

21.1 Doubly Stochastic Maps

Chapter 17 §08 defined majorization and Chapter 19 §07 proved one half of the fact that makes it useful: a doubly stochastic matrix can only even a vector out, $\mathbf{S}\mathbf{y} \prec \mathbf{y}$. The other half was promised twice. Chapter 19 §07 ended by saying that “this converse of the proposition is part of the theorem of Hardy, Littlewood, and Pólya, which Chapter 21 takes up”, and Chapter 19 §09 listed among its unfinished business “the theorem of Hardy, Littlewood, and Pólya, that $\mathbf{x} \prec \mathbf{y}$ exactly when $\mathbf{x} = \mathbf{D}\mathbf{y}$ for a doubly stochastic $\mathbf{D}$ ”. This section pays both. The proof builds $\mathbf{D}$ out of one very simple move, repeated: average two coordinates and leave the rest alone.

Throughout, vectors are real, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and $\prec$ and $\prec_w$ are Chapter 17 §08’s Definition 17.8.1. Recall $\mathbf{x}^\downarrow$, the decreasing rearrangement, and Ω_n , the set of doubly stochastic matrices of Definition 18.8.4: nonnegative entries, every row and every column adding up to 1. Recall also Proposition 19.7.6, the half already proved: if $\mathbf{S} \in \Omega_n$ and $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{S}\mathbf{y} \prec \mathbf{y}$.

21.1.1 What is still missing

Chapter 19 got the easy direction from Birkhoff’s theorem in three lines. A doubly stochastic $\mathbf{S}$ is an average of permutation matrices, a permutation only rearranges $\mathbf{y}$, and majorization cannot see a rearrangement; so $\mathbf{S}\mathbf{y}$ is an average of vectors all majorized by $\mathbf{y}$, and the set of such vectors is convex.

Read backwards, that argument gives nothing. It starts from a matrix and produces a majorization. We now want to start from the majorization $\mathbf{x} \prec \mathbf{y}$, which is a list of $n-1$ inequalities about sorted partial sums together with one equality of totals, and manufacture a matrix. Nothing in Definition 17.8.1 mentions a matrix, so one has to be built.

One such matrix is in fact already available. Chapter 19 §07 observed that Theorem 17.8.5 realizes any $\mathbf{x} \prec \mathbf{y}$ as the diagonal of a Hermitian $\mathbf{A}$ with $\lambda(\mathbf{A}) = \mathbf{y}^\downarrow$, while Lemma 19.7.5 writes that diagonal as $\mathbf{S}\mathbf{y}^\downarrow$ for a doubly stochastic $\mathbf{S}$; composing with the permutation matrix that sorts $\mathbf{y}$ gives $\mathbf{x} = \mathbf{D}\mathbf{y}$ with $\mathbf{D} \in \Omega_n$. What that route does not give is any grip on $\mathbf{D}$, which arrives as the squared moduli of the entries of a unitary matrix supplied by an existence theorem. The construction below builds $\mathbf{D}$ by hand instead, as a product of at most $n-1$ factors, each of which averages a single pair of coordinates and fixes the rest.

Here is the smallest case, done by hand, to see what a construction might look like. Let $n = 2$ and $\mathbf{x} \prec \mathbf{y}$ with both decreasing, so $x_1 + x_2 = y_1 + y_2$ and $x_1 \leq y_1$. Also $x_2 \geq y_2$, since subtracting $x_1 \leq y_1$ from the equal totals reverses the inequality. So $y_2 \leq x_1 \leq y_1$, and x_1 is a weighted average of y_1 and y_2 : if $y_1 = y_2$ then $\mathbf{x} = \mathbf{y}$, and otherwise

$$x_1 = (1-t)y_1 + ty_2, \quad t = \frac{y_1 - x_1}{y_1 - y_2} \in [0, 1].$$

The equal totals then force $x_2 = ty_1 + (1-t)y_2$. In matrix form,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1-t & t \\ t & 1-t \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

and the matrix is doubly stochastic. So for $n = 2$ and sorted vectors the converse holds, with a $\mathbf{D}$ of a very special shape.

