An active-set based recursive approach for solving convex isotonic regression with generalized order restrictions

Xuyu Chen[*], Xudong Li[†], Yangfeng Su[‡]

April 4, 2023

Abstract

This paper studies the convex isotonic regression with generalized order restrictions induced by a directed tree. The proposed model covers various intriguing optimization problems with shape or order restrictions, including the generalized nearly isotonic optimization and the total variation on a tree. Inspired by the success of the pool-adjacent-violator algorithm and its active-set interpretation, we propose an active-set based recursive approach for solving the underlying model. Unlike the brute-force approach that traverses an exponential number of possible active-set combinations, our algorithm has a polynomial time computational complexity under mild assumptions.

Keywords: Active set methods; convex isotonic regression; generalized order restrictions
AMS subject classifications: 90C25, 90C30

1 Introduction

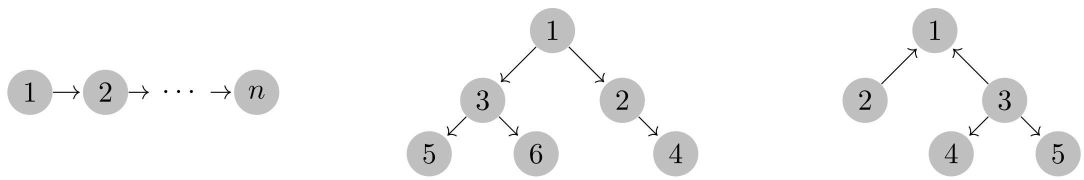
Given a directed tree $G = (V, E)$, we consider the following convex isotonic regression problem with generalized order restrictions:

$$\min_{x\in\Re^{|V|}} \sum_{i\in V} f_i(x_i) + \sum_{(i,j)\in E} \lambda_{i,j}(x_i - x_j)_+ + \sum_{(i,j)\in E} \mu_{i,j}(x_j - x_i)_+, \tag{1}$$

where for each $i \in V$, $f_i : \Re \to \Re$ is a convex loss function, $\lambda_{i,j}$ and $\mu_{i,j}$ for $(i,j) \in E$, are possibly infinite nonnegative scalars, i.e., $0 \le \lambda_{i,j}, \mu_{i,j} \le +\infty$, and $(x)_+ = \max(0, x)$ is the nonnegative part of x for any $x \in \Re$. In (1), when $\lambda_{i,j} = +\infty$ (respectively, $\mu_{i,j} = +\infty$), the corresponding term $\lambda_{i,j}(x_i - x_j)_+$ (respectively, $\mu_{i,j}(x_j - x_i)_+$) should be understood as the indicator function $\delta(x_i, x_j \mid x_i - x_j \le 0)$ (respectively, $\delta(x_i, x_j \mid x_i - x_j \ge 0)$), or equivalently the constraint $x_i - x_j \le 0$ (respectively, $x_i - x_j \ge 0$). See Figure 1 for some simple examples of directed trees.

As one can observe, the involvement of the directed tree G makes problem (1) a rather general model containing many interesting variants as special cases. Here, for simplicity, we only mention

[*]School of Mathematical Sciences, Fudan University, Shanghai, 200433, China, chenxy18@fudan.edu.cn
[†]School of Data Science, Fudan University, Shanghai, 200433, China, lixudong@fudan.edu.cn
[‡]School of Mathematical Sciences, Fudan University, Shanghai, 200433, China, yfsu@fudan.edu.cn



(a) chain (b) arborescence (c) general directed tree

Figure 1: Examples of directed trees. A directed tree is a directed graph whose underlying graph is a tree, and the directed trees are also referred to as directed acyclic graphs.

two of them. The first one is the *generalized nearly isotonic optimization* (GNIO) problem proposed in [26]:

$$\min_{x\in\Re^n} \sum_{i=1}^{n} f_i(x_i) + \sum_{i=1}^{n-1} \lambda_i(x_i - x_{i+1})_+ + \sum_{i=1}^{n-1} \mu_i(x_{i+1} - x_i)_+, \tag{2}$$

which is clearly a special case of (1) with G chosen as a chain, as illustrated in Figure 1a. As is mentioned in [26], model (2) recovers, as special cases, many classic problems in shape restricted statistical regression, including isotonic regression [6, 7], unimodal regression [12, 21], and nearly isotonic regression [22]. The second one is the *total variation on a tree* considered in [15]:

$$\min_{x\in\Re^{|V|}} \sum_{i\in V} f_i(x_i) + \sum_{(i,j)\in E} w_{i,j}|x_i - x_j|, \tag{3}$$

where $G = (V, E)$ is a directed tree and each f_i is assumed to be piecewise linear or piecewise quadratic. Other special cases of model (1) have also been examined in the literature, for example, [8, 25] studied the isotonic regression problems with partial order restrictions induced by an arborescence. These special cases, as well as their applications in statistic inference [20], operations research [1], signal processing [17, 9], medical prognosis [19], and traffic and climate data analysis [16, 24], reveal the importance and necessity of studying model (1).

To the best of our knowledge, there is currently no efficient algorithm available for directly solving the general model (1). However, certain special cases of the model can be solved by existing algorithms. For example, the GNIO problem (2) can be efficiently solved by employing a dynamic programming approach designed in [26]. Moreover, assuming boundedness of the decision variables, the KKT based fast algorithm proposed in [13] can also solve the GNIO problem. However, both algorithms rely heavily on the underlying chain structure, and therefore cannot be applied to solve the general model (1) that involves a directed tree. If G is a chain and each f_i is quadratic, the total variation problem (3) reduces to the well-known ℓ_2 total variation denoising problem, which has been extensively studied in signal processing [10, 14]. The direct algorithm [10] and the taut-string algorithms [2] are considered to be the state-of-the-art for solving the ℓ_2 total variation denoising problem. Meanwhile, if G is assumed to be a directed tree and each f_i is assumed to be continuous piecewise linear or piecewise quadratic with a finite number of breakpoints in (3), the message passing algorithm studied in [15] can be applied. However, these algorithms can not handle problem (1) with general convex loss functions f_i involved.

There is also another line of work dedicated to solving special cases of problem (1). In the 1950s, Ayer in [1] proposed the famous *Pool-Adjacent-Violator algorithm* (PAVA) for solving the following isotonic regression problem:

$$\begin{aligned} \min_{x \in \Re^n} \quad & \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2, \\ \text{s.t.} \quad & x_1 \leq x_2 \leq \ldots \leq x_n, \end{aligned} \tag{4}$$

which is clearly a special case of problem (1). The PAVA has been widely regarded as the state-of-the-art technique for solving the isotonic regression problem since its inception. Later in [4], Best and Chakravarti discovered that the PAVA is, in fact, a dual feasible active set method for solving (4). In [5], the PAVA was generalized to handle (4) but with the least squares objectives replaced by general separable convex loss functions. In [25], Yu and Xing further generalized the PAVA to solve convex separable minimization with order constraints induced by an arborescence. However, the generalized regularizers present in the objective of model (1) were not studied in [25]. As far as we know, it remains unclear whether the ideas behind the PAVA can be adopted to solve the more general model (1).

Encouraged by the successes of the PAVA and its variants in solving special cases of the generalized convex isotonic regression problem (1), we propose a novel active-set based algorithm in this paper. Our approach differs from the brute-force method that explores a potentially exponential number of different active sets. Instead, a recursive approach is proposed to accelerate the search for the desired active sets. We show that problem (1) can be tackled via recursively solving a sequence of smaller subproblems. For these subproblems, special recursive structures of the corresponding Karush-Kuhn-Tucker (KKT) conditions are carefully examined, which further allows us to design a novel active-set based recursive approach (ASRA). In particular, this approach enables us to derive semi-closed formulas of the optimal solutions to the aforementioned recursive subproblems. Under mild assumptions, we further show that the ASRA enjoys a polynomial time computational complexity for solving problem (1).

The subsequent sections of this paper are organized as follows. Section 2 covers the necessary preliminaries associated with problem (1), including fundamental concepts in graph theory and the corresponding KKT conditions. In addition, we describe a naive active-set method to solve (1). Our recursive approach, the ASRA, is described in detail in Section 3. Finally, we conclude the paper in Section 4. The Appendix includes an example of how to apply the ASRA to solve a simple instance of (1).

2 Preliminaries

We start with some relevant preliminaries in graph theory. A directed tree $G = (V, E)$ is a directed graph whose underlying graph is a tree, and an arborescence (also known as rooted directed tree) [11, 23] is a directed tree with exactly one node of zero in-degree. The node is also referred to as the *root* of the arborescence. Let $G = (V, E)$ and $B = (V_B, E_B)$ be two directed trees. If $V_B \subseteq V$ and $E_B \subseteq E$, then we say that B is a *subtree* of G, denoted by $B \subset G$. Two subtrees are *disjoint* if their node sets are disjoint. Given $P = \{B_k\}_{k=1}^K$ as a collection of disjoint subtrees of a certain directed tree $G = (V, E)$, if $V = \cup_{k=1}^K V_{B_k}$, then P is said to be a *partition* of G.

For a given directed tree $G = (V, E)$, we can choose any node $l \in V$ as the *ancestor* of G. Then, for any $i, j \in V$, we say that j is a child of i, denoted by $j \triangleleft i$, if the undirected path connecting l and i is strictly contained in the one connecting l and j. For example, if we pick the node 2 as the ancestor in the directed tree presented in Figure 1c, then we have $3 \lhd 1 \lhd 2$. Now, let $D \in \Re^{|V|\times|E|}$ be the node-arc incidence matrix associated with G. We know from [3] that $\text{rank}(D) = |E|$ and the matrix $\tilde{D}_l \in \Re^{|E|\times|E|}$ obtained by deleting the l-th row from D is invertible. Given a vector $b \in \Re^{|E|}$, we obtain in the following lemma a closed-form formula for the solution to the linear system $\tilde{D}_l z = b$.

Lemma 1. *For any given $b \in \Re^{|E|}$, the unique solution $z^* = (z_{i,j})_{(i,j)\in E} \in \Re^{|E|}$ to the linear system $\tilde{D}_l z = b$ takes the following form:*

$$z^*_{i,j} = \begin{cases} \sum_{k\in C_i} b_k, & \text{if } i \lhd j, \\ -\sum_{k\in C_j} b_k, & \text{if } j \lhd i, \end{cases} \quad \forall\, (i,j) \in E,$$

where for any node i, C_i consists of i and all its children, i.e., $C_i := \{j \in V \mid j \lhd i\} \cup \{i\}$.

Proof. This result is a simple consequence of the special structure of the node-arc incidence matrix and can be verified directly. □

Next, we state the blanket assumption on the loss functions f_i, $i \in V$, and derive the KKT conditions associated with problem (1). To express our main ideas clearly, we put strong assumptions on f_i, such as strong convexity and differentiability. However, as can be observed, these strong assumptions could be removed if more subtle analysis is employed.

Assumption 1. *Each $f_i : \Re \to \Re$, $i \in V$ in (1) is differentiable and strongly convex.*

From the strong convexity of each f_i, we know that the objective function in problem (1) is also strongly convex and therefore level-set bounded. Moreover, by [18, Theorems 27.1 and 27.2], problem (1) has a unique solution. We also note that Assumption 1 holds in some statistical and machine learning problems [4, 10, 22]. Under Assumption 1, we know from [18] that each f_i^* is also a strongly convex differentiable function. Moreover, both f_i' and $(f_i^*)'$ are strictly increasing on $\Re$, and for any given $x, y \in \Re$, $y = f_i'(x)$ if and only if $x = (f_i^*)'(y)$.

