A polynomial algorithm for the earthwork allocation problem with borrow and waste site selection

Hüseyin Güden¹ and Haldun Süral^{2*}

¹Department of Industrial Engineering, Eastern Mediterranean University,

Famagusta, North Cyprus via Mersin 10, Turkey; and ²Department of Industrial Engineering, Middle East Technical University, 06800 Ankara, Turkey

In road construction projects, earthwork is planned together with horizontal and vertical alignments. This study focuses on earthwork operations that basically include cutting the hills and filling the holes on the road path. The candidate borrow and waste sites can also be used to obtain or heap earth when the available cut and fill amounts are not balanced or operating these sites reduces the total earthwork cost. Total earthwork cost contains the transportation cost and the overall cost related to opening the candidate sites. The problem is to determine which borrow and waste sites to operate, and the earth flows between cut, fill, waste, and borrow sites such that the total cost is minimized. It is shown that the problem is a generalization of the well-known lot-sizing problem. A fixed charge network flow problem formulation is presented, and a polynomial time dynamic programming algorithm is developed for solving the problem.

Journal of the Operational Research Society (2017) **68**(9), 1085–1093. doi:10.1057/s41274-016-0140-0;

published online 1 December 2016

Keywords: road construction; dynamic programming; location; lot sizing; logistics

1. Introduction

The earthwork allocation problem with borrow and waste site selection (EAPS) addresses a practical location-allocation decision problem whose workflow and capacity management characteristics are quite different from those in the related logistics literature.

A road construction project mainly requires horizontal and vertical alignments and earthwork planning. Earthwork basically includes cutting the hills (cut/supply sites) and filling the holes (fill/demand sites) on the road path. Exact amounts of earth must be removed from the cut sites and filled into the fill sites. If the distances between cut and fill sites are long or the amounts of available earth supply and demand are not balanced, the open candidate sites are operated in order to match demand with supply and/or decrease the earth transportation costs. Figure 1 shows an illustration of possible sites for a road construction project. Note that candidate sites are called waste (borrow) under the over-supply (undersupply) case.

Opening a waste or borrow site incurs a fixed cost and a variable cost. Fixed cost includes all expenditures for preparations necessary to make the site usable and construction of slip roads necessary to reach the site. Variable cost includes loading/unloading expenditures and transportation costs on the slip road.

We formulate the EAPS on a line network $L = (V, E)$ where V is the node set $\{1, 2, 3, \dots, n\}$ and E is the edge set $\{(i, i + 1) | i = 1, 2, 3, \dots, n - 1\}$. Nodes in V represent cut, fill, waste, and borrow sites. Note that a waste (or borrow) site is represented as a node in V at the connection point of its slip road to the main road in construction. Although there may be limitations on the capacities of borrow and waste sites in practical cases, this study assumes that they are uncapacitated. Edge $(i, i + 1) \in E$ represents the road segment between nodes $i \in V$ and $(i + 1) \in V$. Let $C \subseteq V$ be the set of cut nodes, $F \subseteq V$ the set of fill nodes, $D \subseteq V$ the set of candidate waste nodes, and $Q \subseteq V$ the set of candidate borrow nodes. Let c_i be the amount of material that must be sent from cut node i and f_i be the amount of material that must be sent to fill node i . Let q_i be the fixed cost and p_i be the variable cost for node $i \in D \cup Q$. Let r_{ij} be the unit transportation cost between $i \in V$ and $j \in V$. Although different dipper dredgers and trucks can be used for loading/unloading operations and short-haul/long-haul transportation activities, which may require differentiating the related cost items, this study assumes that the variable transportation cost is a linear function of distances between nodes i and j .

The problem on L is to determine the set of waste nodes $D' \subseteq D$, the set of borrow nodes $Q' \subseteq Q$, and the amount of material flows between node pairs so that (a) the total material sent from Q' and C to $i \in F$ is equal to f_i , (b) the total material sent from $i \in C$ to D' and F is equal to c_i , and (c) the total of fixed and variable costs is minimized.

*Correspondence: Haldun Süral, Department of Industrial Engineering, Middle East Technical University, 06800 Ankara, Turkey.
E-mail: hsural@metu.edu.tr

Figure 1 An illustration for a road construction project.

Considering the diverse nature of cut and fill operations, stocking function of waste sites, and supplying functions of borrow sites, we assume that C , F , D , and Q are disjoint sets. All cost, distance, and supply/demand parameters are assumed to be nonnegative for simplicity.

The EAPS can be considered as a mixture of two oppositely structured location problems on a line. If $D = C = \emptyset$, i.e., there are no waste and cut nodes, it includes only fill nodes with known requirements on a line and candidate uncapacitated borrow nodes with fixed and variable costs. This is exactly the same problem as the uncapacitated facility location problem (UFLP) on a line in terms of matching demand with supply, where fill nodes represent customer locations with known demands and borrow nodes represent uncapacitated candidate facility locations. The problem is to determine the number and locations of facilities and allocations of customers to facilities such that the total cost is minimized. Now, consider the opposite case in which $Q = F = \emptyset$, i.e., there are no borrow and fill nodes. In this case, the EAPS contains only cut nodes and candidate waste nodes. The problem is almost the same as the previous one except that the flows are from customers to facilities.

