Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$ be real sequences with $y_i > 0$, for all i. The tuple (x_i, y_i) will be called a *record* and denoted by R_i. The set of records will be denoted by $\mathcal{D}$. We will generally regard the sequences X, Y as sets and study orderings on the records.

Definition 1. *A priority function is a function $G\colon \mathbf{R} \times \mathbf{R}^+ \to \mathbf{R}$ that induces an ordering on the tuples $R_i = (x_i, y_i)$. We refer to $G(x, y) = \frac{x}{y}$ as the standard priority function. The ordering on the sequences X, Y induced by G is indicated by $R_{(1)}, R_{(2)}, \ldots R_{(n)}$, representing, in order, the highest priority record, next highest, etc.*

Definition 2. *A score function is a function $F\colon \mathbf{R} \times \mathbf{R}^+ \to \mathbf{R}$, $(x, y) \mapsto F(x, y)$ that is increasing in x. If F is of the form $F(x, y) = \frac{x^\gamma}{y}$ for some $\gamma > 0$, then F is a power score function.*

We extend the functions F, G to subsets of $X_i \in X$, $Y_i \in Y$ by defining $F(X_i, Y_i) = F(\sum_{x_i \in X_i} x_i, \sum_{y_i \in Y_i} y_i)$. It is often implicit that the tuple (x_i, y_i) is associated with the record R_i.

Let $\mathcal{P} = \{P_1, \ldots, P_T\}$ be a partition of the set $\{1, \ldots, n\}$, so that $P_1 \cup \cdots \cup P_T = \{1, \ldots, n\}$, with the P_i's pairwise disjoint. In this case $\mathcal{P}$ is a partition of N of size T and we write $|\mathcal{P}| = T$. We will sometimes consider $\mathcal{P}$ as a partition of N, or one or both of the sequences X, Y, or of the records $\{R_i\}$, and write $j \in P_i$, f or $x_i \in \mathcal{P}$, as fits the situation. In this way a subset $X_i \in X$ can be considered as an element of a partition determined by the indices present in the subset, etc.

We are interested in solutions to the optimization problem

$$\mathcal{P} = \max_{\mathcal{P}, |\mathcal{P}|=T} \sum_{j=1}^{T} \frac{(\sum_{R_i \in P_j} x_i)^\gamma}{\sum_{R_i \in P_j} y_i} \tag{1}$$

For $\gamma \geq 0$. For example, given a set of quadratic polynomials $f_i(x) = \frac{1}{2} h_i x^2 + g_i x + c_i$, with $h_i > 0$ for all i, the x values of the vertices are given by $\frac{-g_i}{h_i}$, with minimum values $\frac{-g_i^2}{2h_i}$. The standard priorty function puts an ordering on the vertices $\left(\frac{-g_i}{h_i}, \frac{-g_i^2}{2h_i}\right)$ by considering their x-valus. For $T = 1$, since the sum of all f_i is a quadratic, we can minimize the sum $\sum_i f_i(x)$. For $T >= 2$ the optimizaiton in 1 finds the partitioning of the f_i that minimizes the sum $\sum_{j=1}^{T} \sum_{i \in P_j} f_i$. This can also be viewed in the context of boosting, in which the maximal partition represents optimal leaf values for a classifier that can take at most T values.

It will be necessary to constrain the optimization in 1 and restrict our attention to subsets of the set of all partitions of n, itself a large set.

Definition 3. *A partition $\mathcal{P}$ such that each P_i is an ordered subset of records, so tbat $P_i = \{R_{(j)}, \ldots, R_{(j+l_i)}\}$, for some j, is called an ordered partition. If in addition the partition satisfies*

$$\mathcal{P} = \operatorname*{argmax}_{\mathcal{P}, |\mathcal{P}|=T} \sum_{j=1}^{T} \frac{(\sum_{R_i \in P_j} x_i)^\gamma}{\sum_{R_i \in P_j} y_i}$$

then $\mathcal{P}$ is a maximal ordered partition for the power score function $F(x, y) = \frac{x^\gamma}{y}$, or a maximal ordered partition of power γ.

Theorem 1. *For the dataset $\mathcal{D} = \{R_1, \ldots, R_N\}$, standard priority function $G(R_i) = \frac{x_i}{y_i}$, let F be a score function $F(x_i, y_i) = \frac{x_i^\gamma}{y_i}$, for $\gamma > 0$. The for any fixed $T \in \mathbf{N}, T < N$, there is a maximal ordered partition $\mathcal{P}$ of size T for F iff $\gamma = 2$.*

The set of partitions is greatly reduced by the requirement that $\mathcal{P}$ be ordered. The set of all partitions of n is the Bell number of order T, exponentially increasing with n for any $T > 1$. The set of all size T partitions is a Stirling number of the second kind $\left\{ {n \atop T} \right\}$. The nth Bell number B_n is given by the identity

$$B_n = \sum_{k=0}^{n} \binom{n}{k} B_k$$

The Stirling numbers follow the recursion

$$\left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} = k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} + \left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\}$$

and have asymptotic growth rate

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} \sim \frac{k^n}{k!}$$

So, for example, for $(n, T) = (20, 10)$, we have $B_n \approx 5.9e12$, $\mathcal{O} = 92378$, while for $(n, T) = (30, 10)$, we have $B_n \approx 1.73e22$, $\mathcal{O} = 10,015,005$.