Now, we are ready to write down the KKT conditions associated with problem (1). For $0 \le \lambda, \mu \le +\infty$, let

$$\begin{cases} h^-_\lambda(x) := \delta(x \mid x \ge 0), & \text{if } \lambda = +\infty, \\ h^-_\lambda(x) := \begin{cases} -\lambda x, & x < 0, \\ 0, & x \ge 0, \end{cases} & \text{if } 0 \le \lambda < +\infty, \end{cases} \quad \text{and} \quad \begin{cases} h^+_\mu(x) := \delta(x \mid x \le 0), & \text{if } \mu = +\infty, \\ h^+_\mu(x) := \begin{cases} 0, & x \le 0, \\ \mu x, & x > 0, \end{cases} & \text{if } 0 \le \mu < +\infty. \end{cases}$$

For $(i,j) \in E$, we define $h_{i,j} : \Re \to [0, +\infty]$ by

$$h_{i,j}(x) := h^-_{\lambda_{i,j}}(x) + h^+_{\mu_{i,j}}(x), \quad \forall\, x \in \Re.$$

Clearly, for each $(i,j) \in E$, $h_{i,j}$ is convex and its subdifferential at $x \in \Re$ takes the following form:

$$\partial h_{i,j}(x) = \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x < 0, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x = 0, \\ \{\mu_{i,j}\}, & \text{if } x > 0. \end{cases} \tag{5}$$

Here, $\partial h_{i,j}(x) = \{+\infty\}$ or $\partial h_{i,j}(x) = \{-\infty\}$ should be understood as $\partial h_{i,j}(x) = \emptyset$. We also adopt the conventions in (5) that $[-\infty, +\infty] = (-\infty, +\infty)$, $[-\infty, \alpha] = (-\infty, \alpha]$, and $[\alpha, +\infty] = [\alpha, +\infty)$ for some $\alpha \in \Re$.

Define $H(z) := \sum_{(i,j)\in E} h_{i,j}(z_{i,j})$ for $z \in \Re^{|E|}$, and $F(x) := \sum_{i\in V} f_i(x_i)$ for $x \in \Re^{|V|}$. Let $M = -D^T \in \Re^{|E|\times|V|}$, where D is the node-arc incidence matrix associated with G. That is, for $e = (i,j) \in E$, $M(e,i) = -1$ and $M(e,j) = 1$ and all other entries of M are zero. Let $H_M(x) := H(Mx)$ for $x \in \Re^{|V|}$. Then, it can be easily verified that problem (1) can be equivalently rewritten as

$$\min_{x\in\Re^{|V|}} \; F(x) + H_M(x).$$

Then, we have the following lemma on the KKT conditions associated with problem (1).

Lemma 2. *Problem* (1) *has a unique minimizer* $x^* \in \Re^{|V|}$*. Moreover,* x^* *solves problem* (1) *if and only if there exists a unique multiplier* $z^* \in \Re^{|E|}$*, such that* (x^*, z^*) *satisfies the following KKT system:*

$$\begin{aligned} &\sum_{k:(i,k)\in E} z^*_{i,k} - \sum_{k:(k,i)\in E} z^*_{k,i} = f_i'(x^*_i), \quad \forall\, i \in V, \\ &z^*_{i,j} \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x^*_i > x^*_j, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x^*_i = x^*_j, \\ \{\mu_{i,j}\}, & \text{if } x^*_i < x^*_j, \end{cases} \quad \forall\, (i,j) \in E. \end{aligned} \tag{6}$$

Proof. The existence and the uniqueness of the optimal solution to problem (1) follows from the the strong convexity of F. Since F is differentiable, we know from [18, Theorem 23.8] that

$$0 \in F'(x^*) + \partial H_M(x^*).$$

From [18, Theorem 23.9], it can be seen that $\partial H_M(x^*) = M^T \partial H(Mx^*)$. Thus, there exists $z^* \in \partial H(Mx^*)$, such that

$$F'(x^*) + M^T z^* = F'(x^*) - Dz^* = 0. \tag{7}$$

Since the e-th entry of Mx^* is given by $x^*_j - x^*_i$, we have from (5) that

$$z^*_{i,j} \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x^*_j - x^*_i < 0, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x^*_j - x^*_i = 0, \\ \{\mu_{i,j}\}, & \text{if } x^*_j - x^*_i > 0, \end{cases} \quad \forall\, (i,j) \in E.$$

Thus, we obtain the KKT conditions (6). The uniqueness of z^* follows from (7) and the fact that $\mathrm{rank}(D) = |E|$. We thus complete the proof. □

Next, we investigate a naive active set method for solving problem (1). For each edge $(i,j) \in E$, we can associate it with a sign $\# \in \{<, =, >\}$ to obtain a triple $(i,j,\#)$ representing the relation $x_i \# x_j$. For the consistency, when dealing with edges (i,j) with $\lambda_{i,j} = +\infty$ (or $\mu_{i,j} = +\infty$), the corresponding sign $\#$ can only be chosen from $\{<, =\}$ (or $\{>, =\}$). We denote by $\mathcal{A}$ the collection of all these triples and term it as an *active set* associated with problem (1). Then, the active set $\mathcal{A}$ induces the following $\mathcal{A}$-reduced problem from (1):

$$\begin{aligned} \min_{x\in\Re^{|V|}} \quad & \sum_{i\in V} f_i(x_i) + \sum_{(i,j)\in\mathcal{A}_>} \lambda_{i,j}(x_i - x_j) + \sum_{(i,j)\in\mathcal{A}_<} \mu_{i,j}(x_j - x_i), \\ \text{s.t.} \quad & x_i = x_j, \quad \forall\, (i,j) \in \mathcal{A}_=, \end{aligned} \tag{8}$$

where $\mathcal{A}_\# := \{(i,j) \mid (i,j,\#) \in \mathcal{A}\}$. If $\mathcal{A}_= = \emptyset$, then (8) reduces to an unconstrained optimization problem, which can be efficiently solved since its objective function is separable, smooth and strongly convex. For $i,j \in V$, we say they are $\mathcal{A}$-*connected* if and only if there exists an undirected path in $\mathcal{A}_=$, which is obtained by treating all edges in $\mathcal{A}_=$ as undirected edges, that connects i and j. Let $P_\mathcal{A}$ be the collection of all $\mathcal{A}$-connected components of G. Then, it is not difficult to observe that $P_\mathcal{A}$ is naturally a partition of G. We thus term $P_\mathcal{A}$ as the *partition induced by* $\mathcal{A}$. Without loss of generality, assume $P_\mathcal{A} = \{B_k\}_{k=1}^K$ with each B_k being a subtree of G, we see that the $\mathcal{A}$-reduced problem (8) can be decoupled into K independent subproblems as follows:

$$\min_{x \in \Re^{|V_{B_k}|}} \left\{ \sum_{i \in V_{B_k}} \hat{f}_i(x_i) \mid x_i = x_j, \, \forall\, (i,j) \in E_{B_k} \right\}, \quad 1 \le k \le K, \tag{9}$$

where for each $i \in V_{B_k}$,

$$\hat{f}_i(x_i) := f_i(x_i) + (\sum_{j:(i,j)\in\mathcal{A}_>} \lambda_{i,j} - \sum_{j:(i,j)\in\mathcal{A}_<} \mu_{i,j})x_i + (\sum_{l:(l,i)\in\mathcal{A}_<} \mu_{l,i} - \sum_{l:(l,i)\in\mathcal{A}_>} \lambda_{l,i})x_i.$$

Clearly, the simple constraints in problem (9) can be eliminated. The resulting unconstrained optimization problem has a univariate smooth and strongly convex objective function and thus can be efficiently solved. In this way, we obtain the optimal solution to the $\mathcal{A}$-reduced problem (8).

Unfortunately, there can be up to $3^{|E|}$ different choices for the active set $\mathcal{A}$. Thus, the naive method of exploring all the possible choices of different active sets needs to solve exponential number of $\mathcal{A}$-reduced problems. In order to reduce this prohibitive computational costs, we introduce a novel active-set based recursive algorithm in the next section.

3 An recursive algorithm for solving problem (1)

In this section, we present our recursive algorithm for solving problem (1). We first claim that, without loss of generality, the directed tree G in (1) can be assumed to be an arborescence with the node 1 to be its root. Moreover, we can decompose G into a sequence of subtrees $\{G_m = (V_m, E_m)\}_{m=1}^n$, where $G_1 \subset G_2 \subset \cdots \subset G_n = G$ and $V_m = \{1,2,\ldots,m\}$ for $1 \le m \le n$, and the set of edges $E_{m+1} \setminus E_m$ contains exactly one edge $(i_m, m+1)$, where $i_m \in V_m$. Further details are deferred to the Appendix.

For each $1 \le m \le n$, problem (1), when restricted to the the subtree G_m, takes the following form:

$$\min_{x\in\Re^{|V_m|}} \sum_{i\in V_m} f_i(x_i) + \sum_{(i,j)\in E_m} \lambda_{i,j}(x_i - x_j)_+ + \sum_{(i,j)\in E_m} \mu_{i,j}(x_j - x_i)_+. \tag{10}$$

From Lemma 2, it is not difficult to see that the unique primal-dual optimal pair to problem (10), denote by $(x^{(m)}, z^{(m)}) \in \Re^{|V_m|} \times \Re^{|E_m|}$, satisfies the following KKT system:

$$\begin{aligned} &\sum_{k:(i,k)\in E_m} z_{i,k} - \sum_{k:(k,i)\in E_m} z_{k,i} = f_i'(x_i), \quad \forall\, i \in V_m, \\ &z_{i,j} \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x_i > x_j, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x_i = x_j, \\ \{\mu_{i,j}\}, & \text{if } x_i < x_j, \end{cases} \quad \forall\, (i,j) \in E_m. \end{aligned} \tag{11}$$

The unique optimal pair $(x^{(m)}, z^{(m)})$ is also referred to as the G_m-*optimal pair* for convenience. By carefully exploiting the special structures in the KKT conditions (11), we propose to solve problem (1) in a recursive fashion. Specifically, we will recursively generate the G_{m+1}-optimal pair $(x^{m+1}, z^{(m+1)})$ from the G_m-optimal pair $(x^{(m)}, z^{(m)})$ for $m = 1, \ldots, n-1$.

We summarize the detailed steps of the above recursive approach in Algorithm 1. In the algorithm, the *generate* subroutine is designed to generate the G_{m+1}-optimal pair from the G_m-optimal pair. In the next subsection, we will show that this procedure is accomplished via a novel active-set searching scheme. Hence, it is natural for us to call Algorithm 1 an active-set based recursive approach (ASRA).