The UFLP on a line network is equivalent to the uncapacitated lot-sizing problem (ULSP) with backlogging (ULSPB), which is polynomially solvable (Pochet and Wolsey, 1988). Although each of the above special cases of the EAPS is equivalent to the ULSPB, these two separate problems come together and merge on the same line in the EAPS setting so that one can move material from cut and borrow nodes to fill and waste nodes. However, even if we consider cut and borrow nodes as candidate facility locations, we actually do not make an opening decision about a cut node, but determine the material flows from these nodes. On the other hand, even if we consider fill and waste nodes as customer locations, we

actually decide to open a waste node and determine the material flow to these nodes.

The earthwork allocation problem without borrow and waste site selections is a polynomially solvable linear programming problem. In the EAPS, however, there are fixed costs related to borrow and waste sites so that a site can operate only if its fixed cost is paid. We present the fixed charge network flow problem formulation of the EAPS in which borrow and waste site selection decisions are represented by integer variables. It is shown that the EAPS is the generalization of the ULSP. To the best of our knowledge, this version of the ULSP is not studied before in the literature. Using the properties of the optimal solution to the EAPS, we develop a polynomial time dynamic programming algorithm and show that the EAPS is polynomially solvable.

There are various studies related to the EAPS in the road construction, location, lot sizing, distribution, and transportation literature. However, to the best of our knowledge, the relation between the EAPS and the known optimization problems is only *partially* examined. We review and present these relations in the second section. The third section gives the fixed charge network flow problem formulation of the EAPS. Optimal solution properties of the problem are discussed in the fourth section. Using the properties in the fourth section, the fifth section presents a polynomial time dynamic programming algorithm. The last section concludes the paper.

2. Literature review

There are several studies on the EAPS and its variants in the construction literature. According to Jayawardane and Harris

(1990), one of the common methods used in the literature is the mass-haul diagram and the earliest linear models in the literature are developed by Mayer and Stark (1981) and Nandgaonkar (1981). Easa (1987) and Easa (1988) consider stepwise and linear unit cost functions, respectively. Jayawardane and Harris (1990) consider equipment fleet combinations that affect the project duration and costs. Moreb (1996) considers the vertical alignment and the earthwork simultaneously. Hare *et al* (2011) consider removal of physical blocks such as rivers and trees so that removals and earthworks can be performed simultaneously at different sites on the road being constructed. Lima *et al* (2013) consider the earthwork and paving issues for the cases in which cut and borrow sites have different material properties, fill nodes demand a mix of different materials, and there are material mixing plants. Hare *et al* (2014) consider the vertical alignment and earthwork allocation problems simultaneously on a dynamic setting with features such as side slopes and physical blocks. They propose two mathematical models (i.e., transportation and network flow type models) in order to compute the cost and show that the flow type model is computationally faster. Burdett and Kozan (2014) and Burdett *et al* (2015) consider three-dimensional blocks in the EAP instead of considering sections. Son *et al* (2005) consider the earthwork allocation problem on a plane, rather than on a line. The earthwork area is not limited by roads. They consider the earthwork area in two dimensions (width and length) with smaller rectangular cells. The earth transportation among these cells is determined in order to minimize the total transportation distance. Moreb and Bafail (1994) consider a similar earthwork allocation problem with additional decisions about land leveling.

There are a few studies on the facility location problem on a line in the location literature, some of which are formulated as the p -median and fixed charged facility location problems. There are two main properties that make a location problem on a line easier and lead to efficient algorithms. The first one is the eligibility of non-fractional allocations of demands to facilities, and the second one is to set identical capacities at all facilities. Uncapacitated cases guarantee the validity of the first property. Love (1976) proposes a dynamic programming algorithm to solve the p -median problem. Brimberg and Reville (1998) consider the UFLP and the p -median problem and show that the linear relaxations of their mixed integer models give the integer optimal solution. Hsu *et al* (1997) propose an $O(pn^2)$ algorithm for solving the p -facility location problem with n candidate location sites, fixed location costs, and unimodal service cost structures. Brimberg and Mehrez (2001) solve the location and sizing problems of facilities by dynamic programming. Facilities may reach any capacity level at the expense of a fixed cost, which is a continuous non-decreasing function of the capacity. As a result of this capacity-cost relation, the first property is guaranteed. Brimberg *et al* (2001) investigate the effect of capacity constraints on the continuous location-allocation problem with at most p facilities. The objective is to minimize the total fixed and

transportation costs. They propose a dynamic programming algorithm when the unit transportation cost is an increasing convex function of the distance and the second property is valid. They show that the problem is NP-hard under more general cost structures. Eben-Chaime *et al* (2002) consider the fixed charged capacitated location-allocation problem and use heuristic solution methods to solve the problem. Mirchandani *et al* (1996) consider the capacitated facility location problem where facilities must be located within a given neighborhood of customers. Fixed and service costs depend upon their locations on the line. They develop dynamic programming algorithms for (i) locating minimum cost facilities to serve all customers and (ii) maximizing the profit by locating at most p facilities that may not serve all customers. None of these studies consider the supply-demand flows similar to the EAPS.