Proof. For sufficiency, let $\gamma = 2$, and suppose the partition $\mathcal{P} = \{P_1, \ldots, P_T\}$ is the argmax solution to 1. Let $R_1 = (X_1, Y_1)$ be the set of records in $\mathcal{P}$ that contains the maximal element $R_{(1)}$ of $\mathcal{D}$, and define $R_1^{in} = \text{argmin}_{R_j \in P_1} G(x_i, y_i)$, $R_1^{out} = \text{argmax}_{R_j \notin P_1} G(x_i, y_i)$. Note that X_1 is an ordered subset if and only if $R_1^{in} <= R_1^{out}$, so that there are no "holes" in X_1. This can be made precise by defining $I_1^{in} = j\text{such that} R_{(j)} = R_1^{in}$, $I_1^{out} = j\text{such that} R_{(j)} = R_1^{out}$, and $D_1 = I_1^{in} - I_1^{out}$. It is then the case that $D_1 \geq 0$, and X_1 is ordered if and only iff $D_1 = 0$.

Without loss of generality, assume $R_1^{out} \in P_2$. We will assume that $G(x_1^{in}, y_1^{in}) < G(x_1^{out}, y_1^{out})$, and show that we can obtain an improvement in the sum $F(X_1, Y_1) + F(X_2, Y_2)$ by exchanging elements of X_1, X_2. In this way the elements of the subset X_1 are sucessively swapped out until it is ordered. Since X_1 is the partition with maximal element, we can remove it from consideration, and find the maximal remaining element, and apply the same procedure. In this way we obtain a partition all of whose subsets are ordered.

Assume the tuples R_1^{in}, R_1^{out} are composed of $R_1^{in} = \left(x_1^{in}, y_1^{in}\right), R_1^{out} = (x_1^{out}, y_1^{out})$.

Define

$$\begin{aligned} X'(\lambda) &= \lambda \left(X_1 - x_1^{in}\right) + (1-\lambda)\left(X_1' + x_1^{out}\right) \\ Y'(\lambda) &= \lambda \left(Y_1 - y_1^{in}\right) + (1-\lambda)\left(Y_1' + y_1^{out}\right) \end{aligned}$$

for $\lambda \in [0, 1]$. For $\lambda_* = \frac{y_1^{out}}{y_1^{in} + y_1^{out}}$, we have $Y'(\lambda_*) = Y_1$, and

$$X'(\lambda_*) = X_1 + \frac{y_1^{in} x_1^{out} - y_1^{out} x_1^{in}}{y_1^{in} + y_1^{out}}$$

Since $G(x_1^{in}, y_1^{in}) < G(x_1^{out}, y_1^{out})$, $\frac{y_1^{in} x_1^{out} - y_1^{out} x_1^{in}}{y_1^{in} + y_1^{out}} > 0$ and $X'(\lambda_*) > 0$. We therefore have

$$F(X_1, Y_1) \leq F(X'(\lambda), Y_1) \leq \lambda \left(F(X_1 - x_1^{in}, Y_1 - y_1^{in})\right) + (1+\lambda)\left(F(X_1 + x_1^{out}, Y_1 + y_1^{out})\right) \tag{2}$$

where the second inequality is from quasiconvexity of F for $\gamma = 2$. From 2 it follows that

$$F(X_1, Y_1) \leq \max\left(F(X_1 - x_1^{in}, Y_1 - y_1^{in}), F(X_1 + x_1^{out}, Y_1 + y_1^{out})\right) \tag{3}$$

To get a similar result for the sets X_2, Y_2, define the transformed sequences $\bar{x} = \{x_1, \ldots, x_n\}$, $\bar{y} = \{y_1, \ldots, y_n\}$. The sets $\bar{X_1}, \bar{X_2}, \bar{Y_1}, \bar{Y_2}$ are similarly defined, and we can define $\bar{R_2}^{in} = \text{argmin}_{\bar{R_j} \in \bar{P_2}} G(\bar{x_i}, \bar{y_i})$, $\bar{R_1}^{out} = \text{argmax}_{\bar{R_j} \notin \bar{P_1}} G(\bar{x_i}, \bar{y_i})$. Assuming the correspondence between record and underlying sequences $\bar{R_i} = (\bar{x_i}, \bar{y_i})$, we then have $x_1^{in} = -\bar{x}_2^{out}$, $y_1^{in} = \bar{y_2}^{out}$, and $x_1^{out} = -\bar{x}_2^{in}$, $y_1^{out} = \bar{y_2}^{in}$.

Define

$$\begin{aligned} \bar{X}'(\lambda) &= \lambda \left(\bar{X_2} - \bar{x_2}^{in}\right) + (1-\lambda)\left(\bar{X_2} + \bar{x_2}^{out}\right) \\ \bar{Y}'(\lambda) &= \lambda \left(\bar{Y_2} - \bar{y_2}^{in}\right) + (1-\lambda)\left(\bar{Y_2} + \bar{y_2}^{out}\right) \end{aligned}$$

for $\lambda \in [0,1]$. For $\bar{\lambda}_* = \frac{\bar{y_2}^{out}}{\bar{y_2}^{in}+\bar{y_2}^{out}}$, we have $\bar{Y}'\left(\bar{\lambda}_*\right) = \bar{Y}_1$, and

$$\bar{X}'\left(\bar{\lambda}_*\right) = \bar{X}_2 + \frac{\bar{y_2}^{in}\bar{x_2}^{out} - \bar{y_2}^{out}\bar{x_2}^{in}}{\bar{y_2}^{in} + \bar{y_2}^{out}}$$

