a_k - p_i\|$ for N_j points a_k , then the multiplicity of p_i with $p_i = p_j$ is also N_j .

The (possibly collinear) case with $\lambda \in (0, 1/2)$. Let then p_j minimise $i \mapsto \|a_k - p_i\| > 0$. We may then alter \bar{p} by assigning $p_j \mapsto a_k$, actually decreasing the cost. This follows from the following two observations. Firstly, (a) by the previous alterations, if p_j is a minimiser of the distance for another $a_\ell \neq a_k$, then there is also another $p_i = p_j$ for which this holds. Therefore $\min_i \|a_\ell - p_i\|$ is not increased. Secondly, (b) for $\lambda \in (0, 1/2)$, we have

$$\lambda \|p_{j-1} - a_k\| < \lambda \|p_{j-1} - p_j\| + \frac{1}{2} \|p_j - a_k\| \tag{2.4}$$

and similarly for p_{j+1} . Thus the increase in the length of the path (p_1, \dots, p_n, p_1) is consumed by the decrease of $\min_j \|a_k - p_j\|$ to zero.

We have therefore showed that for $\lambda \in (0, 1/2)$, only the points \bar{a}_σ for permutations σ can be global minimisers. Obviously the actual global minimisers correspond to the permutations that minimise f_{TSP} .

However, $0 \in \text{int } \partial(f_{\text{KM}}(\cdot; \bar{a}) + \lambda f_{\text{TSP}})(\bar{a}_\sigma)$ because $\partial f_{\text{KM}}(\bar{a}_\sigma; \bar{a}) = \prod_{i=1}^n B(0, 1)$ as already noted, and

$$\nabla_{p_i} f_{\text{TSP}}(\bar{a}_\sigma) = \nabla_{p_i} (\|p_i - a_{\sigma(i-1)}\| + \|p_i - a_{\sigma(i+1)}\|)(a_{\sigma i}) = \frac{a_{\sigma i} - a_{\sigma(i+1)}}{\|a_{\sigma i} - a_{\sigma(i+1)}\|} + \frac{a_{\sigma i} - a_{\sigma(i-1)}}{\|a_{\sigma i} - a_{\sigma(i-1)}\|}$$

for $\lambda \in (0, 1/2)$. By the local convexity of f_{KM} in a neighbourhood of \bar{a}_σ , strict local optimality follows.

When $\lambda = 1/2$, we still have $0 \in \partial(f_{\text{KM}}(\cdot; \bar{a}) + \lambda f_{\text{TSP}})(\bar{a}_\sigma)$. Thus local optimality follows from local convexity. For an optimal permutation $\hat{\sigma}$, by Lemma 2.4 we must in fact have $\|\nabla_i f_{\text{TSP}}(\bar{a}_{\hat{\sigma}})\| < 2$, wherefore strict local optimality still holds. It remains to prove global optimality for this case.

The non-collinear case with $\lambda = 1/2$. Let again p_j minimise $i \mapsto \|a_k - p_i\| > 0$. The inequality (2.4) still holds as non-strict. In fact, when it holds as equality for both $j - 1$ and $j + 1$, all the points p_j, p_{j-1}, p_{j+1} and a_k must lie on a line L , such that in one of the natural orders \prec of L , $a_k \prec p_j, p_j \prec p_{j+1}$, and $p_j \prec p_{j-1}$. As before, we may then move p_j to $p'_j \triangleq a_k$, not increasing the cost. Since $\|a_k - p_j\| > 0$ was minimal, p'_j can equal neither p_{j-1} nor p_{j+1} . Therefore, there's a degenerate angle in the altered path at p'_j . Now, if some p_i is not on L , Lemma 2.4 applied to the points $p_1, \dots, p'_j, \dots, p_n$ (duplicates removed) shows that the path can not be optimal.

The possibility then remains that all the points p_i are on L . By the non-collinearity assumption, there's some a_k that is not on L . But now (2.4) holds strictly for the p_j minimising $i \mapsto \|a_k - p_i\| > 0$. Therefore the cost can be decreased as before. \square

Corollary 2.6. *Finding a point arbitrarily close to a minimiser of problem (2.3) is NP-hard for $\lambda \in (0, 1/2]$ when the vertices are non-collinear (with rational coordinates). Consequently, we have another proof that the family of diff-convex problems is NP-hard.*

Proof. We can always assume that $\|a_k - a_\ell\| \geq 1$ ($k \neq \ell$), because scaling does not alter $\hat{\sigma}$. Suppose then that for problem (2.3) and a given $\epsilon > 0$, we were able to find in time polynomial in n (but not in ϵ), a point \hat{p} with $\|\hat{p}_i - a_{\hat{\sigma}_i}\| < \epsilon$ ($i = 1, \dots, n$) for some $\hat{\sigma}$. Then, taking $\epsilon = 1/3$, we could uniquely assign each \hat{p}_i to $a_{\hat{\sigma}_i}$ in polynomial time. But this means we could solve the original NP-hard Euclidean TSP problem (2.1) in polynomial time¹. \square

For small enough λ , a good enough approximate solution should therefore identify the solution of (2.1), there being a unique distance-minimising assignment of each p_j to a_k . For parameters greater than the threshold value of λ , one could look for a permutation σ for which \bar{a}_σ closely matches \bar{p} , for example by following the method used in the proof of Theorem 2.5. Deciding how to optimally assign equal points p_j to the corresponding vertices in that method, can of course be expensive in itself.

The benefit from using a bigger λ comes from the local minima starting to disappear as the objective function becomes “more convex”, and therefore possibly easier to minimise. For very big λ , the global minimisers also drift far from the sought solution, however: the study of this sensitivity is the topic of Section 4.

By the diff-convexity, one could thus try to solve problem (2.1) by (approximately) solving a penalised version (2.3) by methods of global optimisation, such as outer approximation methods; see e.g. [12]. As stated, we are, however, interested in applying the somewhat more lightweight perturbed Weiszfeld method of [23] to the problem. Unfortunately, the present model does not exactly fit within the class of problems considered in [23], and for which we have partial convergence proofs. The problem is that $f_{\text{TSP}}(\bar{p}) - \nu_{\text{KM}}(\bar{p}; \bar{a})$ is not concave.

3. SECOND REFORMULATION

We are thus led to seek for another way to formulate the condition $p_i = a_{\sigma_i}$, that would fit within the above-mentioned class of problems. Given the observed relationship to the K -means clustering problem, a natural candidate is based on the multi-objective clustering problem formulated in [23]. The problem then becomes

$$\min_{\bar{p}} f_{\text{MO}}(\bar{p}; \bar{a}) + \lambda f_{\text{TSP}}(\bar{p}), \quad \lambda > 0, \tag{3.1}$$

where

$$f_{\text{MO}}(\bar{p}; \bar{a}) \triangleq f(\bar{p}; \bar{a}) - \nu_{\text{MO}}(\bar{p})$$

is in structure similar to f_{KM} : the function ν_{KM} has merely been replaced with

$$\nu(\bar{p}) \triangleq \nu_{\text{MO}}(\bar{p}) \triangleq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|p_i - p_j\|.$$

We have fixed the factor $1/2$ already at this point for simplicity; in [23], this may vary up to $n/(2s - 2)$, with $s = n$ in our case, while ensuring level-boundedness of the objective function.

This time, the function $f_{\text{TSP}}(\bar{p}) - \nu(\bar{p})$ is concave for $\lambda \in [0, 1]$, because $\nu(\bar{p})$ contains all the terms $\|p_i - p_{i+1}\|$ in the sum expression.

¹We may also consider approximate solutions in value space as well, since a good enough solution will have unique assignment to some $\hat{\sigma}$ in f_{KM} .

The permutations \bar{a}_σ are strict local minimisers of $f_{\text{MO}}(\cdot; \bar{a})$, as

$$\nabla \nu(\bar{a}_\sigma) \in \text{int } \partial f(\bar{a}_\sigma; \bar{a}) = \nabla \nu(\bar{a}_\sigma) + \prod_{i=1}^n \text{int } B(0, 1). \tag{3.2}$$

This follows from

$$\nabla_{p_i} \nu(\bar{a}_\sigma) = \sum_{j \neq i} \nabla_{p_i} \|p_i - a_{\sigma j}\| (a_{\sigma i}), \quad \partial_{p_i} f(\bar{a}_\sigma) = \sum_k \partial_{p_i} \|p_i - a_k\| (a_{\sigma i}),$$

with the difference $B(0, 1)$ coming from $\sigma i = k$.

