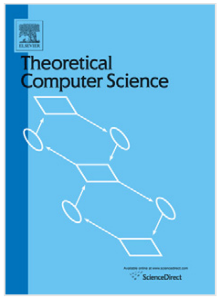
Accepted Manuscript



A rounding algorithm for approximating minimum Manhattan networks

Victor Chepoi, Karim Nouioua, Yann Vaxès

PII: S0304-3975(07)00765-7
DOI: 10.1016/j.tcs.2007.10.013
Reference: TCS 6686

To appear in: *Theoretical Computer Science*

Received date: 17 September 2004
Revised date: 17 July 2007
Accepted date: 6 October 2007

A rounding algorithm for approximating minimum Manhattan networks[1]

Victor Chepoi, Karim Nouioua, Yann Vaxès

Laboratoire d'Informatique Fondamentale de Marseille,
Université de la Méditerranée, Faculté des Sciences de
Luminy, F-13288 Marseille Cedex 9, France,
{*chepoi,nouioua,vaxes*}*@lif.univ-mrs.fr*

Abstract. For a set T of n points (terminals) in the plane, a *Manhattan network* on T is a network $N(T) = (V, E)$ with the property that its edges are horizontal or vertical segments connecting points in $V \supseteq T$ and for every pair of terminals, the network $N(T)$ contains a shortest l_1-path between them. A *minimum Manhattan network* on T is a Manhattan network of minimum possible length. The problem of finding minimum Manhattan networks has been introduced by Gudmundsson, Levcopoulos, and Narasimhan (APPROX'99) and its complexity status is unknown. Several approximation algorithms (with factors 8,4, and 3) have been proposed; recently Kato, Imai, and Asano (ISAAC'02) have given a factor 2 approximation algorithm, however their correctness proof is incomplete. In this paper, we propose a rounding 2-approximation algorithm based on a LP-formulation of the minimum Manhattan network problem.

Keywords. l_1-distance, network design, linear programming, approximation algorithms.

Corresponding author: Victor Chepoi

[1]An extended abstract of this paper appeared in the proceedings of the 8th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX'05, LNCS 3624, pp. 40-51. This research was supported in part by the CNRS MathSTIC project "Combinatoire et géométrie des polyèdres: convexité, géométrie discrète et intégralité".

1 Introduction

A *rectilinear path* P between two points p, q of the plane $\mathbb{R}^2$ is a path connecting p and q and consisting of only horizontal and vertical line segments. More generally, a *rectilinear network* $N = (V, E)$ consists of a finite set V of points of $\mathbb{R}^2$ (the vertices of N) and of a finite set of horizontal and vertical segments connecting pairs of points of V (the edges of N). The *length* $l(P)$ (or $l(N)$) of a rectilinear path P (or of a rectilinear network N) is the sum of lengths of its edges. Analogously, the *length* $l(N)$ of a rectilinear network N is the sum of lengths of its edges. The l_1-distance between two points $p = (p^x, p^y)$ and $q = (q^x, q^y)$ in the plane $\mathbb{R}^2$ is $d(p, q) := ||p - q||_1 = |p^x - q^x| + |p^y - q^y|$. An l_1-*path* between two points $p, q \in \mathbb{R}^2$ is a rectilinear path connecting p, q and having length $d(p, q)$.

Given a set $T = \{t_1, \ldots, t_n\}$ of n points (*terminals*) in the plane, a *Manhattan network* [5] on T is a rectilinear network $N(T) = (V, E)$ such that $T \subseteq V$ and for every pair of points in T, the network $N(T)$ contains an l_1-path between them. A *minimum Manhattan network* on T is a Manhattan network of minimum possible length and the Minimum Manhattan Network problem (*MMN problem*) is to find such a network.

The minimum Manhattan network problem has been introduced by Gudmundsson, Levcopoulos, and Narasimhan [5] in connection with the construction of sparse geometric spanners. Given a set T of n points in the plane endowed with a norm $\|\cdot\|$, and a real number $t \geq 1$, a geometric network N is a t-*spanner* for T if for each pair of points $p, q \in T$, there exists a pq-path in N of length at most t times the distance $\|p - q\|$ between p and q. In the Euclidian plane (and more generally, for l_p-planes with $p \geq 2$), the line segment is the unique shortest path between two endpoints, and therefore the unique 1-spanner of T is the trivial complete graph on T. On the other hand, if the unit ball of the norm is a polygon (in particular, for l_1 and l_∞), the points are connected by several shortest paths, therefore the problem of finding the sparsest 1-spanner becomes non trivial. In this connection, minimum Manhattan networks are precisely the optimal 1-spanners for the l_1 (or l_∞) plane. Sparse geometric spanners have applications in VLSI circuit design, network design, distributed algorithms and other areas, see for example the survey [6] and the book [9]. Finally, Lam, Alexandersson, and Pachter [8] suggested the use of minimum Manhattan networks to design efficient search spaces for pair hidden Markov model (PHMM) alignment algorithms.

The complexity status of the minimum Manhattan network problem is unknown. Gudmundsson et al. [5] proposed an $O(n^3)$-time 4-approximation algorithm, and an $O(n \log n)$-time 8-approximation algorithm. They also conjectured that there exists a 2-approximation algorithm for this problem. Kato, Imai, and Asano [7] presented a 2-approximation algorithm, however, their correctness proof is incomplete. Following [7], Benkert, Shirabe, Wolff, and Widmann [1, 2, 3] described an $O(n \log n)$-time 3-approximation algorithm and presented a mixed-integer programming formulation of the MMN problem. Notice that all four mentioned algorithms are geometric and some of them employ results from computational geometry. Nouioua [10] presented another $O(n \log n)$-time 3-approximation algorithm based on the primal-dual method from linear programming. In this paper we present a rounding method applied to the optimal solution of the flow based linear program described in [1, 10] which leads to a 2-approximation algorithm for the minimum Manhattan network problem (for approximation algorithms based on rounding techniques, see the book by Vazirani [15]). For this, we define two subsets of pairs of terminals, called *strips* and *staircases*, and for each of them, we describe a specific rounding procedure. Each rounded up edge is paid by a group of parallel edges which together support at least one-half unit of fractional flow. Finally, we prove that a rectilinear network containing l_1-paths between all the pairs belonging to strips and staircases is a Manhattan network and thus, we end-up with an integer feasible solution whose cost is at most twice the fractional optimum. After the revised version of our paper had been submitted, Seibert and Unger [11] announced a 1.5-approximation algorithm for the MMN problem, however the conference format of their paper does not permit to understand the description of the algorithm and to check its claimed performance guarantee.

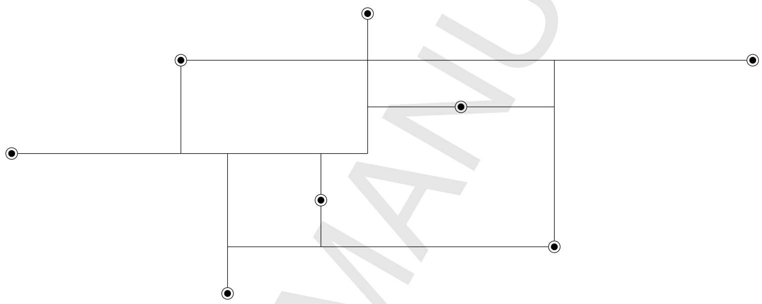


Figure 1: A minimum Manhattan network

2 Properties and LP-formulation

In this section, we present several properties of minimum Manhattan networks. First, we define our notation. For a point $p = (p^x, p^y)$ of $\mathbb{R}^2$, we denote by $\mathcal{Q}_1(p)$ the first (closed) quadrant with respect to the origin p, i.e., $\mathcal{Q}_1(p) = \{q \in \mathbb{R}^2 : q^x \geq p^x, q^y \geq p^y\}$. The remaining closed quadrants are labelled $\mathcal{Q}_2(p), \mathcal{Q}_3(p)$, and $\mathcal{Q}_4(p)$ in counterclockwise order around p. Denote by $\mathcal{Q}_1^\circ, \mathcal{Q}_2^\circ(p), \mathcal{Q}_3^\circ(p)$, and $\mathcal{Q}_4^\circ(p)$ the open quadrants with origin p; for example, $\mathcal{Q}_1^\circ(p) = \{q \in \mathbb{R}^2 : q^x > p^x, q^y > p^y\}$. Denote by $[p, q]$ the line segment having p and q as end-points. The set of all points of $\mathbb{R}^2$ lying on l_1-paths between p and q constitute the smallest axis-parallel closed rectangle $R(p, q)$ containing the points p, q. For two terminals $t_i, t_j \in T$, let $R_{i,j} := R(t_i, t_j)$. This rectangle is *degenerate* if t_i and t_j have the same x- or y-coordinate. We say that $R_{i,j}$ is an *empty rectangle* if $R_{i,j} \cap T = \{t_i, t_j\}$. The *complete grid* is obtained by drawing in the smallest axis-parallel rectangle containing the set T a horizontal segment and a vertical segment through every terminal which span the entire length and width of the rectangle. The following result can be easily proven using standard methods for establishing Hanan grid-type results.