Algorithm 1 ASRA: An active-set based recursive approach for solving problem (1)

1: **Initialize:** $x_1^{(1)} = (f_1^*)'(0) \in \Re$, and $z^{(1)} = \emptyset$
2: **for** $m = 1, \ldots, n-1$ **do**
3: $\quad (x^{(m+1)}, z^{(m+1)}) = \textit{generate}(x^{(m)}, z^{(m)}, G_{m+1})$
4: **end for**
5: **Return:** $(x^{(n)}, z^{(n)}) \in \Re^n \times \Re^{n-1}$

3.1 The *generate* subroutine

To efficiently obtain the G_{m+1}-optimal pair from the given G_m-optimal pair, we shall investigated the KKT conditions associated with the subproblem induced by the subtree G_{m+1}. Specially, it takes the following form:

$$\sum_{k:(i,k)\in E_m} z_{i,k} - \sum_{k:(k,i)\in E_m} z_{k,i} = f_i'(x_i), \quad \forall\, i \in V_m \backslash \{i_m\}, \tag{12}$$

$$z_{i,j} \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x_i > x_j, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x_i = x_j, \quad \forall\, (i,j) \in E_m, \\ \{\mu_{i,j}\}, & \text{if } x_i < x_j, \end{cases} \tag{13}$$

$$\sum_{k:(i_m,k)\in E_m} z_{i_m,k} - \sum_{k:(k,i_m)\in E_m} z_{k,i_m} + z_{i_m,m+1} = f_{i_m}'(x_{i_m}), \tag{14}$$

$$- z_{i_m,m+1} = f_{m+1}'(x_{m+1}), \tag{15}$$

$$z_{i_m,m+1} \in \begin{cases} \{-\lambda_{i_m,m+1}\}, & \text{if } x_{i_m} > x_{m+1}, \\ [-\lambda_{i_m,m+1}, \mu_{i,j}], & \text{if } x_{i_m} = x_{m+1}, \\ \{\mu_{i_m,m+1}\}, & \text{if } x_{i_m} < x_{m+1}. \end{cases} \tag{16}$$

As one can observe, instead of writing the KKT conditions as a whole set of equations, we have singled out those, namely (14), (15) and (16), associated with the dual variable $z_{i_m,m+1}$, which corresponds to the newly added edge $\{(i_m, m+1)\} = E_{m+1} \setminus E_m$. Based on the above KKT conditions, we have the following proposition regarding the sign of $z_{i_m,m+1}$.

Proposition 1. *It holds that* $z_{i_m,m+1}^{(m+1)} f_{m+1}'(x_{i_m}^{(m)}) \leq 0$, *where* $(x^{(m)}, z^{(m)})$ *and* $(x^{(m+1)}, z^{(m+1)})$ *are the* G_m*-optimal pair and the* G_{m+1}*-optimal pair, respectively.*

Proof. Note that when $f'_{m+1}(x^{(m)}_{i_m}) = 0$, the desired result naturally holds. For the remaining parts, we only prove the case where $f'_{m+1}(x^{(m)}_{i_m}) > 0$, since the proof for the case with $f'_{m+1}(x^{(m)}_{i_m}) < 0$ can be easily modified from the arguments here.

Suppose that $f'_{m+1}(x^{(m)}_{i_m}) > 0$, then we shall prove that $z^{(m+1)}_{i_m,m+1} \leq 0$. Assume on the contrary that $z^{(m+1)}_{i_m,m+1} > 0$. Then, from (15), we have $x^{(m+1)}_{m+1} = (f^*_{m+1})'(-z^{(m+1)}_{i_m,m+1}) < (f^*_{m+1})'(0)$. Moreover, (16) implies that $x^{(m+1)}_{m+1} \geq x^{(m+1)}_{i_m}$. Thus, we have from the strict monotonicity of $(f^*_{m+1})'$ the following inequality:

$$x^{(m)}_{i_m} > (f^*_{m+1})'(0) > x^{(m+1)}_{m+1} \geq x^{(m+1)}_{i_m}. \tag{17}$$

Now, from (12), (13), and (14), we see that $\widetilde{x} \in \Re^{|V_m|}$ with $\widetilde{x}_i = x^{(m+1)}_i$ for $i \in V_m$ is the optimal solution to the following optimization problem:

$$\min_{x\in\Re^{|V_m|}} F_1(x) := \sum_{i\in V_m} f_i(x_i) + \sum_{(i,j)\in E_m} \{\lambda_{i,j}(x_i - x_j)_+ + \mu_{i,j}(x_j - x_i)_+\} - z^{(m+1)}_{i_m,m+1} x_{i_m}.$$

Meanwhile, since $(x^{(m)}, z^{(m)})$ is the G_m-optimal pair, $x^{(m)}$ is the optimal solution to the following optimization problem:

$$\min_{x\in\Re^{|V_m|}} F_0(x) := \sum_{i\in V_m} f_i(x_i) + \sum_{(i,j)\in E_m} \{\lambda_{i,j}(x_i - x_j)_+ + \mu_{i,j}(x_j - x_i)_+\}.$$

Then, it holds that

$$0 \geq F_1(\widetilde{x}) - F_1(x^{(m)}) = F_0(\widetilde{x}) - F_0(x^{(m)}) + z^{(m+1)}_{i_m,m+1}(x^{(m)}_{i_m} - \widetilde{x}_{i_m}).$$

Since $F_0(\widetilde{x}) - F_0(x^{(m)}) \geq 0$, $z^{(m+1)}_{i_m,m+1} > 0$, and $\widetilde{x}_{i_m} = x^{(m+1)}_{i_m}$, we have $x^{(m)}_{i_m} - x^{(m+1)}_{i_m} \leq 0$, which contradicts to (17). Thus, we have $z^{(m+1)}_{i_m,m+1} \leq 0$ and $z^{(m+1)}_{i_m,m+1} f'_{m+1}(x^{(m)}_{i_m}) \leq 0$, and complete the proof. □

From Proposition 1, we can determine the sign of $z^{(m+1)}_{i_m,m+1}$ by the value of $f'_{m+1}(x^{(m)}_{i_m})$. Moreover, if $f'_{m+1}(x^{(m)}_{i_m}) = 0$, we can easily construct the G_{m+1}-optimal pair as follows:

$$x^{(m+1)}_i = \begin{cases} x^{(m)}_i, \ \forall\, i \in V_m, \\ x^{(m)}_{i_m}, \ i = m+1, \end{cases} \quad \text{and} \quad z^{(m+1)}_{i,j} = \begin{cases} z^{(m)}_{i,j}, \ \forall\, (i,j) \in E_m, \\ 0, \ (i,j) = (i_m, m+1). \end{cases}$$

Hence, we focus on the case with $f'_{m+1}(x^{(m)}_{i_m}) \neq 0$ in the subsequent discussions. For this purpose, we consider the following parametric optimization problem with the parameter $t \in \Re$:

$$\min_{x\in\Re^{|V_{m+1}|}} \sum_{i\in V_{m+1}} f_i(x_i) + \sum_{(i,j)\in E_m} \{\lambda_{i,j}(x_i - x_j)_+ + \mu_{i,j}(x_j - x_i)_+\} - t(x_{i_m} - x_{m+1}), \tag{18}$$

whose KKT conditions are presented below:

$$
\begin{aligned}
&\sum_{k:(i,k)\in E_m} z_{i,k} - \sum_{k:(k,i)\in E_m} z_{k,i} + 1_{\{i=i_m\}} t = f_i'(x_i), \quad \forall\, i \in V_m, \\
&z_{i,j} \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x_i > x_j, \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x_i = x_j, \\ \{\mu_{i,j}\}, & \text{if } x_i < x_j, \end{cases} \quad \forall\, (i,j) \in E_m, \\
&-t = f_{m+1}'(x_{m+1}).
\end{aligned} \tag{19}
$$

Since each f_i is strongly convex, problem (18) has a unique optimal solution, denoted by $x^*(t)$, for each $t \in \Re$. Moreover, using the Fenchel-Rockafellar duality theorem [18] and the differentiability of each f_i, we know that there exists a unique dual optimal solution to problem (18), denoted by $z^*(t)$, which together with $x^*(t)$ satisfies the KKT conditions (19). If for certain $t^* \in \Re$, it holds that

$$
t^* \in \begin{cases} \{-\lambda_{i_m,m+1}\}, & \text{if } x_{i_m}^*(t^*) > x_{m+1}^*(t^*), \\ [-\lambda_{i_m,m+1}, \mu_{i_m,m+1}], & \text{if } x_{i_m}^*(t^*) = x_{m+1}^*(t^*), \\ \{\mu_{i_m,m+1}\}, & \text{if } x_{i_m}^*(t^*) < x_{m+1}^*(t^*). \end{cases} \tag{20}
$$

Then, by comparing the equations (19) and (20) and the KKT conditions in equations (12) to (16), we can obtain the G_{m+1}-optimal pair based on $(x^*(t^*), z^*(t^*))$. Indeed, the G_{m+1}-optimal pair $(x^{(m+1)}, z^{(m+1)})$ can be constructed via

$$
x^{(m+1)} = x^*(t^*), \text{ and } z_{i,j}^{(m+1)} = z_{i,j}^*(t^*) \text{ for } (i,j) \in E_m, \text{ and } z_{i_m,m+1}^{(m+1)} = t^*.
$$

This observation also indicates that one can determine the sign of t^* using Proposition 1.

To find the desired t^*, we start from the initial guess $t_0 = 0$. We note that when $t_0 = 0$, the corresponding primal-dual optimal pair $(x^*(t_0), z^*(t_0))$ is readily known with $x_i^*(t_0) = x_i^{(m)}$ for $i \in V_m$ and $x_{m+1}^*(t_0) = (f_{m+1}^*)'(-t_0)$, and $z^*(t_0) = z^{(m)}$. Then, we can easily check if $t_0 = 0$ satisfies (20) by comparing $x_{m+1}^*(t_0)$ and $x_{i_m}^*(t_0)$. If $x_{m+1}^*(t_0) \neq x_{i_m}^*(t_0)$, we can use Proposition 1 to determine if t should be decreased or increased. Assume without loss of the generality that $f_{m+1}'(x_{i_m}^*(t_0)) = f_{m+1}'(x_{i_m}^{(m)}) > 0$. From the above discussions and Proposition 1, we see that $t^* < 0$. Then, we rely on an active-set strategy to iteratively update our guess of t^*.

Starting from the initial guess $t_0 = 0$, we denote the active set corresponding to E_m in (18) by

$$
\mathcal{A}^0 = \{(i,j,\#) \mid (i,j) \in E_m,\ x_i^*(t_0) \,\#\, x_j^*(t_0)\}, \text{ where } \# \in \{<,=,>\}. \tag{21}
$$

Then, we add the equality constraints induced by edges in $\mathcal{A}^0_=$ to problem (18) and obtain the $\mathcal{A}^0$-reduced problem of problem (18). The key observation is that the primal-dual optimal solution pair to the $\mathcal{A}^0$-reduced problem can be written in a semi-closed form as functions of the parameter t, denoted by $(x^0(t), z^0(t))$. Then, we construct a dual candidate $\tilde{z}^0(t)$ to problem (18) as follows:

$$
\tilde{z}_{i,j}^0(t) = \begin{cases} z_{i,j}^0(t), & \text{if } (i,j) \in \mathcal{A}^0_=, \\ z_{i,j}^*(t_0), & \text{otherwise}, \end{cases} \quad \forall\, (i,j) \in E_m.
$$

We will show that if $(x^0(t), \tilde{z}^0(t))$ satisfies the complementarity conditions in (19), i.e.,

$$\tilde{z}^0_{i,j}(t) \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x^0_i(t) > x^0_j(t), \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x^0_i(t) = x^0_j(t), \quad \forall\, (i,j) \in E_m, \\ \{\mu_{i,j}\}, & \text{if } x^0_i(t) < x^0_j(t), \end{cases}$$

then $(x^0(t), \tilde{z}^0(t))$ is the primal-dual optimal solution pair to problem (18).

