

The computational complexity of the pooling problem

Dag Haugland¹

Received: 11 December 2014 / Accepted: 2 July 2015 / Published online: 10 July 2015
© Springer Science+Business Media New York 2015

Abstract The pooling problem is an extension of the minimum cost flow problem defined on a directed graph with three layers of nodes, where quality constraints are introduced at each terminal node. Flow entering the network at the source nodes has a given quality, at the internal nodes (pools) the entering flow is blended, and then sent to the terminal nodes where all entering flow streams are blended again. The resulting flow quality at the terminals has to satisfy given bounds. The objective is to find a cost-minimizing flow assignment that satisfies network capacities and the terminals' quality specifications. Recently, it was proved that the pooling problem is NP-hard, and that the hardness persists when the network has a unique pool. In contrast, instances with only one source or only one terminal can be formulated as compact linear programs, and thereby solved in polynomial time. In this work, it is proved that the pooling problem remains NP-hard even if there is only one quality constraint at each terminal. Further, it is proved that the NP-hardness also persists if the number of sources and the number of terminals are no more than two, and it is proved that the problem remains hard if all in-degrees or all out-degrees are at most two. Examples of special cases in which the problem is solvable by linear programming are also given. Finally, some open problems, which need to be addressed in order to identify more closely the borderlines between polynomially solvable and NP-hard variants of the pooling problem, are pointed out.

Keywords Pooling problem · Network flow · Computational complexity · Polynomial reduction

1 Introduction

In many industrial applications, conventional network flow models, such as the minimum cost flow and the maximum flow problems, need to be extended in order to have practical relevance.

✉ Dag Haugland
dag.haugland@ii.uib.no

¹ Department of Informatics, University of Bergen, P.O. Box 7803, 5020 Bergen, Norway

An extension often seen in the process industry, is that the network sources supply flow of unequal qualities, while flow satisfying quality requirements is demanded at the terminals (sinks). Flow streams of unequal qualities are blended when they enter a node jointly, such that the quality of the flow leaving the node is averaged over the entering qualities.

Qualities are typically defined as the proportions of various components in the flow, for instance an undesired contaminant or a desired nutrition. With qualities being such proportions, the formulas for computing quality updates become simple: The quality of the flow leaving a node is the convex combination of the entering qualities, where the coefficients are proportional to the flow values. Despite this simplicity, early industrial applications of optimization revealed that, compared with the pure network flow models, the extension in question represents a significant increase in computational challenge.

Referring to the above extension as the *pooling problem*, Haverly [16] recognized the difficulty of solving it in terms of linear programming (LP). Haverly's work was motivated by problems related to optimal mixing of crude oils. Later, the industrial relevance of the problem, notably in petroleum refining [6,8], the food industry [18], and in waste water processing [12,22], has been acknowledged by many authors. The pooling problem survey by Misener and Floudas [21] has a comprehensive list of references to work in this area.

Adding quality constraints to the traditional network flow models leads very naturally to a *bilinear* model. Early works on inexact solution methods [6,8,16] take advantage of bilinearity, utilizing the advantage of facing an LP when the values of a subset of the variables are fixed. The idea underlying *successive linear programming* (SLP) is to solve this LP, use the solution as the base point of another linearization, and to repeat this procedure until a fixpoint is reached. By means of small numerical examples, Haverly [16,17] demonstrates that the solutions produced by different SLP-algorithms depend heavily on their initial linearization point.

Floudas and Visweswaran [10] gave the first exact solution algorithm for the pooling problem. Lower bounds (in a minimization problem) are computed by solving a Lagrangian relaxation, and upper bounds are computed by solving a projection of the original problem. By proving that the two bounds converge, convergence to the optimal solution is also proved. Lagrangian relaxation is also used in various global optimization algorithms [1,4,7] subsequent to [10], whereas the branch-and-bound algorithm by Foulds et al. [11] is based on convex relaxations of each bilinear term [3,20]. A running time analysis of all these exact methods will reveal exponential worst-case behavior. The above exact algorithms have later been improved [5,9,23–25], and new and innovative methods, such as the mixed integer programming techniques by Dey and Gupte [9] and Gupte et al. [15], are emerging. Although these are capable of computing the global optimum of larger instances than the earliest algorithms could solve, their exponential running times still confine their applicability to instances of moderate size.

Based on practical experience, a general conception of the pooling problem as inapproachable by LP was early established. This has recently been confirmed in terms of a proof of NP-hardness [2], and that the hardness persists even in network instances with only one internal node (pool). Despite such a proof, many questions concerning the complexity of the problem are left open. The goal of the current work is to enhance the understanding of the intractability of the pooling problem further. To this end, the complexity result in [2] is complemented by several new theorems. As the main contributions, it is proved in the current text that the problem remains NP-hard even if there is only one quality parameter (contaminant), and even if there are only two sources and two sinks. Further, NP-hardness of the problem in the case of networks where all nodes have in-degree at most two is proved. By a similar proof, it is also demonstrated that the problem remains NP-hard in the case where

all nodes have at most two out-neighbors. Thus, several questions left open in a recent study [9] are answered in the current work.

The remainder of the text is organized as follows: In Sect. 2, the pooling problem is defined in rigorous terms, necessary mathematical notation is introduced, and some preliminary results are given. Section 3 is devoted to instances with only one pool, and for this instance class, results extending [2] are given. In Sect. 4, hardness theorems for instances with only one quality parameter are proved, while the complexity of instances with bounded node degree is the topic of Sect. 5. Conclusions and some open problems that merit further investigation are summarized in Sect. 6.

2 Preliminaries

2.1 Definitions

Let $D = (S, P, T, A)$ be a directed graph, where the disjoint node sets $S, P,$ and T consist of *sources, pools* and *terminals*, respectively, and the arc set $A \subseteq (S \times P) \cup (P \times T)$ links sources with pools and pools with terminals. Let K be a finite set, and refer to its elements as *quality parameters*. In the sequel, components of vectors and matrices defined over the sets $S, P, T,$ and $K,$ are identified by a subscript if they correspond to $S, P,$ or $T,$ and by a superscript if they correspond to $K.$ Define the vectors of unit cost $c \in \mathbb{Q}^A,$ node capacity $b \in \mathbb{Q}_+^{S \cup P \cup T},$ quality $q \in \mathbb{Q}^{S \times K},$ and quality bounds $u \in \mathbb{Q}^{T \times K}.$ For all sources $s \in S$ and quality parameters $k \in K, q_s^k$ denotes the quality, as defined by $k,$ of the flow entering D at $s,$ and for all terminals $t \in T, u_t^k$ denotes the upper bound on the quality (smaller numerical values indicate better quality) of the flow leaving D at $t.$

For each pool $p \in P,$ the sets of neighbor sources and terminals are denoted by $S_p = \{s \in S: (s, p) \in A\}$ and $T_p = \{t \in T: (p, t) \in A\},$ respectively. For each source $s \in S$ and each terminal $t \in T,$ respectively, the sets of neighbor pools are denoted by $P_s = \{p \in P: (s, p) \in A\},$ and $P_t = \{p \in P: (p, t) \in A\}.$ The arc set A has a partition $(A_S, A_T),$ where $A_S = A \cap (S \times P)$ and $A_T = A \cap (P \times T)$ denote the sets of arcs incident to some source and to some terminal, respectively.