Lemma 2.1 *[5, 17] The complete grid contains at least one minimum Manhattan network on T.*

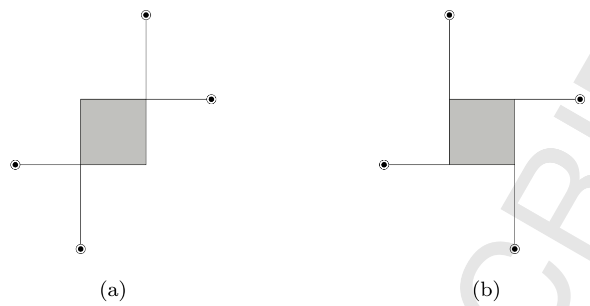


Figure 2: Pareto envelope of four points

For any two points $p, q \in \mathbb{R}^2$, we say that q *dominates* p if and only if (i) for each $t_i \in T$, $d(q, t_i) \leq d(p, t_i)$, and (ii) for at least one $t_j \in T$, $d(q, t_j) < d(p, t_j)$. A point $p \in \mathbb{R}^2$ is said to be an *efficient* point of T [4, 16] if there does not exist any point $q \in \mathbb{R}^2$ that dominates p. Denote the set of all efficient points by $\mathcal{P}$, called the *Pareto envelope* of T. An optimal $O(n \log n)$-time algorithm to compute the Pareto envelope of n points in the l_1-plane is presented in [4]. Its correctness uses the following characterization of $\mathcal{P}$ presented in [4]: $\mathcal{P} = \bigcap_{i=1}^{n} \bigcup_{j=1}^{n} R(t_i, t_j)$. For other properties of $\mathcal{P}$ and an $O(n^2)$-time algorithm see also [16]. In particular, it is known that $\mathcal{P}$ is *ortho-convex*, i.e. the intersection of $\mathcal{P}$ with any vertical or horizontal line is convex, and that every two points of $\mathcal{P}$ can be joined in $\mathcal{P}$ by an l_1-path; Fig. 2 presents two generic forms of the Pareto envelope of four points. Denote by $\Gamma = (V, E)$ the part of the complete grid contained in the Pareto envelope $\mathcal{P}$.

Lemma 2.2 *The graph Γ contains at least one minimum Manhattan network on T.*

Proof. By Lemma 2.1, the complete grid contains a minimum Manhattan network N on T. Denote by $\mathcal{R}(N)$ the union of all inner faces of N. Suppose that N is selected so that the number of faces of the complete grid which belong to $\mathcal{R}(N) \setminus \mathcal{P}$ is minimized. If all vertices of N belong to $\mathcal{P}$, then all edges of N also belong to $\mathcal{P}$ (and therefore to Γ), because the Pareto envelope $\mathcal{P}$ is ortho-convex. Therefore, assume by way of contradiction that u_0 is a vertex of N located outside Γ (and $\mathcal{P}$). Since $\mathcal{P} = \bigcap_{i=1}^{n} \bigcup_{j=1}^{n} R(t_i, t_j)$ and $u_0 \notin \mathcal{P}$, there exists a terminal t_i such that $u_0 \notin \bigcup_{j=1}^{n} R(t_i, t_j)$. Suppose without loss of generality that $u_0^x \leq t_i^x$ and $t_i^y \leq u_0^y$. Then, the fact that $u_0 \notin \bigcup_{j=1}^{n} R(t_i, t_j)$ implies that the closed quadrant $\mathcal{Q}_2(u_0)$ does not contain any terminal of T. Therefore $\mathcal{Q}_2(u_0) \bigcap (\bigcup_{j=1}^{n} R(t_i, t_j)) = \emptyset$, yielding $\mathcal{Q}_2(u_0) \cap \mathcal{P} = \emptyset$.

Let u be the highest vertex of $N \setminus \Gamma$ belonging to $\mathcal{Q}_2(u_0)$ (such a vertex u always exists because $u_0 \in \mathcal{Q}_2(u_0)$). If there are several such vertices, then we break ties by taking the leftmost one. The closed quadrant $\mathcal{Q}_2(u)$ does not contain terminals or Pareto points because $\mathcal{Q}_2(u) \subseteq \mathcal{Q}_2(u_0)$. Since u is a vertex of the complete grid, the horizontal line l_h passing via u contains some terminal t' (necessarily located on the right of u). Analogously, the vertical line l_v passing via u contains some terminal t'' (necessarily located below u). Since $\mathcal{Q}_2(u) \cap \Gamma = \emptyset$, from the choice of u we infer that $\mathcal{Q}_2(u) \cap N = \{u\}$. Therefore the vertex u has exactly two

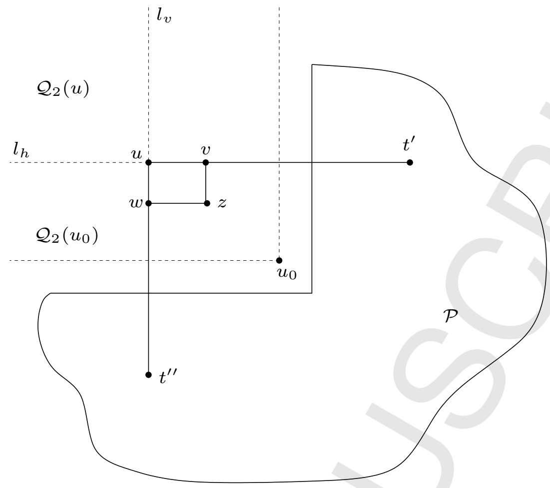
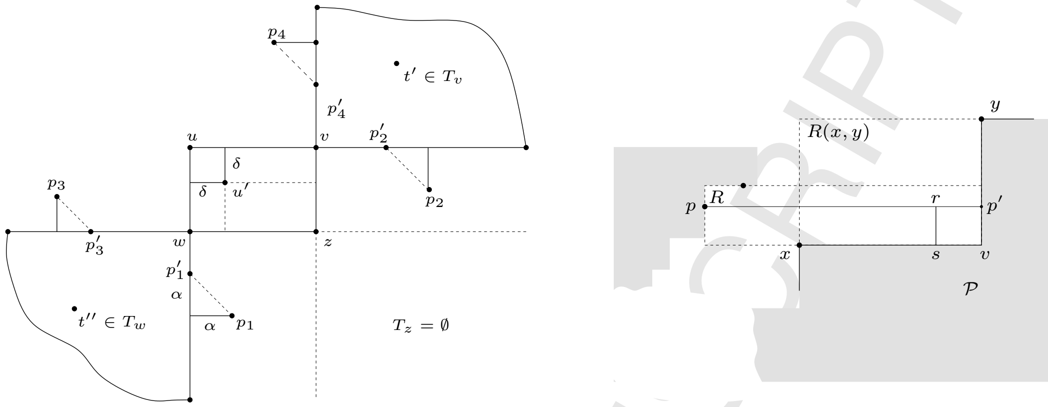


Figure 3: For the proof of Lemma 2.2

neighbors v and w in N : v is to the right of u and w is below u (see Fig. 3). Since $u \in N \setminus \Gamma$, the edges uv and uw do not belong to Γ (however, one or both v, w maybe vertices of this graph). Pick the point $z = (v^x, w^y)$. Since v and w are vertices of the complete grid, z is also a vertex of this grid. Denote by N' the rectilinear network (of length at most $l(N)$) which is obtained from N by removing the edges uv and uw and adding the vertical edge vz and the horizontal edge wz. We claim that N' is a Manhattan network on T. Indeed, since all points of T are located inside or on the boundary of $\mathcal{P}$, the removed vertex u is not a terminal. Additionally, since the degree of u in N is two, any l_1-path L connecting two terminals and passing via u uses both edges uv and uw. Therefore the path L' obtained from L by replacing the edges vu and uw of N by the edges vz and zw of N' is an l_1-path between the same pair of terminals. All this shows that N' is also a minimum Manhattan network contained in the complete grid. Since the rectangle $uvzw$ is a face of the complete grid contained in $\mathcal{R}(N) \setminus \mathcal{P}$ but not in $\mathcal{R}(N')$, we get a contradiction with the choice of N. □

The Pareto envelope $\mathcal{P}$, being ortho-convex, is a union of ortho-convex (possibly degenerate) rectilinear polygons, called *blocks*, glued together along the cut-vertices of the graph Γ. The blocks of $\mathcal{P}$ are obtained from the (graph-theoretic) blocks of the planar graph Γ by replacing every rectilinear face by a rectangle and every cut-edge by a segment. Denote by C the set of cut-vertices of Γ. We call a block $\mathcal{B}$ *trivial* if $\mathcal{B}$ is a single rectangular face of Γ such that $\mathcal{B} \cap (T \cup C)$ consists of two opposite corners of $\mathcal{B}$; the rectangular block from Fig. 2(a) is trivial. Notice that a block containing at least three vertices from $T \cup C$ is necessarily non-trivial. Denote by $\partial\mathcal{B}$ the boundary of a block $\mathcal{B}$ (if $\mathcal{B}$ is a cut-edge, then $\partial\mathcal{B}$ is this edge itself). The boundary $\partial\mathcal{P}$ of the Pareto envelope $\mathcal{P}$ is the union of the boundaries of the blocks of $\mathcal{P}$, and the non-trivial boundary $\partial^{\circ}\mathcal{P}$ of $\mathcal{P}$ is the union of the boundaries of its non-trivial blocks.