By similar arguments, and since $\bar{y_2}^{in}\bar{x_2}^{out} - \bar{y_2}^{out}\bar{x_2}^{in} = x_1^{out}y_1^{in} - x_1^{in}y_1^{out} \geq 0$,

$$\begin{aligned}
F(X_2, Y_2) = F(\bar{X_2}, \bar{Y_2}) &\leq \max\left(F(\bar{X_2} - \bar{x_2}^{in}, \bar{Y_2} - \bar{y_2}^{in}), F(\bar{X_2} + \bar{x_2}^{out}, \bar{Y_2} + \bar{y_2}^{out})\right) \\
&= \max\left(F(\bar{X_2} - \bar{x_2}^{in}, Y_2 - y_1^{our}), F(\bar{X_2} + \bar{x_2}^{out}, \bar{Y_2} + y_2^{out})\right) \\
&= \max\left(F(\bar{X_2} + x_1^{out}, Y_2 - y_1^{out}), F(\bar{X_2} - x_1^{in}, Y_2 + y_2^{out})\right) \\
&= \max\left(F(X_2 - x_1^{out}, Y_2 - y_1^{out}), F(X_2 + x_1^{in}, Y_2 + y_2^{out})\right)
\end{aligned}$$

So

$$F(X_2, Y_2) \leq \max\left(F(X_2 - x_1^{out}, Y_2 - y_1^{out}), F(X_2 + x_1^{in}, Y_2 + y_1^{in})\right) \tag{4}$$

If we form the table

$$\begin{pmatrix} F(X_1 - x_1^{in}, Y_1 - y_1^{in}) & F(X_2 + x_1^{in}, Y_2 + y_1^{in}) \\ F(X_1 + x_1^{out}, Y_1 + y_1^{out}) & F(X_2 - x_1^{out}, Y_2 - y_1^{out}) \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

then the results in 2, 4 imply that $F(X_1, Y_1) \leq \max(A_{11}, A_{21})$ and $F(X_2, Y_2) \leq \max(A_{12}, A_{22})$. What we would like to show is that $F(X_1, Y_1) + F(X_2, Y_2) \leq \max(A_{11}, A_{12})$ or $F(X_1, Y_1) + F(X_2, Y_2) \leq \max(A_{21}, A_{22})$, as those operations represent a swap of records between the two sets X_1, X_2. To this end assume that the maximum values down columns occur in different rows, e.g. $\max(A_{11}, A_{21}) = A_{11}$, $\max(A_{12}, A_{22}) = A_{22}$. The case for which the maximums occur on the opposite diagonal is handled similarly. We can then assume that

$$F(X_1 - x_1^{in}, Y_1 - y_1^{in}) - F(X_1, Y_1) \geq 0 \tag{5}$$
$$F(X_2 - x_1^{out}, Y_2 - y_1^{out}) - F(X_2, Y_2) \geq 0 \tag{6}$$
$$F(X_1 + x_1^{out}, Y_1 + y_1^{out}) - F(X_1, Y_1) \leq 0 \tag{7}$$
$$F(X_2 + x_1^{in}, Y_2 + y_1^{in}) - F(X_2, Y_2) \leq 0 \tag{8}$$

Expand

$$F(X - \alpha, Y - \beta) - F(X, Y) = \frac{\beta X^2 - 2\alpha XY + \alpha^2 Y}{Y\left(Y - \beta\right)}$$

and note that for $Y > 0$, $0 < \beta < Y$,

$$F(X - \alpha, Y - \beta) - F(X, Y) \geq 0 \iff \beta \geq \frac{\alpha Y\left(2X - \alpha\right)}{X^2} \tag{9}$$

We write

$$\begin{aligned}
F(X_1 - x_1^{in}, Y_1 - y_1^{in}) + F(X_2 + x_1^{in}, Y_2 + y_1^{in}) - \left(F(X_1, Y_1) + F(X_2, Y_2)\right) = \\
\left(F(X_1 - x_1^{in}, Y_1 - y_1^{in}) - F(X_1 - x_1^{in}, Y_1)\right) + \left(F(X_1 - x_1^{in}, Y_1) - F(X_1, Y_1)\right) + \\
\left(F(X_2 + x_1^{in}, Y_2 + y_1^{in}) - F(X_2 + x_1^{in}, Y_2)\right) + \left(F(X_2 + x_1^{in}, Y_2) - F(X_2, Y_2)\right)
\end{aligned}$$

and

$$\begin{aligned}
F(X_1 + x_1^{out}, Y_1 + y_1^{out}) + F(X_2 - x_1^{out}, Y_2 - y_1^{out}) - \left(F(X_1, Y_1) + F(X_2, Y_2)\right) = \\
\left(F(X_1 + x_1^{out}, Y_1 + y_1^{out}) - F(X_1 + x_1^{out}, Y_1)\right) + \left(F(X_1 + x_1^{out}, Y_1) - F(X_1, Y_1)\right) + \\
\left(F(X_2 - x_1^{out}, Y_2 - y_1^{out}) - F(X_2 - x_1^{out}, Y_2)\right) + \left(F(X_2 - x_1^{out}, Y_2) - F(X_2, Y_2)\right)
\end{aligned}$$

For ease of notation designate $\alpha = x_1^{in}$, $\beta = y_1^{in}$, $a = x_1^{out}$, $b = y_1^{out}$.
The summands for the top equation can then be written

$$\begin{aligned} s_1 &= \frac{(X_1-\alpha)^2\beta}{Y_1(Y_1-\beta)} \\ s_2 &= \frac{\alpha(\alpha-2X_1)}{Y_1} \\ s_3 &= \frac{-(X_2+\alpha)^2\beta}{Y_2(Y_2+\beta)} \\ s_4 &= \frac{\alpha(\alpha+2X_2)}{Y_2} \end{aligned}$$

and the bottom

$$\begin{aligned} t_1 &= \frac{-(X_1+a)^2 b}{Y_1(Y_1+b)} \\ t_2 &= \frac{a(a+2X_1)}{Y_1} \\ t_3 &= \frac{(X_2-a)^2 b}{Y_2(Y_2-b)} \\ t_4 &= \frac{a(a-2X_2)}{Y_2} \end{aligned}$$

We will show that one of the rows of our table give in improvement to the sum over partitions, i.e., that one of

$$F(X_1-\alpha, Y_1-\beta) + F(X_2+\alpha, Y_2+\beta) - F(X_1,Y_1) - F(X_2,Y_2) \geq 0 \tag{10}$$
$$F(X_1+a, Y_2+b) + F(X_2-a, Y_2-b) - F(X_1,Y_1) - F(X_2,Y_2) \geq 0 \tag{11}$$

Case 1: $X_1 >= 0$, $X_2 >= 0$.

