Appendix to Generalized Isotonic Regression published in the Journal of Computational and Graphical Statistics

Ronny Luss[*] Saharon Rosset[†]

1 Appendix

We need the following additional terminology: A group X *majorizes* (*minorizes*) another group Y if $X \succeq Y$ ($X \preceq Y$). A group X is a *majorant* (*minorant*) of $X \cup A$ where $A = \cup_{i=1}^{k} A_i$ if $X \not\prec A_i$ ($X \not\succ A_i$) $\forall i = 1 \ldots k$.

Proof of Theorem 1:
We prove by contradiction. Assume there exists a union of K blocks in V in the optimal solution labeled $\mathcal{M} = M_1 \cup \ldots \cup M_K$ that get broken by the cut, with M_1 and M_K as the minorant and majorant block in $\mathcal{M}$, and M_k^L and M_k^U as the groups in M_k below and above the cut. Define $\mathcal{L}$ as the union of all blocks in V that lie "below" the algorithm cut, $\mathcal{U}$ as the union of all blocks in V that lie "above" the algorithm cut. Further define $A_K^L \subseteq \mathcal{L}$ ($A_1^U \subseteq \mathcal{U}$) as the union of blocks along the algorithm cut such that $A_K^L \succ M_K^L$ ($A_1^U \prec M_1^U$). Figure 1 depicts an example of these definitions where $A_1^U = A_1^L = A_K^U = A_K^L = \{\}$ for simplicity.

We first prove that $w_{M_1} > w_V$. First, consider the case $A_1^U = \{\}$. By convexity of $f_i(\cdot)$ and summing over group M_1^U, we have

$$\sum_{i \in M_1^U} f_i(w_{M_1^U}) \geq \sum_{i \in M_1^U} f_i(w_V) + (w_{M_1^U} - w_V) \sum_{i \in M_1^U} \frac{\partial f_i(\hat{y}_i)}{\partial \hat{y}_i}\bigg|_{w_V}.$$

[*]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel. ronnyluss@gmail.com

[†]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel. saharon@post.tau.ac.il

Definition of the weight operator gives

$$\sum_{i \in M_1^U} f_i(w_{M_1^U}) \leq \sum_{i \in M_1^U} f_i(w_V) \Rightarrow (w_{M_1^U} - w_V) \sum_{i \in M_1^U} \frac{\partial f_i(\hat{y}_i)}{\partial \hat{y}_i}\bigg|_{w_V} \leq 0.$$

Finally, by the definition of the algorithm cut in (11) since no block exists below M_1^U to affect isotonicity,

$$\sum_{i \in M_1^U} \frac{\partial f_i(\hat{y}_i)}{\partial \hat{y}_i}\bigg|_{w_V} \leq 0 \tag{1}$$

so that $w_{M_1^U} \geq w_V$. Since M_1 is a block, we have $w_{M_1^L} > w_{M_1^U}$, and then

$$w_{M_1^L} > w_{M_1^U} > w_V \Rightarrow w_{M_1} > w_V.$$

For the case, $A_1^U \neq \{\}$, we have $w_{M_1} > w_{A_1^U} > w_V$ with the first inequality due to optimality and the second follows directly the proof above replacing M_1^U by A_1^U. A proof for $w_{M_K} < w_V$ follows a similar argument focusing on M_K^L. Putting this together gives $w_{M_1} > w_V > w_{M_K}$, which contradicts that M_1 and M_K are blocks in the global solution, since by assumption then $w_{M_1} < w_{M_K}$. The case $K = 1$ is also trivially covered by the above arguments. We conclude that the algorithm cannot cut any block. ■

Proof of Theorem 2:

The proof is by induction. The base case, i.e., first iteration, where all points form one group is trivial. The first cut is made by solving linear program (11) which constrains the solution to maintain isotonicity.

Assuming that iteration k (and all previous iterations) provides an isotonic solution, we prove that iteration $k+1$ must also maintain isotonicity. Figure 2 helps illustrate the situation described here. Let G be the group split at iteration $k+1$ and denote A (B) as the group under (over) the cut. Let $\mathcal{A} = \{X : X$ is a group at iteration $k+1, \exists i \in X$ such that $(i,j) \in \mathcal{I}$ for some $j \in A\}$ (i.e., $X \in \mathcal{A}$ border A from below).