As we discussed above, the lot-sizing problem called ULSPB is polynomially solvable (Zangwill, 1969; Pochet and Wolsey, 1988, 2006). Zangwill (1969) is the first to formulate the ULSPB as the network flow problem and proposes an $O(n^3)$ dynamic programming algorithm, where n is the number of periods. Pochet and Wolsey (1988) provide a shortest path formulation with $O(n)$ nodes, requiring $O(n^2)$ operations for computing arc weights, which make the ULSPB solvable in $O(n^2)$. The lot-sizing literature also contains several extensions of the problem including capacities, multiple products, multi levels, and start-ups (e.g., Agra and Constantino, 1999; Cheng *et al*, 2001; Miller *et al*, 2003; Denzel and Sürat, 2005). A few extensions of the core problem include product returns from the customers and disposals of excess inventory. The generic version of this extension is a special case of the EAPS (see Beltran and Krass, 2002). All other studies add features to the problem such as remanufacturing of returned products, production capacities, or multiple products. Note that when backlogging is included in the lot-sizing problem with returns and disposals, the resulting problem is equivalent to the EAPS. To the authors' best knowledge, there is no such study in this literature.

There are several studies on the transportation problem with fixed costs of flows between supply and demand sites (see Adlakha and Kowalski, 1999). However, to the best of our knowledge, there is no study on the transportation problem that incorporates supply and demand location decisions in addition to the initially given sites.

3. The fixed charge network flow problem formulation

The EAPS can be represented as the fixed charge network flow problem (FNFP) on a directed graph $G = (N, A)$ where $N = \{0\} \cup V$ and $A = \{(i, i+1) \mid i = 1, 2, \dots, n-1\} \cup \{(i+1, i) \mid i = 1, 2, \dots, n-1\} \cup \{(0, i) \mid i \in Q\} \cup \{(i, 0) \mid i \in D\}$. Fixed and variable charges of using arcs $(0, i)$ and $(i, 0)$ in G are equal to q_i and p_i , respectively, where node i corresponds to a candidate borrow node for the former case

and waste node for the latter case. Variable charges of using both arcs $(i, i + 1)$ and $(i + 1, i)$ for $i = 1, 2, \dots, n - 1$ are equal to $r_{i,i+1}$. The amount of supply (demand) of $i \in C$ ($i \in F$) is equal to c_i (f_i). Supply c_i and demand f_i are zero for $i \in D \cup Q$. For node 0, the supply amount is equal to $\max\{0, \sum_{i \in C} c_i - \sum_{i \in F} f_i\}$ and the demand amount is equal to $\max\{0, \sum_{i \in F} f_i - \sum_{i \in C} c_i\}$. Decision variables are defined as follows:

- u_i is the amount of forward material flow from node i to $(i + 1)$ for $i = 1, \dots, (n - 1)$. Similarly, v_i is the amount of backward material flow from node $(i + 1)$ to i for $i = 1, \dots, (n - 1)$.
- γ_i is the amount of material obtained from borrow node i for $i \in Q$. Similarly, β_i is the amount of material heaped to waste node i for $i \in D$.
- y_i is equal to 1 if the candidate waste (borrow) node i is used, 0 otherwise for $i \in D$ ($i \in Q$).

Figure 2 illustrates the graph of an EAPS instance with $n = 9$, $C = \{3, 8\}$, $F = \{1, 6, 9\}$, $D = \{4, 7\}$, and $Q = \{2, 5\}$. The directions of arcs show the directions of corresponding flows. For example, u_4 represents the amount of forward material moved from node 4 to node 5, while v_3 represents the amount of reversed material moved from node 4 to node 3. Note that the arc with demand f_i (supply c_i) is a downward (upward) arc.

3.1. FNFP formulation

$$\text{Min} \sum_{i=1}^{n-1} r_{i,i+1} (u_i + v_i) + \sum_{i \in D \cup Q} q_i y_i + \sum_{i \in D} p_i \beta_i + \sum_{i \in Q} p_i \gamma_i \quad (1)$$

s.t.