An $x \in \mathbb{R}_+^A$ is said to be a *flow* in D if it satisfies the capacity and flow conservation constraints:

$$\sum_{p \in P_s} x_{sp} \leq b_s \quad s \in S, \tag{1}$$

$$\sum_{s \in S_p} x_{sp} \leq b_p \quad p \in P, \tag{2}$$

$$\sum_{p \in P_t} x_{pt} \leq b_t \quad t \in T, \tag{3}$$

$$\sum_{s \in S_p} x_{sp} - \sum_{t \in T_p} x_{pt} = 0 \quad p \in P. \tag{4}$$

Let a flow polytope be defined as $F(D, b) = \{x \in \mathbb{R}_+^A: (1-4) \text{ are satisfied}\}.$ For each $x \in F(D, b),$ define $w \in \mathbb{R}^{P \times K}$ to be a *quality matrix* of x if $w_p^k = \sum_{s \in S_p} q_s^k x_{sp} / \sum_{t \in T_p} x_{pt}$ for all $k \in K$ and for all $p \in P$ for which $\sum_{t \in T_p} x_{pt} > 0.$ If $\sum_{t \in T_p} x_{pt} = 0, w_p^k$ can take an arbitrary value. That is, the quality vector $(w_p^k)_{k \in K}$ of the flow blended at pool $p,$ is a convex combination of the source qualities $(q_s^k)_{k \in K},$ with coefficients $x_{sp} / \sum_{t \in T_p} x_{pt}$ ($s \in S_p$). This definition of w is referred to as the *linear blending* assumption. At a terminal

$t \in T$ where $\sum_{p \in P_t} x_{pt} > 0$, the upper quality bounds u_t^k apply, and linear blending at t yields $\sum_{p \in P_t} w_p^k x_{pt} / \sum_{p \in P_t} x_{pt} \leq u_t^k$. Therefore, x is said to be a *feasible* flow in D if $\sum_{p \in P_t} w_p^k x_{pt} \leq u_t^k \sum_{p \in P_t} x_{pt}$ ($t \in T, k \in K$) for any quality matrix w of x . This leads to the problem definition:

Problem 1 The POOLING PROBLEM: Find a feasible flow x in D minimizing $\sum_{(i,j) \in A} c_{ij}x_{ij}$.

For a rational constant ζ , the *decision version* of Problem 1 is to decide whether there exists a feasible flow x in D such that $\sum_{(i,j) \in A} c_{ij}x_{ij} \leq \zeta$.

2.2 Preliminary results

The definition of Problem 1 gives directly a bilinear formulation in flow and quality variables [16], often referred to as the P-formulation:

$$\min_{x,w} \sum_{(i,j) \in A} c_{ij}x_{ij}, \tag{5}$$

$$\sum_{s \in S_p} q_s^k x_{sp} - w_p^k \sum_{t \in T_p} x_{pt} = 0 \quad p \in P, k \in K, \tag{6}$$

$$\sum_{p \in P_t} w_p^k x_{pt} - u_t^k \sum_{p \in P_t} x_{pt} \leq 0 \quad t \in T, k \in K, \tag{7}$$

$$x \in F(D, b). \tag{8}$$

It is readily observed that for a pool $p \in P$ where $\sum_{t \in T_p} x_{pt} = 0$ (a terminal $t \in T$ where $\sum_{p \in P_t} x_{pt} = 0$), constraints (6) [constraints (7)] are satisfied regardless of the value of w .

The assumption that there are no direct arcs from sources to terminals can be made without loss of generality, because any such arc can be replaced by a new pool along with two arcs connecting the pool to the source and the terminal, respectively.

Lower quality bounds, that is constraints on the form $\sum_{p \in P_t} w_p^k x_{pt} \geq \ell_t^k \sum_{p \in P_t} x_{pt}$, where ℓ_t^k are given constants for all $t \in T$ and $k \in K$, do not represent an extension of (5)–(8). For all quality parameters $k \in K$ subject to some lower bound, it is possible to extend K by a new quality parameter k' , to let $q_s^{k'} = -q_s^k$ for all $s \in S$, and to define the upper bounds $u_t^{k'} = -\ell_t^k$ for all $t \in T$. Whenever relevant, instances of the problem where (7) is replaced by $\ell_t^k \sum_{p \in P_t} x_{pt} \leq \sum_{p \in P_t} w_p^k x_{pt} \leq u_t^k \sum_{p \in P_t} x_{pt}$ are therefore referred to as instances of the POOLING PROBLEM. When the number of quality parameters is an issue, however, all quality parameters subject also to lower bounds have to be counted twice.

In later sections, it is made use of a bilinear formulation of Problem 1 that does not involve quality variables. Let (P_S, P_T) be a partition of P , and define $A'_S = A_S \cap (S \times P_S)$, $A'_T = A_T \cap (P_T \times T)$, and $A' = A'_S \cup A'_T$. For each $(s, p) \in A'_S$, let variable y_{sp} be the proportion of the flow through pool p that enters along (s, p) , and for each $(p, t) \in A'_T$, let y_{pt} be the proportion of the flow through pool p that leaves along (p, t) .

Below, it is proved formally that the POOLING PROBLEM can be formulated as follows:

Problem 2

$$\min_{x,y} \sum_{(i,j) \in A} c_{ij}x_{ij}, \tag{9}$$

$$y_{sp} \sum_{t \in T_p} x_{pt} - x_{sp} = 0 \quad (s, p) \in A'_S, \tag{10}$$

$$y_{pt} \sum_{s \in S_p} x_{sp} - x_{pt} = 0 \quad (p, t) \in A'_T, \tag{11}$$

$$\sum_{s \in S_p} y_{sp} = 1 \quad p \in P_S, \tag{12}$$

$$\sum_{t \in T_p} y_{pt} = 1 \quad p \in P_T, \tag{13}$$

$$\sum_{p \in P_T \cap P_S} \sum_{s \in S_p} (q_s^k - u_t^k) y_{sp} x_{pt} + \sum_{p \in P_T \cap P_T} \sum_{s \in S_p} (q_s^k - u_t^k) y_{pt} x_{sp} \leq 0 \quad t \in T, k \in K, \tag{14}$$

$$x \in F(D, b), y \in \mathbb{R}_+^{A'} \tag{15}$$

Constraints (12)–(13) are redundant in the sense that they are implied by the conservation of flow constraints (4) and (10)–(11) when the flow through p is non-zero. They are included in Problem 2 in order to simplify some proofs below.

In the case where $P_S = P$ and $P_T = \emptyset$, eliminating (12)–(13) from Problem 2 yields the Q-formulation [7] of the POOLING PROBLEM. Keeping the redundant constraints, adding the redundant inequalities $x_{sp} \leq b_p y_{sp}$ $((s, p) \in A_S)$, and choosing $(P_S, P_T) = (P, \emptyset)$, renders Problem 2 identical to the PQ-formulation [24]. Likewise, $(P_S, P_T) = (\emptyset, P)$ gives a formulation based uniquely on terminal proportions as suggested in [2]. Problem 2 thus generalizes previously suggested formulations based exclusively on either variables y_{sp} $((s, p) \in A_S)$ [24] or variables y_{pt} $((p, t) \in A_T)$ [2], in that it uses source proportions for the pools in P_S and terminal proportions for the pools in P_T .

Lemma 1 Any $x \in F(D, b)$ is a feasible flow in D if and only if there exists a $y \in \mathbb{R}_+^{A'}$ such that (x, y) is feasible in Problem 2.

Proof Let $x \in F(D, b)$. We first prove that if $w \in \mathbb{R}^{P \times K}$ satisfies (6) and $y \in \mathbb{R}_+^{A'}$ satisfies (10)–(13), then inequalities (7) and (14) are identical. Let $P^+ = \{p \in P: \sum_{t \in T_p} x_{pt} > 0\}$, $P_S^+ = P^+ \cap P_S$, and $P_T^+ = P^+ \cap P_T$. Utilizing $y_{sp} = \frac{x_{sp}}{\sum_{s' \in S_p} x_{s'p}}$ $(p \in P_S^+, s \in S_p)$, and $y_{pt} = \frac{x_{pt}}{\sum_{s' \in S_p} x_{s'p}}$ $(p \in P_T^+, t \in T_p)$, yields $y_{sp} x_{pt} = \frac{x_{sp} x_{pt}}{\sum_{s' \in S_p} x_{s'p}}$ $(p \in P_S^+, s \in S_p)$, and $x_{sp} y_{pt} = \frac{x_{sp} x_{pt}}{\sum_{s' \in S_p} x_{s'p}}$ $(p \in P_T^+, t \in T_p)$, respectively. Plugging these products into (14) gives

$$\sum_{p \in P_T \cap P^+} \frac{x_{pt}}{\sum_{s' \in S_p} x_{s'p}} \sum_{s \in S_p} (q_s^k - u_t^k) x_{sp} \leq 0 \quad t \in T, k \in K. \tag{16}$$

Because $\frac{\sum_{s \in S_p} q_s^k x_{sp}}{\sum_{s' \in S_p} x_{s'p}} = w_p^k$ $(p \in P^+, k \in K)$, (16) is identical to (7).