Consider iteration $k+1$. Denote $\mathcal{X} = \{X \in \mathcal{A} : w_A < w_X\}$ (i.e., $X \in \mathcal{X}$ violates isotonicity with A). The split in G causes the fit in nodes in A to decrease. Proof that

$$\sum_{i \in A} \frac{\partial f_i(\hat{y}_i)}{\partial \hat{y}_i}\bigg|_{w_G} \geq 0$$

Figure 1: Illustration of proof of Theorem 1. Black lines separate blocks. The diagonal red line through the center demonstrates a cut of Algorithm 1. $\mathcal{L}$ is the union of blue blocks below the cut and $\mathcal{U}$ is the union of green blocks above the cut. White blocks are blocks that are potentially split by Algorithm 1. These blocks are split into $M_1^L, \ldots, M_5^L$ below the cut and $M_1^U, \ldots, M_5^U$ above the cut. In the proof, $M_i = M_i^L \cup M_i^U$ $\forall i = 1 \ldots 5$. The proof shows, for example, that if the algorithm splits M_1 into M_1^L and M_1^U according to the defined cut in (11), then there must be no isotonicity violation when creating blocks from M_1^L and M_1^U. However, since M_1 is assumed to be a block, there must exist an isotonicity violation between M_1^L and M_1^U, providing a contradiction.

follows the proof of (15) in Theorem 1 above so that $w_A \geq w_G$. We will prove that when the fits in A decrease, there can be no groups below A that become violated by the new fits to A, i.e., the decreased fits in A cannot be such that $\mathcal{X} \neq \{\}$.

We first prove that $\mathcal{X} = \{\}$ by contradiction. Assume $\mathcal{X} \neq \{\}$. Denote $k_0 < k+1$ as the iteration at which the last of the groups in $\mathcal{X}$, denoted D, was split from G and suppose at iteration k_0, G was part of a larger group H and D was part of a larger group F. It is important to note that $X \bigcap (F \bigcup H) = \{\}$ $\forall X \in \mathcal{X} \setminus D$ at iteration k_0 because by assumption all groups in $\mathcal{X} \setminus D$ were separated from A before iteration i. Thus, at iteration k_0, D is the only group bordering A that violates isotonicity.

Let D_U denote the union of D and all groups in F that majorize D. By construction, D_U is a majorant in F. Hence $w_{D_U} < w_{F \cup H}$ by Algorithm 1 and $w_A < w_{D_U}$ by definition since $w_{D_U} > w_D > w_A$. Also by construction, any set $X \in H$ that minorizes A has $w_X < w_A$ (each set X that minorizes A besides D such that $w_X < w_A$ has already been split from A). Hence we can denote A_L as the union of A and all groups in H that minorize A and we have $w_A > w_{A_L}$ and A_L

is a minorant in H. Since $A_L \subseteq H$ at iteration i, we have

$$w_{F \cup H} < w_{A_L} < w_A < w_{D_U} < w_{F \cup H}$$

which is a contradiction, and hence the assumption $\mathcal{X} \neq \{\}$ is false. The first inequality is because the algorithm left A_L in H when F was split from H, and the remaining inequalities are due to the above discussion. Hence the split at iterations $k+1$ could not have caused a break in isotonicity.

A similar argument can be made to show that the increased fit for nodes in B does not cause any isotonic violation. The proof is hence completed by induction. ∎

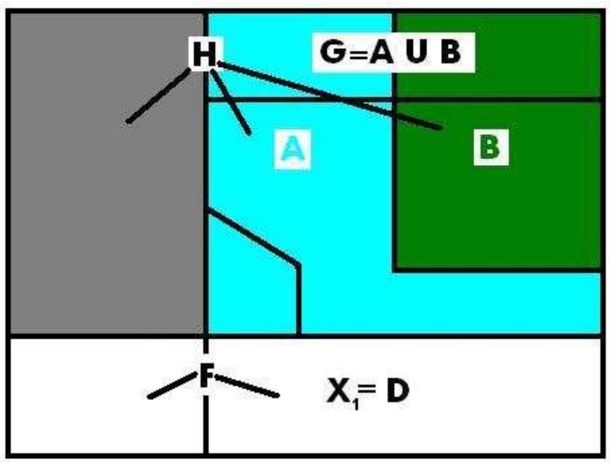

Figure 2: Illustration of proof of Theorem 2 showing the defined sets at iteration $k+1$. G is the set divided at iteration $k+1$ into A (all blue area) and B (all green area). The group bordering A from below denoted by X_1 (also referred to as D in the proof) is in violation with A. At iteration k_0, G is part of the larger group H and X_1 is part of the larger group F. At iteration k_0, groups F and H are separated. The proof shows that when A and B are split at iteration $k+1$, no group such as X_1 where $w_{X_1} > w_A$ could have existed. In the picture, X_1 must have been separated at an iteration $k_0 < k+1$, but the proof, through contradiction, shows that this cannot occur.