We will show that the top row in 10 is positive.

Claim: $\frac{X_1}{Y_1} \geq \frac{X_2}{Y_2}$, $\frac{X_1}{Y_1} \geq \frac{2a}{b}$, $\frac{X_2}{Y_2} \geq \frac{2\alpha}{\beta}$.
Proof: Since $F(X-\alpha, Y-\beta) - F(X,Y) = \frac{\beta X^2 - 2\alpha XY + \alpha^2 Y}{Y(Y-\beta)}$ is a polynomial in X_1, we have

$$\begin{aligned} F(X_1+a, Y_1+b) - F(X_1,Y_1) \leq 0 &\implies X_1 \notin \left(\frac{a}{b}Y_1 \pm |\frac{a}{b}|\sqrt{Y_1(Y1+b)}\right) \\ &\implies \frac{X_1}{Y_1} \notin \left(\frac{a}{b}\left(1 \pm \frac{\sqrt{Y_1(Y_1+b)}}{Y_1}\right)\right) \end{aligned}$$

Since $\frac{X_1}{Y_1} \geq 0$, it follows that $\frac{X_1}{Y_1} \geq \frac{2a}{b}$. By similar reasoning,

$$\begin{aligned} F(X_2+\alpha, Y_2+\beta) - F(X_2,Y_2) \leq 0 &\implies X_2 \notin \left(\frac{\alpha}{\beta}Y_2 \pm |\frac{\alpha}{\beta}|\sqrt{Y_2(Y_2+\beta)}\right) \\ &\implies \frac{X_2}{Y_2} \notin \left(\frac{\alpha}{\beta}\left(1 \pm \frac{\sqrt{Y_2(Y_2+\beta)}}{Y_2}\right)\right) \end{aligned}$$

so that $\frac{X_2}{Y_2} \geq \frac{2\alpha}{\beta}$. Now since $\frac{2\alpha}{\beta} \leq \frac{X_1}{Y_2}$, $\frac{a}{b} \geq \frac{X_2}{Y_2}$ by definition, and $\frac{\alpha}{\beta} \leq \frac{a}{b}$ by supposition, the result follows. ■

Claim: $s_1 + s_3 \geq 0$
Proof:

$$s_1 = \frac{(X_1-\alpha)^2\beta}{Y_1(Y_1-\beta)} = \frac{\beta}{Y_1-\beta}F(X_1+\alpha, Y_1)$$
$$s_3 = -\frac{(X_2+\alpha)^2\beta}{Y_2(Y_2+\beta)} = \frac{-\beta}{Y_2+\beta}F(X_2+\alpha, Y_2)$$

So

$$\begin{aligned} s_1 + s_3 &= \frac{\beta}{Y_1-\beta}F(X_1+\alpha, Y_1) - \frac{\beta}{Y_2+\beta}F(X_2+\alpha, Y_2) \\ &= \frac{\beta}{Y_1-\beta}\left(F(X_1,Y_1)+s_2\right) - \frac{\beta}{Y_2+\beta}\left(F(X_2,Y_2)+s_4\right) \end{aligned}$$

Since $F(X-\alpha, Y-\beta) - F(X,Y) = \frac{\beta X^2 - 2\alpha XY + \alpha^2 Y}{Y(Y-\beta)}$,

$$F(X_1-\alpha, Y_1-\beta) - F(X_1,Y_1) \geq 0 \implies \beta \geq \frac{\alpha Y_1}{X_1^2}(2X_1-\alpha) \implies s_2 \geq -\frac{\beta}{Y_1}F(X_1,Y_1)$$
$$F(X_2+\alpha, Y_2+\beta) - F(X_2,Y_2) \leq 0 \implies \beta \geq \frac{Y_2}{X_2^2}(2X_2+\alpha) \implies s_4 \leq \frac{\beta}{Y_2}F(X_2,Y_2)$$

So

$$\begin{aligned} s_1 + s_3 &\geq \frac{\beta}{Y_1-\beta}\left(F(X_1,Y_1) - \frac{\beta}{Y_1}F(X_1,Y_1)\right) - \frac{\beta}{Y_2+\beta}\left(F(X_2,Y_2) + \frac{\beta}{Y_2}F(X_2,Y_2)\right) \\ &= \frac{\beta}{Y_1}F(X_1,Y_1) - \frac{\beta}{Y_2}F(X_2,Y_2) \\ &= \beta\left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2\right) \\ &\geq 0, \end{aligned}$$

by the previous claim, and all quantities are positive.