$$u_{i-1} + v_i + c_i = u_i + v_{i-1} \quad \forall i \in C \quad (2)$$

$$u_{i-1} + v_i = f_i + u_i + v_{i-1} \quad \forall i \in F \quad (3)$$

$$u_{i-1} + v_i + \gamma_i = u_i + v_{i-1} \quad \forall i \in Q \quad (4)$$

$$u_{i-1} + v_i = \beta_i + u_i + v_{i-1} \quad \forall i \in D \quad (5)$$

$$\beta_i \leq \sum_{j \in C} c_j y_i \quad \forall i \in D \quad (6)$$

$$\gamma_i \leq \sum_{j \in F} f_j y_i \quad \forall i \in Q \quad (7)$$

$$u_i, v_i \geq 0 \quad i = 1, \dots, n - 1 \quad (8)$$

$$\beta_i \geq 0 \quad i \in D \quad (9)$$

$$\gamma_i \geq 0 \quad i \in Q \quad (10)$$

$$y_i \in \{0, 1\} \quad i \in D \cup Q \quad (11)$$

$$u_i = v_i = 0 \quad i = 0, n \quad (12)$$

The objective function (1) minimizes the total cost. Constraints (2)–(5) are material flow balance equations and guarantee that the exact amount of material is removed (filled) from (to) cut (fill) nodes, and the amount of material is obtained (heaped) from (to) borrow (waste) nodes, if necessary. Constraints (6) and (7) set corresponding binary variable equal to 1 if a shipment occurs from (to) borrow (waste) node. (8)–(12) are set and integrality constraints. Note that variables u_0 , v_0 , u_n , and v_n are fixed at zero, and the balance equation is not written for node 0 because it would be satisfied as a result of constraints (2)–(5).

4. Optimal solution properties of the EAPS

It is known that when the values of y variables in FNFP formulation are given, the FNFP reduces to the minimum cost network flow problem, whose optimal solution satisfies the acyclic graph property. Hence, an extreme solution of the FNFP has a tree structure as shown in Figure 3.

Example 1 in Figure 3a displays a possible extreme solution. In this example, borrow node 2 meets the demand at fill node 1, the demand at fill node 6 is satisfied from the supply at cut nodes 3 and 8, and the demand at fill node 9 is satisfied from the supply at cut node 8. The remaining supply at node 8 is sent to waste node 7. Note that no material flows from borrow site 5 to waste site 4, but a supply (cut) quantity from node 3 is sent to node 6 through nodes 4 and 5 for satisfying a part of demand (fill) at node 6.

Example 2 given in Figure 3b illustrates another extreme solution in which waste node 4 and borrow node 2 are open. Node 2 meets the demand at fill node 1, while the demands at fill nodes 6 and 9 are satisfied from the supply at node 8, and the supply at node 3 is sent to node 4. In this example, waste node 7 and borrow node 5 are not used. A supply (cut) from node 8 is sent to node 6 via node 7 for satisfying the demand (fill) at node 6.

Using the tree structure of extreme solutions of the FNFP, we specify the structure of optimal solutions to the EAPS in Observation 1.

Observation 1 There exists an optimal solution to the EAPS where (1) $u_i v_i = 0$ for all $1 \leq i < n$ and (2) $u_i = v_i = 0$ for at least one $i \in N$, $k \leq i < l$ where $\beta_k > 0$ or $\gamma_k > 0$ and $\beta_l > 0$ or $\gamma_l > 0$ for any $k, l \in D \cup Q$.

Definition If an extreme solution of the EAPS satisfies $u_{a-1} = v_{a-1} = u_b = v_b = 0$ and $u_i + v_i > 0$ for all i where $a \leq i < b$ when $1 \leq a < b \leq n$, then the set of all nodes i for $a \leq i \leq b$ is called *segment* $S[a, b]$.

Note that nodes 1 and 2 in Figure 3a form the first segment $S[1, 2]$, while nodes from 3 to 9 constitute the second segment

Figure 2 G for an EAPS instance with $n = 9$.

(a) Example 1

(b) Example 2

Figure 3 Two extreme solutions for the EAPS instance given in Figure 2.

$S[3,9]$ in the extreme solution of Example 1. $S[1,2]$, $S[3,4]$, $S[5,5]$, and $S[6,9]$ are the other segments in the extreme solution of Example 2 in Figure 3b.

Observation 1 implies that an optimal solution to the EAPS can be expressed as a set of sequential segments, $S[a_1, b_1]$, $S[a_2, b_2]$, $S[a_3, b_3]$, ..., $S[a_k, b_k]$, where $a_e, b_e \in N$, $a_e \leq b_e$ for

$\varepsilon = 1, \dots, k$ and $a_1 = 1, b_k = n, b_\varepsilon = a_{(\varepsilon+1)} - 1$ for $\varepsilon = 1, \dots, (k - 1)$. These segments correspond to “regeneration intervals” of the ULSPB (see Pochet and Wolsey, 2006) and generalize them since the ULSPB is a special case of the EAPS. An extreme solution for a nine-period ULSPB instance is illustrated in Figure 4, where $S[1,4]$ and $S[5,9]$ form the first and second regeneration intervals.

Observation 2 Let $c_{ab} = \sum_{i \in C}^{b=a} c_i$ and $f_{ab} = \sum_{i \in F}^{b=a} f_i$. A segment $S[a,b]$ of the EAPS satisfies the following properties.