Assume now that x is a feasible flow, and choose w such that (6)–(7) are satisfied. Let $y_{sp} = \frac{x_{sp}}{\sum_{s' \in S_p} x_{s'p}}$ for all $p \in P_S^+, s \in S_p$ and $y_{pt} = \frac{x_{pt}}{\sum_{s' \in S_p} x_{s'p}}$ for all $p \in P_T^+, t \in T_p$, implying that (12)–(13) are satisfied for all $p \in P^+$. For $p \in P \setminus P^+$, the values of y_{sp} $((s, p) \in A'_S)$ and y_{pt} $((p, t) \in A'_T)$ can be chosen arbitrarily such that (12)–(13) are satisfied. By the equivalence of (7) and (14), (x, y) is feasible in Problem 2, and the only-if-part of the lemma follows. Finally, assume that (x, y) is feasible in Problem 2 for some $y \in \mathbb{R}_+^{A'}$, and choose $w_p^k = \frac{\sum_{s \in S_p} q_s^k x_{sp}}{\sum_{s \in S_p} x_{sp}}$ for all $p \in P^+$. Hence (6) is satisfied, (7) is satisfied because it is identical to (14), and the if-part follows. \square

Proposition 1 For any partition (P_S, P_T) of P , Problem 2 is a valid formulation for the POOLING PROBLEM.

Proof Follows directly from Lemma 1. □

Proposition 2 *Given any $\alpha^k \in \mathbb{Q}_+$ and $\beta^k \in \mathbb{Q}$ for all $k \in K$, the set of feasible flows in D is invariant under affine transformations $\alpha^k q_s^k + \beta^k$ and $\alpha^k u_t^k + \beta^k$ of q_s^k ($s \in S$) and u_t^k ($t \in T$), respectively.*

Proof Since the suggested operations do not alter the set of solutions satisfying constraints (14), which are the sole constraints in Problem 2 where parameters q and u occur, the proposition follows directly from Proposition 1. □

By constraint (14), it is easily seen that if $x \in F(D, b)$ is feasible, then $\sum_{p \in P_t} x_{pt} = 0$ for all $t \in T$ for which there exists some $k \in K$ where $u_t^k < m^k = \min_{s \in S} q_s^k$. Such terminals can thus be removed from D . Also, if there exists some $k \in K$ such that $u_t^k > M^k = \max_{s \in S} q_s^k$, then x remains feasible if the bound is tightened to have the value M^k . Consequently, it is henceforth assumed that $m^k \leq u_t^k \leq M^k$ for all $t \in T$.

Without further loss of generality, all q_s^k and u_t^k can be assumed to be in $[0, 1]$. This follows immediately by choosing $\alpha^k = 1/(M^k - m^k)$ and $\beta^k = -m^k \alpha^k$ in Proposition 2. Although not adopted generally in the text, this assumption will be made without further justification whenever it is useful.

In cases where $|S| = |T| = 1$, it is straightforward to formulate the POOLING PROBLEM as an LP [9]. By virtue of the valid formulation (9)–(15), a more general statement is proved:

Proposition 3 *If $\min\{|S_p|, |T_p|\} = 1$ for all $p \in P$, then the POOLING PROBLEM can be formulated as a compact linear program.*

Proof Choose $P_S = \{p \in P: |S_p| = 1\}$ and $P_T = P \setminus P_S$. Combined with $y \geq 0$, constraints (12) and (13) then imply $y_{sp} = 1$ for all $(s, p) \in A'_S$ and $y_{pt} = 1$ for all $(p, t) \in A'_T$, respectively. Substituting all y -variables by 1 in constraint (14) renders Problem 2 linear, and the claim follows from Proposition 2 and the observation that the numbers of variables and constraints are polynomial in the input size. □

While Proposition 3 shows that restricting either the in- or the out-degree of all pools to 1 leaves us with an LP, it is shown in the next section that restricting the degrees of all sources and terminals does not have such favorable computational benefits.

Subsequent sections make extensive references to the POOLING PROBLEM defined over the set of instances satisfying some condition C on D or its parameters. Such special cases of the problem will be referred to as ‘the POOLING PROBLEM with C ’.

3 Single pool networks

Problem 3 MAXIMUM INDEPENDENT SET (MIS): Given a graph $G = (V, E)$, find a vertex set $V' \subseteq V$ of maximum cardinality such that all pairs of vertices in V' are non-neighbors in G .

For an integer z , the decision version of Problem 3 asks whether G has an independent vertex set of cardinality z .

Theorem 1 *The POOLING PROBLEM with $|P| = 1$ is strongly NP-hard.*

Proof See [2]. □

For consistency, all hardness proofs in this article are cast in terms of polynomial reductions from an NP-complete decision problem, to the decision version of the POOLING PROBLEM. Such a reduction provides a proof of NP-hardness of the original optimization version (Problem 1). The proof of Theorem 1 [2] is by a polynomial reduction from MIS. Its underlying idea is to define the network D , with associated parameters, such that there is a one-to-one correspondence between independent vertex sets V' in G and feasible flows of cost $-|V'|$ in D . That is, $\zeta = -z$.

For every vertex in V , the digraph D has one source and one terminal, both of which have an arc to/from the unique pool. The sources and terminals corresponding to vertices in V' send/receive unit flow in the feasible solution, whereas the flow at other sources and terminals is zero. To this end, there is also one quality parameter for each vertex, and the quality at source $s \in S$ is the unit vector with a 1-entry in the position of $k \in K$ if s and k correspond to the same vertex.

At the terminals, there are both lower and upper quality bounds. If $t \in T$ and $k \in K$ correspond to the same vertex v , a lower bound $\ell_t^k = 1/|V|$ is imposed, implying that t receives non-zero flow only if the source corresponding to v also supplies non-zero flow. Further, terminal t cannot receive any flow originating from source s if s and t correspond to $v_1 \in V$ and $v_2 \in V$, respectively, where $\{v_1, v_2\} \in E$. This is accomplished by the upper quality bound $u_t^{k_1} = 0$, where k_1 is the quality parameter corresponding to v_1 . Finally, unit flow capacities are imposed at sources and terminals, and the arc costs are defined consistently with flow maximization.

Theorem 2 *If $|P| = 1$, then the POOLING PROBLEM can be solved in terms of $\min \{(|T| + 1)^{|K|}, 2^{|T|}\}$ compact linear programs.*

Proof To show that $(|T| + 1)^{|K|}$ compact LPs are sufficient, see the proof of Proposition 2 in [2]. Consider any non-zero flow $x \in F(D, b)$, and let $w_p^k = \frac{\sum_{s \in S} q_s^k x_{sp}}{\sum_{s \in S} x_{sp}}$, where $P = \{p\}$ and $k \in K$. Since all terminals receive flow uniquely from the pool p , x is feasible if and only if, for all $t \in T$, $x_{pt} = 0$ or $w_p^k \leq u_t^k$ for all $k \in K$. Thus, the solution to Problem 1 is $\min \{z(T^+) : T^+ \subseteq T\}$, where

$$z(T^+) = \min_x \begin{aligned} & \sum_{(i,j) \in A} c_{ij} x_{ij}, \\ & x \in F(D, b), \\ & \sum_{s \in S} (q_s^k - u_t^k) x_{sp} \leq 0 \quad t \in T^+, k \in K, \\ & x_{pt} = 0 \quad t \in T \setminus T^+. \end{aligned}$$

Since the numbers of variables and constraints in all $2^{|T|}$ linear programs hence defined are polynomial in the input size, the proof is complete. □

Although the single-pool instances in general are hard, Theorem 2 shows that algorithms with polynomial running time exist for the subclasses where the size of either K or T is bounded. This follows directly from the compactness of the corresponding LP, and the existence of LP-algorithms with polynomial running time.

It is left as an open question whether polynomial algorithms exist when $|P| = 1$ and $|S| \leq s_{\max}$ for any $s_{\max} \geq 2$.