■

Claim: $\sum_{i=1}^4 s_i \geq 0$
Proof:

$$\begin{aligned} s_2 + s_4 &= F(X_1+\alpha, Y_1) - F(X_1,Y_1) + F(X_2+\alpha, Y_2) - F(X_2,Y_2) \\ &= \frac{\alpha(\alpha-2X_1)}{Y_1} + \frac{\alpha(\alpha+2X_2)}{Y_2} \end{aligned}$$

As a polynomial in α we have

$$s_2 + s_4 = p(\alpha) = \alpha\left(g + \frac{1}{2}h\alpha\right),$$

where

$$g = \frac{2(X_2Y_1 - X_1Y_2)}{Y_1Y_2}$$
$$h = \frac{Y_1+Y_2}{Y_1Y_2}$$

So p has two real roots, one at $\alpha_1 = 0$ and one for $\alpha_2 \geq 0$. The graph is an upward parabola so that $p \geq 0$ for $\alpha \leq 0$, so the remaining case is $\alpha > 0$. We have, as above

$$F(X_1 - \alpha, Y_1 - \beta) - F(X_1, Y_1) \geq 0 \implies \alpha \notin \left(X_1 \pm |X_1| \sqrt{\frac{Y_1 - \beta}{Y_1}} \right)$$
$$F(X_2 + \alpha, Y_2 + \beta) - F(X_2, Y_2) \leq 0 \implies \alpha \notin \left(-X_2 \pm |X_2| \sqrt{\frac{Y_2 + \beta}{Y_2}} \right)$$

so that $\alpha > 0$ means that $\alpha \in \left[0, \min X_1 \left(1 - \sqrt{\frac{Y_1 - \beta}{Y_1}}\right), X_2 \left(\sqrt{\frac{Y_2+\beta}{Y_2}} - 1\right)\right]$. The idea is that α is small relative to β so that the polynomial $p(\alpha) = s_2 + s_4$ never goes negative enough to violate $s_1 + s_2 \leq -(s_1 + s_3)$. The proof is technical and is contained in the Appendix. ■

Case 2: X_1, $X_2 \leq 0$

Writing the transformed sets $\bar{X_1} = -X_1$, $\bar{Y_1} = Y_1$, $\bar{X_2} = -X_2$, $\bar{Y_2} = Y_2$, and defining $\eta = \bar{x_2}^{in}$, $\theta = \bar{y_2}^{i}n$, we have $a = -\eta$, $b = \theta$ by definition. Assume that $\bar{X_2}$ has the maximal element; if it is contained in $\bar{X_1}$ the reasoning is similar. We proceed as in case 1:

$$\begin{aligned} F(X_1, Y_1) + F(X_2, Y_2) = F(\bar{X_2}, \bar{Y_2}) + F(\bar{X_1}, \bar{Y_1}) &\leq F(\bar{X_2} - \eta, \bar{Y_2} - \theta) + F(\bar{X_1} + \eta, \bar{Y_1} + \theta) \\ &= F(-\left(\bar{X_2} - \eta\right), Y_2 - \theta) + F(-\left(\bar{X_1} + \eta\right), Y + 2 + \theta) \\ &= F(X_2 + \eta, Y_2 - \theta) + F(X_1 - \eta, Y_2 + \theta) \\ &= F(X_2 - a, Y_2 - b) + F(X_1 + a, Y_1 + b) \end{aligned}$$

so that the bottom row represents an improvement to the original partition.

Case 3: $X_1 \geq 0$, $X_2 \leq 0$

<u>Claim:</u> One of $s_1 + s_3$, $t_1 + t_3$ is positive.
<u>Proof:</u> From the claim above,

$$s_1 + s_3 \geq \beta \left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2 \right) \tag{12}$$

In this case we don't necessarily know that $F(X_1, Y_1) \geq F(X_2, Y_2)$. We have

$$t_1 = \frac{-\left(X_1 + a\right)^2 b}{Y_1 \left(Y_1 + b\right)} = \frac{b}{Y_1 + b} F(X_1 + a, Y_1)$$
$$t_3 = -\frac{\left(X_2 - a\right)^2 b}{Y_2 \left(Y_2 - b\right)} = \frac{b}{Y_2 - b} F(X_2 - a, Y_2)$$

So

$$\begin{aligned} t_1 + t_3 &= \frac{b}{Y_2 - b} F(X_2 - a, Y_2) - \frac{-b}{Y_1 + b} F(X1 + a, Y_1) \\ &= \frac{b}{Y_2 - b} \left(F(X_2, Y_2) + t_2\right) - \frac{-b}{Y_1 + b} \left(F(X_1, Y_1) + t_4\right) \end{aligned}$$

Since $F(X - \alpha, Y - \beta) - F(X, Y) = \frac{\beta X^2 - 2\alpha XY + \alpha^2 Y}{Y(Y - \beta)}$,

$$F(X_2 - a, Y_2 - b) - F(X_2, Y_2) \geq 0 \implies b \geq \frac{aY_2}{X_2^2} \left(2X_2 - a\right) \implies t_4 \geq \frac{-b}{Y_2} F(X_2, Y_2)$$
$$F(X_1 + a, Y_1 + b) - F(X_1, Y_1) \leq 0 \implies b \geq \frac{aY_1}{X_1^2} \left(2X_1 + a\right) \implies t_2 \leq \frac{b}{Y_1} F(X_1, Y_1)$$

So

$$\begin{aligned}
t_1 + t_3 &\geq \frac{b}{Y_2 - b}\left(F(X_2, Y_2) - \frac{b}{Y_2}F(X_2, Y_2)\right) - \frac{b}{Y_1 + b}\left(F(X_1, Y_1) + \frac{b}{Y_1}F(X_1, Y_1)\right) \\
&= \frac{b}{Y_2}F(X_2, Y_2) - \frac{b}{Y_1}F(X_1, Y_1) \\
&= b\left(\left(\frac{X_2}{Y_2}\right)^2 - \left(\frac{X_1}{Y_1}\right)^2\right)
\end{aligned}$$

This along with 12 proves the claim. ■

Claim: One of $\sum_{i=1}^4 s_i$ or $\sum_{i=1}^4 t_i$ is positive.