1. One of the following three cases is valid:

Case 1 If $c_{ab} = f_{ab}$, then $\sum_{i \in D}^{b=a} y_i = \sum_{i \in Q}^{b=a} y_i = 0$,

Case 2 If $c_{ab} > f_{ab}$, then $\sum_{i \in D}^{b=a} y_i = 1$ and $\sum_{i \in Q}^{b=a} y_i = 0$, or

Case 3 If $c_{ab} < f_{ab}$, then $\sum_{i \in D}^{b=a} y_i = 0$ and $\sum_{i \in Q}^{b=a} y_i = 1$

2. Cut and fill nodes can partially or entirely satisfy each other's demand.
3. Any $i \in C \cup F$, $a \leq i \leq b$, can receive one or both of the forward and backward flows. For example, node 6 receives both forward and backward flows in Figure 3a.
4. Forward and backward flows occur in any order. Considering $S[3,9]$ in Figure 3a, it starts with three steps of forward flows, continues with two steps of backward flows, and ends with a forward flow.
5. Assume that the second or third case in Observation 2.1 is valid for $S[a,b]$. Let us divide $S[a,b]$ into two

parts, called left and right parts, assuming an open waste node or a borrow node in the center of two parts. The material on $S[a,b]$ flows from one part to the other. Considering Figure 3a, the material on $S[3,9]$ flows from the right to the left as the waste node 7 is open.

Below we match the five properties for the EAPS in Observation 2 with the corresponding properties of a regeneration interval $R[a,b]$ of the ULSPB in order to highlight the similarities and differences between two problems.

1. There is exactly one production period.
2. All demands are satisfied from the production period.
3. Demand in a period is entirely satisfied by only inventory, production, or backlogging.
4. Inventory and backlogging flows are separated on the subparts of the regeneration interval. Backlogging flows occur only on the left of the production period, and inventory flows occur only on the right of the production period.
5. There is no flow from the left of the production period to its right or vice versa.

5. A dynamic programming algorithm for the EAPS

For any segment $S[a,b]$ of an extreme solution to the EAPS, one can compute the associated amounts of material flows between (i) cut and fill nodes, (ii) cut nodes and open waste node, and (iii) fill nodes and open borrow node in $S[a,b]$ plus (iv) the corresponding total cost TC_{ab} of these flows in $S[a,b]$. Let D^{ab} and Q^{ab} be the sets of candidate waste and borrow nodes in $S[a,b]$, respectively. We know that one of the following three cases occur according to Observation 2.1.

Figure 4 An extreme solution example for an ULSPB instance with the demand d_i in period i .

Case 1 If $c_{ab} = f_{ab}$, then there are no open borrow and waste nodes in $S[a, b]$. The associated amounts of material flows in $S[a, b]$ can be computed for each j , where $a \leq j \leq b$ due to Observation 1.1 as $v_j = \max\{0, f_{aj} - c_{aj}\}$ and $u_j = \max\{0, c_{aj} - f_{aj}\}$. The corresponding total cost is

$$TC_{ab} = \sum_{j=a}^{b-1} r_{j,j+1}(u_j + v_j).$$

Case 2 If $c_{ab} > f_{ab}$, then there is only one open waste node (say at node m). The amount of material sent to waste node m is equal to the difference between the cut and fill amounts in $S[a, b]$. Thus, we have $\beta_m = c_{ab} - f_{ab}$, $v_j = \max\{0, f_{aj} - c_{aj}\}$ for $a \leq j \leq m-1$, $v_j = \max\{0, f_{aj} + \beta_m - c_{aj}\}$ for $m \leq j \leq b$, $u_j = \max\{0, c_{aj} - f_{aj}\}$ for $a \leq j \leq m-1$, and $u_j = \max\{0, c_{aj} - f_{aj} - \beta_m\}$ for $m \leq j \leq b$. The corresponding total cost is $TC_{ab}^m = q_m + p_m(\beta_m) + \sum_{j=a}^{b-1} r_{j,j+1}(u_j + v_j)$. Note that locating a waste node in $S[a, b]$ can be decided by computing TC_{ab}^m values for all $m \in D^{ab}$ and selecting the minimum cost location among them. If $D^{ab} = \emptyset$, then $S[a, b]$ is not a valid segment of a basic feasible solution. We therefore redefine the total cost term as

$$TC_{ab} = \begin{cases} \min_{m \in D^{ab}} \{TC_{ab}^m\} & \text{If } D^{ab} \neq \emptyset \\ M & \text{If } D^{ab} = \emptyset, \end{cases}$$

where M is a very large positive number.