4 Instances with a single quality parameter

Recently, Dey and Gupte [9] raised the question whether the POOLING PROBLEM remains NP-hard if there is only one quality parameter. Below, their question is answered affirmatively. Moreover, it is proved that the NP-hardness for $|K| = 1$ persists even if there are only two sources and two terminals. If $|K| = 1$ and the conditions on S and T is relaxed such that there are either two sources or two terminals, the NP-hardness becomes strong.

Problem 4 PACKING IN TWO BINS: Given positive integers d, a_1, \dots, a_n , does there exist a set $M \subseteq \{1, \dots, n\}$ such that $\sum_{i \in M} a_i \leq d$ and $\sum_{i \notin M} a_i \leq d$?

PACKING IN TWO BINS is known to be NP-complete [13].

Theorem 3 The POOLING PROBLEM with $|K| = 1$ and $|S| = |T| = 2$ is NP-hard.

Proof We prove that any instance of PACKING IN TWO BINS can be reduced to an instance of the decision version of the POOLING PROBLEM with $\zeta = -\sum_{i=1}^n a_i$, $|K| = 1$ and $|S| = |T| = 2$. Since $|K| = 1$, all superscripts denoting quality parameters are omitted. Let $S = \{s_0, s_1\}$, $T = \{t_0, t_1\}$, $P = \{p_1, \dots, p_n\}$, $A = (S \times P) \cup (P \times T)$, $b_{s_0} = b_{s_1} = b_{t_0} = b_{t_1} = d$, $q_{s_0} = u_{t_0} = 0$, and $q_{s_1} = u_{t_1} = 1$. For all $i = 1, \dots, n$, let $b_{p_i} = a_i$, $c_{s_0 p_i} = 3$, $c_{p_i t_0} = -4$, $c_{s_1 p_i} = 1$, and $c_{p_i t_1} = -2$.

Assume $x \in F(D, b)$ is a feasible flow with cost $-\sum_{i=1}^n a_i$, and define $c[p] = \sum_{s \in S_p} c_{sp} x_{sp} + \sum_{t \in T_p} c_{pt} x_{pt}$ as the corresponding cost of the flow through pool $p \in P$. Let $P^0 = \{i = 1, \dots, n: x_{p_i t_0} > 0\}$ be the indices of pools sending positive flow to terminal t_0 , and let $P^1 = \{1, \dots, n\} \setminus P^0$. For $i \in P^0$, the quality constraint at t_0 implies that p_i receives all its flow from s_0 . Because $c_{s_0 p_i} + c_{p_i t_0} = -1$ and $c_{s_0 p_i} + c_{p_i t_1} > -1$, we get $\sum_{i \in P^0} c[p_i] \geq -\sum_{i \in P^0} a_i$, and $\sum_{i \in P^0} c[p_i] = -\sum_{i \in P^0} a_i$ only if $x_{p_i t_0} = a_i$ and $x_{p_i t_1} = 0$ for all $i \in P^0$. For $i \in P^1$, $c_{s_0 p_i} + c_{p_i t_1} > -1$ and $c_{s_1 p_i} + c_{p_i t_1} = -1$ imply $\sum_{i \in P^1} c[p_i] \geq -\sum_{i \in P^1} a_i$, with equality only if $x_{s_0 p_i} = 0$ and $x_{p_i t_1} = a_i$ for all $i \in P^1$. Because the total costs equal $\zeta = -\sum_{i=1}^n a_i = \sum_{i \in P^0} c[p_i] + \sum_{i \in P^1} c[p_i]$, it is shown that $x_{p_i t_0} = a_i$ ($i \in P^0$) and $x_{p_i t_1} = a_i$ ($i \in P^1$), and the capacity constraints at T prove $\sum_{i \in P^0} a_i \leq d$ and $\sum_{i \in P^1} a_i \leq d$. It follows that (d, a_1, \dots, a_n) is a yes-instance.

Conversely, if (d, a_1, \dots, a_n) is a yes-instance, then there exists a partition (P^0, P^1) of $\{1, \dots, n\}$ such that $\sum_{i \in P^0} a_i \leq d$ and $\sum_{i \in P^1} a_i \leq d$, and a feasible flow with cost $\sum_{(s,p) \in A_S} c_{sp} x_{sp} + \sum_{(p,t) \in A_T} c_{pt} x_{pt} = -\sum_{i=1}^n a_i$ is obtained by letting $x_{s_0 p_i} = x_{p_i t_0} = a_i$ for all $i \in P^0$ and $x_{s_1 p_i} = x_{p_i t_1} = a_i$ for all $i \in P^1$.

Thus, (d, a_1, \dots, a_n) is a yes-instance if and only if there exists a feasible flow with cost no more than ζ in the POOLING PROBLEM, which completes the proof. □

Since PACKING IN TWO BINS is not proved to be strongly NP-complete, Theorem 3 is also restricted to NP-hardness in the weak sense. When the condition on either the source or the terminal set is relaxed, however, strong NP-hardness persists. We prove this by polynomial reductions from the following problem:

Problem 5 EXACT COVER BY 3-SETS (X3C): Given positive integers m and n , where $3 \mid n$, and a set M of subsets $X_1, \dots, X_m \subseteq N = \{1, \dots, n\}$, where $|X_1| = \dots = |X_m| = 3$, is there a subset $M' \subseteq M$ such that $|M'| = \frac{n}{3}$ and $\bigcup_{X_i \in M'} X_i = N$?

Problem 5 is proved to be NP-complete [13].

The idea behind the first reduction is to introduce a pair of terminals, referred to as the restrictive and the tolerant terminal, respectively, for each of the elements $1, \dots, n$ to be

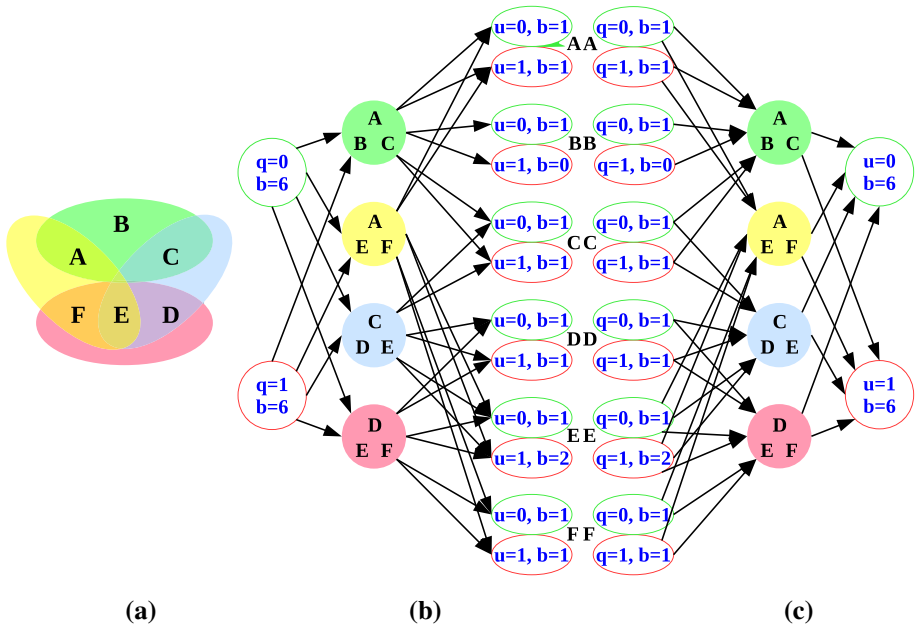


Fig. 1 POOLING PROBLEM instances corresponding to the X3C-instance with $M = \{\{A, B, C\}, \{A, E, F\}, \{C, D, E\}, \{D, E, F\}\}$. **a** X3C-instance, **b** reduction to $|S| = 2$, **c** reduction to $|T| = 2$

covered. The restrictive terminals demand perfectly clean flow, whereas any flow is accepted at the tolerant terminals. For each 3-set, there is a pool, from which there are arcs to both terminals in all three pairs corresponding to elements in the 3-set. All pools have capacity 3, all restrictive terminals have unit capacity, and the capacity of the tolerant terminal corresponding to element $j \in N$ is one less than the number of 3-sets in which j occurs. Flow from a pool to the restrictive terminal is to indicate that the corresponding 3-set covers the corresponding element, whereas flow to the tolerant terminal indicates the converse (see Fig. 1a, b for an example).