Proof: Define $S = \sum_{i=1}^4 s_i$, $T = \sum_{i=1}^4 t_i$. We first examing the case $\alpha \leq 0$. By the claim above, one of $s_1 + s_3$, $t_1 + t_3$ is positive. Since

$$s_2 + s_4 = \frac{(X_1 - \alpha)^2 - X_1^2}{Y_1} + \frac{(X_2 + \alpha)^2 - X_2^2}{Y_2}$$

it is clear that if $s_1 + s_3$ is positive, then S is. So suppose $s_1 + s_3 \leq 0$ and $t_1 + t_3 \geq 0$. Then since

$$t_2 + t_4 = \frac{(X_1 + a)^2 - X_1^2}{Y1} + \frac{(X_2 - a)^2 - X_2^2}{Y_2}$$

we have $T \geq 0$ if $a \leq 0$. We have

$$\begin{aligned}
F(X_1 + a, Y_1 + b) - F(X_1, Y_1) \leq 0 &\implies a \in \left[-X_1 \pm |X_1|\sqrt{\frac{Y_1 + b}{Y_1}}\right) \\
F(X_2 - a, Y_2 - b) - F(X_2, Y_2) \geq 0 &\implies a \notin \left(X_2 \pm |X_2|\sqrt{\frac{Y_2 - b}{Y_2}}\right)
\end{aligned}$$

so that $a > 0$ means that $a \in \left[0, \min X_1\left(\sqrt{\frac{Y_1+b}{Y_1}} - 1\right), X_2\left(1 - \sqrt{\frac{Y_2-b}{Y_2}}\right)\right]$. Again we show that $t_2 + t_4$ is a polynomial in a, with real roots at $a = 0$ and $a > 0$, and that with this constraint on a, we never violate $t_2 + t_4 \leq -(t_1 + t_3)$. The proof is technical and is given in the appendix. For $\alpha \geq 0$. note that this condition implies that all elements of X_1 are nonnegative. We can again replace X_1, X_2 with $\bar{X_1} = -X_1$, $\bar{X_2} = -X_2$, with $\bar{\alpha} = \text{argmin}_{R_j \in \bar{X_2}} G(x_i, y_i) \leq 0$. The previous subcase for $\alpha \leq 0$ can then be invoked. Alternatively, we could argue along similar lines using X_1, X_2, noting that $\frac{\alpha}{\beta} \leq \frac{a}{b}$ implies that $a \geq 0$. Since $t_2 + t_4 \geq 0$ in this case, we have that $t_1 + t_3 \geq 0$ forces $T \geq 0$. If $s_1 + s_3 \geq 0$, then one of S, T is positive as in the previous case. ■

Case 4: $X_1 \leq 0$, $X_2 \geq 0$

Defining $\bar{X_1} = -X_1$, $\bar{Y_1} = Y_1$, and $\bar{X_2} = -X_2$, $\bar{Y_2} = Y_2$ will allow us to use the previous case, if the minimal element $\bar{R_{(n)}}$ lies in $\bar{X_1}$, so that the maximal element is in the positive partition in that case. This is true if and only if the minimal element of the original partition, $m = R_{(n)}$, does not lie in X_2. If it did, then direct computation of $F(X_1 + m, Y_1 + m) - F(X_1, Y_1)$, $F(X_2 - m, Y_2 + m) - F(X_2, Y_2)$ shows that the sum $F(X_1 + m, Y_1 + m) + F(X_2 - m, Y_2 - m)$ is positive, and represents an improvement to the original partition. So we can assume that $m \notin X_2$, and the positive partition $\bar{X_1}$ contains the maximal element, and the previous case can be applied.

For the necessity, choose $\gamma = 2 + \epsilon$, for $\epsilon > 0$. Set

$$X = [1 - \delta, \delta, 1 + \delta], Y = [1, \delta, 1]$$

for some $\delta > 0$. There are 3 partitions of $\{1, 2, 3\}$, namely $\mathcal{P}_1 = \{[0], [1, 2]\}$, $\mathcal{P}_2 = \{[0, 1], [2]\}$, and $\mathcal{P}_3 = \{[0, 2], [1]\}$. Only the first 2 are ordered. The scores can be computed

$$\begin{aligned}
\text{Score}(\mathcal{P}_1) &= (1-\delta)^\gamma + \frac{(1+2\delta)^\gamma}{1+\delta} \\
\text{Score}(\mathcal{P}_2) &= \frac{1}{1+\delta} + (1+\delta)^\gamma \\
\text{Score}(\mathcal{P}_3) &= 2^{1+\epsilon}
\end{aligned}$$

The last expression is independent of δ, so can choose δ small so that the last score dominates. For $\gamma = 2 - \epsilon$, $\epsilon > 0$, and the sequences

$$X = \left[1, \frac{1}{\delta}, 1\right], Y = \left[\frac{1}{1+\delta}, \frac{1}{\delta}, \frac{1}{1-\delta}\right]$$

we have

$$\begin{aligned}
\text{Score}(\mathcal{P}_1) &= \frac{1^\gamma}{\frac{1}{1+\delta}} + \frac{\left(1+\frac{1}{\delta}\right)^\gamma}{\left(\frac{1}{\delta} + \frac{1}{1-\delta}\right)} = \frac{\left(1+\frac{1}{\delta}\right)^\gamma}{\left(\frac{1}{\delta} + \frac{1}{1-\delta}\right)} + (1+\delta) \\
\text{Score}(\mathcal{P}_2) &= \frac{1^\gamma}{\frac{1}{1-\delta}} + \frac{\left(1+\frac{1}{\delta}\right)^\gamma}{\left(\frac{1}{1+\delta} + \frac{1}{\delta}\right)} = \frac{\left(1+\frac{1}{\delta}\right)^\gamma}{\left(\frac{1}{1+\delta} + \frac{1}{\delta}\right)} + (1-\delta) \\
\text{Score}(\mathcal{P}_3) &= \frac{(1+1)^\gamma}{\left(\frac{1}{1+\delta} + \frac{1}{1-\delta}\right)} + \frac{\left(\frac{1}{\delta}\right)^\gamma}{\frac{1}{\delta}} = \frac{\left(\frac{1}{\delta}\right)^\gamma}{\frac{1}{\delta}} + 2^{\gamma-1}\left(1-\delta^2\right)
\end{aligned}$$

and letting $\delta \to 0$ gives the result, as the first summands in each score can be made arbitrarily close to each other.