As before, we have the inclusion $\partial f_{\text{TSP}}(\bar{a}_\sigma) \subset B(0, 1)$, strict at $\sigma = \hat{\sigma}$ under Assumption 2.3 by Lemma 2.4. Therefore, all the points \bar{a}_σ are strict local minimisers for $\lambda \in (0, 1/2)$, and $\bar{a}_{\hat{\sigma}}$ for $\lambda = 1/2$ as well. As for global optimality, we have:

Theorem 3.1. *Suppose the points a_i ($i = 1, \dots, n$) are distinct. Then,*

- (i) *The global minimisers of $f_{\text{MO}}(\cdot; \bar{a})$ are exactly \bar{a}_σ for all σ .*
- (ii) *There exists a $\hat{\lambda} > 0$, such that the minimisers of $f_{\text{MOTSP}}^\lambda \triangleq f_{\text{MO}}(\cdot; \bar{a}) + \lambda f_{\text{TSP}}$ are exactly the optimal TSP paths $\bar{a}_{\hat{\sigma}}$ for $\lambda \in (0, \hat{\lambda})$.*

We begin the proof with a few lemmas. For the case $m > 1$, we will use the following extension (to strict inequalities) of a reduction theorem of Levi; cf. [16], p. 175.

Lemma 3.2. *Let $k_i \in \mathbb{R}$, and $\rho_{ij} \in \mathbb{R}$, $i = 1, \dots, K$, $j = 1, \dots, N$. Suppose that for all $\bar{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$, we have*

$$\sum_{i=1}^K k_i |\rho_{i1}x_1 + \dots + \rho_{iN}x_N| \geq 0, \tag{3.3}$$

and let C be the cone of \bar{x} s, on which this inequality is strict. Then for all $\bar{y} = (y_1, \dots, y_N) \in \mathbb{R}^{mN}$,

$$\sum_{i=1}^K k_i \|\rho_{i1}y_1 + \dots + \rho_{iN}y_N\| \geq 0.$$

Furthermore, this inequality is strict on the cone C' where

$$A(\bar{y}) \triangleq \{b \in \mathbb{R}^m \mid \|b\| = 1, (b^T y_1, \dots, b^T y_N) \in C\}$$

has positive Lebesgue measure on the unit sphere.

Proof. Let $\xi_i \triangleq \rho_{i1}y_1 + \dots + \rho_{iN}y_N$. Then for some constant $C_m > 0$,

$$\begin{aligned} \sum_{i=1}^K k_i \|\xi_i\| / C_m &= \sum_{i=1}^K k_i \|\xi_i\| \int_{\|b\|=1} |b^T \xi_i / \|\xi_i\|| \, db = \int_{\|b\|=1} \sum_{i=1}^K k_i |b^T \xi_i| \, db \\ &= \int_{\|b\|=1} \sum_{i=1}^K k_i |\rho_{i1}x_j(b) + \dots + \rho_{iN}x_N(b)| \, db \geq 0, \end{aligned}$$

where $x_j(b) \triangleq b^T y_j$. As the area integrated over includes $A(\bar{y})$, the claim on strictness of the inequality follows. □

Lemma 3.3. *Suppose the points a_i ($i = 1, \dots, n$) are distinct, and define $r_\sigma(\bar{p}) \triangleq 2 \sum_i \|p_i - a_{\sigma_i}\|$. Then for all $\lambda \in [0, 1/2)$, and permutations σ , there exist neighbourhoods D_σ^λ of \bar{a}_σ , where*

$$f_{\text{MO}}(\bar{p}; \bar{a}) - f_{\text{MO}}(\bar{a}_\sigma; \bar{a}) \geq \lambda r_\sigma(\bar{p}), \quad \bar{p} \in D_\sigma^\lambda. \tag{3.4}$$

When $m = 1$, $\bigcup_\sigma D_\sigma^\lambda = \mathbb{R}^{nm}$ (i.e. the whole space), and when $\lambda = 0$, $D_\sigma^0 = \mathbb{R}^{nm}$. In both of these cases, the inequality holds strictly when $\bar{p} \neq \bar{a}_\sigma$ for all σ .

Proof. We have

$$\begin{aligned} f(\bar{p}; \bar{a}) - f(\bar{a}_\sigma; \bar{a}) &\geq \max \partial f(\bar{a}_\sigma; \bar{a})^T (\bar{p} - \bar{a}_\sigma) = \max \left[\prod_{i=1}^n B(0, 1) + \nabla \nu(\bar{a}_\sigma) \right]^T (\bar{p} - \bar{a}_\sigma) \\ &= (1/2)r_\sigma(\bar{p}) + \nabla \nu(\bar{a}_\sigma)^T (\bar{p} - \bar{a}_\sigma), \end{aligned}$$

where the subdifferential is calculated as for (3.2), and the last equality follows from the expression $\|x\| = \max\{z^T x \mid z \in B(0, 1)\}$ for $x \in \mathbb{R}^m$. Likewise,

$$\nu(\bar{a}_\sigma) - \nu(\bar{p}) \geq \partial \nu(\bar{p})^T (\bar{a}_\sigma - \bar{p}).$$

Therefore

$$\begin{aligned} f_{\text{MO}}(\bar{p}) - f_{\text{MO}}(\bar{a}_\sigma) &= [f(\bar{p}; \bar{a}) - f(\bar{a}_\sigma; \bar{a})] + [\nu(\bar{a}_\sigma) - \nu(\bar{p})] \\ &\geq (1/2)r_\sigma(\bar{p}) - \min[\partial \nu(\bar{p}) - \nabla \nu(\bar{a}_\sigma)]^T (\bar{p} - \bar{a}_\sigma). \end{aligned}$$

By monotonicity $[\partial \nu(\bar{p}) - \nabla \nu(\bar{a}_\sigma)]^T (\bar{p} - \bar{a}_\sigma) \geq 0$. The problem is now to bound

$$L \triangleq \min[\partial \nu(\bar{p}) - \nabla \nu(\bar{a}_\sigma)]^T (\bar{p} - \bar{a}_\sigma) \leq (1/2 - \lambda)r_\sigma(\bar{p}). \tag{3.5}$$

But, since the a_i are distinct, ν is continuously differentiable² in some neighbourhood of each a_σ . Now, we approximate

$$L \leq \sum_{i=1}^n \|[\nabla \nu(\bar{p}) - \nabla \nu(\bar{a}_\sigma)]_i\| \|p_i - a_{\sigma_i}\| \leq \max_j \|\nabla_j \nu(\bar{p}) - \nabla_j \nu(\bar{a}_\sigma)\| r_\sigma(\bar{p})/2.$$

From this we see that some neighbourhoods D_σ^λ of \bar{a}_σ can be found, where the maximum term is small enough for (3.5) to hold.

Now, if $m = 1$, there actually exists for each \bar{p} a permutation σ , for which the left hand side of (3.5) is zero. Therefore (3.5) and then (3.4) are true for all $\lambda \leq 1/2$, and $\bigcup_\sigma D_\sigma^\lambda = \mathbb{R}$. To see this, recall that where ν is differentiable (i.e. $p_i \neq p_j$ for $i \neq j$),

$$\nabla_{p_i} \nu(\bar{p}) = \sum_{j \neq i}^n \frac{p_i - p_j}{\|p_i - p_j\|}.$$

In the $m = 1$ case the terms summed over are ± 1 , indicating the direction p_j faces from p_i on the real line. But the set of these numbers over all i then uniquely determines the order of the p_i on the real line, and consequently a permutation σ , for which $\nabla \nu(\bar{p}) = \nabla \nu(\bar{a}_\sigma)$. In the non-differentiable case, $p_i = p_j$ for some $i \neq j$. In this case we can arbitrarily decide on the order, and choose the corresponding signs ± 1 from $\partial_{(p,p')} \|p - p'\| (p, p) = \{(z, -z) \mid z \in B(0, 1)\}$.

The claim on strictness of the inequality (3.4) in the $m = 1$ case follows from the non-strict variant, since $D_\sigma^{1/2}$ cover the whole space, and $(1/2)r_\sigma(\bar{p}) > \lambda r_\sigma(\bar{p})$ when $\bar{p} \neq a_\sigma$ and $\lambda < 1/2$.

²Twice actually, so we could alternatively apply the mean value theorem.

Now, if $\lambda = 0$ (and still $m = 1$), the right hand side of (3.4) is zero, and independent of σ . We have also previously shown that for every \bar{p} , the inequality holds for some σ . But since $2\nu(\bar{a}_\sigma) = f(\bar{a}_\sigma; \bar{a})$ and $f_{\text{MO}}(\bar{a}_\sigma) = f(\bar{a}_\sigma; \bar{a}) - \nu(\bar{a}_\sigma) = \nu(\bar{a}_\sigma) = \nu(\bar{a})$ does not depend on σ , actually

$$f_{\text{MO}}(\bar{p}; \bar{a}) - f_{\text{MO}}(\bar{a}_\sigma; \bar{a}) = f(\bar{p}; \bar{a}) - \nu(\bar{p}) - \nu(\bar{a}) \geq 0 \quad \text{for all } \sigma \text{ and } \bar{p}. \tag{3.6}$$

Therefore, in the $m = 1$ case, $D_\sigma^0 = \bigcup_{\sigma'} D_{\sigma'}^0 = \mathbb{R}$, for all σ .