This is the same computation as Lemma 17.8.3, where the weights were c^2 and s^2 for a rotation and the averaged numbers were two eigenvalues. There the averaging was performed by conjugating a diagonal matrix; here it is performed on the vector directly. Both say: *any number between two others is an average of them*, and that single fact is the engine of this section.

21.1.2 Averaging two coordinates

The 2×2 matrix above can be planted inside $\mathbf{I}_n$, at any pair of positions. That gives the move we will repeat.

A *T-transform* mixes two coordinates and leaves every other coordinate alone.

Definition 21.1.1. T-transform

Let $n \geq 2$, let $1 \leq j < k \leq n$, and let $\mathbf{Q} \in M_n(\mathbb{R})$ be the permutation matrix (Definition 3.4.2) of the transposition interchanging j and k . For a real number t with $0 \leq t \leq 1$, the matrix

$$\mathbf{T} = (1 - t)\mathbf{I} + t\mathbf{Q}$$

is called a **T-transform**, and we say that it **acts on the pair** $\{j, k\}$.

In words: $\mathbf{T}$ is the convex combination, with weights $1 - t$ and t , of doing nothing and swapping two coordinates. Applied to a vector,

$$\begin{aligned} (\mathbf{T}\mathbf{y})_j &= (1 - t)y_j + ty_k, \\ (\mathbf{T}\mathbf{y})_k &= ty_j + (1 - t)y_k, \\ (\mathbf{T}\mathbf{y})_i &= y_i \quad (i \neq j, k). \end{aligned} \tag{21.1.1}$$

The two moved entries are averages of the old y_j and y_k , with the **same** pair of weights used in both slots, exchanged. In particular $(\mathbf{T}\mathbf{y})_j + (\mathbf{T}\mathbf{y})_k = y_j + y_k$: a T-transform conserves the total, and indeed conserves the total of the pair it touches.

The first thing to check is that these matrices belong in the discussion at all.

Proposition 21.1.1. T-transforms Are Doubly Stochastic

Every T-transform lies in Ω_n .

Proof. Let $\mathbf{T} = (1 - t)\mathbf{I} + t\mathbf{Q}$ with $0 \leq t \leq 1$, and let $\{j, k\}$ be the pair $\mathbf{Q}$ interchanges. In the two positions (j, j) and (k, k) the entry of $\mathbf{T}$ is $1 - t$, in the two positions (j, k) and (k, j) it is t , in the remaining $n - 2$ diagonal positions both $\mathbf{I}$ and $\mathbf{Q}$ carry a 1, so the entry is 1, and everywhere else it is 0. Each entry is therefore 1, $1 - t$, t , or 0, hence nonnegative, since $t \geq 0$ and $1 - t \geq 0$. A permutation matrix has exactly one 1 in each row and each column and zeros elsewhere (the remark after Definition 3.4.2), so $\mathbf{Q}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\top \mathbf{Q} = \mathbf{1}^\top$; the same is true of $\mathbf{I}$, which is the permutation matrix of the identity permutation. Hence

$$\mathbf{T}\mathbf{1} = (1 - t)\mathbf{1} + t\mathbf{1} = \mathbf{1}, \quad \mathbf{1}^\top \mathbf{T} = (1 - t)\mathbf{1}^\top + t\mathbf{1}^\top = \mathbf{1}^\top,$$

which says that every row and every column of $\mathbf{T}$ adds up to 1. By Definition 18.8.4, $\mathbf{T} \in \Omega_n$. ■

Example. Three T-transforms, and two impostors

Take $n = 3$ and $\mathbf{y} = (6, 3, 0)$.

- (a) Compute $\mathbf{T}\mathbf{y}$ for the T-transform on $\{1, 3\}$ with $t = \frac{1}{3}$, and for the ones with $t = 0$ and $t = 1$.
- (b) Is the matrix $\frac{1}{2}(\mathbf{I} + \mathbf{P})$, with $\mathbf{P}$ the permutation matrix of the 3-cycle $1 \mapsto 2 \mapsto 3 \mapsto 1$, a T-transform?
- (c) Is $-\mathbf{I} + 2\mathbf{Q}$, with $\mathbf{Q}$ the transposition matrix of $\{1, 2\}$, a T-transform?

Solution.