To extend these examples to larger $|\mathcal{D}| = M > N$, $S \geq T$, simply add $M - N$ identical records of the form $R_i = (0, y)$, for any $y > 0$. Index the new elements by $N+1, \ldots, M$. Since $S \geq T$, an unordered candidate partition is formed by adjoining to $\mathcal{P}' = [[0, 2], [1]]$ an arbitrary partition $\mathcal{P}''$ of size $S - T$ of the indices $N+1, \ldots, M$, the new partition is $\mathcal{P} = \mathcal{P}' \cup \mathcal{P}''$. We have $\text{Score}(\mathcal{P}) = \text{Score}(\mathcal{P}')$. It is clear that inserting any record in any subset of $\mathcal{P}''$ into any subset from a partition in $\mathcal{P}'$ only decreases the total score, while inserting a record from $\mathcal{P}'$ into $\mathcal{P}''$ also represents a decrease from one of the above partitions scores for $\mathcal{P}_1$, $\mathcal{P}_2$, if the element of index 1 is switched, nonetheless the new partition won't be ordered. Of the elements 0, 2, the only one that can be switched while retaining an ordered partition is 2, in which case the new partition is of the form $[[0], [1], [2, \ldots]]$. But the score of any such partition must be less than

$$(1-\delta)^\gamma + \frac{\delta^\gamma}{\delta} + (1+\delta)^\gamma$$

which is dominated by the unordered partition $\mathcal{P}_3 = \{[0, 2], [1]\}$ above. The argument for $\gamma < 2$ is similar. In this way we can generate optimal, unordered partitions for any $\gamma > 0$ and (N, T). □

The membership of the maximal element in the last case is essential. Namely, for partitions $X_1 \leq 0$, $X_2 \geq 0$ for which the minimal element $m = R_{(n)}$ is in X_2, we reject as suboptimal out of hand. If the minimal element lies in X_2, it becomes maximal in $\bar{X}_2$, and we cannot directly apply any of the substitutions of elements across partition subsets that are used in the proof. For example, consider the case for $(N, T) = (4, 2)$, with $X = [-5.64, -5.12, 10.0, 1.94]$, $Y = [0.077, 1.23, 3.36, 0.029]$, and improvements to the suboptimal, unordred partition $[[1, 2], [0, 3]] = [X_2, X_1]$. The sequences are already sorted according to the standard priority. There are 6 partitions of $\{1, 2, 3, 4\}$, and in this case

R_1^{in}, R_1^{out} correspond to the indices 0, 2, respectively. The substitutions considered in the previous cases are

$$\left[X_2 + R_1^{in}, X_1 - R_1^{in}\right] = \left[[0,1,2]\right], [3]]$$
$$\left[X_2 - R_1^{out}, X_1 + R_1^{out}\right] = \left[[1], [0,2,3]\right]$$

In this case neither of the two partitions represent an improvement. The optimal partition is $\{[0], [1,2,3]\}$ and represents the only improvement over the partition considered:

```
SEQUENCES:
x = array([-5.64, -5.12, 10.0,  1.94])
y = array([0.077, 1.23 , 3.36, 0.029])
x/y = array([-73.24675325,  -4.16260163,   2.97619048,  66.89655172])

INDEX: 0 PARTITION: [[0, 1, 2], [3]]
    SUBSET: [0, 1, 2] SCORE: 0.12376258838654375
    SUBSET: [3]   SCORE: 129.77931034482756
    FINAL SCORE: 129.9030729332141
INDEX: 1 PARTITION: [[0, 2], [1, 3]]
    SUBSET: [0, 2]    SCORE: 5.530869944719233
    SUBSET: [1, 3]    SCORE: 8.032088959491661
    FINAL SCORE: 13.562958904210895
INDEX: 2 PARTITION: [[0], [1, 2, 3]]
SUBSET: [0]       SCORE: 413.1116883116883
    SUBSET: [1, 2, 3] SCORE: 10.069798657718122
    FINAL SCORE: 423.1814869694064
INDEX: 3 PARTITION: [[0, 1], [2, 3]]
    SUBSET: [0, 1]    SCORE: 88.58270849273144
    SUBSET: [2, 3]    SCORE: 42.06656830923576
    FINAL SCORE: 130.6492768019672
INDEX: 4 PARTITION: [[0, 1, 3], [2]]
    SUBSET: [0, 1, 3] SCORE: 58.227844311377254
    SUBSET: [2]       SCORE: 29.761904761904763
    FINAL SCORE: 87.98974907328201
INDEX: 5 PARTITION: [[0, 3], [1, 2]]
    SUBSET: [0, 3]    SCORE: 129.1509433962264
    SUBSET: [1, 2]    SCORE: 5.188322440087146
    FINAL SCORE: 134.33926583631356
INDEX: 6 PARTITION: [[0, 2, 3], [1]]
    SUBSET: [0, 2, 3] SCORE: 11.451240623196773
    SUBSET: [1]       SCORE: 21.312520325203252
    FINAL SCORE: 32.76376094840003
MAX_SUM: 423.1814869694064, MAX_PARTITION: [[0], [1, 2, 3]]
```

It is not clear how to achieve the optimal partition in one step. Even the substitution

$$\left[X_2 + R_1^{in} - R_1^{out}, X_1 - R_1^{in} + R_1^{out}\right] = \left[[0,1]\right], [2,3]]$$

does not represent an improvemnt. The improvement is obtained by substition to obtain $[[0], [1,2,3]]$, from the original $[[1,2], [0,3]]$, by moving the maximal item from X_1 to X_2, and doesn't touch R_1^{in} nor R_1^{out}.