Finally, suppose $m > 1$ and $\lambda = 0$. Since (3.6) is of the form (3.3) with $\bar{x} = (\bar{p}, \bar{a})$ when $p_i, a_k \in \mathbb{R}$, we may apply Lemma 3.2 with $\bar{y} = (\bar{p}, \bar{a})$ when $p_i, a_k \in \mathbb{R}^m$ to obtain that (3.6) holds generally. For the strict inequality, to show that $A(\bar{p}, \bar{a})$ has positive measure, choose the projection b in Lemma 3.2 so that (3.4) holds strictly, *i.e.* at least for some i , $b^T p_i \neq b^T a_k$ for all k . This can be done if $p_i \neq a_k$ for all k , because the set of projections with $b^T p_i = b^T a_k$ is then finite. By continuity, the same holds in a neighbourhood of positive measure of the chosen points and projection. Therefore $A(\bar{p}, \bar{a})$ has positive measure. \square

Proof of Theorem 3.1. Lemma 3.3 with $\lambda = 0$ proves claim (i).

As for claim (ii), since f_{MO} is continuous and level-bounded (as noted above), the cluster points of minimisers \hat{p}_λ of f_{MOTSP}^λ with $\lambda \searrow 0$, must be those of $f_{\text{MO}} = f_{\text{MOTSP}}^0$ [21], Theorem 1.17. Since there are finitely many permutations σ , there is a constant c dependent on \bar{a} , such that $f_{\text{TSP}}(\bar{a}_\sigma) \geq c + f_{\text{TSP}}(\bar{a}_{\hat{\sigma}})$ for non-optimal σ . Therefore the cluster points must be the optimal TSP paths $\bar{a}_{\hat{\sigma}}$.

To show the existence of the threshold on λ , choose an arbitrary $\tilde{\lambda} \in (0, 1/2)$. There must now exist $\hat{\lambda} \leq \tilde{\lambda}$, such that $\hat{p}_\lambda \in D \triangleq \bigcup_{\hat{\sigma}} D_{\hat{\sigma}}^{\tilde{\lambda}}$ for $\lambda \in (0, \hat{\lambda})$. If this were not so, we could find a cluster point outside D , in contradiction to previously established results.

Now, apply

$$\begin{aligned} \|a_{\sigma i} - a_{\sigma(i+1)}\| &= \|a_{\sigma i} - p_i + p_i - p_{i+1} + p_{i+1} - a_{\sigma(i+1)}\| \\ &\leq \|a_{\sigma i} - p_i\| + \|p_i - p_{i+1}\| + \|p_{i+1} - a_{\sigma(i+1)}\|, \end{aligned} \tag{3.7}$$

to yield

$$f_{\text{TSP}}(\bar{a}_\sigma) - f_{\text{TSP}}(\bar{p}) = \sum_i \left(\|a_{\sigma i} - a_{\sigma(i+1)}\| - \|p_i - p_{i+1}\| \right) \leq 2 \sum_i \|p_i - a_{\sigma i}\|. \tag{3.8}$$

Combined with (3.4), we therefore have

$$f_{\text{MOTSP}}^\lambda(\bar{p}) - f_{\text{MOTSP}}^\lambda(\bar{a}_{\hat{\sigma}}) = f_{\text{MO}}(\bar{p}) - f_{\text{MO}}(\bar{a}_{\hat{\sigma}}) + \lambda(f_{\text{TSP}}(\bar{p}) - f_{\text{TSP}}(\bar{a}_{\hat{\sigma}})) \geq (\tilde{\lambda} - \lambda)r_{\hat{\sigma}}(\bar{p}),$$

whenever $\lambda \in [0, \tilde{\lambda})$, and $\bar{p} \in D_{\hat{\sigma}}^{\tilde{\lambda}}$. This says that for $\lambda \in (0, \hat{\lambda})$, we must have $\hat{p}_\lambda = \bar{a}_{\hat{\sigma}}$ for some $\hat{\sigma}$. \square

Corollary 3.4. *Either (or both) the calculation of $\hat{\lambda}$ is NP-hard, or the problem (3.1) is NP-hard for $\lambda \in (0, \hat{\lambda})$ (and non-collinear vertices \bar{a}).*

Proof. Identical to Corollary 2.6. \square

Remark 3.5. Actually, the upper bound $\hat{\lambda}$ is not strict under Assumption 2.3, for (3.7) is strict for some $i \in \{1, \dots, n\}$. Suppose it weren't. Then all the vectors $a_{\sigma i} - p_i$, $p_i - p_{i+1}$, and $p_{i+1} - a_{\sigma(i+1)}$ would point in the same direction, for all i . But this can not be unless both p_i and p_{i+1} are collinear with $a_{\sigma i}$ and $a_{\sigma(i+1)}$. Therefore, all the four points are collinear. But likewise p_{i+1} and $a_{\sigma(i+1)}$ are collinear also with p_{i+2} and $a_{\sigma(i+2)}$. By extension, all the points p_1, \dots, p_n and, in particular, a_1, \dots, a_n are collinear, which violates our assumptions.

Lemma 3.3 also contains the following interesting special cases, obtained with $\lambda = 0$, stated here separately:

Corollary 3.6. *For any points $a_1, \dots, a_n \in \mathbb{R}^m$, and $p_1, \dots, p_n \in \mathbb{R}^m$, it holds that*

$$\sum_{i=1}^n \sum_{k=1}^n \|p_i - a_k\| \geq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|p_i - p_j\| + \frac{1}{2} \sum_{k=1}^n \sum_{\ell=1}^n \|a_k - a_\ell\|.$$

In particular, when $p_i = -a_i$,

$$\sum_{k=1}^n \sum_{\ell=1}^n \|a_k + a_\ell\| \geq \sum_{k=1}^n \sum_{\ell=1}^n \|a_k - a_\ell\|.$$

Proof. This follows from the equivalence and inequality in (3.6). □

4. SENSITIVITY ANALYSIS

Here we provide some sensitivity results for our penalised reformulations of the Euclidean TSP. This is in order to understand how the solutions vary, as the penalty parameter λ varies above $1/2$ or $\hat{\lambda}$, and to justify the use of values higher than this threshold. Recall that ϵ -argmin $f \triangleq \{x \mid f(x) \leq \min f + \epsilon\}$, and that the polar of a convex set C is $C^\circ \triangleq \{z \mid z^T x \leq 1 \text{ for all } x \in C\}$.

Define $f_{\text{KMTSP}}^\lambda \triangleq f_{\text{KM}}(\cdot; \bar{a}) + \lambda f_{\text{TSP}}$. Note that this function is locally convex at $\bar{a}_{\hat{\sigma}}$, so that the convex subdifferential is defined there. For the reformulation based on the multisource Weber problem, we then get the following theorem. The first steps of the argument are based on the epigraphical methods of [21], Section 7.J, and [6], but incorporate various optimisations and generalisations for improved bounds. (Recall that the ‘‘auxiliary ρ -epi-distance’’ of f_0 and g is defined as the infimum of $\eta \geq 0$ that satisfy $\min_{B(x, \eta)} g \leq \max\{f_0(x), -\rho\} + \eta$ and $\min_{B(x, \eta)} f_0 \leq \max\{g(x), -\rho\} + \eta$ for all $x \in B(0, \rho)$, also known as the Kenmochi conditions [5]. We replace $B(0, \rho)$ with an arbitrary set D , and translate our functions in order to not need the first inequality to bound $\min_D g$ from above. For the second inequality, we use a poor maximum value difference estimate (4.2), as a consequence of which we do not have to consider the ball $B(x, \eta)$, and minimise the gauge ψ over it.)

Theorem 4.1. *Suppose Assumption 2.3 holds. Let $\epsilon \geq 0$, $\lambda_0 \in (0, 1/2]$, and $\lambda \geq \lambda_0$. Suppose $D \cap \text{argmin } f_{\text{KMTSP}}^{\lambda_0} \neq \emptyset$, and that $\hat{p} \in \epsilon$ -argmin $D f_{\text{KMTSP}}^\lambda$. Denote $\eta \triangleq f_{\text{TSP}}(\bar{a}_{\hat{\sigma}}) - \min_D f_{\text{TSP}}$ (this value does not depend on the choice of $\hat{\sigma}$), and $C_{\hat{\sigma}} \triangleq \partial f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}})$. Then,*

(i) *If for some $\hat{\sigma}$, also $\hat{p} \in \bar{a}_{\hat{\sigma}} + \prod_{i=1}^n B(0, \delta_{\hat{\sigma}i})$ for $\delta_i \triangleq \min_{j \neq i} \|a_i - a_j\| / 2$, we actually have*

$$\hat{p} \in \bar{a}_{\hat{\sigma}} + ((\lambda - \lambda_0)\eta + \epsilon)C_{\hat{\sigma}}^\circ, \tag{4.1}$$

with the set on the right bounded.