Further, there is one perfectly clean and one totally contaminated source, both of which are linked with all pools. The capacity of the clean (contaminated) source equals the total capacity of all restrictive (tolerant) terminals. In case of full capacity utilization, all the flow from the clean source thus has to reach only restrictive terminals. To this end, the pools must be partitioned into two subsets. Pools in the first subset transit flow from the clean source to restrictive terminals, while the others pass contaminated flow on to tolerant terminals. Below, it is shown formally that partitioning the pools this way, and such that all terminals receive flow at their full capacity, is feasible if and only if M has an exact cover of N .

Theorem 4 *The POOLING PROBLEM with $|K| = 1$ and $|S| = 2$ is strongly NP-hard.*

Proof Let $S = \{s_0, s_1\}$ and $T = T^0 \cup T^1$, where $T^v = \{t_j^v : j \in N\}$, let $P = \{p_1, \dots, p_m\}$, and let $A_S = S \times P$ and $A_T = A_T^0 \cup A_T^1$, where $A_T^v = \{(p_i, t_j^v) : j \in X_i\}_{i=1}^m$ ($v = 0, 1$). Let $|K| = 1$, and for the purpose of simplified notation, all superscripts referring to $k \in K$ are omitted in this proof. Let $b_p = 3$ for all $p \in P$, let $b_{s_0} = n$, $b_{s_1} = 3m - n$, $b_{t_j^0} = 1$ and

$b_{t_j^1} = |M_j| - 1$ for all $j \in N$, where $M_j = \{X_i \in M : j \in X_i\}$. Finally, let $c_{sp} = 0$ for all $(s, p) \in A_S$, $c_{pt} = -1$ for all $(p, t) \in A_T$, $q_{s_0} = 0$, $q_{s_1} = 1$, and $u_t = \begin{cases} 0, & t \in T^0 \\ 1, & t \in T^1. \end{cases}$

By the definition of the costs, the objective in this POOLING PROBLEM instance is to maximize $f(x) = \sum_{(p,t) \in A_T} x_{pt}$, and for a given ζ , the decision version is to decide whether there exists a feasible flow x such that $f(x) \geq \zeta$. Since the total flow capacity at the pools equals $3m$, $f(x) \leq 3m$ for all feasible flows x in D . We let $\zeta = 3m$, and prove that $f(x) = \zeta$ is achievable if and only if (M, N) is a yes-instance of X3C.

If: Assume $M' \subseteq M$ covers N exactly. For all $X_i \in M'$, define the arc flows $x_{s_0 p_i} = 3$, $x_{s_1 p_i} = 0$, $x_{p_i t_j^0} = 1$, $x_{p_i t_j^1} = 0$, and for all $X_i \in M \setminus M'$, let $x_{s_0 p_i} = 0$, $x_{s_1 p_i} = 3$, $x_{p_i t_j^0} = 0$, $x_{p_i t_j^1} = 1$ ($j \in X_i$). The pool qualities, w_p ($p \in P$), are determined uniquely such that (6) is satisfied. Obviously, $x \in F(D, b)$, $f(x) = 3m$, and the quality bounds at T^1 are satisfied since $w_p \leq 1$ for all $p \in P$. It remains to prove that the quality bounds at all t_j^0 are satisfied. Since M' is an exact cover of N , there is a unique $X_i \in M'$ such that $j \in X_i$. Correspondingly, t_j^0 receives positive flow uniquely from p_i , which receives all its flow from s_0 . It follows that $w_{p_i} = 0$, proving that the quality bound at t_j^0 is satisfied. Hence, x is a feasible flow in D .

Only if: Assume x is a feasible flow with $f(x) = 3m$, and define $P^0 = \left\{ p \in P : \sum_{t \in T_p \cap T^0} x_{pt} > 0 \right\}$ as the set of pools sending flow to T^0 . Since the total flow capacity at the terminals equals $\sum_{t \in T} b_t = \sum_{t \in T^0} 1 + \sum_{t_j^1 \in T^1} (|M_j| - 1) = n + 3m - n = 3m = b_{s_0} + b_{s_1}$, $f(x) = 3m$ implies that the capacity is fully utilized at every node. Consequently, all terminals in T^0 receive positive flow, and $M' = \{X_i : p_i \in P^0\}$ is a cover of N . It only remains to prove that $|M'| \leq \frac{n}{3}$. Because the zero quality bounds at all $t \in T^0$ are satisfied, the pool qualities, w_p , are zero for all $p \in P^0$, and thereby, $x_{s_1 p} = 0$ for all $p \in P^0$. The capacity constraints at $P \setminus P^0$ and full capacity utilization at s_1 thus yield $3|P \setminus P^0| = \sum_{p \in P \setminus P^0} b_p \geq \sum_{p \in P \setminus P^0} x_{s_1 p} = b_{s_1} = 3m - n$. Hence, $|M'| = |P^0| = |P| - |P \setminus P^0| \leq m - \frac{3m-n}{3} = \frac{n}{3}$.

To conclude that the problem is NP-hard in the strong sense, it suffices to observe that the absolute values of all parameters in the instance sets in question, are integers bounded by $3|M|$, which is polynomial in the input size. □

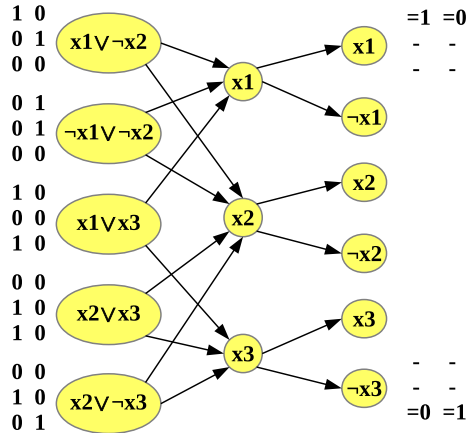
The idea of the above proof is easily carried over to a reduction from X3C to the decision version of the POOLING PROBLEM with $|K| = 1$ and $|T| = 2$ (see Fig. 1c for an example). Throughout the proof, the roles of the sources and the terminals are swapped straightforwardly, and the result is therefore stated without proof.

Theorem 5 *The POOLING PROBLEM with $|K| = 1$ and $|T| = 2$ is strongly NP-hard.*

Corollary 1 *The POOLING PROBLEM with $\max\{|S_p|, |T_p|\} \leq 6$ for all $p \in P$, $\min\{|S|, |T|\} = 2$, and $|K| = 1$, is strongly NP-hard.*

A comparison of Proposition 3, Theorem 3, and Corollary 1 reveals an interesting evolution in the intractability of the POOLING PROBLEM as the in- and out-degrees of the pools grow. From being solvable in polynomial time when $\min\{|S_p|, |T_p|\} = 1$, the problem becomes NP-hard (at least) in the weak sense when $|S_p| = |T_p| = 2$, and finally strongly NP-hard when $\min\{|S_p|, |T_p|\} = 2$.

Fig. 2 Polynomial reduction from MAXIMUM 2-SATISFIABILITY to the POOLING PROBLEM



5 Degree-bounded instances

The results of the previous section show that the POOLING PROBLEM remains NP-hard even if the number of arcs incident to any pool is bounded. In this section, it is proved that if either all sources and pools have two leaving arcs, or all pools and terminals have two entering arcs, then the problem also remains strongly NP-hard. By virtue of the theorems proved here, another question addressed by Dey and Gupte [9] is answered with affirmation.

5.1 Bounded out-degree

Problem 6 MAXIMUM 2-SATISFIABILITY: Given a set $X = \{X_1, \dots, X_n\}$ of Boolean variables, and a set $C = \{C_1, \dots, C_m\}$ of disjunctive clauses of two literals of X , find a truth assignment to X such that a maximum number of the clauses in C are satisfied.

For an integer z , the decision version of Problem 6 asks whether there exists a truth assignment to X such that at least z of the clauses in C are satisfied.