(a) (b)

Figure 4: For the proof of Lemma 2.3

Lemma 2.3 *Let $\mathcal{B}$ be a non-trivial block of the Pareto envelope $\mathcal{P}$ of T. Then (i) every convex vertex u of $\mathcal{B}$ is either a terminal of T or a cut-vertex of Γ and (ii) the subpath $P(x, y)$ of $\partial\mathcal{B}$ between two consecutive convex vertices x, y of $\partial\mathcal{B}$ is the unique l_1-path connecting x and y inside $\mathcal{P}$.*

Proof. In order to prove (i), suppose by way of contradiction that u is a convex vertex of $\mathcal{B}$ which is neither a terminal nor a cut-vertex. Then u has exactly two neighbors v, w in Γ, both belonging to $\mathcal{B}$. Suppose without loss of generality that v is to the right of u and w is below u. Let $z = (v^x, w^y)$. Since u is a convex vertex of $\mathcal{B}$, the face $uvzw$ of the complete grid belongs to $\mathcal{B}$. Pick $0 < \delta \leq \min\{d(u, v), d(u, w)\}$ and let $u' = (u^x + \delta, u^y - \delta)$. Now, we partition T into three subsets: T_z consists of all terminals $t \in T$ which can be connected to u using an l_1-path passing via z, T_v consists of all terminals $t \in T \setminus T_z$ such that any l_1-path connecting t to u passes via v, and, finally, T_w consists of all terminals $t \in T \setminus T_z$ such that any l_1-path connecting t to u passes via w (see Fig. 4(a)); the definition of v, w implies that there are no terminals in the open vertical half-band bounded by the edge uv and the vertical lines passing through u and v, and similarly, there are no terminals in the open horizontal half-band bounded by the edge uw and the horizontal lines passing through u and w. Notice also that the point u' has the same distance as u to any terminal in $T_v \cup T_w$ and u' is closer than u to any terminal in T_z. Since u' cannot dominate u because $u \in \mathcal{P}$, we conclude that $T_z = \emptyset$. Therefore any terminal t of T either is located above and to the right of v (i.e., $t \in T_v$) or is located below and to the left of w (i.e., $t \in T_w$). Since w is a vertex of Γ, there exists a terminal of T_w on the horizontal half-line $\{(q^x, w^y) : q^x \leq w^x\}$. In this case, it can be easily seen that every point p_1 lying in the open quadrant $\mathcal{Q}_4^\circ(w)$ is dominated by the point $p_1' = (p_1^x - \alpha, p_1^y + \alpha)$ where $\alpha = \min\{p_1^x - w^x, w^y - p_1^y\}$ (see Fig. 4(a)). Analogously, it can be shown that every point p lying in the three open quadrants $\mathcal{Q}_4^\circ(v)$, $\mathcal{Q}_2^\circ(w)$, and $\mathcal{Q}_2^\circ(v)$ is dominated by an appropriately chosen point p' (see Fig. 4(a)). Therefore $\mathcal{P}$ consists of a rectangular block $uvzw$, one or several blocks located above and to the right of v, and one or several blocks located below and to the left of w. This shows that $\mathcal{B}$ coincides with the rectangle $uvzw$. Since v, w are cut-vertices, the set T_z is empty, and there are no terminals in the open edges uv and uw, we conclude that $\mathcal{B} \cap (T \cup C) = \{v, w\}$. Thus $\mathcal{B}$ is a trivial block, contrary to our assumption. This establishes the property (i).

To show (ii), note that the path $P(x, y)$ between two consecutive convex vertices of a non-trivial block $\mathcal{B}$ is either a single vertical or horizontal segment or it consists of two segments, one vertical and another horizontal. In the first case, $P(x, y) = R(x, y)$ and we are done. In the second case, the segments $[x, v]$ and $[v, y]$ constituting $P(x, y)$ are incident sides of the rectangle $R(x, y)$; let $[x, v]$ be the bottom side and $[v, y]$ be the rightmost side of $R(x, y)$. We assert that $\mathcal{P} \cap R(x, y) = P(x, y)$. Indeed, since v is a concave vertex of $\mathcal{B}$, there exists a small rectangle R_0 contained in $R(x, y)$, having v as a corner, and such that $R_0 \cap \mathcal{P} \subseteq P(x, y)$. Let R be a maximal by inclusion rectangle containing R_0 and such that the interior of R is disjoint from $\mathcal{P}$ (notice that R may be unbounded). If $R(x, y) \subseteq R$, then we are done. Otherwise, there exists a point $p \in \partial R \cap \mathcal{P}$ such that the horizontal or the vertical line passing via p intersects $P(x, y)$. Suppose without loss of generality that the horizontal line passing via p intersects $[v, y]$ at a point p' (see Fig. 4(b)). Since $\mathcal{P}$ is ortho-convex, the segment $[p, p']$ belongs to $\mathcal{P}$, yielding that all vertical segments having one end on $[p', p]$ and another end on $[v, x]$ belong to $\mathcal{P}$, contrary to the assumption that the interior of R is disjoint from $\mathcal{P}$. This contradiction shows that $\mathcal{P} \cap R(x, y) = P(x, y)$, thus $P(x, y)$ is the unique l_1-path between x and y in the Pareto envelope $\mathcal{P}$. □

Lemma 2.4 *Any Manhattan network on T contained in Γ is a Manhattan network on $T \cup C$. In particular, the edges of $\partial^\circ \mathcal{P}$ belong to any minimum Manhattan network on T located inside Γ.*

Proof. Pick a cut-vertex v of Γ. First, we will show that the union $\mathcal{P}'$ of all blocks induced by each connected component of $\Gamma \setminus \{v\}$ contains at least one terminal of T. Suppose by way of contradiction that $\mathcal{P}' \cap T = \emptyset$. We assert that any point $p \in \mathcal{P}'$ is dominated by v. Since $\mathcal{P}$ is ortho-convex, p and any terminal t_i can be joined in $\mathcal{P}$ by an l_1-path. Since p and t_i belong to different connected components of $\mathcal{P} \setminus \{v\}$, this path necessarily passes via v, thus $d(v, t_i) < d(p, t_i)$ for any $t_i \in T$, showing that p is dominated by v. This contradicts the assumption that p belongs to $\mathcal{P}$.

From the previous assertion we immediately conclude that every cut-vertex and every cut-edge of Γ belong to all minimum Manhattan networks N on T which are subgraphs of Γ. Pick two cut-vertices x and y of Γ. Let A_x be a connected component of $\Gamma \setminus \{x\}$ not containing y and let A_y be a connected component of $\Gamma \setminus \{y\}$ not containing x. By the previous assertion, there exists a terminal t_i in A_x and a terminal t_j in A_y. Any l_1-path connecting t_i and t_j in the graph Γ passes via the vertices x and y, therefore x and y are connected in N by an l_1-path. In the same way, one can show that any terminal and any cut-vertex are joined by an l_1-path in N. Hence N is a Manhattan network on $T \cup C$ in Γ. This also shows that if x and y are consecutive convex vertices of a non-trivial block, then the unique l_1-path $P(x, y)$ connecting x and y in $\mathcal{P}$ belongs to all Manhattan networks on T and on $T \cup C$ located inside

Γ. To conclude the proof, notice that $x, y \in T \cup C$ by Lemma 2.3 and that the boundary of any non-trivial block $\mathcal{B}$ is covered by the l_1-paths $P(x, y)$ between consecutive convex vertices x, y of $\partial\mathcal{B}$. This shows that the edges of $\partial^\circ\mathcal{P}$ belong to all Manhattan networks on T contained in Γ. □

By this result, any minimum Manhattan network on $T \cup C$ contained in Γ is a minimum Manhattan network on T; thus, in order to solve the MMN problem on T, it suffices to add to the set of terminals T the set C of cut-points of the Pareto envelope and to solve a MMN problem on each non-trivial block $\mathcal{B}$ of $\mathcal{P}$ with respect to new and old terminals located inside or on its boundary. On the other hand, for each trivial block $\mathcal{B}$, the MMN problem is trivial: if $\mathcal{B} \cap (T \cup C) = \{x, y\}$, then there exist two terminals t_i, t_j of T such that x and y belong to all l_1-paths of Γ connecting t_i and t_j, whence any Manhattan network in Γ will employ an l_1-path between x and y; therefore, for each trivial block $\mathcal{B}$, it suffices to include in the resulting Manhattan network one of the l_1-paths between x and y consisting of two incident sides of the rectangle $\mathcal{B}$. Due to this decomposition of the MMN problem into smaller subproblems, further in this paper we can assume without loss of generality that

– $\mathcal{B}$ is a non-trivial block (ortho-convex polygon) of $\mathcal{P}$ and

– $T^+ := (T \cup C) \cap \mathcal{B}$.

With some abuse of notation, we will denote by $\Gamma = (V, E)$ the part of the complete grid contained in $\mathcal{B}$ and by $t_1, \ldots, t_n$ the terminals of T^+. Moreover, let $\partial\mathcal{B}$ denote the boundary of $\mathcal{B}$.

Two edges of the graph Γ are called *twins* if they are opposite edges of a rectangular face of Γ. Two edges e, f of Γ are called *congruent* if there exists a sequence $e = e_1, e_2, \ldots, e_{m+1} = f$ of edges such that, for $i = 1, \ldots, m$, the edges e_i, e_{i+1} are twins. By definition, any edge e is congruent to itself and all edges congruent to e have the same length. Notice also that from the ortho-convexity of $\mathcal{B}$ follows that exactly two edges congruent to a given edge e belong to $\partial\mathcal{B}$.

We continue with the notion of a generating set introduced in [7] and used in approximation algorithms described in [1, 10]. A *generating set* is a subset F of pairs of terminals (or, more compactly, of their indices) with the property that a rectilinear network containing l_1-paths for all pairs in F is a Manhattan network on T^+. For example, $F_\varnothing$ consisting of all pairs i, j with $R_{i,j}$ empty is a generating set [7]. In the next section, we will describe a generating set which is a subset of $F_\varnothing$.