- (a) With $j = 1$, $k = 3$, and $t = \frac{1}{3}$, Eq. 21.1.1 gives

$$\mathbf{T}\mathbf{y} = \left(\frac{2}{3} \cdot 6 + \frac{1}{3} \cdot 0, 3, \frac{1}{3} \cdot 6 + \frac{2}{3} \cdot 0\right) = (4, 3, 2).$$

At $t = 0$, $\mathbf{T} = \mathbf{I}$ and $\mathbf{T}\mathbf{y} = \mathbf{y}$; at $t = 1$, $\mathbf{T} = \mathbf{Q}$ and $\mathbf{T}\mathbf{y} = (0, 3, 6)$. So the endpoints of the family are “do nothing” and “swap”.

- (b) No. It is doubly stochastic, by the computation in Proposition 21.1.1 with $\mathbf{Q}$ replaced by $\mathbf{P}$, but Definition 21.1.1 asks for the permutation matrix of a **transposition**. Concretely, $\frac{1}{2}(\mathbf{I} + \mathbf{P})$ changes all three coordinates of $\mathbf{y}$, while a T-transform fixes at least $n - 2$ of them.
- (c) No: $t = 2$ is outside $[0, 1]$, and the matrix has the entry -1 in position $(1, 1)$, so it is not even doubly stochastic. Applied to $\mathbf{y}$ it gives $(0, 9, 0)$, whose first two entries are further apart than 6 and 3: with $t \notin [0, 1]$ the “average” is not between the two numbers, and the move spreads rather than evens.

Quick check. Let $\mathbf{T}$ be the T-transform on $\{1, 2\}$ with $t = \frac{1}{4}$, and let $\mathbf{y} = (8, 4, 1)$. Compute $\mathbf{T}\mathbf{y}$, and verify $\mathbf{T}\mathbf{y} \prec \mathbf{y}$ directly from Definition 17.8.1.

Solution. $(\mathbf{T}\mathbf{y})_1 = \frac{3}{4} \cdot 8 + \frac{1}{4} \cdot 4 = 7$ and $(\mathbf{T}\mathbf{y})_2 = \frac{1}{4} \cdot 8 + \frac{3}{4} \cdot 4 = 5$, so $\mathbf{T}\mathbf{y} = (7, 5, 1)$. Both vectors are decreasing with total 13, and the running totals are 7, 12 against 8, 12. Since $7 \leq 8$ and $12 \leq 12$, (M1) holds, and (M2) is the equal totals. So $\mathbf{T}\mathbf{y} \prec \mathbf{y}$, as Proposition 19.7.6 already promised, since $\mathbf{T} \in \Omega_n$.

Warning. A product of T-transforms is usually not a T-transform. T-transforms are doubly stochastic, and so are their products (this is checked inside the proof of the main theorem below), but the product of two of them can move every coordinate. Averaging coordinates 1 and 2, then coordinates 2 and 3, changes all three. The theorem below produces a product, not a single T-transform.

21.1.3 One averaging step

Now the engine. Given $\mathbf{x} \prec \mathbf{y}$, we want to push $\mathbf{y}$ toward $\mathbf{x}$ by one T-transform, and we want a measure of progress that cannot stall. The measure is the number of coordinates in which the two vectors already agree.

Lemma 21.1.2. One Averaging Step

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be **decreasing**, with $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$. Then there is a T-transform $\mathbf{T}$ such that $\mathbf{y}' = \mathbf{T}\mathbf{y}$ satisfies

- (a) $\mathbf{y}'$ is decreasing;
- (b) $\mathbf{x} \prec \mathbf{y}'$;
- (c) $y'_i = x_i$ for strictly more indices i than $y_i = x_i$ does.

Idea. Both vectors are sorted, so compare them entry by entry. Somewhere $\mathbf{y}$ is too big and somewhere it is too small. Let j be the **last** place where $\mathbf{y}$ is too big and k the **first** place after j where it is too small; between them the two vectors already agree. Averaging coordinates j and k moves y_j down and y_k up by the same amount, and we move by exactly as much as the tighter of the two deficits allows, so that one of the two positions lands on $\mathbf{x}$ exactly. Two things then need checking, and they are not equally hard. The partial sums are the easy one: the only ones that drop are those between j and k , and they have slack at least $y_j - x_j$ to give away, while we give away no more than that. The delicate one is that $\mathbf{y}'$ must come out **decreasing**, since otherwise the lemma cannot be applied to it again; that is what taking j last buys, because it makes the entries between j and k agree with $\mathbf{x}$'s.