Based on this observation, a new guess of t^* is constructed by searching for the smallest possible t_1 such that $-\lambda_{i_m,m+1} \le t^* \le t_1 \le t_0 = 0$ and $(x^0(t_1), \tilde{z}^0(t_1))$ still satisfies the above complementarity conditions. Then, we have $(x^*(t_1), z^*(t_1)) = (x^0(t_1), \tilde{z}^0(t_1))$ and we can check if t_1 satisfies the system (20). If not, then a new active set $\mathcal{A}^1$ is constructed and the above process continues until t^* is found. In a nutshell, our approach is summarized in the following flowchart:

$$(t_0, x^*(t_0), z^*(t_0), \mathcal{A}^0) \Rightarrow \cdots \Rightarrow (t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q) \Rightarrow \cdots \Rightarrow (t^*, x^*(t^*), z^*(t^*), \mathcal{A}^*).$$

In what follows, we shall discuss the detailed steps of our procedure and we will prove that the search process of t^* terminates in at most $2m-1$ steps.

At t_q with $t^* < t_q \le t_0$, we assume that $(x^*(t_q),\, z^*(t_q))$, and the corresponding active set $\mathcal{A}^q$ are available. Then, we construct the following $\mathcal{A}^q$-reduced parametric optimization problem with parameter $t \in \Re$:

$$\begin{aligned} \min_{x\in\Re^{|V_{m+1}|}} \quad & \sum_{i\in V_{m+1}} f_i(x_i) + \sum_{(i,j)\in\mathcal{A}^q_>} \lambda_{i,j}(x_i - x_j) + \sum_{(i,j)\in\mathcal{A}^q_<} \mu_{i,j}(x_j - x_i) - t(x_{i_m} - x_{m+1}), \\ \text{s.t.} \quad & x_i = x_j, \quad \forall\, (i,j) \in \mathcal{A}^q_=, \end{aligned} \tag{22}$$

whose unique primal-dual optimal pair is denoted by $(x^q(t), z^q(t))$. If $\mathcal{A}^q_= = \emptyset$, then we set $z^q(t) = \emptyset$. Here, we require the following compatibility conditions between $\mathcal{A}^q$ and $(x^*(t_q), z^*(t_q))$, which also servers as an induction hypothesis.

Assumption 2. *The active set $\mathcal{A}^q$ and the primal-dual pair $(x^*(t_q), z^*(t_q))$ are compatible. That is, $x^*(t_q)$ is the optimal solution to the problem (22) at $t = t_q$, i.e., $x^q(t_q) = x^*(t_q)$ and the corresponding dual optimal solution $z^q(t_q)$ can be constructed via $z^q_{i,j}(t_q) = z^*_{i,j}(t_q)$ for $(i,j) \in \mathcal{A}^q_=$. Moreover, it holds that $x^*_{i_m}(t_q) - x^*_{m+1}(t_q) > 0$.*

We shall emphasize that according to the construction of $\mathcal{A}^0$, it is not difficult to observe that the active set $\mathcal{A}^0$ and the primal-dual pair $(x^*(t_0), z^*(t_0))$ are compatible, and $x^*_{i_m}(t_0) - x^*_{m+1}(t_0) > 0$. Next, we focus on obtaining $(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1})$ from $(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q)$.

We start by investigating the optimal primal-dual solution pair corresponding to problem (22). Particularly, instead of solving problem (22) for each $t \neq t_q$, we derive in the following proposition the semi-closed formulas for $(x^q(t), z^q(t))$ under Assumption 2. We also show that the optimal primal-dual solution pair of problem (18) can be obtained from $(x^q(t), z^q(t))$ provided that some complementarity conditions hold.

Proposition 2. *Let $P_{\mathcal{A}^q}$ be the partition of G_m induced by $\mathcal{A}^q$ and $B^q \in P_{\mathcal{A}^q}$ be the subtree such that $i_m \in B^q$. Then, under Assumption 2, for any $t \in \Re$, the primal optimal solution $x^q(t)$ takes* *the following form:*

$$\begin{cases} x_i^q(t) = x_i^*(t_q), \quad \forall\, i \in V_m \backslash V_{B^q}, \\ x_i^q(t) = \big((\sum_{i \in V_{B^q}} f_i)^*\big)'\Big(t + \beta^q\Big), \quad \forall\ i \in V_{B^q}, \\ x_{m+1}^q(t) = (f_{m+1}^*)'(-t), \end{cases} \tag{23}$$

where

$$\beta^q = \sum_{\substack{(i,k)\in E_m \\ i \in V_{B^q}, k \notin V_{B^q}}} z_{i,k}^*(t_q) - \sum_{\substack{(k,i)\in E_m \\ k \notin V_{B^q}, i \in V_{B^q}}} z_{k,i}^*(t_q).$$

Pick i_m as the ancestor of B^q. Then, for any $t \in \Re$, $z^q(t)$ is given by

$$\begin{cases} z_{i,j}^q(t) = z_{i,j}^*(t_q), \quad \forall (i,j) \in \mathcal{A}_=^q \backslash E_{B^q}, \\ z_{i,j}^q(t) = \begin{cases} \sum_{l \in C_i} f_l'(x_l^q(t)) - \alpha_{i,j}^q, & \text{if } i \lhd j, \\ \sum_{l \in C_j} -f_l'(x_l^q(t)) + \alpha_{i,j}^q, & \text{if } i \lhd j, \end{cases} \quad \forall (i,j) \in E_{B^q}, \end{cases} \tag{24}$$

where $C_i := \{j \in V_{B^q} \mid j \lhd i\} \cup \{i\}$ for any $i \in V_{B^q}$, and

$$\alpha_{i,j}^q = \begin{cases} \sum_{\substack{(l,k)\in E_m \\ l \in C_i, k \notin V_{B^q}}} z_{l,k}^*(t_q) - \sum_{\substack{(k,l)\in E_m \\ l \in C_i, k \notin V_{B^q}}} z_{k,l}^*(t_q), & \text{if } i \lhd j, \\ \sum_{\substack{(l,k)\in E_m \\ l \in C_j, k \notin V_{B^q}}} z_{l,k}^*(t_q) - \sum_{\substack{(k,l)\in E_m \\ l \in C_j, k \notin V_{B^q}}} z_{k,l}^*(t_q), & \text{if } j \lhd i, \end{cases} \quad \forall (i,j) \in E_{B^q}.$$

Let $\Omega^q = \{(i,j) \in E_m \backslash E_{B^q} \mid$ exactly one of i and j is in $V_{B^q}\}$. If

$$z_{i,j}^q(t) \in [-\lambda_{i,j}, \mu_{i,j}], \quad \forall\, (i,j) \in E_{B^q}, \tag{25}$$

$$z_{i,j}^*(t_q) \in \begin{cases} \{-\lambda_{i,j}\}, & \text{if } x_i^q(t) > x_j^q(t), \\ [-\lambda_{i,j}, \mu_{i,j}], & \text{if } x_i^q(t) = x_j^q(t), \\ \{\mu_{i,j}\}, & \text{if } x_i^q(t) < x_j^q(t), \end{cases} \quad \forall\, (i,j) \in \Omega^q, \tag{26}$$

then $(x^q(t), \tilde{z}^q(t))$ solves the KKT system (19), where

$$\tilde{z}_{i,j}^q(t) = \begin{cases} z_{i,j}^q(t), & \text{if } (i,j) \in \mathcal{A}_=^q, \\ z_{i,j}^*(t_q), & \text{otherwise}, \end{cases} \quad \forall\, (i,j) \in E_m. \tag{27}$$

Proof. Without loss of generality, we can assume that $P_{\mathcal{A}^q} = \{B_k\}_{k=1}^K \cup B^q$ where B_k, $1 \le k \le K$, and B^q are subtrees of G_m. Then, problem (22) can be decomposed into $K+2$ independent subproblems on each subtree B_k and B^q and the singleton $\{m+1\}$. Note that the parameter t only appears in the subproblems corresponding to the subtree B^q and the singleton $\{m+1\}$. Hence, from Assumption 2, it is not difficult to deduce that for any $t \in \Re$,

$$x_i^q(t) = x_i^*(t_q),\ i \in V_m \backslash V_{B^q}, \quad \text{and} \quad z_{i,j}^q(t) = z_{i,j}^*(t_q),\ (i,j) \in \mathcal{A}_{=}^q \setminus E_{B^q}.$$

The subproblem associated with $\{m+1\}$ is easily solved via $x_{m+1}^q(t) = (f_{m+1}^*)'(-t)$. Therefore, we only need to focus on the subproblem associated with the subtree B^q:

$$\min_{x \in \Re^{|V_{B^q}|}} \left\{ \sum_{i \in V_{B^q}} \hat{f}_i(x_i) - t x_{i_m} \mid x_i = x_j,\ \forall\, (i,j) \in E_{B^q} \right\}, \tag{28}$$

where

$$\hat{f}_i(x_i) := f_i(x_i) + \sum_{\substack{k \notin V_{B^q} \\ (k,i) \in E_m}} z_{k,i}^*(t_q) x_i - \sum_{\substack{k \notin V_{B^q} \\ (i,k) \in E_m}} z_{i,k}^*(t_q) x_i, \quad \forall\, i \in V_{B^q}.$$

Let $\mathcal{L}$ be the Lagrangian function associated with problem (28)

$$\mathcal{L}(x; z) = \sum_{i \in V_{B^q}} \hat{f}_i(x_i) - t x_{i_m} - \sum_{(i,j) \in E_{B^q}} z_{i,j}(x_i - x_j), \quad \forall\, (x,z) \in \Re^{|V_{B^q}|} \times \Re^{|E_{B^q}|}.$$

Then, the optimal primal-dual solution pair to problem (28) satisfies the following KKT system:

$$\begin{cases} x_i = x_j, \quad \forall\, (i,j) \in E_{B^q}, \\ f_i'(x_i) + \displaystyle\sum_{\substack{k \notin V_{B^q} \\ (k,i) \in E_m}} z_{k,i}^*(t_q) + \sum_{\substack{k \in V_{B^q} \\ (k,i) \in E_{B^q}}} z_{k,i} - \sum_{\substack{k \notin V_{B^q} \\ (i,k) \in E_m}} z_{i,k}^*(t_q) - \sum_{\substack{k \in V_{B^q} \\ (i,k) \in E_{B^q}}} z_{i,k} - 1_{\{i = i_m\}} t = 0, \quad \forall\, i \in V_{B^q}. \end{cases} \tag{29}$$

Summing over all $i \in V_{B^q}$, we deduce from the above system that

$$\sum_{i \in V_{B^q}} f_i'(x_i^q(t)) = - \sum_{\substack{(k,i) \in E_m \\ k \notin V_{B^q}, i \in V_{B^q}}} z_{k,i}^*(t_q) + \sum_{\substack{(i,k) \in E_m \\ i \in V_{B^q}, k \notin V_{B^q}}} z_{i,k}^*(t_q) + t,$$

i.e.,

$$x_i^q(t) = ((\sum_{i \in V_{B^q}} f_i)^*)'(t + \sum_{\substack{(i,k) \in E_m \\ i \in V_{B^q}, k \notin V_{B^q}}} z_{i,k}^*(t_q) - \sum_{\substack{(k,i) \in E_m \\ k \notin V_{B^q}, i \in V_{B^q}}} z_{k,i}^*(t_q)), \quad \forall\ i \in V_{B^q}.$$

Next, we obtain from the above KKT system (29) the following linear system corresponding to $z_{i,j}$ for $(i,j) \in E_{B^q}$:

$$\sum_{k:(i,k) \in E_{B^q}} z_{i,k} - \sum_{k:(k,i) \in E_{B^q}} z_{k,i} = f_i'(x_i^q(t)) + \sum_{\substack{k \notin V_{B^q} \\ (k,i) \in E_m}} z_{k,i}^*(t_q) - \sum_{\substack{k \notin V_{B^q} \\ (i,k) \in E_m}} z_{i,k}^*(t_q), \quad \forall\, i \in V_m \setminus \{i_m\}.$$

Since i_m is the ancestor of the subtree B, we obtain from Lemma 1 the updated formula for $z_{i,j}^q(t)$, $(i,j) \in E_{B^q}$. Thus, we proved (24).