- (ii) *Suppose (4.1) holds and $\lambda \leq \lambda_0 + \min_i (\delta_{\hat{\sigma}i} - \epsilon M_i)(\eta M_i)^{-1}$ with $M_i = \max_{\bar{x} \in C_{\hat{\sigma}}^\circ} \|x_i\|$. Then $\hat{p} \in \bar{a}_{\hat{\sigma}} + \prod_{i=1}^n B(0, \delta_{\hat{\sigma}i})$.*
- (iii) *There exists a finite index set \mathcal{T} , closed sets A_t , compact sets C_t , points \bar{q}_t , and constants $c_t \in [0, f_{\text{KMTSP}}^{\lambda_0}(\bar{q}_t) - f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}})]$, such that C_t° is bounded, and for some $t \in \mathcal{T}$,*

$$\hat{p} \in A_t \cap (\bar{q}_t + ((\lambda - \lambda_0)\eta + \epsilon - c_t)C_t^\circ).$$

(iv) *For $\hat{p} \in A_t$ in (iii), we must have $(\lambda - \lambda_0)\eta + \epsilon \geq f_0(\bar{q}_t) - \min f_0$.*

Proof. (i) Let $f_0 \triangleq f_{\text{KMTSP}}^{\lambda_0} - f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}})$ and $g \triangleq f_{\text{KMTSP}}^\lambda - f_{\text{KMTSP}}^\lambda(\bar{a}_{\hat{\sigma}})$ (the choice of $\hat{\sigma}$ not affecting the result). Then f_0 and g have the same minimisers as $f_{\text{KMTSP}}^{\lambda_0}$ and f_{KMTSP}^λ , respectively, and $\min f_0 = 0$.

We now have

$$\max_D f_0 - g = \max_{p \in D} (\lambda - \lambda_0)(f_{\text{TSP}}(\bar{a}_{\hat{\sigma}}) - f_{\text{TSP}}(p)) = \eta' \triangleq (\lambda - \lambda_0)\eta. \tag{4.2}$$

Thus for $\hat{p} \in \epsilon\text{-arg min}_D g$, we get

$$f_0(\hat{p}) - \min f_0 \leq g(\hat{p}) + \eta' - \min f_0 \leq \eta' + \epsilon, \tag{4.3}$$

since $g(\hat{p}) \leq \min_D g + \epsilon \leq \epsilon$ by the construction of g , employing the assumption $(\{\bar{a}_{\hat{\sigma}} \mid \hat{\sigma}\} = \arg \min f_0) \cap D \neq \emptyset$, and the fact that $g(\bar{a}_{\hat{\sigma}})$ has equal value for all $\hat{\sigma}$.

When $\hat{p} \in \bar{a}_{\hat{\sigma}} + \prod_{i=1}^n B(0, \delta_i)$, the choice of δ_i forces the distance $\|p_i - a_{\hat{\sigma}i}\|$ to be minimal for both p_i and $a_{\hat{\sigma}i}$ (against alternatives of the other), so that both \hat{p} and $\bar{a}_{\hat{\sigma}}$ belong to a neighbourhood on which f_{KM} is locally convex. Therefore f_0 is also convex in this neighbourhood, and we have the estimate

$$f_0(\hat{p}) - \min f_0 = f_{\text{KMTSP}}^{\lambda_0}(\hat{p}) - f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}}) \geq \max(\hat{p} - \bar{a}_{\hat{\sigma}})^T \partial f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}}). \tag{4.4}$$

The right hand side of this equation is the support function of $C_{\hat{\sigma}} = \partial f_{\text{KMTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}})$ applied to $\hat{p} - \bar{a}_{\hat{\sigma}}$. Since $0 \in C_{\hat{\sigma}}$ (being the subdifferential at a minimiser), by [20], Theorem 14.5, the support function of $C_{\hat{\sigma}}$ is the gauge $\psi_{C_{\hat{\sigma}}^\circ}(x) \triangleq \inf\{t \geq 0 \mid x \in tC_{\hat{\sigma}}^\circ\}$ of the polar of $C_{\hat{\sigma}}$. Combining with (4.3), we therefore have

$$\eta' + \epsilon \geq f_0(\hat{p}) - \min f_0 \geq \psi_{C_{\hat{\sigma}}^\circ}(\hat{p} - \bar{a}_{\hat{\sigma}}), \tag{4.5}$$

and hence $\hat{p} - \bar{a}_{\hat{\sigma}} \in (\eta' + \epsilon)C_{\hat{\sigma}}^\circ$, as claimed.

Since $C_{\hat{\sigma}}$ is bounded, $0 \in \text{int } C_{\hat{\sigma}}^\circ$ (so that $M_i > 0$); see [20]. It remains to prove that $C_{\hat{\sigma}}^\circ$ is bounded. This likewise follows if $0 \in \text{int } C_{\hat{\sigma}}$. Because $\partial f_{\text{KM}}(\bar{a}_{\hat{\sigma}}) = \prod_{i=1}^n B(0, 1)$, it suffices to have $D_i < 2$ for

$$D_i \triangleq \left\| \frac{a_{\hat{\sigma}i} - a_{\hat{\sigma}(i-1)}}{\|a_{\hat{\sigma}i} - a_{\hat{\sigma}(i-1)}\|} + \frac{a_{\hat{\sigma}i} - a_{\hat{\sigma}(i+1)}}{\|a_{\hat{\sigma}i} - a_{\hat{\sigma}(i+1)}\|} \right\| = \|\nabla_i f_{\text{TSP}}(\bar{a}_{\hat{\sigma}})\|.$$

But $D_i \leq 2$, and equality can only happen when there's a degenerate angle between $a_{\hat{\sigma}(i-1)}$, $a_{\hat{\sigma}i}$, and $a_{\hat{\sigma}(i+1)}$. By Lemma 2.4 and Assumption 2.3 this can not happen for optimal permutations $\hat{\sigma}$.

(ii) Approximating the gauge $\psi_{C_{\hat{\sigma}}^\circ}$, we get

$$\begin{aligned} \psi_{C_{\hat{\sigma}}^\circ}(x) &\geq \inf\{t \geq 0 \mid x_i = z_i^i \text{ for } \bar{z}^i \in tC_{\hat{\sigma}}^\circ, \text{ for all } i = 1, \dots, n\} \\ &= \max_{i=1, \dots, n} \inf\{t \geq 0 \mid x_i = z_i^i \text{ for } \bar{z}^i \in tC_{\hat{\sigma}}^\circ\} \\ &\geq \max_{i=1, \dots, n} \inf\{t \geq 0 \mid \|x_i\| \leq tM_i\} = \max_i \|x_i\| / M_i. \end{aligned}$$

Since the condition in the statement of the claim is equivalent to $(\eta' + \epsilon) \leq \delta_{\hat{\sigma}i} / M_i$ for all i , claim (ii) follows from applying this information to (4.5).

(iii) Notice that f_{KM} and then f_0 is convex on a finite family $\{A_t \mid t \in \mathcal{T}\}$ of closed sets – corresponding to different associations of the p_i to a_k (possibly multiple/empty) – that fill the entire space. On these regions f_{KM} is equal to some convex function $f^t : \bar{p} \mapsto \sum_{k=1}^n \|a_k - p_{i(k,t)}\|$ for some association $i(k,t)$ (not necessarily a permutation). Let \bar{q}_t be a minimiser of $f_{\text{TSP}}^t \triangleq f^t + \lambda_0 f_{\text{TSP}}$, not necessarily in A_t . The subdifferential of f_{TSP}^t may be a singleton at \bar{q}_t , and thus not provide much information. But we can use a more informative approximate subdifferential containing 0 in its interior and thus with bounded polar, as follows.

The function f_{TSP}^t is level-bounded: suppose $\|\bar{z}\| = 1$. Then the triangle inequality gives $f_{\text{TSP}}^t(\bar{q}_t + \alpha\bar{z}) / \alpha \geq \sum_{k=1}^n \|z_{i(k,t)}\| + \lambda_0 f_{\text{TSP}}(\bar{z}) - f_{\text{TSP}}^t(\bar{q}_t) / \alpha$. If $f_{\text{TSP}}(\bar{z}) < \delta$, we must have $\|z_k - z_i\| < \delta$ for all $k, i = 1, \dots, n$. But then for small enough $\delta > 0$, by $\|\bar{z}\| = 1$ each $\|z_i\|$ must be close to $1/\sqrt{n}$. Thus $\sum_{k=1}^n \|z_{i(k,t)}\| + f_{\text{TSP}}(\bar{z})$ is bounded from below on $\{\|\bar{z}\| = 1\}$ by some value greater than zero. Therefore, for big enough $\alpha > 0$, $f_{\text{TSP}}^t(\bar{q}_t + \alpha\bar{z}) / \alpha$ is greater than some constant, and hence f_{TSP}^t is level-coercive and then level-bounded. Thus by Lemma A.2 in Appendix 7, $0 \in \text{int } \partial_{\epsilon'} f_{\text{TSP}}^t(\bar{q}_t)$ for $\epsilon' > 0$.