With an instance of MAXIMUM 2-SATISFIABILITY, a POOLING PROBLEM network D^{\max} is associated such that there are one-to-one correspondences between clauses and sources, between Boolean variables and pools, and between literals and terminals. If the clause corresponding to source s contains a literal of the variable corresponding to pool p , then there is an arc from s to p . Each pool is connected to the two terminals corresponding to the literals of the variable corresponding to the pool. Consequently, the out-degrees of all sources and pools are 2 (see Fig. 2 for an example).

Further, the quality parameters are defined such that there is one $k \in K$ for each literal. For reasons of brevity, and whenever appropriate, the expression ‘Boolean variable p ’ is used when referring to the Boolean variable corresponding to pool p . Analogous nomenclature for literals (corresponding to terminals or quality parameters) and clauses (corresponding to sources) is adopted.

Let $q_s^k = 1$ if clause s contains literal k , and $q_s^k = 0$, otherwise. Quality bounds are defined as follows: If terminal t and quality parameter k correspond to the same literal, then $\ell_t^k = u_t^k = 1$. If they correspond to different literals of the same variable, then $\ell_t^k = u_t^k = 0$, and if they correspond to literals of different variables, then $\ell_t^k = 0$ and $u_t^k = 1$. All source capacities equal 1 ($b_s = 1, s \in S$), whereas pools and terminals have capacities equal to

the number of clauses ($b_i = m, i \in P \cup T$). Arc costs are defined consistently with flow maximization, e.g., $c_{sp} = -1$ if $(s, p) \in A_S$ and $c_{pt} = 0$ if $(p, t) \in A_T$.

Let t^+ and t^- (k^+ and k^-) denote the terminals (quality parameters) corresponding to, respectively, the positive and the negative literal of the Boolean variable p . In the following, we establish a one-to-one correspondence between sets of z satisfied clauses and feasible flows in D^{\max} with costs $\zeta = -z$.

Lemma 2 *If x is a feasible flow in D^{\max} , then at least one of x_{pt^+} and x_{pt^-} is zero.*

Proof Assume $x_{pt^+}x_{pt^-} > 0$. Because t^+ and k^+ correspond to the same literal, and k^+ and t^- correspond to different literals of the same variable, the definitions of the quality bounds at t^+ and t^- , respectively, yield $\ell_{t^+}^{k^+} = 1$ and $u_{t^-}^{k^+} = 0$. Since t^+ and t^- both receive non-zero flow, but uniquely from pool p , the pool quality $w_p^{k^+}$ respects the quality bounds at both t^+ and t^- . Hence, $\ell_{t^+}^{k^+} \leq w_p^{k^+} \leq u_{t^-}^{k^+}$, which is a contradiction. \square

Lemma 3 *If $\sum_{s \in S_p} x_{sp} > 0$ for a feasible flow x in D^{\max} , then, for any quality matrix w of x , $w_p^{k^+} = 1$ and $w_p^{k^-} = 0$ if $x_{pt^+} > 0$, and $w_p^{k^+} = 0$ and $w_p^{k^-} = 1$ if $x_{pt^-} > 0$.*

Proof By Lemma 2, all flow leaving p enters exactly one of the terminals t^+ and t^- . Because $w_p^{k^+}$ satisfies the quality constraints at the receiving terminal, $1 = \ell_{t^+}^{k^+} \leq w_p^{k^+} \leq u_{t^+}^{k^+} = 1$ and $0 = \ell_{t^+}^{k^-} \leq w_p^{k^-} \leq u_{t^+}^{k^-} = 0$ if t^+ receives the flow. The proof in the case where t^- receives the flow is analogous. \square

For any flow $x \in F(D, b)$, let $S[x] = \left\{s \in S: \sum_{p \in P_s} x_{sp} > 0\right\}$ denote the set of sources supplying non-zero flow.

Lemma 4 *If x is a feasible flow in D^{\max} , then all clauses in $S[x]$ can be satisfied simultaneously.*

Proof Let P^+ (P^-) consist of all pools from which the leaving flow enters a terminal corresponding to a positive (negative) literal. By Lemma 2, $P^+ \cap P^- = \emptyset$. We prove that by assigning the value `true` to all Boolean variables in P^+ and the value `false` to those in P^- , all clauses in $S[x]$ are satisfied. To this end, we prove that each clause in $S[x]$ contains either the positive literal of some Boolean variable in P^+ , or the negative literal of some Boolean variable in P^- . For all $p \in P^+$, Lemma 3 shows that $w_p^{k^+} = 1$, and pool p receives flow exclusively from sources $s \in S[x]$ with quality $q_s^{k^+} = 1$. Thus, all corresponding clauses contain the positive literal k^+ of Boolean variable p . Analogously, for all $p \in P^-$, $w_p^{k^-} = 1$, and p receives flow exclusively from sources $s \in S[x]$ with quality $q_s^{k^-} = 1$, and the corresponding clauses contain the negative literal k^- of Boolean variable p . By conservation of flow (4), $P^+ \cup P^-$ covers all pools receiving flow from $S[x]$, and the proof is complete. \square

Lemma 5 *If the clauses $C^* \subseteq C$ can be satisfied simultaneously, then there exists a feasible flow x in D^{\max} such that $\sum_{(s,p) \in A_S} x_{sp} = |C^*|$.*

Proof Let P^+ and P^- be disjoint subsets of P such that the clauses C^* are satisfied by assigning the value `true` to all Boolean variables $p \in P^+$ and `false` to all $p \in P^-$. For all clauses $C_i \in C^*$, define the pool $p_i \in P^+ \cup P^-$ such that C_i is satisfied by the Boolean variable p_i (break ties arbitrarily).

Arcs incident to pool $p \in P^+ \cup P^-$ are assigned flow as follows: For sources $s \in S$ corresponding to clauses C_i such that $p_i = p$, the definition of D^{\max} implies that $(s, p) \in A_S$. Let $x_{sp} = 1$ for such arcs (s, p) . If $p \in P^+$, let $x_{pt^+} = |\{C_i \in C^*: p_i = p\}|$ and $x_{pt^-} = 0$, and let $x_{pt^-} = |\{C_i \in C^*: p_i = p\}|$ and $x_{pt^+} = 0$ if $p \in P^-$. Arcs incident to $P \setminus P^+ \setminus P^-$ are assigned zero flow. Consequently, $x \in F(D, b)$ and $\sum_{(s,p) \in A_S} x_{sp} = |C^*|$.

It remains to prove that the quality constraints at terminals receiving non-zero flow are satisfied. The flow through a pool $p \in P^+$ comes exclusively from sources s corresponding to clauses C_i such that $p_i = p$, and goes exclusively to terminal t^+ . Assigning `true` to Boolean variable p satisfies clause C_i , and hence the clause contains the positive literal k^+ . Therefore, $q_s^{k^+} = 1$ and $q_s^{k^-} = 0$, yielding the same qualities at pool p , which means that $w_p^{k^+} = 1 = \ell_{t^+}^{k^+} = u_{t^+}^{k^+}$ and $w_p^{k^-} = 0 = \ell_{t^+}^{k^-} = u_{t^+}^{k^-}$. For all $k \notin \{k^+, k^-\}$, $q_s^k \in [0, 1]$, and consequently, $\ell_{t^+}^k = 0 \leq w_p^k \leq 1 = u_{t^+}^k$. Thus, $\ell_{t^+}^k \leq w_p^k \leq u_{t^+}^k$ for all $k \in K$, and since t^+ receives flow exclusively from p , all quality constraints (7) at t^+ are satisfied.