Finally, it is not difficult to see that if the assumed conditions (25) and (26) are satisfied, then $x^q(t)$ and $\tilde{z}^q(t)$ satisfy the complementarity conditions in the KKT system (19). The rest equations in (19) hold automatically by noting (27) and the KKT system (29). □

Using the semi-closed formulas in Proposition 2, we compute the following lower bound $\Delta t_q \leq 0$:

$$\Delta t_q := \min\left\{\Delta t \mid \text{ (25) and (26) hold for all } t \in [t_q + \Delta t, t_q]\right\}.$$

The computations are divided into two parts. Firstly, we focus on the value of $z^q_{i,j}(t)$ for $(i,j) \in E_{B^q}$. For any $(i,j) \in E_{B^q}$, we note that $z^q_{i,j}(t_q) \in [-\lambda_{i,j}, \mu_{i,j}]$ and $z^q_{i,j}(t)$ is increasing if $i \triangleleft j$ and is decreasing if $j \triangleleft i$ with respect to t from (24). We define the threshold $\Delta(E_{B^q})$ as follows:

$$\Delta(E_{B^q}) := \begin{cases} \max\limits_{(i,j)\in E_{B^q}} \Delta t_{i,j}, & \text{if } E_{B^q} \neq \emptyset, \\ -\infty, & \text{otherwise.} \end{cases} \tag{30}$$

Here, each $\Delta t_{i,j} \leq 0$ solves

$$z^q_{i,j}(t_q + \Delta t_{i,j}) = -\lambda_{i,j}, \quad \text{if } i \triangleleft j, \quad \text{and} \quad z^q_{i,j}(t_q + \Delta t_{i,j}) = \mu_{i,j}, \quad \text{if } j \triangleleft i. \tag{31}$$

Next, the relations in (26) corresponding to the edges in Ω^q are examined. For this purpose, we divide Ω^q into two parts, namely,

$$\Omega^q_+ = \{(i,j) \in \Omega^q \mid i \in V_{B^q},\, j \in V_m \backslash V_{B^q}\} \text{ and } \Omega^q_- = \{(i,j) \in \Omega^q \mid i \in V_m \backslash V_{B^q},\, j \in V_{B^q}\}, \tag{32}$$

and handle them separately. From (23), we know that $x^q_i(t)$ takes the same value for all $i \in V_{B^q}$ and is increasing with respect to t. Hence, we can simply denote $x_{B^q}(t) = x^q_i(t)$ for any $i \in V_{B^q}$. Then, we compute the threshold $\Delta(\Omega^q) := \max\{\Delta(\Omega^q_+), \Delta(\Omega^q_-)\}$, where

$$\Delta(\Omega^q_+) := \begin{cases} \Delta\bar{t} \text{ satisfying } x_{B^q}(t_q + \Delta\bar{t}) = \max\limits_{(i,j)\in\Omega^q_+ \cap \mathcal{A}^q_>} x^*_j(t_q), & \text{if } \Omega^q_+ \cap \mathcal{A}^q_> \neq \emptyset, \\ -\infty, \text{ otherwise,} \end{cases} \tag{33}$$

and

$$\Delta(\Omega^q_-) := \begin{cases} \Delta\bar{t} \text{ satisfying } x_{B^q}(t_q + \Delta\bar{t}) = \max\limits_{(i,j)\in\Omega^q_- \cap \mathcal{A}^q_<} x^*_i(t_q), & \text{if } \Omega^q_- \cap \mathcal{A}^q_< \neq \emptyset, \\ -\infty, \text{ otherwise.} \end{cases} \tag{34}$$

It can be easily verified that

$$\Delta t_q = \max\{\Delta(E_{B^q}), \Delta(\Omega^q)\}. \tag{35}$$

Thus, using Proposition 2, we can obtain the semi-closed form for the optimal solution $x^*(t)$, as well as its corresponding dual optimal solution $z^*(t)$, to problem (18) for any $t \in [t_q + \Delta t_q, t_q]$.

Now, we are ready to discuss the search of t_{q+1}. Note that according to Assumption 2, we have

$$x^q_{i_m}(t_q) - x^q_{m+1}(t_q) = x^*_{i_m}(t_q) - x^*_{m+1}(t_q) > 0.$$

Using the closed-form formulas in Proposition 2, we know that $x^q_{i_m}(t) - x^q_{m+1}(t)$ is strictly increasing with respect to t, and we can obtain a unique $\Delta\widetilde{t}_q < 0$ via solving the following univariate nonlinear equation:

$$x^q_{i_m}(t_q + \Delta\widetilde{t}_q) - x^q_{m+1}(t_q + \Delta\widetilde{t}_q) = 0,$$

which is nothing but the optimality condition associated with the following univariate strongly convex optimization problem:

$$t_q + \Delta\widetilde{t}_q = \operatorname*{argmin}_t \left\{ (\sum_{i \in V_{B^q}} f_i)^*(t + \beta^q) + (f_{m+1}^*)(-t) \right\}.$$

The existence of $\Delta\widetilde{t}_q$ is thus guaranteed. Then, we set

$$t_{q+1} = \max\{t_q + \Delta t_q, t_q + \Delta\widetilde{t}_q, -\lambda_{i_m,m+1}\}. \tag{36}$$

As one can observe, it always holds that $t_{q+1} \in [t_q + \Delta t_q, t_q]$ and

$$\begin{aligned} x_{i_m}^*(t_{q+1}) - x_{m+1}^*(t_{q+1}) &= x_{i_m}^q(t_{q+1}) - x_{m+1}^q(t_{q+1}) \\ &\geq x_{i_m}^q(t_q + \Delta\widetilde{t}_q) - x_{m+1}^q(t_q + \Delta\widetilde{t}_q) = 0. \end{aligned} \tag{37}$$

Then, we reveal the relation between t_{q+1} and t^* in the following lemma.

Lemma 3. *It holds that $-\lambda_{i_m,m+1} \leq t^* \leq t_{q+1} \leq t_q \leq 0$. Moreover, $t_{q+1} = t^*$ if and only if $x_{i_m}^*(t_{q+1}) - x_{m+1}^*(t_{q+1}) = 0$ or $t_{q+1} = -\lambda_{i_m,m+1}$.*

Proof. If $t^* > t_{q+1}$, we have from (36) that $t^* > t_{q+1} \geq -\lambda_{i_m,m+1}$. It then follows from (20) that

$$x_{i_m}^q(t^*) - x_{m+1}^q(t^*) = x_{i_m}^*(t^*) - x_{m+1}^*(t^*) = 0.$$

However, we know from (37) and the strict monotonicity of $x_{i_m}^q(t) - x_{m+1}^q(t)$ that

$$x_{i_m}^q(t^*) - x_{m+1}^q(t^*) > x_{i_m}^q(t_{q+1}) - x_{m+1}^q(t_{q+1}) \geq 0.$$

We arrive at a contradiction. Thus, $t^* \leq t_{q+1}$.

Next, if $x_{i_m}^*(t_{q+1}) - x_{m+1}^*(t_{q+1}) = 0$ or $t_{q+1} = -\lambda_{i_m,m+1}$, one can easily verify that t_{q+1}, $x_{i_m}^*(t_{q+1})$ and $x_{m+1}^*(t_{q+1})$ satisfy (20), i.e., $t^* = t_{q+1}$. Conversely, if $t^* = t_{q+1}$, we have $t_{q+1} \geq -\lambda_{i_m,m+1}$. If $t_{q+1} > -\lambda_{i_m,m+1}$, it follows directly from (20) that $x_{i_m}^*(t^*) - x_{m+1}^*(t^*) = 0$. We thus complete the proof of the lemma. □

Remark 1. *It is only necessary to compute $\Delta\widetilde{t}_q$ at most once during the entire search process for t^*. Indeed, let*

$$\Delta_* := \begin{cases} x_{i_m}^q(t_q + \Delta t_q) - x_{m+1}^q(t_q + \Delta t_q), & \text{if } \Delta t_q > -\infty, \\ -\infty, & \text{otherwise.} \end{cases}$$

If $\Delta_* \geq 0$, then by the strict monotonicity of $x_{i_m}^q(t) - x_{m+1}^q(t)$, we must have $\Delta\widetilde{t}_q \leq \Delta t_q$. In this case, we can directly set

$$t_{q+1} = \max\{t_q + \Delta t_q, -\lambda_{i_m,m+1}\},$$

without computing $\Delta\widetilde{t}_q$. Only when $\Delta_* < 0$, we shall compute $\Delta\widetilde{t}_q$ and set

$$t_{q+1} = \max\{t_q + \Delta\widetilde{t}_q, -\lambda_{i_m,m+1}\}.$$

Then, from Lemma 3, it holds that $t_{q+1} = t^*$. Therefore, $\Delta\widetilde{t}_q$ only needs to be computed at most once.