We may assume that $\bar{q}_t \neq \bar{a}_{\hat{\sigma}}$, for otherwise claim (i) provides the result. Let then $C_t \triangleq \partial_{\epsilon'} f_{\text{TSP}}^t(\bar{q}_t)$ for small enough $\epsilon' > 0$ that $c_t \triangleq f_0(\bar{q}_t) - \min f_0 - \epsilon' \geq 0$ holds. Similarly to (4.4) and (4.5), we may then approximate

$$\eta' + \epsilon \geq f_0(\hat{p}) - f_0(\bar{q}_t) + f_0(\bar{q}_t) - \min f_0 \geq \psi_{C_t^{\circ}}(\hat{p} - \bar{q}_t) + c_t \quad \text{for } \hat{p} \in A_t.$$

The claim follows from the definition of the gauge $\psi_{C_t^{\circ}}$. Note that we must have $\eta' + \epsilon \geq c_t$ if $\hat{p} \in A_t$, since $\psi_{C_t^{\circ}} \geq 0$. Expanding this condition, and letting $\epsilon' \searrow 0$, we get claim (iv). \square

Suppose that \hat{p} is an (approximate) minimiser of the perturbed problem $f_{\text{KMTSP}}^{\lambda}$ in a predetermined neighbourhood D of any $\bar{a}_{\hat{\sigma}}$. The first two claims of Theorem 4.1 then say that for small $\lambda > \lambda_0 = 1/2$, \hat{p} actually belongs to a smaller set that behaves quite well with respect to λ and $\epsilon \geq 0$. The fourth claim says that for \hat{p} to not belong to the predetermined neighbourhood of some $\bar{a}_{\hat{\sigma}}$, λ or ϵ must be large enough (since $f_0(\bar{q}_t) > \min f_0$ for $\bar{q}_t \neq \bar{a}_{\hat{\sigma}}$). Therefore for small enough $\lambda > 1/2$, the minimisers of $f_{\text{KMTSP}}^{\lambda}$ stay within a linearly-scaled region around $\bar{a}_{\hat{\sigma}}$.

The optimal solution appears in the local bound through $C_{\hat{\sigma}}^{\circ}$. However, applying the argument proving the boundedness of this set, we can approximate it by considering all the possible non-degenerate angles between the points a_k , and choosing the smallest ones. That will, of course, increase the bound. Computing the global bound is much more complicated.

Note that claim (i) of Theorem 4.1 provides a necessary condition for a local minimisers (or, in fact, any point for either the KM or MO reformulation) to be close to a real solution of the Euclidean TSP: if the point \hat{p} can be unambiguously morphed into \bar{a}_{σ} for some, not necessarily optimal σ – which is the case *e.g.* when $\hat{p} \in \prod_{i=1}^n B(a_{\sigma i}, \delta_{\sigma i})$ – the condition (4.1) must hold for the permutation σ to be an optimal path. It suffices to take $\eta = f_{\text{TSP}}(\bar{a}_{\sigma}) - f_{\text{TSP}}(\hat{p})$ and $D = \{\bar{a}_{\sigma}, \hat{p}\}$, for if η becomes negative this way, we know a better minimiser and test point.

It remains to discuss the sensitivity in λ of solutions to $f_{\text{MOTSP}}^{\lambda}$. Similarly to the KM case, by Lemma 3.3 and (3.8), we could for $\lambda_1 \in (\lambda_0, 1/2)$, obtain a gauge approximating $f_{\text{MOTSP}}^{\lambda_0}(\hat{p}) - f_{\text{MOTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}})$ from below in $D_{\hat{\sigma}}^{\lambda_1}$. This happens by replacing (4.4) by

$$\begin{aligned} f_{\text{MOTSP}}^{\lambda_0}(\hat{p}) - f_{\text{MOTSP}}^{\lambda_0}(\bar{a}_{\hat{\sigma}}) &= f_{\text{MO}}(\hat{p}; \bar{a}) - f_{\text{MO}}(\bar{a}_{\hat{\sigma}}; \bar{a}) - \lambda_0(f_{\text{TSP}}(\bar{a}_{\hat{\sigma}}) - f_{\text{TSP}}(\hat{p})) \\ &\geq \lambda_1 r_{\hat{\sigma}}(\hat{p}) - \lambda_0 r_{\hat{\sigma}}(\hat{p}) = \max(p_i - a_{\hat{\sigma} i})^T C_{\hat{\sigma}}, \end{aligned}$$

where $C_{\hat{\sigma}} \triangleq 2(\lambda_1 - \lambda_0) \prod_{i=1}^n B(0, 1)$. However, as the neighbourhoods $D_{\hat{\sigma}}^{\lambda_1}$ are merely proved to exist (when $m > 1$, and $\lambda_1 > 0$, which we require), the bounds so obtained would be rather poor compared to $f_{\text{KMTSP}}^{\lambda}$.

5. HEURISTICS

As we shall see in Section 6, the performance of our basic algorithm is not all that great for larger instances of the Euclidean TSP. Therefore, in this section, we consider various heuristic approaches that could be used to speed up the algorithm or improve the results otherwise. As a first task, however, the association heuristic demands some clarification.

5.1. The association heuristic

The proof of Theorem 2.5 provides a conceptual algorithm for obtaining a permutation σ from any sequence of points $\bar{p} = (p_1, \dots, p_n)$:

- (1) Assign the points a_k to the closest p_j , forming the cluster C_j (handling ambiguous cases arbitrarily).
- (2) Remove all the points p_j with empty clusters.
- (3) Re-insert points in the path, at any $a_k \in C_j$, $a_k \neq p_j$ (the closest in our implementations), before or after p_j (depending on which seems to provide shorter path).
- (4) Repeat steps (1)–(3) while there's something to be done.

Note that when C_j consists of a_k alone (and there were no ambiguous assignments), these steps amount to moving p_j at a_k , as p_j would be removed after the new point has been placed at a_k .

Any reinsertion may change the clusters, the new (reinserted) point assimilating points from clusters of p_i , for $i \neq j$ as well as j . If we ignore this fact for $i \neq j$, we may construct σ locally in a hierarchic fashion, “splitting” each cluster until it consist of a single a_k . Otherwise we need to recalculate/shuffle the clusters after each reinsertion. Some improvements to the resulting path length can sometimes be obtained this way, but the method is quite dependent on the order of processing.

5.2. Number of cluster centres

A straightforward heuristic derived from our reformulations in the earlier sections, is to reduce the number s of the points p_i used in the minimisation method. After the “shape” of the path has been obtained with a reduced number of points, it can then be refined by adding more points using the already described rules for associating (unassociated and duplicate) points with cities. In case of the MO variant, when $s < n$, we have to alter the factor of the function ν_{MO} , in order to keep the objective function level-bounded, and for reasonable results. Our somewhat arbitrary but obvious choice of factor is $n/(2s)$, which is below the $n/(2s - 2)$ upper bound from [23]:

$$\nu_{\text{MO}}^s(\bar{p}) \triangleq \frac{n}{2s} \sum_{i=1}^s \sum_{j=1}^s \|p_i - p_j\|.$$

Notice that the upper bound for λ ensuring that $\lambda f_{\text{TSP}} - \nu_{\text{MO}}^s$ is concave, increases similarly, and we have indeed used $\lambda = n/s$ in our experiments.

5.3. Hierarchical clustering

An obvious refinement of the previous heuristic is analogous to hierarchical clustering:

- (1) Run our path-length perturbed clustering algorithm on the whole data, with a small number s of clusters.
- (2) Assign each a_k to the closest p_i , producing the cluster C_i .
- (3) Run the algorithm again on C_i with a new set of “cluster centres”, of size $s_i \leq \#C_i$. Continue this subdivision until the size of the cluster C_i is small enough to merit choosing $s_i = \#C_i$.
- (4) Construct the full path by combining the paths of the lowest-level clusters along the paths formed by the higher-level clusters centres.

There’s a small problem with this approach as such: the paths are closed, so combining them will produce unnecessary detours. However, this is no big problem: we just have to alter f_{TSP} to not attract the first and last points of the open path we want. We can do more: we can attract the endpoints to points in the previous cluster:

$$f_{\text{TSP}}^{\text{open}}(\bar{p}; a_{\text{prev}}, a_{\text{next}}) \triangleq \|p_1 - a_{\text{prev}}\| + \|p_s - a_{\text{next}}\| + \sum_{j=1}^{s-1} \|p_j - p_{j+1}\|. \tag{5.1}$$

There are various potential choices for a_{prev} and a_{next} . One is the points p_{i-1} and p_{i+1} in the higher-level path (when we’re working on C_i). Another would be the points $a_{\text{prev}} \in C_{i-1}$ and $a_{\text{next}} \in C_{i+1}$ that minimise the distance to C_i . In the experiments of Section 6 we have chosen the former. Based on a limited number of tests, the latter more complex approach does not seem to improve the results. Note that the first two terms of $f_{\text{TSP}}^{\text{open}}$ are Euclidean distances from fixed points. They can therefore be included in the convex part of the objective function, when we choose to minimise it with the perturbed Weiszfeld method.