Case 3 If $c_{ab} < f_{ab}$, then there is only one open borrow node (say at node m). Although the computations for this case are similar to Case 2, we present all necessary steps for the sake of completeness. The amount of material obtained from borrow node m is equal to the difference between the cut and fill amounts in $S[a, b]$. So, we have $\gamma_m = f_{ab} - c_{ab}$, $v_j = \max\{0, f_{aj} - c_{aj}\}$ for $a \leq j \leq m-1$, $v_j = \max\{0, f_{aj} - c_{aj} - \gamma_m\}$ for $m \leq j \leq b$, $u_j = \max\{0, c_{aj} - f_{aj}\}$ for $a \leq j \leq m-1$, and $u_j = \max\{0, c_{aj} + \gamma_m - f_{aj}\}$ for $m \leq j \leq b$. The total cost is $TC_{ab}^m = q_m + p_m(\gamma_m) + \sum_{j=a}^{b-1} r_{j,j+1}(u_j + v_j)$. Locating or not

locating a borrow node in this segment can be decided as it has been defined in Case 2. First, compute TC_{ab}^m values for all $m \in Q^{ab}$ and choose the location node with the minimal cost value. Since $Q^{ab} = \emptyset$ makes the segment $S[a, b]$ invalid as discussed for Case 2, we have again

$$TC_{ab} = \begin{cases} \min_{m \in Q^{ab}} \{TC_{ab}^m\} & \text{If } Q^{ab} \neq \emptyset \\ M & \text{If } Q^{ab} = \emptyset \end{cases}$$

Consider the EAPS instance in Figure 2, where $f_1 = 10$, $f_6 = f_9 = 5$, $c_3 = 12$, $c_8 = 8$, $q_2 = q_4 = q_5 = q_7 = 100$, $p_2 = p_4 = p_5 = p_7 = 10$ and $r_{i,i+1} = 2$ for $i = 1, \dots, 8$. Below we illustrate necessary computations for the associated flows and costs of three candidate segments. Let us consider $S[1, 9]$ in Table 1, where $c_{19} = f_{19} = 20$. It implies that Case 1 is satisfied and there will be no open waste and borrow nodes. The corresponding total cost $TC_{1,9}$ equals to $2(10 + 10 + 2 + 2 + 2 + 3 + 3 + 5) = 74$.

Now, we consider $S[1, 8]$ in Table 2, where $c_{18} = 20 > f_{18} = 15$, which corresponds to Case 2 that opens a waste node. Let us assume that the waste node is opened at node 4, i.e., $m = 4$, and $\beta_4 = 5$. The resulting total cost $TC_{1,8}^4$ is obtained as $100 + (10)(5) + 2(10 + 10 + 2 + 3 + 3 + 8 + 8) = 238$.

Finally, we consider $S[1, 6]$ in Table 3, where $c_{16} = 12 < f_{16} = 15$, to illustrate opening a borrow node (Case 3). We assume that the borrow node is opened at node 2, i.e., $m = 2$, and $\gamma_2 = 3$. The total cost $TC_{1,6}^2$ is computed as $100 + 10 \times 3 + 2(10 + 7 + 5 + 5 + 5) = 194$.

Note that computational complexity of finding u and v values (forward and backward flow amounts) for a fixed a and b pair and any m is $O(b-a)$. Similarly, complexity of computing the total cost for a fixed a and b pair and any m is $O(b-a)$. Since these computations must be repeated for all $m \in D^{ab}$ and $m \in Q^{ab}$ if cases 2 and 3 are satisfied, respectively, the complexity of computing TC_{ab} is $O((b-a)(\max\{|D^{ab}|, |Q^{ab}|\}))$.

Table 1 Computations for the segment $S[1, 9]$

j	1	2	3	4	5	6	7	8	9
c_{1j}	0	0	12	12	12	12	12	20	20
f_{1j}	10	10	10	10	10	15	15	15	20
u_j	0	0	2	2	2	0	0	5	0
v_j	10	10	0	0	0	3	3	0	0

Table 2 Computations for the segment $S[1, 8]$

j	1	2	3	4	5	6	7	8
c_{1j}	0	0	12	12	12	12	12	20
f_{1j}	10	10	10	10	10	15	15	15
u_j	0	0	2	0	0	0	0	0
v_j	10	10	0	3	3	8	8	0

Table 3 Computations for the segment $S[1,6]$

j	1	2	3	4	5	6
c_{1j}	0	0	12	12	12	12
f_{1j}	10	10	10	10	10	15
u_j	0	0	5	5	5	0
v_j	10	7	0	0	0	0

Let G_b be the optimal objective function value of the subproblem including only the first b nodes. Let $G_0 = 0$. Compute $G_b = \min_{1 \leq a \leq b} \{G_{a-1} + TC_{ab}\}$ for $b = 1, 2, 3, \dots, n$ sequentially. G_n gives the value of an optimal solution. The optimal solution can be constructed by backtracking.

Since the complexity of computing TC_{ab} is $O((b-a)(\max\{|D^{ab}|, |Q^{ab}|\}))$ and these computations are done for all pairs of a and b ($1 \leq a \leq b \leq n$) in the above forward DP algorithm, the total computational complexity results in $O(n^3 \max\{|D|, |Q|\})$, or simply $O(n^4)$.