If $t_{q+1} \neq t^*$, we know from (36), (37), and Lemma 3 that $t^* < t_{q+1}$ and

$$t_{q+1} = t_q + \Delta t_q, \quad \text{and} \quad x^*_{i_m}(t_{q+1}) - x^*_{m+1}(t_{q+1}) > 0. \tag{38}$$

Then, we give the details of the construction of $\mathcal{A}^{q+1}$. Let $\mathcal{M}(E_{B^q}) = \mathcal{M}(E^+_{B^q}) \cup \mathcal{M}(E^-_{B^q})$ with

$$\begin{cases} \mathcal{M}(E^+_{B^q}) &= \{(i,j) \in E_{B^q} \mid \Delta t_{i,j} = \Delta t_q, \text{ and } i \triangleleft j\}, \\ \mathcal{M}(E^-_{B^q}) &= \{(i,j) \in E_{B^q} \mid \Delta t_{i,j} = \Delta t_q, \text{ and } j \triangleleft i\}, \end{cases} \tag{39}$$

and $\mathcal{M}(\Omega^q) = \mathcal{M}(\Omega^q_+) \cup \mathcal{M}(\Omega^q_-)$ with

$$\begin{cases} \mathcal{M}(\Omega^q_+) = \{(i,j) \in \Omega^q_+ \cap \mathcal{A}(t_q)_> \mid x^q_i(t_q + \Delta t_q) = x^*_j(t_q)\}, \\ \mathcal{M}(\Omega^q_-) = \{(i,j) \in \Omega^q_- \cap \mathcal{A}(t_q)_< \mid x^q_j(t_q + \Delta t_q) = x^*_i(t_q)\}. \end{cases} \tag{40}$$

The active set $\mathcal{A}^{q+1}$ is constructed via

$$\begin{cases} \mathcal{A}^{q+1}_= = \big(\mathcal{A}^q_= \cup \mathcal{M}(\Omega^q)\big) \backslash \mathcal{M}(E_{B^q}), \\ \mathcal{A}^{q+1}_> = \big(\mathcal{A}^q_> \cup \mathcal{M}(E^+_{B^q})\big) \backslash \mathcal{M}(\Omega^q_+), \\ \mathcal{A}^{q+1}_< = \big(\mathcal{A}^q_< \cup \mathcal{M}(E^-_{B^q})\big) \backslash \mathcal{M}(\Omega^q_-). \end{cases} \tag{41}$$

Similar to (27), we can construct $\widetilde{z}^q(t_{q+1})$ from $z^q(t_{q+1})$ as follows:

$$\widetilde{z}^q_{i,j}(t_{q+1}) = \begin{cases} z^q_{i,j}(t_{q+1}), & \text{if } (i,j) \in \mathcal{A}^q_=, \\ z^*_{i,j}(t_q), & \text{otherwise}, \end{cases} \quad \forall\, (i,j) \in E_m.$$

Then, we obtain the optimal primal-dual solution pair $(x^*(t_{q+1}), z^*(t_{q+1})) = (x^q(t_{q+1}), \widetilde{z}^q(t_{q+1}))$ to problem (18) with $t = t_{q+1}$.

Next, it can be easily verified from the construction of $\mathcal{A}^{q+1}$ in (41), and the computation steps of t_{q+1} in (36) that the new active set $\mathcal{A}^{q+1}$ and the primal-dual pair $(x^*(t_{q+1}), z^*(t_{q+1}))$ are compatible. This, together with (38), allows us to perform induction on $q \in \mathbb{N}$ and obtain that for all $q \in \mathbb{N}$, as long as $t_q \neq t^*$, it always holds that $\mathcal{A}^q$ and $(x^*(t_q), z^*(t_q))$ are compatible and

$$x^*_{i_m}(t_q) - x^*_{m+1}(t_q) > 0.$$

Therefore, we can iteratively repeat the above searching process, i.e., from $(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q)$ to $(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1})$, until t^* is obtained. The details of the search process are summarized in Algorithm 2. We name it the *update*$^-$ subroutine, since in this case $t^* < 0$. The procedure corresponding to the case with $t^* > 0$, which we termed as the *update*$^+$ subroutine, can be easily adapted from the *update*$^-$ subroutine. Details of the *update*$^+$ subroutine can be found in the Appendix.

Before presenting the details of the *generate* subroutine, we make some key observations about the active set $\mathcal{A}^{q+1}$ in the following lemma.

Lemma 4. *For any given $q \in \mathbb{N}$, the following propositions hold:*

(a) If $t_{q+1} \neq t^*$, then $\mathcal{A}^{q+1}_= \neq \mathcal{A}^q_=$;

Algorithm 2 $(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*) = \textbf{update}^-(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q, \lambda)$

Input: $(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q)$, $\lambda \geq 0$;
Compute $\Delta(E_{B^q}), \Delta(\Omega_+^q), \Delta(\Omega_-^q)$ via definitions (30), (33) and (34)
$\Delta(\Omega^q) = \max\{\Delta(\Omega_-^q), \Delta(\Omega_+^q)\}$
$\Delta t_q = \max\{\Delta(E_{B^q}), \Delta(\Omega^q)\}$
$\Delta^* = x_{i_m}^q(t_q + \Delta t_q) - x_{m+1}^q(t_q + \Delta t_q)$
if $\Delta^* \geq 0$ **then**
  $t_{q+1} = \max\{t_q + \Delta t_q, -\lambda\}$
else
  $\Delta\widetilde{t}^q = -t_q + \underset{t}{\operatorname{argmin}} \left\{ (\sum_{i \in V_{B^q}} f_i)^*(t + \beta^q) + (f_{m+1}^*)(-t) \right\}$
  $t_{q+1} = \max\{t_q + \Delta\widetilde{t}^q, -\lambda\}$
end if
$(x^*(t_{q+1}), z^*(t_{q+1})) = (x^q(t_{q+1}), \widetilde{z}^q(t_{q+1}))$
if $t_{q+1} = -\lambda$ **or** $x_{i_m}^*(t_{q+1}) = x_{m+1}^*(t_{q+1})$ **then**
  $t^* = t_{q+1}$
  Let $\mathcal{A}^{q+1} = \{(i, j, \#) \mid (i, j) \in E_m,\ x_i^*(t_{q+1}) \# x_j^*(t_{q+1})\}$
else
  $t^* = \emptyset$
  Update $\mathcal{A}^{q+1}$ from $\mathcal{A}^q$ via (41)
end if
Output: $(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*)$

(b) If $(i,j) \in \mathcal{M}(E_{B^q})$, then for any $\widehat{q} \in \mathbb{N}$ with $\widehat{q} > q$ and $t_{\widehat{q}} \neq t^*$, $(i,j) \notin \mathcal{A}_=^{\widehat{q}}$.

Proof. We prove (a) first. If $t_{q+1} \neq t^*$, from (38), we have $t_{q+1} = t_q + \Delta t_q > t^*$. Hence, at least one of the two sets, $\mathcal{M}(E_{B^q})$ and $\mathcal{M}(\Omega^q)$, is nonempty. The desired result thus follows since $\mathcal{A}_=^{q+1} = \big(\mathcal{A}^{q+1} \cup \mathcal{M}(\Omega^q)\big) \backslash \mathcal{M}(E_{B^q})$ and $\mathcal{M}(E_{B^q}) \cap \mathcal{M}(\Omega^q) = \emptyset$.

Next, we prove (b). We first consider the case where $i \triangleleft j$. If $(i,j) \in \mathcal{M}(E_{B^q})$ and $i \triangleleft j$, we see from (31), (39) and (41) that

$$z_{i,j}^q(t_q + \Delta t_q) = -\lambda_{i,j}, \text{ and } (i,j) \in \mathcal{M}(E_{B^q}^+) \subseteq \mathcal{A}_>^{q+1}.$$

Since $(i,j) \in \mathcal{A}_>^{q+1}$, then at least one of i and j is not in B^{q+1}, i.e., $(i,j) \notin E_{B^{q+1}}$. Since $i \triangleleft j$, we have the following two possible cases:

(i) $j \in B^{q+1}, i \notin B^{q+1}$. In this case we have $(i,j) \in \Omega_-^{q+1}$. Since $(i,j) \in \mathcal{A}_>^{q+1}$, it holds from (40) that $(i,j) \notin \mathcal{M}(\Omega_+^{q+1})$. Thus, (41) implies that $(i,j) \in \mathcal{A}_>^{q+2}$.

(ii) $j \notin B^{q+1}, i \notin B^{q+1}$. From (32), we know that $(i,j) \notin \Omega^{q+1}$. Hence, (40) and (41) imply that $(i,j) \in \mathcal{A}_>^{q+2}$.

Therefore, in both cases, we have $(i,j) \notin \Omega_+^{q+2}$ and $(i,j) \in \mathcal{A}_>^{q+2}$. By induction, we can prove that $(i,j) \notin \Omega_+^{\widehat{q}}$ and $(i,j) \in \mathcal{A}_>^{\widehat{q}}$ for all $\widehat{q} > q$.

Similarly, for the case with $j \triangleleft i$, we can obtain that $(i,j) \notin \Omega_-^{\widehat{q}}$ and $(i,j) \in \mathcal{A}_<^{\widehat{q}}$ for all $\widehat{q} > q$. We thus complete the proof. ∎

With the two subroutines *update*$^-$ and *update*$^+$ at hand, we are ready to present the details of the *generate* subroutine in Algorithm 3. As one can easily observe, the complexity of the *generate* subroutine depends critically on the number of executions of the while-loops (i.e., lines 9-12 and lines 15-18 in Algorithm 3).

Algorithm 3 The *generate* subroutine: $(x^{(m+1)}, z^{(m+1)}) = \textbf{generate}(x^{(m)}, z^{(m)}, G_{m+1})$

1: **Input:** $x^{(m)} \in \Re^m, z^{(m)} \in \Re^{m-1}, G_{m+1} = (V_{m+1}, E_{m+1})$
2: Let $x_i^*(0) = x_i^{(m)}$ for $i \in V_m$ and $x_{m+1}^*(0) = (f_{m+1}^*)'(0)$
3: Let $z_{i,j}^*(0) = z_{i,j}^{(m)}$ for $(i,j) \in E_m$ and $t^* = \emptyset$
4:
5: **if** $f'_{m+1}(x^*_{i_m}(0)) = 0$ **then**
6: $\quad t^* = 0$
7: **else if** $f'_{m+1}(x^*_{i_m}(0)) > 0$ **then**
8: $\quad$ Let $t_0 = 0$, $q = 0$ and $\mathcal{A}^0$ be the active set constructed from $x^*(0)$ as in (21)
9: $\quad$ **while** $t^* = \emptyset$ **do**
10: $\quad\quad (t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*) = \textit{update}^-(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q, \lambda_{i_m,m+1})$
11: $\quad\quad q = q + 1$
12: $\quad$ **end while**
13: **else**
14: $\quad$ Let $t_0 = 0$, $q = 0$ and $\mathcal{A}^0$ be the active set constructed from $x^*(0)$ as in (21)
15: $\quad$ **while** $t^* = \emptyset$ **do**
16: $\quad\quad (t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*) = \textit{update}^+(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q, \mu_{i_m,m+1})$
17: $\quad\quad q = q + 1$
18: $\quad$ **end while**
19: **end if**
20: Let $x^{(m+1)} = x^*(t^*)$, $z_{i,j}^{(m+1)} = z_{i,j}^*(t^*)$ for $(i,j) \in E_m$, and $z_{i_m,m+1}^{(m+1)} = t^*$
21: **Return:** $(x^{(m+1)}, z^{(m+1)}) \in \Re^{m+1} \times \Re^m$

Lemma 5. *The while-loops executed in the generate subroutine will find t^* in at most $2m-1$ iterations.*

Proof. Without loss of generality, we only consider the case $f'_{m+1}(x^*_{i_m}(0)) > 0$, i.e., $t^* < 0$. Assume that after $2m-2$ times executions of the while-loops, t^* has not been found. That is, the algorithm generates $\{(t_i, x^*(t_i), z^*(t_i), \mathcal{A}^i)\}_{i=1}^{2m-2}$ and $t_i > t^*$ for all $i = 0, \dots, 2m-2$. From Lemma 4(a), we know that

$$\mathcal{A}^q_{=} \neq \mathcal{A}^{q+1}_{=}, \quad \forall\, q = 0, \dots, 2m-3. \tag{42}$$

Next, we note from Lemma 4(b) that if some edge $(i,j) \in E_m$ is removed from $\mathcal{A}^q_{=}$ for some q, then $(i,j) \notin \mathcal{A}^{\widehat{q}}_{=}$ for all $2m-2 \geq \widehat{q} \geq q \geq 0$. Therefore, for each edge $(i,j) \in E_m$, it can be added to and removed from $\mathcal{A}^q_{=}$ for at most once. This, together with (42) and the fact that $|E_m| = m-1$, implies that at t_{2m-2}, every edge in E_m has been added to and removed from some $\mathcal{A}^q_{=}$. Thus, $\mathcal{A}^{2m-2}_{=} = \emptyset$, and the sets $\mathcal{A}^{2m-2}_{>}$ and $\mathcal{A}^{2m-2}_{<}$ remain unchanged in the next iterations, i.e., $E_{B^q} = \emptyset$, $\Omega^{2m-2}_{+} \cap A^{2m-2}_{>} = \emptyset$ and $\Omega^{2m-2}_{-} \cap A^{2m-2}_{<} = \emptyset$. Therefore, we have $\Delta t_{2m-2} = -\infty$ from its definition in (35). By (36) and Lemma 3, we have $t_{2m-1} = t^*$ and complete the proof. $\square$

Lemma 5 guarantees that t^* can be found by the *generate* subroutine efficiently. Along with t^*, the G_{m+1}-optimal pair $(x^{(m+1)}, z^{(m+1)})$, i.e., the output of the *generate* subroutine, is also obtained. We thus naturally obtain the correctness of our Algorithm 1.