A few more choices remain in the hierarchical algorithm: At which point to run the association heuristic: for the whole path, or for the lowest-level clusters? In the experiments to follow, we have chosen the former combined with the local association heuristic, as this seems to provide the best ratio of time spent to quality of results. Another available choice is the number of points s_i to use in each cluster. Our somewhat arbitrary choice has been to specify a maximum number $M \geq 2$, but instead of greedily choosing $s_i = \min\{M, \#C_i\}$,

we try to do this bottom-up: we try to predictively assign the largest number of points to the lowest-level clusters, by choosing

$$s_i = \lceil \#C_i / M^{\lceil \log_M \#C_i \rceil} \rceil \quad \text{when} \quad \#C_i > M. \quad (5.2)$$

This appears to provide better results than the greedy approach, based on a limited number of tests.

5.4. Clustering for initial iterate

This approach consists of running the previous heuristic without the association step to obtain an initial iterate for the basic algorithm, that we perform only a few steps of.

5.5. Path-following

Yet another approach would be to calculate an approximate solution to a penalised version of the problem for some λ , and then with a smaller one starting from the previous result. Unfortunately, at least with the limitation $\lambda \leq 1$ inherent in the perturbed Weiszfeld method, this does not appear to provide considerably improved results.

6. EXPERIMENTS

We have implemented our algorithms (in Haskell [17]) and tested our method on some problems from TSPLIB [19], on an Athlon64 3200+ tabletop computer. In each case, we have used the step size $\omega = 1.4$ in the perturbed Weiszfeld algorithm³ of [23]: of the values we've tried, it seems to provide the most consistently best results, largely in agreement with experimental results for the plain Weiszfeld algorithm [7], Appendices 2–3. Although each $\omega \in [1, 2)$ does provably provide a descending sequence of iterates, it would be possible to do a line search step in the algorithm as well. The initial iterate has likewise in each case been with the cities equally distributed on a circle, centred and scaled to fit in the problem data. Such a choice seems to provide generally better results than a (totally) random initial iterate, which may contain self-crossings of the path on a large scale, that our method seems poor at removing.

6.1. The basic algorithm

In summary, the basic algorithm consists of

- (1) Choose an initial iterate \bar{p}_0 , step length $\omega \in [1, 2)$, penalty parameter $\lambda \in (0, 1]$, and maximum iterations count or other stopping criterion.
- (2) Apply the perturbed Weiszfeld method to (3.1), to get \tilde{p} .
- (3) Use the association heuristic to find a permuted path \tilde{a}_σ from \tilde{p} .

The results for this method may be found in Tables 1–4. Furthermore, Figure 1 shows results for some simple instances from the first series of tests. In most of the test cases, we have used $\lambda = 1.0$, as it is the upper limit at which the TSP penalty term can certainly be “absorbed” into the concave part of the diff-convex objective, and thus that our algorithm can handle. Lower values also do not appear to provide better results. In each of these tests of the basic algorithm, we have used the “semi-global” variant of this association heuristic discussed in Section 5, to obtain a permutation of the points a_1, \dots, a_n from the results of the Weiszfeld algorithm; cf. Figure 1(c). In two problems, Eil101' and PR1002', some of the parameters have been varied to offer points of comparison: the problem Eil101' uses $\lambda = 2.0$, although our algorithm is not entirely applicable for such a choice. In the problem PR1002' we have used only $s = 50$ p_i s in the perturbed Weiszfeld method and added the rest later, as again discussed in Section 5.

³In summary, the method consists of choosing an initial iterate $\bar{p}^{(0)}$, step length $\omega \in [1, 2)$, and on the simplifying assumption that $f(\cdot; \bar{a})$ is differentiable at \bar{p} , iterating until a stopping criterion is satisfied, the mapping $\bar{p} \mapsto \bar{p} - \omega S^t(\bar{p})(\nabla f(\bar{p}; \bar{a}) - v)$ with arbitrary $v \in \partial(\nu_{\text{MO}}(\bar{p}) - \lambda f_{\text{TSP}}(\bar{p}))$ and $S(\bar{p}) \triangleq (\sum_{k=1}^n I/\|a_k - p_1\|, \dots, \sum_{k=1}^n I/\|a_k - p_s\|)$. When $f(\cdot; \bar{a})$ is non-differentiable at \bar{p} , the method requires more space to explain than available here. However, numerical difficulties aside, this could be avoided by altering the step length ω within the mentioned bounds.

TABLE 1. Results for $\text{max_iters} = 1000$, $\tau = 10^{-5}$, and $\omega = 1.4$.

Problem	Berlin52	Eil101	TS225	PR1002	Eil101'	PR1002'
s	n	n	n	n	n	50
λ	1.0	1.0	1.0	1.0	2.0	1.0
Weiszfeld iterations	1000	1000	1000	1000	1000	1000
Weiszfeld time	3.0	10.9	53.1	1315.0	11.5	46.4
Total time	3.1	11.3	56.0	1604.0	11.9	472.9
TSPLIB path length	7542	629	126 643	259 045	629	259 045
Result path length	8951.6	726.0	207 730.3	370 184.2	706.8	375 395.6

TABLE 2. Results for $\text{max_iters} = 10\,000$, $\tau = 10^{-2}$, and $\omega = 1.4$.

Problem	Berlin52	Eil101	TS225	PR1002	Eil101'	PR1002'
s	n	n	n	n	n	50
λ	1.0	1.0	1.0	1.0	2.0	1.0
Weiszfeld iterations	396	201	797	10 000	10 000	1875
Weiszfeld time	1.2	2.3	42.6	12 877.3	114.1	88.2
Total time	1.3	2.6	45.5	13 146.1	114.5	518.7
TSPLIB path length	7542	629	126 643	259 045	629	259 045
Result path length	8951.6	719.7	207 730.3	363 456.1	702.9	365 239.2

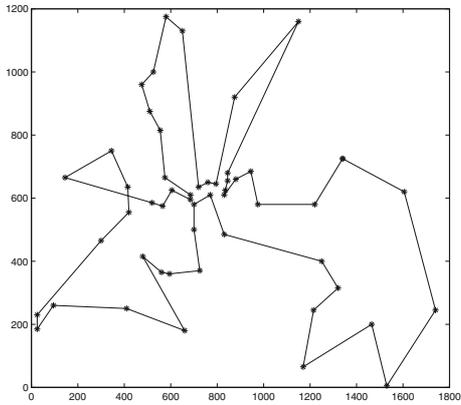
TABLE 3. Results for $\text{max_iters} = 10\,000$, $\tau = 10^{-5}$, and $\omega = 1.4$.

Problem	Berlin52	Eil101	TS225	PR1002'
s	n	n	n	50
λ	1.0	1.0	1.0	1.0
Weiszfeld iterations	1531	2573	2376	2344
Weiszfeld time	4.7	28.3	125.0	112.9
Total time	4.7	28.6	127.8	554.12
TSPLIB path length	7542	629	126 643	259 045
Result path length	8951.6	706.7	207 730.3	365 239.2

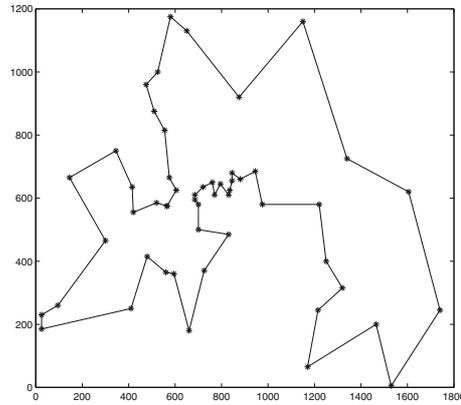
TABLE 4. Results for $\text{max_iters} = 10 \log_2 n$, $\tau = 10^{-5}$, and $\omega = 1.4$.