To give an LP-formulation of the minimum Manhattan network problem, let $\overrightarrow{F}$ be an arbitrary generating set whose pairs are ordered in an arbitrary way; for each ordered pair $(i, j) \in \overrightarrow{F}$, let $\Gamma_{i,j} := \Gamma \cap R(t_i, t_j)$ and set $\Gamma_{i,j} = (V_{i,j}, E_{i,j})$. Orient each network $\Gamma_{i,j}, (i, j) \in \overrightarrow{F}$ so that each l_1-path between t_i and t_j is oriented from t_i to t_j (notice that this orientation is not overall consistent in the sense that the same edge may be oriented in different ways in different networks to which it belongs). For a vertex $v \in V_{i,j}$ denote by $\Gamma^-_{i,j}(v)$ the oriented edges of $\Gamma_{i,j}$ entering v and by $\Gamma^+_{i,j}(v)$ the oriented edges of $\Gamma_{i,j}$ out of v (notice that $\Gamma^-_{i,j}(t_i) = \Gamma^+_{i,j}(t_j) = \emptyset$). We formulate the MMN problem as a cut covering problem using an exponential number of constraints, which we further convert into an equivalent formulation that employs only a polynomial number of variables and constraints. In both formulations, l_e will denote the length of an edge e of the network $\Gamma = (V, E)$ and x_e will be a 0-1 decision variable associated with e. A subset of edges C of $E_{i,j}$ is called a (t_i, t_j)-*cut* if every l_1-path between t_i and t_j in $\Gamma_{i,j}$ shares an edge with C. Let $\mathcal{C}_{i,j}$ denote the collection of all (t_i, t_j)-cuts and set $\mathcal{C} := \bigcup_{(i,j)\in\overrightarrow{F}} \mathcal{C}_{i,j}$. Then the minimum Manhattan networks can be viewed as the optimal solutions of the following integer linear program (the dual of the relaxation of this program is a packing problem of the cuts from $\mathcal{C}$):

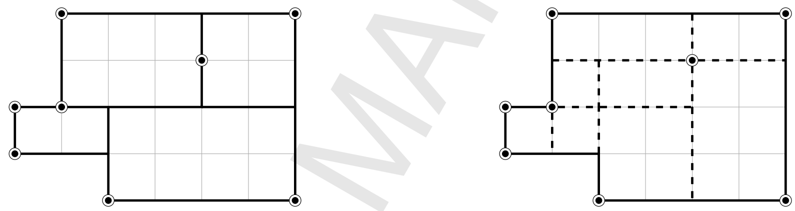
$$\begin{aligned} \text{minimize} \quad & \sum_{e\in E} l_e x_e && && (1)\\ \text{subject to} \quad & \sum_{e\in C} x_e \geq 1, && C \in \mathcal{C} \\ & x_e \in \{0,1\}, && e \in E \end{aligned}$$

Indeed, every Manhattan network N must contain at least one edge from every cut $C \in \mathcal{C}$, thus N yields a feasible solution of (1). Conversely, let $x_e, e \in E$, be a feasible solution for (1). Considering the x_es as capacities of the edges e of Γ, and applying the covering constraints and the Ford-Fulkerson's theorem to each network $\Gamma_{i,j}, (i,j) \in \overrightarrow{F}$, oriented as described above, we conclude the existence in $\Gamma_{i,j}$ of an integer (t_i, t_j)-flow of value 1, i.e., of an l_1-path between t_i and t_j. As a consequence, we obtain a Manhattan network of the same cost. This observation leads to the second integer programming formulation for the MMN problem (but this time, having a polynomial size). In addition to the variables x_e, we introduce a (flow) variable $f_e^{i,j}$ for each pair $(i,j) \in \overrightarrow{F}$ and each edge $e \in E_{i,j}$. We obtain the following integer program:

$$\begin{aligned} \text{minimize} \quad & \sum_{e\in E} l_e x_e && && (2)\\ \text{subject to} \quad & \sum_{e\in\Gamma^-_{i,j}(v)} f_e^{i,j} = \sum_{e\in\Gamma^+_{i,j}(v)} f_e^{i,j}, && (i,j) \in \overrightarrow{F}, v \in V_{i,j} \setminus \{t_i, t_j\} \\ & \sum_{e\in\Gamma^+_{i,j}(t_i)} f_e^{i,j} = 1, && (i,j) \in \overrightarrow{F} \\ & 0 \leq f_e^{i,j} \leq x_e, && (i,j) \in \overrightarrow{F}, \forall e \in E_{i,j} \\ & x_e \in \{0,1\}, && e \in E \end{aligned}$$

The first two sets of constraints ensure that $f_e^{ij}, e \in E_{i,j}$, is a flow of value 1 for each pair i, j. From the last two sets of constraints, we infer that $x_e = 1$ for all edges such that $f_e^{ij} > 0$. Since all (t_i, t_j)-paths of $\Gamma_{i,j}$ are l_1-paths, we conclude that the edges e with $x_e = 1$ define a Manhattan network.

Denote by $(1')$ and $(2')$ the LP-relaxations of (1) and (2) obtained by replacing the boolean constrains $x_e \in \{0,1\}$ by the linear constraints $x_e \geq 0$. (The constraint $x_e \leq 1$ was omitted because in any optimal solution, for each x_e, at least one constraint $f_e^{i,j} \leq x_e$ is tight and for all i', j', $f_e^{i',j'} \leq 1$ by the first and second constraints.) Since (2′) contains a polynomial number of variables and inequalities, it can be solved in strongly polynomial time using the algorithm of Tardos [13]. The x-part of any optimal solution $(\mathbf{x}, \mathbf{f})$ of (2′) is an optimal solution of (1′). It can be viewed as a "fractional Manhattan network" in the following sense. In the network $\Gamma_{i,j}$ endowed with capacities x_e, $e \in E_{i,j}$, for each pair $\{i,j\}$, there exists one or several l_1-paths carrying together a flow of total value 1. If the optimal solution $\mathbf{x}$ is integral, i.e. $x_e \in \{0,1\}$, $e \in E$, then every such flow uses a unique l_1-path and therefore $\mathbf{x}$ is the characteristic vector of an optimal Manhattan network. Unfortunately, this is not always the case; moreover there exist instances of the MMN problem for which the cost of an optimal (fractional) solution of (1′) or (2′) is smaller than the cost of an optimal (integer) solution of (1) or (2). Fig. 5 shows such an example ($x_e = 1$ for heavy edges and $x_e = \frac{1}{2}$ for dashed edges). Notice also that by Lemma 2.4 in any feasible solution of (1′) and (2′) for any edge $e \in \partial\mathcal{B}$ it holds that $x_e = 1$.



Integer optimum = 28 Fractional optimum = 27.5

Figure 5: Integrality gap

3 Strips and staircases

A degenerate empty rectangle $R_{i,j}$ is called a *degenerate vertical* or *horizontal strip.* A non-degenerate empty rectangle $R_{i,j}$ is called a *vertical strip* if there exists no terminal in T^+ with x-coordinate strictly between the x-coordinates of t_i and t_j and the intersection of $R_{i,j}$ with any degenerate vertical strip is either empty or one of the points t_i or t_j. The first condition means that the x-coordinates of t_i and t_j are consecutive entries of the sorted list of all distinct x-coordinates of the terminals. The second condition means that either t_i is the highest terminal among all terminals having the same x-coordinate as t_i and t_j is the lowest terminal among all terminals having the same x-coordinates as t_j, or, vice-versa, t_i is the lowest terminal and t_j is the highest terminal among respective terminals (in Fig. 6 the vertical strips are exactly the shaded rectangles; note that the rectangles $R_{j,k}$ and $R_{j,l}$ are not vertical strips). Therefore, the second condition ensures that there exists at most one vertical strip between two consecutive x-coordinates of the terminals, thus each subset of pairwise congruent horizontal edges of the grid Γ may intersect at most one vertical strip.

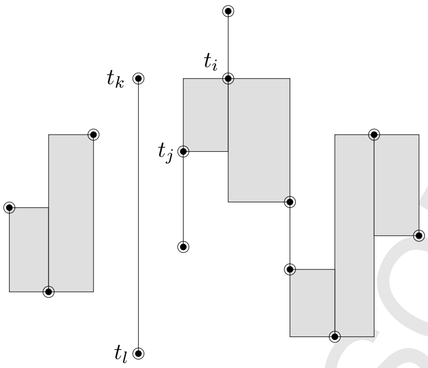


Figure 6: The shaded rectangles are all the vertical strips for the indicated set of terminals.