In this case it is seems necessary to consider the partition $\bar{X}_1$, $\bar{X}_2$. for which we have

```
SEQUENCES:
x = array([ -1.94, -10.0, 5.12, 5.64])
y = array([0.029, 3.36, 1.23, 0.077])
x/y = array([-66.89655172,  -2.97619048,   4.16260163,  73.24675325])

INDEX: 0 PARTITION: [[0, 1, 2], [3]]
    SUBSET: [0, 1, 2] SCORE: 10.06979865771812
    SUBSET: [3]       SCORE: 413.1116883116883
    FINAL SCORE: 423.1814869694064
INDEX: 1 PARTITION: [[0, 2], [1, 3]]
    SUBSET: [0, 2]    SCORE: 8.032088959491661
    SUBSET: [1, 3]    SCORE: 5.530869944719233
    FINAL SCORE: 13.562958904210895
INDEX: 2 PARTITION: [[0], [1, 2, 3]]
SUBSET: [0]       SCORE: 129.77931034482756
    SUBSET: [1, 2, 3] SCORE: 0.12376258838654375
    FINAL SCORE: 129.9030729332141
INDEX: 3 PARTITION: [[0, 1], [2, 3]]
SUBSET: [0, 1]    SCORE: 42.06656830923576
    SUBSET: [2, 3]   SCORE: 88.58270849273144
    FINAL SCORE: 130.6492768019672
INDEX: 4 PARTITION: [[0, 1, 3], [2]]
SUBSET: [0, 1, 3] SCORE: 11.45124062319677
    SUBSET: [2]       SCORE: 21.312520325203252
    FINAL SCORE: 32.76376094840002
INDEX: 5 PARTITION: [[0, 3], [1, 2]]
SUBSET: [0, 3]   SCORE: 129.1509433962264
    SUBSET: [1, 2]    SCORE: 5.188322440087146
    FINAL SCORE: 134.33926583631356
INDEX: 6 PARTITION: [[0, 2, 3], [1]]
SUBSET: [0, 2, 3] SCORE: 58.227844311377254
    SUBSET: [1]       SCORE: 29.761904761904763
    FINAL SCORE: 87.98974907328201
MAX_SUM: 423.1814869694064, MAX_PARTITION: [[0, 1, 2], [3]]
```

The partition becomes $[[1,2],[0,3]] = [X_2, X_1]$ in the new setting, and

$$\left[X_2 + R_1^{in}, X_1 - R_1^{in}\right] = [[0,1,2]],[3]]$$
$$\left[X_2 - R_1^{out}, X_1 + R_1^{out}\right] = [[1],[0,2,3]]$$

which both provide an improvement.

A Appendix

Lemma 1. *For X_1, $X_2 > 0$, alpha > 0, $s1 + s3 \geq 0$, we have* $\sum_i s_i \geq 0$

Proof. We can write

$$\begin{aligned} s_2 + s_4 &= \frac{\alpha\left(\alpha - 2X_1\right)}{Y_1} + \frac{\alpha\left(\alpha + 2X_2\right)}{Y_2} \\ &= \left(\frac{Y_1}{Y_2} + \frac{Y_2}{Y_1}\right)\alpha^2 + 2\left(\frac{X_2}{Y_2} - \frac{X_1}{Y_1}\right)\alpha \end{aligned}$$

By the proof of the claim tha $s_1 + s_3 \geq 0$, we have that $s_1 + s_3 \geq \beta\left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2\right)$, so it is sufficient to show that

$$\left(\frac{Y_1}{Y_2} + \frac{Y_2}{Y_1}\right)\alpha^2 + 2\left(\frac{X_2}{Y_2} - \frac{X_1}{Y_1}\right)\alpha \geq -\beta\left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2\right)$$

Writing the left-hand side as $q\left(\alpha\right) = h\alpha^2 + g\alpha + c$, it is sufficient to show that $|\alpha||g + h\alpha| \leq |\beta\left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2\right)|$. By elementary methods, and the fact that

$$\begin{aligned} F(X_1 - \alpha, Y_1 - \beta) - F(X_1, Y_1) \geq 0 &\implies \alpha \notin \left(X_1 \pm |X_1|\sqrt{\frac{Y_1 - \beta}{Y_1}}\right) \\ F(X_2 + \alpha, Y_2 + \beta) - F(X_2, Y_2) \leq 0 &\implies \alpha \notin \left(-X_2 \pm |X_2|\sqrt{\frac{Y_2 + \beta}{Y_2}}\right) \end{aligned}$$

so that $\alpha > 0$ means that $\alpha \in \left[0, \min X_1\left(1 - \sqrt{\frac{Y_1-\beta}{Y_1}}\right)\right]$, it can be shown that since $h \leq 0$, $g \geq 0$, $|\alpha||g + h\alpha| \leq |\alpha g|$. Finally,

$$\begin{aligned} |\alpha g| = 2\alpha\left(\frac{X_1}{Y_1} - \frac{X_2}{Y_2}\right) &\leq \beta\left(\left(\frac{X_1}{Y_1}\right)^2 - \left(\frac{X_2}{Y_2}\right)^2\right) \\ \iff 2\alpha &\leq \beta\left(\frac{X_1}{Y_1} + \frac{X_2}{Y_2}\right) \end{aligned}$$

By the claim, we have $\frac{X_1}{Y_1} \geq \frac{\alpha}{\beta}$, $\frac{X_2}{Y_2} \geq \frac{2\alpha}{\beta}$, which proves the lemma.

□