Problem	Berlin52	Eil101	TS225	PR1002	Eil101'	PR1002'
s	n	n	n	n	n	50
λ	1.0	1.0	1.0	1.0	2.0	1.0
Weiszfeld iterations	57	67	78	100	67	100
Weiszfeld time	0.18	0.8	4.6	145.5	0.8	4.8
Total time	0.25	1.2	9.6	617.4	1.2	442.5
TSPLIB path length	7542	629	126 643	259 045	629	259 045
Result path length	9087.1	741.9	210 694.5	392 377.7	741.9	372 602.0

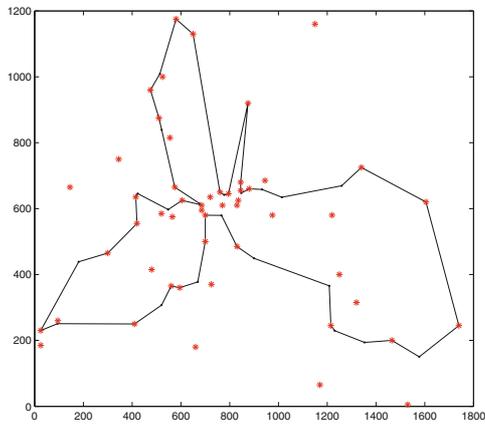
In the first series, in Table 1, the maximum number of iterations of the perturbed Weiszfeld method has been 1000, and the stopping threshold τ (maximum difference in norm between successive iterates) has been 10^{-5} , whereas in Table 2 the values have been 10 000 and 10^{-2} , respectively. In the third series in Table 1, where we have excluded the cases from the second series that used the maximum number of iterations, the values are 10 000 and 10^{-5} , respectively. Finally in Table 4, we have allowed for just $10 \log_2 n$ iterations.



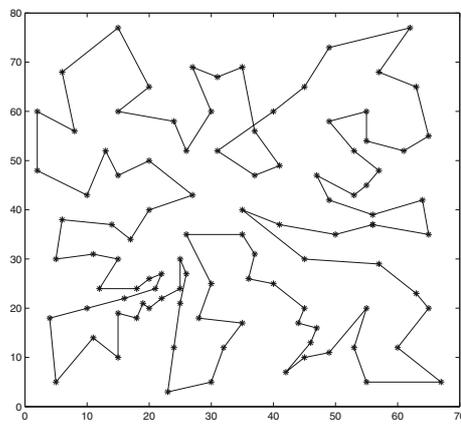
(a) Berlin52: Result, $\lambda = 1.0$



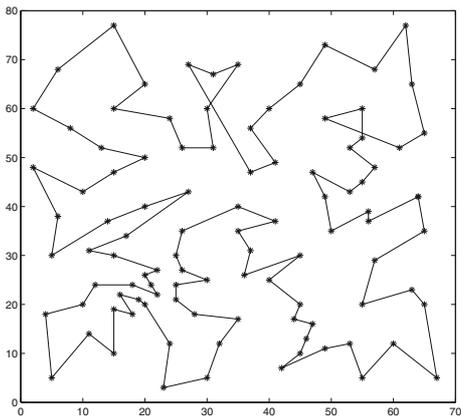
(b) Berlin52: TSPLIB optimal path



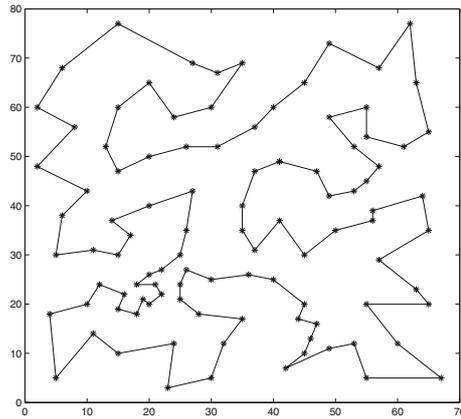
(c) Berlin52: raw Weiszfeld, $\lambda = 1.0$



(d) Eil101: Result, $\lambda = 1.0$



(e) Eil101: Result, $\lambda = 2.0$



(f) Eil101: TSPLIB optimal path

FIGURE 1. Results for Berlin52 and Eil101 from TSPLIB.

In the tables, the “Weiszfeld time” field is the time (in seconds) it took for the perturbed Weiszfeld method to finish, and the field “Weiszfeld iterations” is the number of iterations of this method used. The “Total time” field indicates the time it took in addition to this, to move the resulting points towards the cities, as described above. Such an intermediate result is included in Figure 1(c) for the Berlin52 problem. Note that the “TSPLIB path length” is calculated with the Euclidean metric rounded to nearest integer, instead of the plain Euclidean metric, with which “Our path length” has been calculated. Finally, the instance size (n) is indicated by the TSPLIB problem name itself.

As we can see, the results are not all that great, compared to what is achievable with other methods; *cf.* [13,14]. Some of the run-time can be attributed to our choice of language: Haskell and the compilers available for it, with standard unoptimised data structures, are not presently quite up to par with lower-level languages in speed, but offer much comfort of implementation. As for the quality of the paths, it can clearly be seen that the relative quality of the results degrades as the number of cities grows. Looking at the figures, our results seem to share a lot of the overall structure of the optimal results, however, which would indicate that they could serve as starting points for other methods. Also note comparing Figure 1(a)–1(c), that our naïve association heuristics induce some clear mistakes, such as self-crossings of the path.

That increasing λ for Eil101’ improves the path, seems to be a general trend, although occasionally worse results are obtained as for $\lambda = 1.0$. Decreasing λ below 1.0 also usually seems to degrade the result, as would increasing it too much. In the second series of tests, we also see that the algorithm indeed does not appear to converge.

In the last series with just $10 \log_2 n$ iterations, the performance does not actually decrease relatively that much from series with more iterations. In this series, the result for PR1002 with 50 p_i s actually beats the one for all 1002 p_i s in the Weiszfeld method. In both cases, considerable time is spent in the (quite unoptimised and naïvely implemented) association heuristic.

6.2. The hierarchical algorithm

In summary, this heuristic consists of the steps

- (1) Choose maximum prototype count $M \geq 2$, as well as parameters for the perturbed Weiszfeld method, and initialise the initial cluster $C_0 = \{a_1, \dots, a_n\}$.
- (2) Calculating new prototype count s_i for each cluster from (5.2).
- (3) Apply the Weiszfeld method with the modified penalty (5.1) and s_i new prototypes on each present cluster C_i . The points a_{prev} and a_{next} are the prototypes of the present prototypes with next and previous index.
- (4) Split clusters that did not yet have equally many new prototypes and vertices, by the new prototypes. Recursively continue from Step (2).
- (5) Apply the association heuristic on each completed cluster, and join the in-cluster paths in the order given by the higher-level clusters.

Table 5 lists results for this approach. The number “Total Weiszfeld its.” in the table, is the total number of iterations of the Weiszfeld algorithm at all scales. As already mentioned in Section 5, we have used the local variant of the association heuristic on the full resulting hierarchical Weiszfeld path, to obtain the final permutation.

In this series of experiments, we have used bigger problem instances than in the previous experiments. As can be seen from the results, with this heuristic, the running time becomes noticeably more feasible than that of the basic algorithm, and without degrading the results – improving them, in fact. (For the smallest instances from the other experiments, the heuristic degrades the results, however.) Note that for the biggest instances we only have bounds on the optimal path length from TSPLIB, and for PLA33810 this is, in fact, for the ceiling of the Euclidean distance, instead of rounded.

Using only a small number of iterations has been more our goal in this series of tests, than obtaining the best possible result we can with our algorithms. By using two times as many steps in each cluster ($20 \log_2 \#C$), we could still improve some of the results noticeably, whereas others would simply take longer to compute without

TABLE 5. Clustering heuristic results ($\text{max_iters} = 10 \log_2 \#C$, $\tau = 10^{-5}$, $\omega = 1.4$).

Problem	PR1002			PR2392		
M	50	100	150	50	100	150
Total Weiszfeld its.	1741	1033	835	4472	2606	1743
Weiszfeld time	5.2	7.7	10.6	27.2	24.5	32.4
Total time	5.4	8.0	10.9	28.6	25.8	33.7
TSPLIB path length	259 045			378 032		
Result path length	345 380	346 628	346 902	558 451	535 006	521 040
Problem	RL11849			PLA33810		
M	50	100	150	50	100	150
Total Weiszfeld its.	21 298	12 759	8980	62 609	34 811	24 940
Weiszfeld time	122.3	288.4	454.1	586.9	1270.3	2067.3
Total time	172.2	339.0	504.5	1037.0	1720.6	2518.7
TSPLIB path l. bnd.	[920 847, 923 368]			[65 913 275, 66 116 530]		
Result path length	1 410 087	1 386 317	1 360 373	99 304 887	97 915 373	96 554 643

TABLE 6. Average results for random instances.

Algorithm	Basic		Clustering 50, 100 and 150	
	uniform 1k–3k	clustered 1k–10k	uniform 1k–3k	clustered 1k–10k
# samples	15	15	54	54
Average performance	1.66	1.49	1.42	1.36

much improvement. (More meticulous choice of τ could of course be used to control the number of steps as well.) Likewise, using the hierarchical method with a small number of iterations of the basic Weiszfeld method to obtain an initial iterate, as discussed in Section 5, would slightly improve the results. For larger instances there would be a noticeable increase in time spent, however.