Analogously, a non-degenerate empty rectangle $R_{i,j}$ is called a *horizontal strip* if there exists no terminal in T^+ with y-coordinate between the y-coordinates of t_i and t_j and the intersection of horizontal sides of $R_{i,j}$ with any degenerate horizontal strip is either empty or one of the points t_i or t_j. The *sides* of a vertical (resp., horizontal) strip $R_{i,j}$ are the vertical (resp., horizontal) sides of $R_{i,j}$. Notice that two points t_i, t_j may define both a horizontal and a vertical strip. We say that the rectangles $R_{i,i'}$ and $R_{j,j'}$ (degenerate or not) form a *crossing configuration* if they intersect and the Pareto envelope of the points $t_i, t_{i'}, t_j, t_{j'}$ is of type (a); see Fig. 2. The importance of such configurations resides in the following property whose proof is straightforward:

Lemma 3.1 *If the rectangles $R_{i,i'}$ and $R_{j,j'}$ form a crossing configuration as in Fig. 7, then from the l_1-paths between t_i and $t_{i'}$ and between t_j and $t_{j'}$ one can derive an l_1-path connecting t_i and $t_{j'}$ and an l_1-path connecting $t_{i'}$ and t_j.*

For a crossing configuration defined by the strips $R_{i,i'}, R_{j,j'}$, denote by o and o' the cut points of the rectangular block of the Pareto envelope of $t_i, t_{i'}, t_j, t_{j'}$; assume that the four segments of this envelope connect o with t_i, t_j and o' with $t_{i'}, t_{j'}$. Additionally, suppose without loss of generality, that t_i and t_j belong to $\mathcal{Q}_1(o)$, i.e., to the first quadrant with respect to the origin o. Then $t_{i'}$ and $t_{j'}$ belong to $\mathcal{Q}_3(o')$. Denote by $T_{i,j}$ the set of all terminals $t_k \in (T^+ \setminus \{t_i, t_j\}) \cap \mathcal{Q}_1(o)$ such that (i) $R(t_k, o) \cap T^+ = \{t_k\}$ and (ii) the region $\{q \in \mathcal{Q}_2(o) : q^y \leq t_k^y\} \cup \{q \in \mathcal{Q}_4(o) : q^x \leq t_k^x\}$ does not contain any terminal of T^+. If $T_{i,j}$ is nonempty, then all its terminals belong to the rectangle $R_{i,j}$, more precisely, they are all located on a common shortest rectilinear path between t_i and t_j. When $T_{i,j} \neq \emptyset$, we define the *staircase* $\mathcal{S}_{i,j|i',j'}$ as the non-degenerate block of the Pareto envelope of the set $T_{i,j} \cup \{o, t_i, t_j\}$; see Fig. 7 for an illustration. The point o is called the *origin* of this staircase. Analogously one can define the set $T_{i',j'}$ and the staircase $\mathcal{S}_{i',j'|i,j}$ with origin o'. Two other types of staircases will be defined if t_i, t_j belong to the second quadrant with respect to o and $t_{i'}, t_{j'}$ belong to $\mathcal{Q}_4(o')$. In order to simplify the presentation, further we will prove all results under the assumption that the staircase is located in the first quadrant. By symmetry, all these results also hold for the

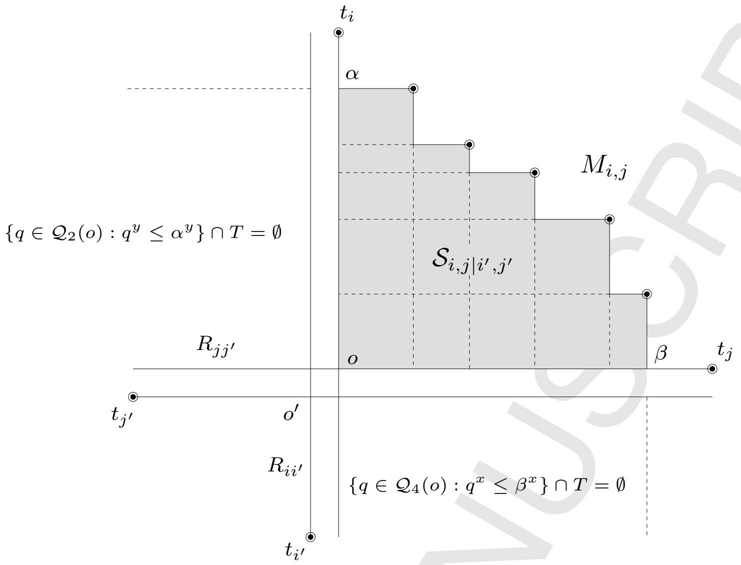


Figure 7: Staircase $\mathcal{S}_{i,j|i',j'}$

other types of staircases. (Notice that our staircases are different from the staircase polygons occurring in the algorithms from [5].)

Let α be the leftmost highest point of the staircase $\mathcal{S}_{i,j|i',j'}$ and let β be the rightmost lowest point of this staircase. By definition, $\mathcal{S}_{i,j|i',j'} \cap T^+ = T_{i,j}$. By the choice of $T_{i,j}$, there are no terminals of T^+ located in the regions $\{q \in \mathcal{Q}_2(o) : q^y \leq \alpha^y\}$ and $\{q \in \mathcal{Q}_4(o) : q^x \leq \beta^x\}$. The following result describes the layout of strips and staircases.

Lemma 3.2

(i) The interiors of any strip and any staircase are disjoint;

(ii) The interiors of any two staircases are disjoint.

Proof. If a strip traverses a staircase located in the first quadrant $\mathcal{Q}_1(o)$, then necessarily one of the terminals defining this strip must be located in the set $\{q \in \mathcal{Q}_2(o) : q^y \leq \alpha^y\} \cup \{q \in \mathcal{Q}_4(o) : q^x \leq \beta^x\}$, which is impossible. This establishes (i). To show (ii), suppose, for contradiction, that the interiors of two staircases $\mathcal{S}$ and $\mathcal{S}'$ intersect, and let p be a point belonging to this intersection. Up to symmetry, suppose that $\mathcal{S}$ is located in the first quadrant $\mathcal{Q}_1(o)$. Let $R(o', t_{k'})$ be a rectangle of $\mathcal{S}'$ whose interior contains p. In view of assertion (i) applied to the staircase $\mathcal{S}'$, the rectangle $R(o', t_{k'})$ cannot traverse the vertical and the horizontal strips defining $\mathcal{S}$. This and the fact that $R(o', t_{k'})$ does not contain any terminal in its interior show that $R(o', t_{k'})$ is entirely located in $\mathcal{Q}_1(o)$. Then the interior of $R(o', t_{k'}) \cap \mathcal{Q}_3(p)$ is a subset of the interior of $\mathcal{S}$. Let e_1 and e_2 be the edges of $R(o', t_{k'})$ intersected by the boundary of $\mathcal{Q}_3(p)$. Because $o \neq o'$ and both the edges of $R(o', t_{k'})$ that are incident on o' are incident on the strips defining the staircase $\mathcal{S}'$, assertion (i) implies that neither e_1 nor e_2 is incident on o' ; therefore, the common endpoint of e_1 and e_2 is $t_{k'}$. This,

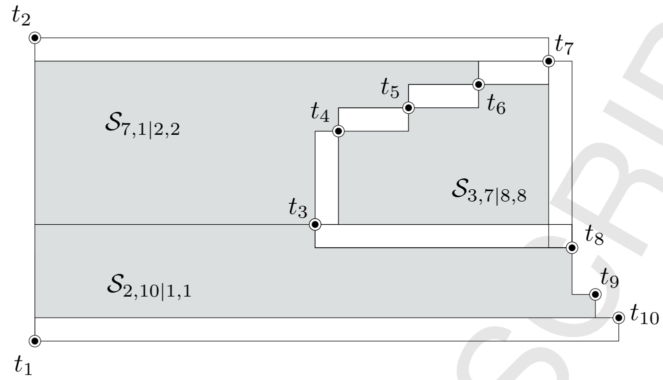


Figure 8: A sample set of terminals and their staircases

however, contradicts the fact that no terminal lies in the interior of a staircase $\mathcal{S}_{i,j|i',j'}$ or on its edges $o\alpha$ and $o\beta$ (see Fig. 7). $\square$

From Lemma 3.2 we infer that two staircases may intersect only on the boundary. In this case, the intersection is either a subset of vertices of both staircases (all these vertices are terminals) or a single edge (see Fig. 8 for a sample of terminals and their staircases). Therefore every edge of Γ either belongs to at most one staircase or to the boundary of exactly two staircases.

Let F' be the set of all pairs $\{i, j\}$ such that $R_{i,j}$ is a strip. Let F'' be the set of all pairs $\{i', k\}$ such that there exists a staircase $\mathcal{S}_{i,j|i',j'}$ such that t_k belongs to the set $T_{i,j}$. From the definition of strips and staircases immediately follows that $F' \cup F'' \subseteq F_{\varnothing}$.

Lemma 3.3 *The set $F := F' \cup F''$ is a generating set.*

Proof. Let N be a rectilinear network containing l_1-paths for all pairs in F. To prove that N is a Manhattan network on T^+, it suffices to establish that for an arbitrary pair $\{k, k'\} \in F_{\varnothing} \setminus F$, the terminals t_k and $t_{k'}$ are joined in N by an l_1-path. Assume without loss of generality that $t^x_{k'} \leq t^x_k$ and $t^y_{k'} \leq t^y_k$. Since $\{k, k'\} \in F_{\varnothing}$, the rectangle $R_{k,k'}$ is empty. The vertical and horizontal lines through the points t_k and $t_{k'}$ partition the plane into the rectangle $R_{k,k'}$, four open quadrants and four closed unbounded half-bands labelled counterclockwise $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$, and $\mathcal{H}_4$ (see Fig. 9). Consider the leftmost terminal t_{i_1} of $\mathcal{H}_1$, breaking ties by minimizing the y-coordinate (this terminal exists because $t_k \in \mathcal{H}_1$). Now, consider the rightmost terminal $t_{i'_1}$ of $\mathcal{H}_3$ such that $t^x_{i'_1} \leq t^x_{i_1}$, breaking ties by maximizing the y-coordinate (again this terminal exists because $t_{k'} \in \mathcal{H}_3$ and $t^x_{k'} \leq t^x_{i_1}$). By the choice of t_{i_1} and $t_{i'_1}$, the rectangle R_{i_1,i'_1} is the leftmost vertical strip crossing the rectangle $R_{k,k'}$. Analogously, by letting t_{i_2}, t_{j_1}, and t_{j_2} be the rightmost terminal of $\mathcal{H}_3$, the highest terminal of $\mathcal{H}_2$, and the lowest terminal of $\mathcal{H}_4$, respectively (minimizing the distance to $R_{k,k'}$ in case of ties), we obtain the rightmost vertical strip R_{i_2,i'_2}, the lowest horizontal strip R_{j_1,j'_1}, and the highest horizontal strip R_{j_2,j'_2} crossing the rectangle $R_{k,k'}$. Notice that the strips R_{j_2,j'_2} and R_{i_2,i'_2} as well as the strips R_{j_1,j'_1} and R_{i_1,i'_1} constitute crossing configurations.