Theorem 1. *The output $x^{(n)}$ of Algorithm 1 is the optimal solution to problem* (1).

At the end of this section, we provide a brief analysis of the worst-case complexity of our Algorithm 1. Here, we assume that for a given strongly convex differentiable function f and $x \in \Re$, the computational complexity of finding t such that $f'(t) = x$ is $\mathcal{O}(1)$. Then, the computational complexity of *update*$^-$ (and *update*$^+$) is $\mathcal{O}(m)$. By Lemma 5, we see that the computational complexity of the *generate* subroutine is $\mathcal{O}(m^2)$. Therefore, the computational complexity of Algorithm 1 is $\mathcal{O}(n^3)$.

4 Conclusion

In this paper, we focus on the convex isotonic regression problem (1) with tree-induced generalized order restrictions. Inspired by the successes of the PAVA, an efficient active-set based recursive approach, ASRA, is carefully designed to solve (1). Under mild assumptions, we show that ASRA has a polynomial time computational complexity.

5 Appendix

5.1 The arborescence assumption on G

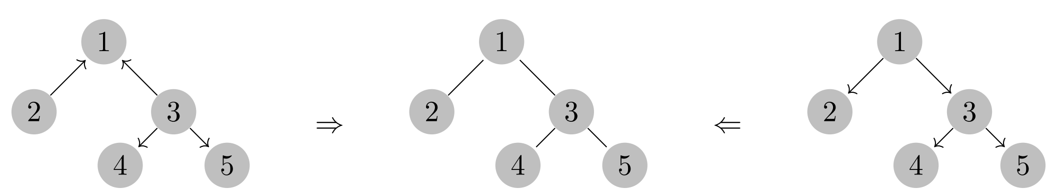
For the given $G = (V, E)$ in the formulation of problem (1), let $\widehat{G} = (V, \widehat{E})$ be an arborescence that shares the same underlying graph with G. Therefore, for any $(i, j) \in \widehat{E}$, we have either $(i, j) \in E$ or $(j, i) \in E$. Then, for any $(i, j) \in \widehat{E}$, let

$$\widehat{\lambda}_{i,j} = \begin{cases} \lambda_{i,j}, & \text{if } (i,j) \in E, \\ \mu_{j,i}, & \text{if } (j,i) \in E, \end{cases} \quad \text{and} \quad \widehat{\mu}_{i,j} = \begin{cases} \mu_{i,j}, & \text{if } (i,j) \in E, \\ \lambda_{j,i}, & \text{if } (j,i) \in E. \end{cases}$$

It can be easily verified that problem (1) is equivalent to the following optimization problem:

$$\min_{x \in \Re^V} \quad \sum_{i \in V} f_i(x_i) + \sum_{(i,j) \in \widehat{E}} \widehat{\lambda}_{i,j}(x_i - x_j)_+ + \sum_{(i,j) \in \widehat{E}} \widehat{\mu}_{i,j}(x_j - x_i)_+.$$

Hence, we can assume that the directed tree G in (1) is an arborescence.



(a) a directed tree G (b) underlying graph of G and $\widehat{G}$ (c) an arborescence $\widehat{G}$

Figure 2: A directed tree G and an arborescence $\widehat{G}$ that share the same underlying graph.

Next, we discuss the decomposition of G. For an arborescence $G = (V, E)$, let $n = |V|$. Without loss of generality, we assume that the node 1 is the root of G, and the nodes in G are arranged such that for any edge $(i, j) \in E$, $i < j$ always holds. Then, we define $G_n = G$, and let $G_{m-1} = (V_{m-1}, E_{m-1})$ be the subgraph of $G_m = (V_m, E_m)$ obtained by deleting the node m and the related edges from G_m, where $n \geq m \geq 2$. Since for any $(i, j) \in E$, it holds that $i < j$, we know that the node m must be a leaf node of G_m, hence, according to [23], G_{m-1} is still a directed tree and $G_{m-1} \subset G_m$ for $m = 2, ..., n$. It's easy to verify that $V_m = \{1, 2, ..., m\}$ for $1 \leq m \leq n$, and $\{(i_m, m + 1)\} = E_{m+1} \backslash E_m$ with $i_m \in V_m$ for $1 \leq m \leq n - 1$.

5.2 The update$^+$ subroutine

We briefly describe the *update*$^+$ subroutine here, which corresponds to the case with $t^* > 0$. Assume that we have obtained a guess t_q of t^* satisfying $0 = t_0 \leq t_q < t^* \leq \mu_{i_m,m+1}$, Meanwhile, the corresponding primal-dual optimal solution pair $(x^*(t_q), z^*(t_q))$ and the active set $\mathcal{A}^q$ are available, such that $\mathcal{A}^q$ and $(x^*(t_q), z^*(t_q))$ are compatible and $x^*_{i_m}(t_q) - x^*_{m+1}(t_q) < 0$. Then, the semi-closed formulas (23) and (24) for the $\mathcal{A}^q$-reduced problem in Proposition 2 still hold.

Here, we need to search

$$\Delta t_q := \max\{\Delta t \mid (25) \text{ and } (26) \text{ hold for all } t \in [t_q, t_q + \Delta t]\}.$$

First, let

$$\Delta(E_{B^q}) = \begin{cases} \min\limits_{(i,j) \in E_{B^q}} \Delta t_{i,j}, & \text{if } E_{B^q} \neq \emptyset, \\ +\infty, & \text{otherwise}, \end{cases} \tag{43}$$

where each $\Delta t_{i,j} \geq 0$ solves:

$$z^q_{i,j}(t_q + \Delta t_{i,j}) = \mu_{i,j}, \quad \text{if } i \triangleleft j, \quad \text{and} \quad z^q_{i,j}(t_q + \Delta t_{i,j}) = -\lambda_{i,j}, \quad \text{if } j \triangleleft i.$$

Next, let $\Delta(\Omega^q) = \min\{\Delta(\Omega^q_+), \Delta(\Omega^q_-)\}$, where

$$\Delta(\Omega^q_+) := \begin{cases} \Delta\bar{t} \text{ satisfying } x_{B^q}(t_q + \Delta\bar{t}) = \min\limits_{(i,j) \in \Omega^q_+ \cap \mathcal{A}^q_<} x^*_j(t_q), & \text{if } \Omega^q_+ \cap \mathcal{A}^q_< \neq \emptyset, \\ +\infty, \text{ otherwise}, \end{cases} \tag{44}$$

and

$$\Delta(\Omega^q_-) := \begin{cases} \Delta\bar{t} \text{ satisfying } x_{B^q}(t_q + \Delta\bar{t}) = \min\limits_{(i,j) \in \Omega^q_- \cap \mathcal{A}^q_>} x^*_i(t_q), & \text{if } \Omega^q_- \cap \mathcal{A}^q_> \neq \emptyset, \\ +\infty, \text{ otherwise}. \end{cases} \tag{45}$$

Then, $\Delta t_q = \min\{\Delta(E_{B^q}), \Delta(\Omega^q)\}$. Compute $\widetilde{\Delta t_q} \geq 0$ via solving $x^q_{i_m}(t_q + \widetilde{\Delta t_q}) - x^q_{m+1}(t_q + \widetilde{\Delta t_q}) = 0$, and set

$$t_{q+1} = \min\{t_q + \Delta t_q, t_q + \widetilde{\Delta t_q}, \mu_{i_m,m+1}\}.$$

If $t_{q+1} < t^*$, we will update the active set $\mathcal{A}^{q+1}$ in the following fashion. Let $\mathcal{M}(E_{B^q}) = \mathcal{M}(E^+_{B^q}) \cup \mathcal{M}(E^-_{B^q})$ with

$$\begin{cases} \mathcal{M}(E^+_{B^q}) = \{(i, j) \in E_{B^q} \mid \Delta t_{i,j} = \Delta t_q, \text{ and } i \triangleleft j\}, \\ \mathcal{M}(E^-_{B^q}) = \{(i, j) \in E_{B^q} \mid \Delta t_{i,j} = \Delta t_q, \text{ and } j \triangleleft i\}, \end{cases}$$

and $\mathcal{M}(\Omega^q) = \mathcal{M}(\Omega^q_+) \cup \mathcal{M}(\Omega^q_-)$ with

$$\begin{cases} \mathcal{M}(\Omega^q_+) &= \{(i,j) \in \Omega^q_+ \cap \mathcal{A}(t_q)_< \mid x^q_i(t_q + \Delta t_q) = x^*_j(t_q)\}, \\ \mathcal{M}(\Omega^q_-) &= \{(i,j) \in \Omega^q_- \cap \mathcal{A}(t_q)_> \mid x^q_j(t_q + \Delta t_q) = x^*_i(t_q)\}. \end{cases}$$

Then, $\mathcal{A}^{q+1}$ is obtained via

$$\begin{cases} \mathcal{A}^{q+1}_= = \big(\mathcal{A}^q_= \cup \mathcal{M}(\Omega^q)\big)\backslash \mathcal{M}(E_{B^q}), \\ \mathcal{A}^{q+1}_> = \big(\mathcal{A}^q_> \cup \mathcal{M}(E^-_{B^q})\big)\backslash \mathcal{M}(\Omega^q_-), \\ \mathcal{A}^{q+1}_< = \big(\mathcal{A}^q_< \cup \mathcal{M}(E^+_{B^q})\big)\backslash \mathcal{M}(\Omega^q_+). \end{cases} \tag{46}$$

We summarize the *update*$^+$ subroutine in Algorithm 4.