Given that the EAPS is generalization of the ULSPB, we now examine whether the complexity of the above algorithm can simply be reduced from $O(n^4)$ to $O(n^3)$ time by using the properties of the regeneration intervals in the ULSPB although they are not valid for the segments in the EAPS (see Observation 2).

Consider nodes $(a-1), a, \dots, b, (b+1) \in V, S[a,b]$, and $S[a, (b+1)]$ ($S[(a-1), b]$) where node $(b+1)$ ($(a-1)$) is added to $S[a,b]$ as the last (first) node instead of b (a) for an EAPS instance. We assume that the material flow amounts and the associated total cost for the segment $S[a,b]$ are given, and we explore $S[a, (b+1)]$ to compute its material flows and the associated total cost with performing a constant number of operations using the knowledge about $S[a,b]$.

Note that $S[a,b]$ and the new segment $S[a, (b+1)]$ may fit different cases in Observation 2.1, which make the information at hand useless for necessary computations as flows, open node types, and their locations are changing. Similarly, if both $S[a,b]$ and $S[a, (b+1)]$ fit the second (or third) case in Observation 2.1, the number of operations needed to compute the flows and cost in the new segment would be affected largely by the size (i.e., the number of nodes) of the part between waste node or borrow node and node $(b+1)$. The reason is that when node $(b+1)$ is added to $S[a,b]$, the flows on this part may increase, decrease, or even change their directions (i.e., some forward flows can turn to backward flows or vice versa). It follows that all flows and the total cost on this part need to be computed from scratch.

Now assume that the original segment $S[a,b]$ fits the first case in Observation 2.1. If $(b+1)$ is a cut or fill node, $S[a, (b+1)]$ cannot fit the first case because the total cut and fill amounts cannot be equal to each other after adding $(b+1)$. If $(b+1)$ is a candidate waste node or borrow node, then node $(b+1)$ would not be opened and no material flow would occur between $[a,b]$ and $(b+1)$ since the total cut and fill amounts on $[a,b]$ are equal to each other. We thus can

conclude that the number of operations needed to compute the flows and cost in a new segment would be affected from the number of nodes and it would not be trivial to perform a constant number of operations using the original information at hand. This remark is also valid for the case of adding node $(a-1)$ to $S[a,b]$ as the first node to have the new segment $S[(a-1), b]$.

Now, consider an ULSPB instance with n periods. Let p_k and d_k be unit production cost and demand in period k , respectively. Consider a regeneration interval $R[a,b]$. Let i be the production period in $R[a,b]$ and TC_{ab} be the total cost associated with $R[a,b]$. For a new regeneration interval $R[a, (b+1)]$, the material flow on $R[a,b]$ would not be affected from adding $(b+1)$ because all periods in $R[a,b]$ remain as they are. The periods before the production period satisfy their demands by backlogging, the production period satisfies its demand from itself by production, and the periods after the production period satisfy their demands by inventory. After adding period $(b+1)$ to the regeneration interval, it satisfies its entire demand from the production period by inventory. Hence, the total cost in the new situation can be computed in a constant number of operations, which is independent of the size of $R[a,b]$ by just adding the term $(d_{(b+1)}(p_i + h_{i,(b+1)}))$, where $h_{i,(b+1)}$ is the unit inventory holding cost from production period i to period $(b+1)$. The case of adding the period $(a-1)$ to $R[a,b]$ as the first period is very similar to the case of adding period $(b+1)$ as the last node. In this case, the additional cost term is $(d_{a-1}(p_i + s_{(a-1),i}))$, where $s_{(a-1),i}$ is the unit backlogging cost from period $(a-1)$ to production period i . Hence, the total cost in the new regeneration interval can also be computed in a constant number of operations, which is independent of the size of $R[a,b]$.

Pochet and Wolsey (1988) use the property of the ULSPB related to Observation 2.5 and consider regeneration interval $R[a,b]$ in three parts in their $O(n^2)$ algorithm for the ULSPB: $[a,i]$, $[i]$, and $(i,b]$, where i is the production period. They show that computing the associated costs for $[a,i]$, $[i]$, and $(i,b]$ requires $O(n^2)$ operations for all (a,i) , $O(n)$ operations for all i , and $O(n^2)$ operations for all (i,b) , respectively. Thus, the ULSPB is converted to a shortest path problem instance solvable in $O(n^2)$ time. However, according to Observation 2.5, a material flow occurs from one part to another part in the segment of the EAPS so that the flow affects the amount of demand on the opposite part, which would be satisfied from the open waste or borrow node, and its total cost. Therefore, such a division of the segments into the parts is not immediately applicable for the segments of the EAPS.

6. Conclusion

The earthwork allocation problem with borrow and waste site selection appears in road construction projects and is studied in different forms in the construction literature. We show that the EAPS is an extension of the well-known uncapacitated lot-sizing problem with backloging and can be solved in $O(n^4)$ time. Extensions of the solution algorithm to the EAPS with capacitated sites and/or nonlinear transportation costs can be considered as future studies. Besides, the current state-of-the-art for road design is to consider vertical alignment and earthwork simultaneously. Future research can explore whether extending the EAPS to include vertical alignment optimization can still be achieved in polynomial time.