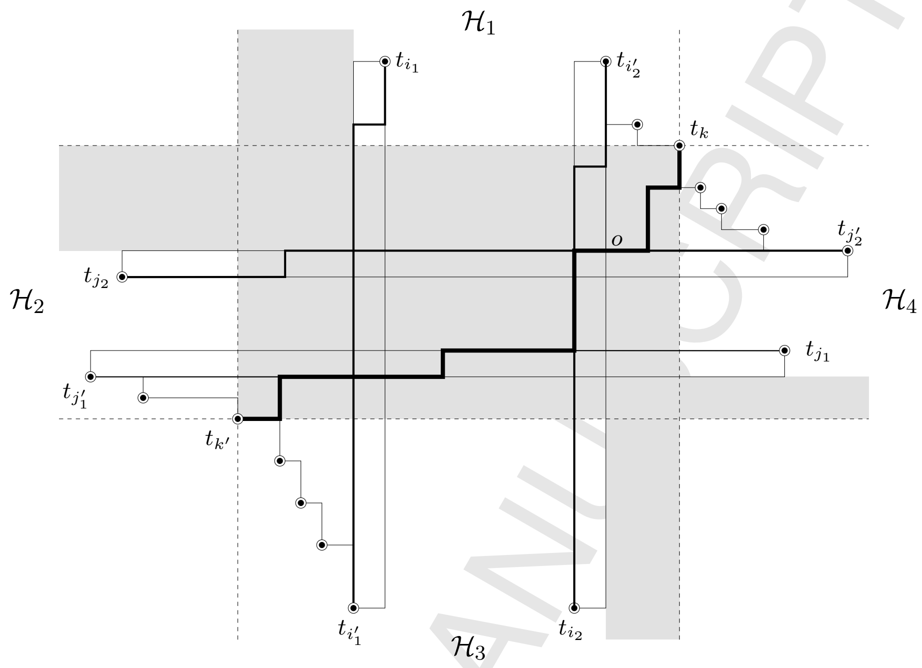


Figure 9: For the proof of Lemma 3.3

Now, we will prove that N contains an l_1-path between t_k and t_{i_2} and an l_1-path between $t_{k'}$ and t_{j_1}. We distinguish three cases. If $t_k = t_{i'_2}$, then $R_{i_2,k} = R_{i_2,i'_2}$ is a strip and thus $\{k, i_2\} \in F$. If $t_k = t_{j'_2}$, then the strips $R_{j_2,k}$ and R_{i_2,i'_2} form a crossing configuration. By Lemma 3.1, from the l_1-paths of N between t_{j_2} and t_k and between t_{i_2} and $t_{i'_2}$, we can derive an l_1-path between t_k and t_{i_2}. Finally, if $t_k \notin \{t_{i'_2}, t_{j'_2}\}$, we assert that the crossing configuration R_{i_2,i'_2} and R_{j_2,j'_2} defines a staircase $\mathcal{S}_{i'_2,j'_2|i_2,j_2}$ such that t_k belongs to $T_{i'_2,j'_2}$. Indeed, let o be the highest leftmost intersection point of the strips R_{i_2,i'_2} and R_{j_2,j'_2} (see Fig. 7). Since $R(t_k, o)$ is contained in the empty rectangle $R(t_k, t_{k'})$, we conclude that $R(t_k, o) \cap T^+ = \{t_k\}$. Moreover, by the choice of t_{i_2} and t_{j_2}, the unbounded half-bands $\{q \in \mathcal{H}_3 : q^x \geq t'_{i_2}\}$ and $\{q \in \mathcal{H}_2 : q^y \geq t_{j'_2}\}$ do not contain terminals (in Fig. 9, the shaded region does not contain terminals), thus establishing our assertion. This implies that $t_k \in T_{i'_2,j'_2}$, whence $\{k, i_2\} \in F$. Therefore, in all three cases the terminals t_k and t_{i_2} are connected in N by an l_1-path. Using a similar analysis, one can show that $t_{k'}$ and t_{j_1} are also connected in N by an l_1-path. By construction, the rectangles R_{k,i_2} and R_{k',j_1} form a crossing configuration and thus, by Lemma 3.1, there is an l_1-path of N between the terminals t_k and $t_{k'}$, concluding the proof. □

4 The rounding algorithm

Let $(\mathbf{x},\mathbf{f})=((x_e)_{e\in E}, (f_e^{i,j})_{e\in E,(i,j)\in F})$ be an optimal solution of the linear program (2′) (in general, this solution is not half-integral). The algorithm rounds up the solution $(\mathbf{x},\mathbf{f})$ in three phases. In **Phase 0**, we insert all edges of $\partial\mathcal{B}$ in the integer solution. In **Phase 1**, the rounding is performed inside every strip $R_{i,i'}$, in order to ensure the existence of an l_1-path $P_{i,i'}$ between the terminals t_i and $t_{i'}$. In **Phase 2**, an iterative rounding procedure is applied to each staircase.

Phase 1. Let $R_{i,i'}$ be a strip. If $R_{i,i'}$ is degenerate, then $[t_i, t_{i'}]$ is the unique l_1-path between t_i and $t_{i'}$, yielding $x_e = f_e^{i,i'} = 1$ for any edge $e \in [t_i, t_{i'}]$. If $R_{i,i'}$ is not degenerate, then any l_1-path in Γ between t_i and $t_{i'}$ has a simple form: it goes along the side of $R_{i,i'}$ containing t_i, then it makes a turn by following an edge of Γ traversing $R_{i,i'}$, which we call further a *switch* of $R_{i,i'}$, and continues its way on the side containing $t_{i'}$ until it reaches $t_{i'}$. It may happen that several such l_1-paths have been used by the fractional flow $f^{i,i'}$ between t_i and $t_{i'}$. However, since any pair e, e' of twins on opposite sides of the strip $R_{i,i'}$ constitutes a cut separating the terminals t_i and $t_{i'}$, the value of the $f^{i,i'}$-flow traversing this cut equals to 1, yielding $x_e + x_{e'} \geq f_e^{i,i'} + f_{e'}^{i,i'} \geq 1$, and therefore $\max\{x_e, x_{e'}\} \geq \frac{1}{2}$.

Let p be the vertex on the side of $R_{i,i'}$ containing t_i such that $x_e \geq \frac{1}{2}$ for every edge e of the segment $[t_i, p]$, and p is farthest away from t_i. Let pp' be the edge of Γ incident to p that traverses the strip $R_{i,i'}$. By the choice of p, the $(t_i, t_{i'})$-flow carried by the l_1-paths which make a turn before p or at the vertex p has total value $\geq \frac{1}{2}$. Since all these paths contain all edges e of the segment $[p', t_{i'}]$, we have $x_e \geq \frac{1}{2}$ for all such edges.

`Procedure RoundStrip`. For each strip $R_{i,i'}$, if $R_{i,i'}$ is degenerate, then include in the integer solution all edges of the segment $[t_i, t_{i'}]$, otherwise include in the integer solution the edges of the segments $[t_i, p]$ and $[p', t_{i'}]$ and the edge pp' (used as a switch of $R_{i,i'}$); in both cases, denote by $P_{i,i'}$ the resulting l_1-path between t_i and $t_{i'}$.

Phase 2. Let $\mathcal{S}_{i,i'|j,j'}$ be a staircase. Denote by ϕ the common point of the l_1-paths $P_{i,i'}$ and $P_{j,j'}$, which is closest to t_i (this point is a corner of the rectangular face of Γ containing the vertices o and o'). Let $P^+_{i,i'}$ and $P^+_{j,j'}$ be the sub-paths of $P_{i,i'}$ and $P_{j,j'}$ between ϕ and the terminals t_i and t_j, respectively. Denote also by $M_{i,j}$ the monotone boundary path of $\mathcal{S}_{i,j|i',j'}$ between α and β and passing via the terminals of $T_{i,j}$ (see Fig. 7). Now we slightly expand the staircase $\mathcal{S}_{i,i'|j,j'}$ by considering as $\mathcal{S}_{i,i'|j,j'}$ the region bounded by the paths $P^+_{i,i'}, P^+_{j,j'}$, and $M_{i,j}$ ($P^+_{i,i'}$ and $P^+_{j,j'}$ are not included in the staircase but $M_{i,j}$ and the terminals from the set $T_{i,j}$ are). Inside $\mathcal{S}_{i,i'|j,j'}$, any flow $f^{k,i'}$ (or $f^{k,j'}$), $t_k \in T_{i,j}$, may be split among several l_1-paths between t_k and $t_{i'}$. Any l_1-path carrying flow between these terminals intersects one of the paths $P^+_{i,i'}$ or $P^+_{j,j'}$, therefore the total $f^{k,i'}$-flow arriving at $P^+_{i,i'} \cup P^+_{j,j'}$ is equal to 1. (This flow can be directed to ϕ via the paths $P^+_{i,i'}$ and $P^+_{j,j'}$, and further, along the path $P_{i,i'}$, to the terminal $t_{i'}$). Therefore it remains to decide how to round up the flow $f^{k,i'}$ inside the expanded staircase $\mathcal{S}_{i,i'|j,j'}$, i.e., to decide which edges from this staircase to include in the integer solution. For this, notice that either the total $f^{k,i'}$-flow carried by the l_1-paths that arrive at $P^+_{i,i'}$ is at least $\frac{1}{2}$ or the total $f^{k,i'}$-flow on the l_1-paths that arrive at $P^+_{j,j'}$ is at least $\frac{1}{2}$.