Algorithm 4 $(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*) = \textbf{update}^+(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q, \mu)$

Input: $(t_q, x^*(t_q), z^*(t_q), \mathcal{A}^q)$, $\mu \geq 0$;
Compute $\Delta(E_{B^q}), \Delta(\Omega^q_+), \Delta(\Omega^q_-)$ via definitions (43), (44) and (45)
$\Delta(\Omega^q) = \min\{\Delta(\Omega^q_-), \Delta(\Omega^q_+)\}$
$\Delta t_q = \min\{\Delta(E_{B^q}), \Delta(\Omega^q)\}$
$\Delta^* = x^q_{i_m}(t_q + \Delta t_q) - x^q_{m+1}(t_q + \Delta t_q)$
if $\Delta^* \leq 0$ **then**
  $t_{q+1} = \min\{t_q + \Delta t_q, \mu\}$
else
  $\Delta\widetilde{t}_q = -t_q + \underset{t}{\operatorname{argmin}} \Big\{(\sum_{i \in V_{B^q}} f_i)^*(t + \beta^q) + (f^*_{m+1})(-t)\Big\}$
  $t_{q+1} = \min\{t_q + \Delta\widetilde{t}_q, \mu\}$
end if
$(x^*(t_{q+1}), z^*(t_{q+1})) = (x^q(t_{q+1}), \widetilde{z}^q(t_{q+1}))$
if $t_{q+1} = \mu$ **or** $x^*_{i_m}(t_{q+1}) = x^*_{m+1}(t_{q+1})$ **then**
  $t^* = t_{q+1}$
  Let $\mathcal{A}^{q+1} = \{(i,j,\#) \mid (i,j) \in E_m,\ x^*_i(t_{q+1}) \,\#\, x^*_j(t_{q+1})\}$
else
  $t^* = \emptyset$
  Update $\mathcal{A}^{q+1}$ from $\mathcal{A}^q$ via (46)
end if
Output:$(t_{q+1}, x^*(t_{q+1}), z^*(t_{q+1}), \mathcal{A}^{q+1}, t^*)$

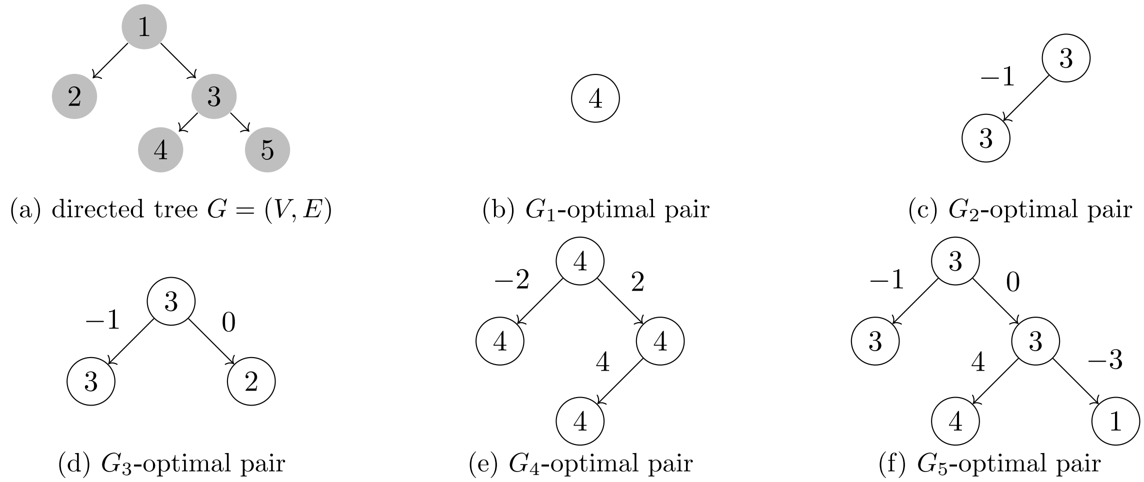
5.3 An illustration of the ASRA

We provide an example of applying the ASRA for solving problem (1). Let $G = (V, E)$ be the directed tree shown in Figure 3a, where $V = \{1,2,3,4,5\}$ and $E = \{(1,2),(1,3),(3,4),(3,5)\}$. Let

$$f_i(x_i) = \frac{1}{2}(x_i - y_i)^2 \text{ for } i = 1,...,4, \text{ where } y = (4,2,2,8) \in \Re^4, \text{ and } f_5(x_5) = x_5^2 + \frac{1}{4}x_5^4,$$

and we set the regularization parameters as follows:

$$(\lambda_{1,2}, \mu_{1,2}) = (+\infty, 0),\ (\lambda_{1,3}, \mu_{1,3}) = (0, +\infty),\ (\lambda_{3,4}, \mu_{3,4}) = (0, 4), \text{ and } (\lambda_{3,5}, \mu_{3,5}) = (3, 3).$$



(a) directed tree $G = (V, E)$ (b) G_1-optimal pair (c) G_2-optimal pair (d) G_3-optimal pair (e) G_4-optimal pair (f) G_5-optimal pair

Figure 3: An example of applying the ASRA for solving problem (1) with given $G = (V, E)$. The first subfigure represents the directed tree $G = (V, E)$, and the remaining five subfigures are the illustrations of the G_m-optimal pairs for $m = 1, 2, 3, 4, 5$, where the values of x_i for $i \in V$ are presented within the circles while the values of $z_{i,j}$ for $(i, j) \in E$ are presented above the edges.

The detailed steps of the ASRA are given below:

(i) First, we initialize with $x_1^{(1)} = 4$.

(ii) Since $(f_2^*)'(x_1^{(1)}) > 0$, it holds that $t^* \leq 0$. We start from $t_0 = 0$ and terminate at $t^* = t_1 = -1$. Then, the G_2-optimal pair $(x^{(2)}, z^{(2)})$ is $x^{(2)} = (3, 3)$ and $z_{1,2}^{(2)} = -1$.

(iii) Since $(f_3^*)'(x_1^{(2)}) > 0$, we have $t^* \leq 0$. Here, we have $t^* = t_0 = -\lambda_{1,3} = 0$. The corresponding G_3-optimal pair $(x^{(3)}, z^{(3)})$ is $x^{(3)} = (3, 3, 2)$, and $z_{1,2}^{(3)} = -1, z_{1,3}^{(3)} = 0$.

(iv) Since $(f_4^*)'(x_3^{(3)}) < 0$, it holds that $t^* \geq 0$. Starting at $t_0 = 0$, we first arrive at $t_1 = 1$, and modify the corresponding active set, i.e., replace $(2, 3, >)$ with $(2, 3, =)$, then continue the searching of t^*. We terminate at $t^* = t_2 = 4$. Therefore, the G_4-optimal pair $(x^{(4)}, z^{(4)})$ is $x^{(4)} = (4, 4, 4, 4)$, and $z_{1,2}^{(4)} = -2, z_{1,3}^{(4)} = 2, z_{3,4}^{(4)} = 4$.

(v) Since $(f_5^*)'(x_3^{(4)}) > 0$, we have $t^* \leq 0$. Starting from $t_0 = 0$, we first arrive $t_1 = 0$ and replace $(3, 4, =)$ with $(3, 4, <)$ in the corresponding active set. Then, we terminate the searching at $t^* = t_2 = -3$, and the G_5-optimal pair $(x^{(5)}, z^{(5)})$ is $x^{(5)} = (3, 3, 3, 4, 1)$, and $z_{1,2}^{(5)} = -1, z_{1,3}^{(5)} = 0, z_{3,4}^{(5)} = 4, z_{3,5}^{(5)} = -3$.

Thus, the optimal solution to problem (1) is $x^* = (3, 3, 3, 4, 1)$. An illustration of the above procedure is presented in Figure 3.

References

[1] M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman, *An empirical distribution function for sampling with incomplete information*, Annals of Mathematical Statistics, 26 (1955), pp. 641–647.

[2] A. Barbero and S. Sra, *Modular proximal optimization for multidimensional total-variation regularization*, Journal of Machine Learning Research 19 (2018), pp. 1–82.

[3] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, MA, 1997.

[4] M. J. Best and N. Chakravarti, *Active set algorithms for isotonic regression: a unifying framework*, Mathematical Programming 47 (1990), pp. 425–439.

[5] M. J. Best, N. Chakravarti, and V. A. Ubhaya, *Minimizing separable convex functions subject to simple chain constraints*, SIAM Journal on Optimization, 10 (2000), pp. 658–672.

[6] H. D. Brunk, *Maximum likelihood estimates of monotone parameters*, Annals of Mathematical Statistics, 26 (1955), pp. 607–616.

[7] N. Chakravarti, *Isotonic median regression: a linear programming approach*, Mathematics of Operation Research, 14 (1989), pp. 303–308.

[8] N. Chakravarti, *Isotonic median regression for orders representable by rooted trees*, Naval Research Logistics, 39 (1992), pp. 599–611.

[9] X. Chang, Y.-L. Yu, Y. Yang, and E. P. Xing, *Semantic pooling for complex event analysis in untrimmed videos*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2016), pp. 1617–1732.

[10] L. Condat, *A direct algorithm for 1D total variation denoising*, IEEE Signal Processing Letters, 20 (2013), pp. 1054–1057.

[11] N. Deo, *Graph theory with applications to engineering and computer science*, Prentice Hall, NJ, 1974.

[12] M. Frisen, *Unimodal regression*, The Statistician, 35 (1986), pp. 479–485.

[13] C. Lu and D. S. Hochbaum, *A unified approach for a 1D generalized total variation problem*, Mathematical Programming, 194 (2022), pp. 415–442.

[14] H. Höefling, *A path algorithm for the fused lasso signal approximator*, Journal of Computational and Graphical Statistics, 19 (2010), pp. 984–1006.

[15] V. Kolmogorov, T. Pock, and M. Rolinek, *Total varaition on a tree*, SIAM Journal of Imaging Sciences, 9 (2016), pp. 605–636.

[16] I. Matyasovszky, *Estimating red noise spectra of climatological time series*, Quarterly Journal of the Hungarian Meteorological Service, 117 (2013), pp. 187–200.

[17] A. Restrepo and A. C. Bovik, *Locally monotonic regression*, IEEE Transactions on Signal Processing, 41 (1993), pp. 2796–2810.

[18] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[19] Y. U. Ryu, R. Chandrasekaran, and V. Jacob, *Prognosis using an isotonic prediction technique*, Management Science, 50 (2004), pp. 777–785.

[20] M. J. Silvapulle and P. K. Sen, *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*, John Wiley & Sons, 2005.

[21] Q. F. Stout, *Unimodal regression via prefix isotonic regression*, Computational Statistics & Data Analysis, 53 (2008), pp. 289–297.

[22] R. Tibshirani, H. Höefling, and R. Tibshirani, *Nearly-isotonic regression*, Technometrics, 53 (2011), pp. 54–61.

[23] D. B. West, *Intorduction to Graph Theory*, 2nd edition, Upper Saddle River: Prentice hall, 2001.

[24] C. Wu, J. Thai, S. Yadlowsky, A. Pozdnoukhov, and A. Bayen, *Cellpath: Fusion of cellular and traffic sensor data for route flow estimation via convex optimization*, Transportation Research Part C: Emerging Technologies, 59 (2015), pp. 111–128.

[25] Y.-L. Yu and E. P. Xing, *Exact algorithms for isotonic regression and related*, Journal of Physics: Conference Series 699, 2016.

[26] Z. Yu, X. Chen, and X. D. Li, *A dynamic programming approach for generalized nearly isotonic regression* , Mathematical Programming Computation, 15 (2023), pp. 195–225.