Notice, nevertheless, that the results appear to fall approximately around 1.5 times the optimal path length (modulo slightly differing distance measures). Further evidence for this is provided in Table 6. There, we have calculated the average performance of our methods for the 1–3k random and 1–10k city random clustered Euclidean instances of the TSP DIMACS challenge problems [14]⁴. The average for the clustering heuristic is further taken over all the parameter values $M = 50, 100, 150$. The performance reported is the proportion of the path length calculated by our algorithm, to the Held-Karp bound for the problem. Our methods appear to perform better for the clustered than non-clustered instances, as can be expected.

6.3. Use as an initial tour

We also tested in a few cases, the use of our method for providing an initial tour for other methods: LKH [10], Concorde [1], and basic 2-Opt. All of these methods improved upon the initial tour from our method. Unfortunately, our method did not significantly improve upon a random or default initial tour: LKH and Concorde did in fact seem to take longer in their computations. The 2-Opt results varied, with the initial tour from our method occasionally providing significant improvements in the final results, and at other times slightly worse results. (The results obviously depend on the processing order in the implementation of the method.) It seems to us that these non-geometrical algorithms fail to exploit the overall shape of the path that our method seems to approximate, with the errors being mostly (but not exclusively) on the small scale.

⁴For the data, see <http://www.research.att.com/~dsj/chtsp/>.

7. CONCLUSIONS

Our computational results are not earth-shattering. Local algorithms seldom perform miracles, after all. The primary contributions of this paper consist more of providing reformulations of the Euclidean TSP, that are closely related to clustering problems. There still remain questions to be answered (and asked) in that area. The application of our method for generating an initial tour for other (possibly new) methods is one possible avenue for further research. It also seems to be worth trying out other optimisation methods that would allow higher penalty parameters. This was chosen to be outside the scope of this paper, however.

APPENDIX A: LEMMAS ON SUBDIFFERENTIALS

In this section we include a few simple results on convex (approximate) subdifferentials needed in the sensitivity analysis of Section 4, that do not seem to appear in the standard literature. For convex $f : \mathbb{R}^m \rightarrow \mathbb{R}$, we denote the range of the subdifferential by $\mathcal{R}(\partial f) \triangleq \bigcup_{p \in \mathbb{R}^m} \partial f(p)$. First we have the rather obvious,

Lemma A.1. *Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be convex, proper and level-bounded. Then $0 \in \text{int } \mathcal{R}(\partial f)$.*

Proof. Since f is proper and level-bounded, it has a finite minimum, and we may assume without loss of generality, that $0 \in \partial f(0)$. Denote $A \triangleq \text{cl } \mathcal{R}(\partial f)$. The set A is then convex [20], Section 24. Suppose $0 \in \text{bd } A$. Then there exists a direction $z \in N_A(0)$, the normal cone to A at 0, with $z \neq 0$. Thus in particular $z^T \partial f(\alpha z) \leq 0$ for all $\alpha \geq 0$. But by monotonicity (abusing notation slightly), $(z - 0)^T (\partial f(\alpha z) - 0) = z^T \partial f(\alpha z) \geq 0$. Thus $z^T \partial f(\alpha z) = 0$ for all $\alpha \geq 0$. But then $f(0) \geq f(\alpha z) + \partial f(\alpha z)^T (0 - \alpha z) = f(\alpha z)$ for all $\alpha \geq 0$ in contradiction to level-boundedness. \square

Lemma A.2. *Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be convex, continuous and level-bounded, achieving its minimum at q . Then $0 \in \text{int } \partial_\epsilon f(q)$ for $\epsilon > 0$.*

Proof. By Lemma A.1, for small $r > 0$, $B(0, r) \subset \text{int } \mathcal{R}(\partial f)$. Let now

$$\epsilon(r) \triangleq - \min_{p \in \mathbb{R}^m} g(p, r) \triangleq - \min_{p \in \mathbb{R}^m} (f(p) - f(q) - r \|p - q\|).$$

The function g is continuous, and since $\text{int } \mathcal{R}(\partial f) \supset B(0, r) = \mathcal{R}(\partial(r \|\cdot - q\|))$ with the latter closed, $g(\cdot, r)$ is level-bounded by [23], Theorem 7. Since g is a decreasing function of r , it is also locally uniformly level-bounded. Thus for small $r > 0$, the function ϵ is continuous by [21], Theorem 1.17, and finite (by the showed properties of g). As, in fact,

$$\epsilon(r) = - \min_{z \in B(0, r)} \min_{p \in \mathbb{R}^m} (f(p) - f(q) - z^T(p - q)),$$

we have (cf. [11], Sect. XI) that $\partial_{\epsilon(r)} f(q) \supset B(0, r)$. Finally, since ϵ is continuous and increasing with $\epsilon(0) = 0$, we can find for small enough $\epsilon > 0$ an $r(\epsilon) > 0$, such that $\partial_\epsilon f(q) \supset B(0, r(\epsilon))$. From this the claim follows for small ϵ , and then for all from the nesting of the approximate subdifferentials. \square

Acknowledgements. The authors would like to thank professor Eugene Stepanov for very helpful feedback on the manuscript, as well as a pseudonymous person on the newsgroup `sci.math.research`, for pointing us towards the reduction theorem of Levi (for the proof of Cor. 3.6).

REFERENCES

- [1] D. Applegate, R. Bixby, V. Chavátal and W. Cook, On the solution of traveling salesman problems, in *Doc. Math., Extra volume ICM 1998 III*, Berlin (1998) 645–656.
- [2] S. Arora, Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM* **45** (1998) 753–782.
- [3] S. Arora, Approximation schemes for NP-hard geometric optimization problems: a survey. *Math. Program.* **97** (2003) 43–69.

- [4] S. Arora, P. Raghavan and S. Rao, Approximation schemes for Euclidean k -medians and related problems, in *ACM Symposium on Theory of Computing* (1998) 106–113.
- [5] H. Attouch and R.J.-B. Wets, Quantitative stability of variational systems: I. The epigraphical distance. *Trans. Amer. Math. Soc.* **328** (1991) 695–729.
- [6] H. Attouch and R.J.-B. Wets, Quantitative stability of variational systems: II. A framework for nonlinear conditioning. *SIAM J. Optim.* **3** (1993) 359–381.
- [7] S. Äyrämö, Knowledge Mining Using Robust Clustering. *Jyväskylän Studies in Computing* **63**. University of Jyväskylä, Ph.D. thesis (2006).
- [8] J.J. Bentley, Fast algorithms for geometric traveling salesman problems. *ORSA J. Comput.* **4** (1992) 887–411.
- [9] G. Buttazzo and E. Stepanov, Minimization problems for average distance functionals, in *Calculus of Variations: Topics from the Mathematical Heritage of Ennio De Giorgi*, D. Pallara Ed., *Quaderni di Matematica, Seconda Università di Napoli, Caserta* **14** (2004) 47–83.
- [10] K. Helsgaun, An effective implementation of the Lin-Kernighan traveling salesman heuristic. *Eur. J. Oper. Res.* **126** (2000) 106–130.
- [11] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms I-II*. Springer (1993).
- [12] R. Horst and P.M. Pardalos Eds., *Handbook of Global Optimization*. Kluwer Academic Publishers (1995).
- [13] D.S. Johnson and L.A. McGeoch, The traveling salesman problem: A case study in local optimization, in *Local Search in Combinatorial Optimization*, E. Aarts and J. Lenstra Eds., John Wiley and Sons (1997) 215–310.
- [14] D.S. Johnson and L.A. McGeoch, Experimental analysis of heuristics for the STSP, in *The Traveling Salesman Problem and Its Variations*, G. Gutin and A.P. Punnen Eds., Springer (2002) 369–443.
- [15] J.D. Litke, An improved solution to the traveling salesman problem with thousands of nodes. *Commun. ACM* **27** (1984) 1227–1236.
- [16] D.S. Mitrinović, *Analytic Inequalities*. Springer-Verlag (1970).
- [17] S. Peyton Jones, *Haskell 98 Language and Libraries: The Revised Report*. Cambridge University Press (2003).
- [18] P. Polak and G. Wolansky, The lazy travelling salesman problem in \mathbb{R}^2 . *ESAIM: COCV* **13** (2007) 538–552.
- [19] G. Reinelt, TSPLIB – A traveling salesman problem library. *ORSA J. Comput.* **3** (1991) 376–384.
- [20] R.T. Rockafellar, *Convex Analysis*. Princeton University Press (1972).
- [21] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*. Springer (1998).
- [22] T. Valkonen, Convergence of a SOR-Weiszfeld type algorithm for incomplete data sets. *Numer. Funct. Anal. Optim.* **27** (2006) 931–952.
- [23] T. Valkonen and T. Kärkkäinen, Clustering and the perturbed spatial median. *Computer and Mathematical Modelling* (submitted).