`Procedure RoundStaircase`. For an expanded staircase $\mathcal{S}_{i,i'|j,j'}$ defined by the l_1-paths $P^+_{i,i'}$ and $P^+_{j,j'}$ and the monotone path $M_{i,j}$, find the lowest terminal $t_m \in T_{i,j}$ such that the $f^{m,i'}$-flow on the l_1-paths between t_m and $t_{i'}$ that arrive first at $P^+_{i,i'}$ is $\geq \frac{1}{2}$. If t_m exists, let t_s

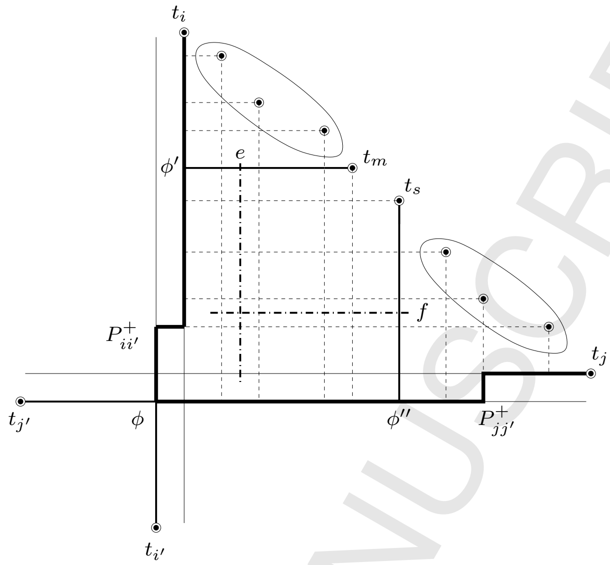


Figure 10: Procedure `RoundStaircase`

be the terminal of $M_{i,j}$ located immediately below t_m. Note that if the terminal t_s does not exist, then t_m is the lowest terminal of $M_{i,j}$. Otherwise, if t_m does not exist, then let t_s be the highest terminal of $M_{i,j}$. Note that in all cases at least one of the terminals t_m, t_s exists. By the choice of t_m, the $f^{m,i'}$-flow on the l_1-paths between t_m and $t_{i'}$ that arrive first at $P^+_{i,i'}$ is $\geq \frac{1}{2}$ and the $f^{s,i'}$-flow on the l_1-paths between t_s and $t_{i'}$ that arrive first at $P^+_{i,i'}$ is $< \frac{1}{2}$. Therefore, the $f^{s,i'}$-flow that arrives first at $P^+_{j,j'}$ is $> \frac{1}{2}$. The same property holds if t_m does not exist.

If t_m exists, then we include in the integer solution the edges of the horizontal segment $[t_m, \phi']$, where ϕ' is the intersection of the horizontal line passing via t_m with the path $P^+_{i,i'}$. Analogously, if t_s exists, then we include in the integer solution the edges of the vertical segment $[t_s, \phi'']$, where ϕ'' is the intersection of the vertical line passing via t_s with the path $P^+_{j,j'}$. If $T_{i,j}$ contains terminals located above the horizontal line (t_m, ϕ'), then recursively call `RoundStaircase` to the expanded sub-staircase defined by $[t_m, \phi']$, the sub-path of $P^+_{i,i'}$ between ϕ' and t_i, and the sub-path $M_{i,m}$ of the monotone path $M_{i,j}$ between t_m and α. Analogously, if $T_{i,j}$ contains terminals located to the right of the vertical line (t_s, ϕ''), then recursively call `RoundStaircase` to the expanded sub-staircase defined by $[t_s, \phi'']$, the sub-path of $P^+_{j,j'}$ comprised between ϕ'' and t_j, and the sub-path $M_{s,j}$ of the monotone path $M_{i,j}$ between t_s and β; see Fig. 10 for an illustration.

Approximation ratio. Let E_0 denote the edges of Γ which belong to the boundary of the Pareto envelope of T^+. Let E_1 be the set of all edges selected by the procedure `RoundStrip` and which do not belong to E_0, and let E_2 be the set of all edges selected by the recursive procedure `RoundStaircase` and which do not belong to $E_0 \cup E_1$. Denote by $N^* = (V^*, E_0 \cup E_1 \cup E_2)$ the resulting rectilinear network. From Lemma 3.3 and the rounding procedures presented above, we infer that N^* is a Manhattan network. Let $\mathbf{x}^*$ be the integer solution of (1) associated with N^*, i.e., $x^*_e = 1$ if $e \in E_0 \cup E_1 \cup E_2$ and $x^*_e = 0$ otherwise. We will show now that the length of the Manhattan network N^* is at most twice the cost of the optimal fractional solution $\mathbf{x}$ of $(1')$, i.e., that

$$\mathrm{cost}(\mathbf{x}^*) = \sum_{e \in E} l_e x^*_e \leq 2 \sum_{e \in E} l_e x_e = 2\mathrm{cost}(\mathbf{x}). \qquad (3)$$

To establish the inequality (3), to every edge $e \in E_1 \cup E_2$ we will assign a set E_e of edges congruent to e such that (i) $\sum_{e' \in E_e} x_{e'} \geq \frac{1}{2}$ and (ii) $E_e \cap E_f = \emptyset$ for any two edges $e, f \in E_1 \cup E_2$ (the edges of E_e will pay for the inclusion of the edge e in N^*). By Lemma 2.4 the equality $x_e = x^*_e = 1$ holds for every edge $e \in E_0$, thus every such edge e can pay one half of x_e for itself. The other half of x_e can be recycled to pay for an edge from $E_1 \cup E_2$, namely it will be used to pay for some switch.

First pick an edge $e \in E_1$, say $e \in P_{i,i'}$ for a strip $R_{i,i'}$. If e belongs to a side of this strip, then $x_e \geq \frac{1}{2}$, and in this case we can set $E_e := \{e\}$. Now, if e is the switch of $R_{i,i'}$, then E_e consists of one of the two edges of $\partial\mathcal{B}$ congruent to e. From the definition of a strip one concludes that no other switch can be congruent to these edges of $\partial\mathcal{B}$. Therefore each edge of ∂P may appear in at most one set E_e for a switch e.

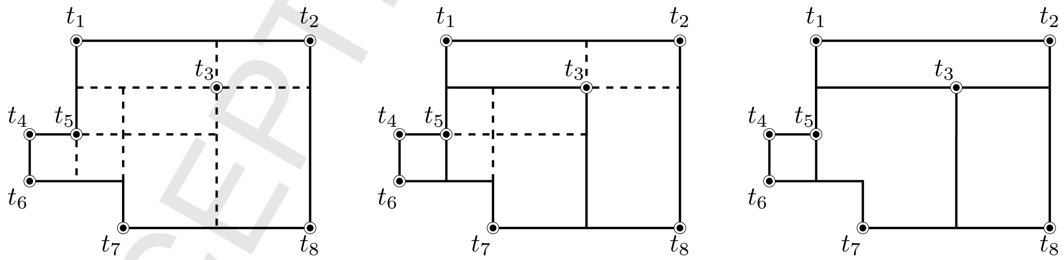
Finally suppose that $e \in E_2$; suppose that e belongs to the expanded staircase $\mathcal{S}_{i,i'|j,j'}$. If e belongs to the segment $[t_m, \phi']$, then E_e consists of e and all edges of $\mathcal{S}_{i,i'|j,j'}$ congruent to e and located below e; see Fig. 10. Since every l_1-path between t_m and $t_{i'}$ intersecting the path $P^+_{i,i'}$ contains an edge of E_e, we infer that the value of the $f^{m,i'}$-flow traversing the set E_e is at least $\frac{1}{2}$, therefore $\sum_{e' \in E_e} x_{e'} \geq \frac{1}{2}$, thus establishing (i). Analogously, if f is an edge of the vertical segment $[t_s, \phi'']$, then E_f consists of f and all edges of $\mathcal{S}_{i,i'|j,j'}$ congruent to f and located to the left of f. Obviously, $E_e \cap E_f = \emptyset$. Since E_e and E_f belong to the region of $\mathcal{S}_{i,i'|j,j'}$ delimited by the segments $[t_m, \phi']$ and $[t_s, \phi'']$ and the recursive calls of the procedure `RoundStaircase` concern the staircases disjoint from this region, we deduce that E_e and E_f are disjoint from the sets $E_{e'}$ for all edges e' selected by the recursive calls of `RoundStaircase` to the staircase $\mathcal{S}_{i,i'|j,j'}$. Now suppose that two distinct edges e and f belong to different staircases $\mathcal{S}_{i,i'|j,j'}$ and $\mathcal{S}_{k,k'|l,l'}$, respectively. Since E_e consists of inner edges of $\mathcal{S}_{i,i'|j,j'}$ except possibly e, and E_f consists of inner edges of $\mathcal{S}_{k,k'|l,l'}$ except possibly f, and the interiors of $\mathcal{S}_{i,i'|j,j'}$ and $\mathcal{S}_{k,k'|l,l'}$ are disjoint as noticed above, we conclude that E_e and E_f are disjoint as well. Finally, since there are no terminals of T^+ located below or to the left of the staircase $\mathcal{S}_{i,i'|j,j'}$, no strip traverses this staircase (a strip intersecting $\mathcal{S}_{i,i'|j,j'}$ either coincides with $R_{i,i'}$ and $R_{j,j'}$, or intersects the staircase along segments of the boundary path $M_{i,j}$). Therefore, no edge from E_1 can be assigned to a set E_e for some $e \in E_2 \cap \mathcal{S}_{i,i'|j,j'}$, thus establishing (ii) and the desired inequality (3). Now, since the cost of an optimal solution of the MMN problem is at least the cost of an optimal fractional solution, we are in position to formulate our main result:

Theorem 4.1 *The rounding algorithm described in this section achieves an approximation guarantee of 2 for the minimum Manhattan network problem.*

Remark 1. The running time of our rounding algorithm is dominated by the time taken to solve the linear program (2′). The number of variables $f_e^{i,j}$ introduced for a given edge e is equal to the number of rectangles $R_{i,j}$, $\{i,j\} \in F$, to which e belongs. Since every edge e belongs to at most three strips and to at most two staircases, the number of rectangles $R_{ij}, \{i,j\} \in F$, to which e belongs, is proportional to the number of convex vertices of the staircase, i.e., $O(n)$. Thus the $O(n^2)$ edges of Γ yield $O(n^3)$ variables $f_e^{i,j}$. A similar analysis shows that the number of constraints is also $O(n^3)$. Therefore, the linear program (2′) can be solved in strongly polynomial time by using the algorithm of Tardos [13].

Remark 2. Given a staircase $\mathcal{S}_{i,i'|j,j'}$ and the paths $P^+_{i,i'}$ and $P^+_{j,j'}$, the problem of constructing a minimum rectilinear network such that every terminal of $T_{i,j}$ can be connected by an l_1-path to $P^+_{i,i'} \cup P^+_{j,j'}$ can be solved in polynomial time using dynamic programming (for example, by adapting the algorithm from [12] for the Rectilinear Steiner Arborescence problem on staircases). However, we do not know how to prove the optimality of this solution via linear programming. Furthermore, we do not have examples of staircases having an integrality gap in (1′).

Remark 3. Fig. 11 illustrates the run of the algorithm on the example with 8 terminals given in Fig. 5 (recall that this is one of the smallest instances having an integrality gap). The edges of $\partial\mathcal{P}$ give rise to l_1-paths between all pairs of terminals defining strips except $\{t_3, t_5\}, \{t_5, t_7\}$, and $\{t_3, t_7\}$. To satisfy these remaining pairs, we round up the optimal fractional solution shown in Fig. 5 and Fig. 11(a). Phase 1 outputs three new segments, two incident to t_3 and one incident to t_5. The rectilinear network returned after Phases 0 and 1 and drawn in Fig. 11(b) satisfies all pairs of the generating set F except $\{t_2, t_3\}$. This pair belongs to the staircase $\mathcal{S}_{1,2|8,2}$. Phase 2 outputs a single segment incident to t_3. The resulting Manhattan network is given in Fig. 11(c). Its length is 29, while an optimal Manhattan network has length 28.



(a) Fractional optimum (b) After Phase 1 (c) Final solution

Figure 11: Example

5 Conclusions and perspectives

In this paper, we presented a simple rounding algorithm for the minimum Manhattan network problem and we established that the length of the Manhattan network returned by this algorithm is at most twice the cost of the optimal fractional solution (and consequently, at most twice the cost of the optimal solution) of the MMN problem. Nevertheless, experiments show that the ratio between the costs of the solution returned by our algorithm and the optimal solution of the linear programs (1′) and (2′) is much better than 2. We do not know the worst integrality gap of (1) (the worst gap obtained by computer experiences is about 1.087). Is this gap smaller or equal than 1.5? Does there exist a gap in the case where the terminals are the origin and the corners of a staircase?

The minimum Manhattan network problem can be compared with the (NP-complete) Rectilinear Steiner Arborescence problem (*RSA problem*) [12]. In this problem, given n terminals (lying in the first quadrant), one searches for a minimum rectilinear network containing an l_1-path between the origin of coordinates and each of the n terminals (clearly, such an optimal network will be a tree). The LP-formulation for the MMN problem can be viewed as a generalization of the LP-formulation of the RSA problem given in [14]. The paper [12] presents an instance of the RSA problem having an integrality gap. To the best of our knowledge, the worst integrality gap for this problem is also not known.

Consider now the following common generalization of the MMN and RSA problems which we call the *F-restricted MMN problem*: given a set of n terminals and a collection F of pairs of terminals, find a minimum rectilinear network $N_F(T)$, such that for every pair $\{t_i, t_j\} \in F$, the network $N_F(T)$ contains an l_1-path between t_i and t_j. If (T, F) is a complete graph, then we obtain the MMN problem and if (T, F) is a star, then we obtain the RSA problem. We can show that there exists a minimum F-restricted Manhattan network contained in the sub-grid of Γ generated by all empty rectangles. Using this grid, one can view (1) and (2) as integer programming formulations for the F-restricted MMN problem.

Notice that the rounding algorithm presented in our paper (as well as all other approximation algorithms for the MMN or RSA problems) cannot be extended in a direct way to get an approximation algorithm for the F-restricted MMN problem. Developing such an algorithm seems to be an interesting question. A simple example shows that the integrality gap in this case is at least 1.5: consider the four corners of a unit square as the set T of terminals, and let F consist of the two pairs of opposite corners of this square. Then, assigning $x_e = \frac{1}{2}$ for every side e of the square gives an optimal solution of (1′) having cost 2, while an optimal integer solution uses three edges of the square and has cost 3.

Acknowledgment. We are grateful to one of the anonymous referees for a very careful reading of all six versions of the manuscript, numerous corrections and insightful comments.

References

[1] M. Benkert, T. Shirabe, and A. Wolff, The minimum Manhattan network problem: approximations and exact solutions. In Proc. 20th European Workshop on Computational Geometry (EWCG'04), pp. 209–212, Sevilla, 24-26 March 2004.

[2] M. Benkert, A. Wolff, and F. Widmann, The minimum Manhattan network problem: a fast factor-3 approximation. In Proc. 8th Japanese Conf. on Discrete and Computational Geometry (JCDCG'04), pp. 85–86, Tokyo, 8-11 October 2004.

[3] M. Benkert, A. Wolff, F. Widmann, and T. Shirabe, The minimum Manhattan network problem: approximations and exact solutions, Computational Geometry 35 (2006) 188–208.

[4] G. Chalmet, L. Francis, and A. Kolen, Finding efficient solutions for rectilinear distance location problems efficiently, European J. Operations Research 6 (1981) 117–124.

[5] J. Gudmundsson, C. Levcopoulos, and G. Narasimhan, Approximating a minimum Manhattan network, Nordic J. Computing 8 (2001) 219–232 and Proc. APPROX'99, 1999, pp. 28–37.

[6] D. Eppstein, Spanning trees and spanners. In J.-R. Sack and J. Urrutia, editors, Handbook of Computational Geometry, pp. 425–461, Elsevier Science Publishers B.V. North-Holland, Amsterdam, 2000.

[7] R. Kato, K. Imai, and T. Asano, An improved algorithm for the minimum Manhattan network problem, ISAAC'02, LNCS, vol. 2518, 2002, pp. 344-356.

[8] F. Lam, M. Alexanderson, and L. Pachter, Picking alignements from (Steiner) trees, J. Computational Biology 10 (2003) 509–520.

[9] G. Narasimhan and M. Smid, Geometric Spanner Networks, Cambridge University Press, 2007.

[10] K. Nouioua, Une approche primale-duale pour le problème du réseau de Manhattan minimal, RAIRO Operations Research (submitted).

[11] S. Seibert and W. Unger, A 1.5-approximation of the minimal Manhattan network, ISAAC'05, Lecture Notes Computer Science, vol. 3827, 2005, pp. 246-255.

[12] S.K. Rao, P. Sadayappan, F.K. Hwang, and P.W. Shor, The rectilinear Steiner arborescence problem, Algorithmica 7 (1992) 277–288.

[13] E. Tardos, A strongly polynomial algorithm to solve combinatorial linear programs, Operations Research 34 (1986) 250–256.

[14] V. Trubin, Subclass of the Steiner problem on a plane with rectilinear metric, Cybernetics 21 (1985) 320–322.

[15] V.V. Vazirani, Approximation Algorithms, Springer-Verlag, Berlin, 2001.

[16] R.E. Wendell, A.P. Hurter, and T.J. Lowe, Efficient points in location theory, AIEE Transactions 9 (1973) 314-321.

[17] M. Zachariasen, A catalog of Hanan grid problems, Networks 38 (2001) 76–83.