Analogous arguments apply to pools in P^- , and by feasibility of x , the proof is complete. □

Theorem 6 *The POOLING PROBLEM with all out-degrees no greater than two is strongly NP-hard.*

Proof The proof is by a polynomial reduction from the decision version of MAXIMUM 2-SATISFIABILITY, which is known to be NP-complete [14]. Let $\zeta = -z$, and assume there exists a feasible flow in D^{\max} with total cost no more than ζ . Then the total flow leaving S in D^{\max} is at least z , and $b_s = 1$ ($s \in S$) implies that it is feasible to assign positive flow to at least z sources. Lemma 4 then shows that there exists a satisfiable $C^* \subseteq C$ with $|C^*| \geq z$. Conversely, Lemma 5 shows that for any satisfiable $C^* \subseteq C$, there exists a flow assignment x to D^{\max} such that $z = |C^*|$ flow units leave S . It follows that there exists a feasible flow in D^{\max} with cost no more than ζ if and only if there exists a satisfiable subset of C with cardinality at least z . Thus, the decision version of MAXIMUM 2-SATISFIABILITY has been polynomially reduced to the decision version of the POOLING PROBLEM in the network D^{\max} . The proof is complete by observing that the absolute values of all parameters are integers bounded by m . □

5.2 Bounded in-degree

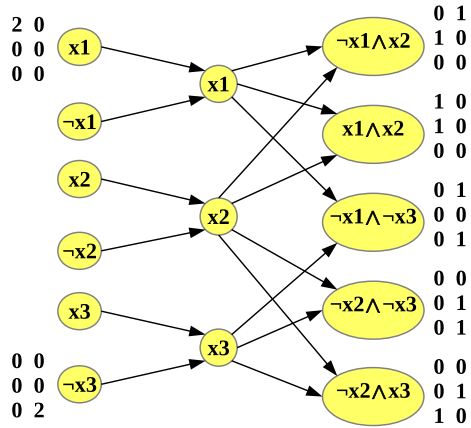
By a reduction similar to the one of the previous section, it is proved in this section that the POOLING PROBLEM remains strongly NP-hard if all nodes have at most two entering arcs. The truth assignments that leave a disjunctive clause of two literals unsatisfied, are exactly those that satisfy the conjunctive clause of the inverses of the same two literals. Consequently, the problem of minimizing the number of satisfied clauses can be stated as the following maximization problem:

Problem 7 MINIMUM 2-SATISFIABILITY: Given a set $X = \{X_1, \dots, X_n\}$ of Boolean variables, and a set $\bar{C} = \{\bar{C}_1, \dots, \bar{C}_m\}$ of conjunctive clauses of two literals of X , find a truth assignment to X such that a maximum number of the clauses in \bar{C} are satisfied.

For an integer z , the decision version of Problem 7 asks whether there exists a truth assignment satisfying at least z conjunctive clauses in \bar{C} .

The POOLING PROBLEM network, D^{\min} , associated with an instance of MINIMUM 2-SATISFIABILITY resembles D^{\max} , but with the roles of sources and terminals interchanged. That is, the bijections are between literals and sources, literals and quality parameters,

Fig. 3 Polynomial reduction from MINIMUM 2-SATISFIABILITY to the POOLING PROBLEM



Boolean variables and pools, and conjunctive clauses and terminals. Source-to-pool and pool-to-terminal connections follow the patterns of, respectively, pool-to-terminal and source-to-pool connections explained in the previous section (see Fig. 3). Consequently, the in-degrees of all pools and terminals are 2.

Following the nomenclature introduced in the previous section, the literal corresponding to source $s \in S$ or quality parameter $k \in K$ will henceforth be referred to as literal s or literal k , and analogous references to Boolean variables and clauses will be used. For any $s \in S$ and $k \in K$, define the source quality $q_s^k = 2$ if source s and quality parameter k correspond to the same literal, and let $q_s^k = 0$, otherwise. If clause $t \in T$ contains literal k , the quality bounds at terminal t are $\ell_t^k = u_t^k = 1$, whereas $\ell_t^k = u_t^k = 0$, otherwise. The POOLING PROBLEM instance is completed by the node capacities $b_t = 2$ ($t \in T$) and $b_i = 2m$ ($i \in S \cup P$), and the arc costs $c_{sp} = 0$ ($(s, p) \in A_S$) and $c_{pt} = -1$ ($(p, t) \in A_T$).

Consider any terminal $t \in T$ with in-neighbors $P_t = \{p, r\}$. By the definition of D^{\min} , the clause t contains a literal of Boolean variable p . Denote by k and k' (s and s') the quality parameters (sources) corresponding to this literal and its inverse, respectively.

Lemma 6 *If D^{\min} is assigned feasible flow such that t receives non-zero flow, then the flow on both arcs entering t is non-zero.*

Proof Assume x is a feasible flow in D^{\min} such that $x_{pt} > 0 = x_{rt}$. Since $\ell_t^k = u_t^k = 1$, the lower and upper quality bounds (7) become $x_{pt}w_p^k + x_{rt}w_r^k = x_{pt} + x_{rt}$, which, by $x_{rt} = 0$, yields $w_p^k = 1$. The flow enters p exclusively from sources s and s' , producing the quality

$$1 = w_p^k = \frac{x_{sp}q_s^k + x_{s'p}q_{s'}^k}{x_{sp} + x_{s'p}} = \frac{2x_{sp}}{x_{sp} + x_{s'p}},$$

since $q_s^k = 2$ and $q_{s'}^k = 0$. Thus, $x_{sp} = x_{s'p} > 0$. For parameter k' , the pool quality becomes

$$w_p^{k'} = \frac{x_{sp}q_s^{k'} + x_{s'p}q_{s'}^{k'}}{x_{sp} + x_{s'p}} = \frac{q_s^{k'} + q_{s'}^{k'}}{2} = 1.$$

Because $\ell_t^{k'} = u_t^{k'} = 0$, the quality bounds (7) corresponding to k' at t read

$$x_{pt}w_p^{k'} + x_{rt}w_r^{k'} = 0. \tag{17}$$

But $x_{pt}w_p^{k'} > 0$ and $x_{rt} = 0$ imply that (17) is violated, contradicting the assumption that x is feasible. \square

Lemma 7 *If $x_{pt} + x_{rt} > 0$ for a feasible flow x in D^{\min} , then $w_p^k = 2$ and $w_p^{k'} = 0$ for any quality matrix w of x .*

Proof As t receives non-zero flow, Lemma 6 implies that $x_{pt} > 0$ and $x_{rt} > 0$. Since non-negativity of the source qualities implies that also the pool qualities are non-negative, the quality constraint (17) thus gives $w_p^{k'} = 0$. Because $q_s^{k'} > 0$, pool p receives all its flow from source s , implying also $w_p^k = q_s^k = 2$. \square

For any flow $x \in F(D, b)$, let $T[x] = \{t \in T: \sum_{p \in P_t} x_{pt} > 0\}$ denote the set of terminals receiving non-zero flow.

Lemma 8 *If x is a feasible flow in D^{\min} , then the clauses $T[x]$ can be satisfied simultaneously.*

Proof We have to prove that there exist no two clauses in $T[x]$ containing literals that are each others' inverse. Assume there exist $t_1, t_2 \in T[x]$ and $k_1, k_2 \in K$ such that literal k_v is in clause t_v ($v = 1, 2$), and such that literal k_1 is the inverse of literal k_2 . Then there exists some $p \in P$ adjacent to both t_1 and t_2 . Lemma 7 applied to t_1 then gives $w_p^{k_1} = 2$, whereas the same lemma applied to t_2 gives $w_p^{k_1} = 0$, which is a contradiction. \square

Lemma 9 *If the clauses $\bar{C}^* \subseteq \bar{C}$ can be satisfied simultaneously, then there exists a feasible flow x in D^{\min} such that $\sum_{(p,t) \in A_T} x_{pt} = 2|\bar{C}^*|$.*

Proof Define the flow as follows: For all $(s, p) \in A_S$, let x_{sp} equal the number of clauses in \bar{C}^* that contain literal s . For all $(p, t) \in A_T$, let $x_{pt} = 1$ if clause t is in \bar{C}^* , and let $x_{pt} = 0$, otherwise.

Obviously, x satisfies all capacity constraints in D^{\min} . Since the total flows entering and leaving pool p both equal the number of clauses in \bar{C}^* containing Boolean variable p or its negation, conservation of flow is also satisfied.