Acknowledgments—Authors would like to thank Cüneyt Uğur who motivated the present study by his excellent term project in Engineering Management Program at the Department of Industrial Engineering of METU, Ankara.

References

- Adlakha V and Kowalski K (1999). On the fixed charge transportation problem. *Omega* **27**(3):381–388.
- Agra A and Constantino M (1999). Lotsizing with backloging and start-ups: The case of Wagner–Whitin costs. *Operations Research Letters* **25**(2):81–88.
- Beltran JL and Krass D (2002). Dynamic lot sizing with returning items and disposals. *IIE Transactions* **34**(5):437–448.
- Brimberg J, Korach E, Eben-Chaime M and Mehrez A (2001). The capacitated p -facility location problem on the real line. *International Transactions in Operational Research* **8**(6):727–738.
- Brimberg J and Mehrez A (2001). Location and sizing of facilities on a line. *TOP* **9**(2):271–280.
- Brimberg J and ReVelle C (1998). Solving the plant location problem on a line by linear programming. *TOP* **6**(2):277–286.
- Burdett RL and Kozan E (2014). An integrated approach for earthwork allocation, sequencing and routing. *European Journal of Operational Research* **238**(3):741–759.
- Burdett RL, Kozan E and Kenly R (2015). Block models for improved allocation planning in linear infrastructure construction. *Engineering Optimization*, **47**(3):347–369.
- Cheng CH, Madan MS, Gupta Y and So S (2001). Solving the capacitated lot-sizing problem with backorder consideration. *The Journal of the Operational Research Society* **52**(8):952–959.
- Denizel M and Süral H (2005). On alternative mixed integer programming formulations and LP-based heuristics for lot-sizing with setup times. *The Journal of the Operational Research Society* **57**(4):389–399.
- Easa SM (1987). Earthwork allocations with nonconstant unit costs. *Journal of Construction Engineering and Management* **113**(1):34–50.
- Easa SM (1988). Earthwork allocations with linear unit costs. *Journal of Construction Engineering and Management* **114**(4):641–655.
- Eben-Chaime M, Mehrez A and Markovich G (2002). Capacitated location-allocation problems on a line. *Computers and Operations Research* **29**(5):459–470.
- Hare LW, Koch VR and Lucet Y (2011). Models and algorithms to improve earthwork operations in road design using mixed integer linear programming. *European Journal of Operational Research* **215**(2):470–480.
- Hare W, Hossain S, Lucet Y and Rahman F (2014). Models and strategies for efficiently determining an optimal vertical alignment of roads. *Computers and Operations Research* **44**(1):161–173.
- Hsu VN, Lowe TJ and Tamir A (1997). Structured p -facility location problems on the line solvable in polynomial time. *Operations Research Letters* **21**(4):159–164.
- Jayawardane A and Harris F (1990). Further development of integer programming in earthwork optimization. *Journal of Construction Engineering and Management* **116**(1):18–34.
- Lima RX, Junior EFN, Prata BA and Weissmann J (2013). Distribution of materials in road earthmoving and paving: Mathematical programming approach. *Journal of Construction Engineering and Management* **139**(8):1046–1054.
- Love RF (1976). One dimensional facility location-allocation using dynamic programming. *Management Science* **22**(5):614–617.
- Mayer R and Stark R (1981). Earthmoving logistics. *Journal of the Construction Division* **107**(2):297–312.
- Miller A, Nemhauser GL, and Savelsbergh MWP (2003). A multi-item production planning model with setup times: Algorithms, reformulations, and polyhedral characterizations for a special case. *Mathematical Programming B* **95**(1):71–90.
- Mirchandani P, Kohli R and Tamir A (1996). Capacitated location problems on a line. *Transportation Science* **30**(1):75–80.
- Moreb AA (1996). Linear programming model for finding optimal roadway grades that minimize earthwork cost. *European Journal of Operational Research* **93**(1):148–154.
- Moreb AA and Bafail AO (1994). A linear programming model combining land levelling and transportation problems. *The Journal of the Operational Research Society* **43**(12):1418–1424.
- Nandgaonkar SM (1981). Earthwork transportation, allocations: Operations research. *Journal of the Construction Division* **107**(1):373–392.
- Pochet Y and Wolsey LA (1988). Lot size models with backloging: Strong formulations and cutting planes. *Mathematical Programming* **40**(1):317–335.
- Pochet Y and Wolsey LA (2006). *Production planning by mixed integer programming*. Springer: New York.
- Son J, Mattila KG and Myers DS (2005). Determination of haul distance and direction in mass excavation. *Journal of Construction Engineering and Management* **131**(3):302–309.
- Zangwill WI (1969). A backloging model and a multi-echelon model of a dynamic economic lot size production system—a network approach. *Management Science* **15**(9):506–527.

Received 7 October 2014;
accepted 18 October 2016