Choose any $t \in T$ corresponding to a clause in \bar{C}^* , and denote its adjacent pools p_1 and p_2 , respectively. We need to show that the quality bounds (7) at t are satisfied, and because $x_{p_1t} = x_{p_2t} = 1$, these constraints read

$$2\ell_t^k \leq w_{p_1}^k + w_{p_2}^k \leq 2u_t^k \quad k \in K. \tag{18}$$

Denote by s_1 and s_2 (k_1 and k_2) the sources (quality parameters) corresponding to the literals in clause t . Adjacent to p_v ($v \in \{1, 2\}$) are the sources s_v and s'_v , where s'_v corresponds to the inverse of literal s_v . Since all clauses in \bar{C}^* are satisfied, and literal s_v is in one such clause, literal s'_v is not in any clause in \bar{C}^* . By definition of x , it follows that $x_{s_p v} = 0$ for $s = s'_v$, and thereby p_v receives flow uniquely from s_v . The quality at p_v is hence identical to the quality at s_v , which means that $w_{p_v}^k = 2$ if $k = k_v$, and $w_{p_v}^k = 0$, otherwise. Consequently, $w_{p_1}^k + w_{p_2}^k = 2$ if $k \in \{k_1, k_2\}$, and $w_{p_1}^k + w_{p_2}^k = 0$, otherwise. By definition, $\ell_t^k = u_t^k = 1$ if $k \in \{k_1, k_2\}$, and $\ell_t^k = u_t^k = 0$, otherwise, proving (18), and thereby feasibility of x . The proof is complete by observing that $|\bar{C}^*| = |T[x]|$ and $\sum_{p \in P_t} x_{pt} = 2$ for all $t \in T[x]$. \square

Theorem 7 *The POOLING PROBLEM with all in-degrees no greater than two is strongly NP-hard.*

Proof The proof is by a polynomial reduction from the decision version of MINIMUM 2-SATISFIABILITY. Let $\zeta = -2z$. It follows from Lemma 9 that if there exists a truth assignment rendering at least z conjunctive clauses in \bar{C} true, then there exists a flow in D^{\min} such that at least $2z$ flow units enter T . Conversely, if $2z$ flow units can be sent from S to T in D^{\min} , at least z terminals receive positive flow since $b_t = 2$ ($t \in T$), and Lemma 8 shows that \bar{C} contains z conjunctive clauses that can be satisfied simultaneously. It follows that at least z conjunctive clauses can be satisfied simultaneously, if and only if there exists a feasible flow in D^{\min} with cost at most ζ . The proof is complete by the NP-hardness of MINIMUM 2-SATISFIABILITY [19], and the observation that all POOLING PROBLEM parameters are integers with absolute values bounded by $2m$. \square

6 Conclusions and open questions

The POOLING PROBLEM is proved to remain NP-hard in the strong sense even for instances with only one quality parameter. Adding the restriction that the networks cannot have more than two sources and two terminals, still leaves us with an NP-hard problem. Strong hardness is however not proved, and it is left as a challenge for further research to determine whether pseudo-polynomial algorithms exist when $|K| = 1$ and $|S| = |T| = 2$. If only one of the sets S and T is restricted to have only two elements, however, the POOLING PROBLEM with a single quality parameter is proved to be strongly NP-hard.

Instances with $|S| = |T| = 2$ belong obviously to the instance class where the minimum of the in- and out-degrees of each pool is at most two. The implied NP-hardness of this class contrasts the observation that the problem is an LP if the minimum degree of each pool is one. Further, it has been proved that the POOLING PROBLEM is strongly NP-hard if all sources and pools have out-degree at most 2, or if all pools and terminals have in-degree at most two.

It is proved that for single-pool networks, the problem is solvable in terms of a polynomial number of compact LPs if either $|K|$ or $|T|$ is bounded. However, the question whether a bound on $|S|$ also makes the problem solvable in polynomial time is left as a challenge for future research. If that turns out to be the case, new related and interesting questions arise: Do there exist polynomial algorithms for $|P| \leq p_{\max}$ in combination with a bound on $\min\{|S|, |T|, |K|\}$ also for any $p_{\max} > 1$? Future research should address questions of this kind, in order to assess closely the border lines between polynomially solvable and NP-hard instance classes of the POOLING PROBLEM.

Acknowledgments This article was written while the author was visiting Department of Computer Architecture, University of Málaga, Spain. Invitation and support from Prof. Eligius M.T. Hendrix are gratefully acknowledged.

References

1. Adhya, N., Tawarmalani, M., Sahinidis, N.V.: A Lagrangian approach to the pooling problem. *Ind. Eng. Chem. Res.* **38**(5), 1965–1972 (1999)
2. Alfaki, M., Haugland, D.: Strong formulations for the pooling problem. *J. Glob. Optim.* **56**(3), 897–916 (2013)
3. Al-Khayyal, F.A., Falk, J.E.: Jointly constrained biconvex programming. *Math. Oper. Res.* **8**(2), 273–286 (1983)
4. Almutairi, H., Elhedhli, S.: A new Lagrangian approach to the pooling problem. *J. Glob. Optim.* **45**(2), 237–257 (2009)

5. Audet, C., Brimberg, J., Hansen, P., Le Digabel, S., Mladenović, N.: Pooling problem: alternate formulations and solution methods. *Manag. Sci.* **50**(6), 761–776 (2004)
6. Baker, T.E., Lasdon, L.S.: Successive linear programming at Exxon. *Manag. Sci.* **31**(3), 264–274 (1985)
7. Ben-Tal, A., Eiger, G., Gershovitz, V.: Global minimization by reducing the duality gap. *Math. Program.* **63**(1–3), 193–212 (1994)
8. DeWitt, C.W., Lasdon, L.S., Waren, A.D., Brenner, D.A., Melham, S.: OMEGA: an improved gasoline blending system for Texaco. *Interfaces* **19**(1), 85–101 (1989)
9. Dey, S., Gupte, A.: Analysis of MILP techniques for the pooling problem. *Oper. Res.* **62**(2), 412–427 (2015)
10. Floudas, C.A., Visweswaran, V.: A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: I. *Theory Comput. Chem. Eng.* **14**(12), 1397–1417 (1990)
11. Foulds, L.R., Haugland, D., Jörnsten, K.: A bilinear approach to the pooling problem. *Optimization* **24**, 165–180 (1992)
12. Galan, B., Grossmann, I.E.: Optimal design of distributed wastewater treatment networks. *Ind. Eng. Chem. Res.* **37**(10), 4036–4048 (1998)
13. Garey, M.R., Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, New York (1979)
14. Garey, M.R., Johnson, D.S., Stockmeyer, L.: Some simplified NP-complete graph problems. *Theor. Comput. Sci.* **1**, 237–267 (1976)
15. Gupte, A., Ahmed, S., Dey, S., Cheon, M.: Pooling problems: an overview. In: Furman, K., Song, J. (eds.) *Optimization and Analytics in the Oil and Gas Industry*, International Series in Operations Research and Management Science. Springer, Berlin (2015)
16. Haverly, C.A.: Studies of the behavior of recursion for the pooling problem. *ACM SIGMAP Bull.* **25**, 19–28 (1978)
17. Haverly, C.A.: Behavior of recursion models—more studies. *ACM SIGMAP Bull.* **26**, 22–28 (1979)
18. Kallrath, J.: Solving planning and design problems in the process industry using mixed integer and global optimization. *Ann. Oper. Res.* **140**(1), 339–373 (2005)
19. Kohli, R., Krishnamurti, R., Mirchandani, P.: The minimum satisfiability problem. *Discrete Math.* **7**, 275–283 (1994)
20. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Math. Program.* **10**(1), 147–175 (1976)
21. Misener, R., Floudas, C.A.: Advances for the pooling problem: modeling, global optimization, and computational studies. *Appl. Comput. Math.* **8**, 3–22 (2009)
22. Misener, R., Floudas, C.A.: Global optimization of large-scale generalized pooling problems: quadratically constrained MINLP models. *Ind. Eng. Chem. Res.* **49**, 5424–5438 (2010)
23. Misener, R., Thompson, J.P., Floudas, C.A.: APOGEE: global optimization of standard, generalized, and extended pooling problems via linear and logarithmic partitioning schemes. *Comput. Chem. Eng.* **35**, 876–892 (2011)
24. Sahinidis, N.V., Tawarmalani, M.: Accelerating branch-and-bound through a modeling language construct for relaxation-specific constraints. *J. Glob. Optim.* **32**(2), 259–280 (2005)
25. Visweswaran, V., Floudas, C.A.: Computational results for an efficient implementation of the GOP algorithm and its variants. In: Grossmann, I.E. (ed.) *Global Optimization in Chemical Engineering*, pp. 111–153. Kluwer, Dordrecht (1996)