Proof. Write $X_m = x_1 + \cdots + x_m$ and $Y_m = y_1 + \cdots + y_m$, with $X_0 = Y_0 = 0$. Since both vectors are decreasing, Definition 17.8.1 says $X_m \leq Y_m$ for $m = 1, \dots, n-1$ and $X_n = Y_n$.

Step 1: the two indices. Some index has $y_i > x_i$: otherwise $y_i \leq x_i$ for every i , and since $\mathbf{x} \neq \mathbf{y}$ at least one inequality is strict, so $Y_n < X_n$, contradicting $X_n = Y_n$. Let j be the **largest** index with $y_j > x_j$.

Next, some index $i > j$ has $y_i < x_i$. Suppose not. By the choice of j , $y_i \leq x_i$ for every $i > j$; so $y_i = x_i$ for every $i > j$, and subtracting these from $X_n = Y_n$ gives $X_j = Y_j$. Then

$$X_{j-1} = X_j - x_j > Y_j - y_j = Y_{j-1},$$

because $x_j < y_j$. For $j \geq 2$ this contradicts $X_{j-1} \leq Y_{j-1}$, and for $j = 1$ it reads $0 > 0$, which is false. So we may let k be the **smallest** index with $k > j$ and $y_k < x_k$.

By the maximality of j we have $y_i \geq x_i$ for $j < i < k$, and by the minimality of k we have $y_i \leq x_i$ there. Hence

$$y_i = x_i \quad (j < i < k). \quad (*)$$

Step 2: the transform. Put

$$\delta = \min\{y_j - x_j, x_k - y_k\} > 0.$$

Since $\mathbf{x}$ is decreasing and $j < k$, $x_j \geq x_k$, so

$$y_j > x_j \geq x_k > y_k,$$

and in particular $y_j - y_k > 0$. Also $\delta \leq y_j - x_j \leq y_j - y_k$, using $x_j \geq x_k > y_k$. So

$$t = \frac{\delta}{y_j - y_k}$$

satisfies $0 < t \leq 1$, and Definition 21.1.1 gives a T-transform $\mathbf{T}$ acting on $\{j, k\}$ with this t . By Eq. 21.1.1,

$$y'_j = y_j - t(y_j - y_k) = y_j - \delta, \quad y'_k = y_k + \delta,$$

and $y'_i = y_i$ for every other i .

Step 3: (c), the count. By the choice of δ , either $\delta = y_j - x_j$, and then $y'_j = x_j$; or $\delta = x_k - y_k$, and then $y'_k = x_k$. So $\mathbf{y}'$ agrees with $\mathbf{x}$ in at least one of the positions j, k . Neither position agreed before, since $y_j > x_j$ and $y_k < x_k$. Every other position is untouched. Hence the number of agreements strictly increases.

Step 4: (a), the sorted order. Only positions j and k change, and $\delta > 0$ lowers y_j and raises y_k . From $\delta \leq y_j - x_j$ and $\delta \leq x_k - y_k$,

$$y'_j \geq x_j, \quad y'_k \leq x_k.$$

If $j \geq 2$, then $y'_{j-1} = y_{j-1} \geq y_j > y'_j$. If $j+1 < k$, then $y'_{j+1} = y_{j+1} = x_{j+1}$ by (*), and $x_{j+1} \leq x_j \leq y'_j$. If $j+1 = k$, then $y'_{j+1} = y'_k \leq x_k \leq x_j \leq y'_j$. Symmetrically, if $k-1 > j$ then $y'_{k-1} = x_{k-1} \geq x_k \geq y'_k$, and if $k \leq n-1$ then $y'_{k+1} = y_{k+1} \leq y_k < y'_k$. Every remaining consecutive pair consists of two untouched entries of the decreasing $\mathbf{y}$. So $\mathbf{y}'$ is decreasing.

Step 5: (b), the majorization. Write $Y'_m = y'_1 + \cdots + y'_m$. Since $\mathbf{y}'$ differs from $\mathbf{y}$ only at j and k , by $-\delta$ and $+\delta$,

$$Y'_m = \begin{cases} Y_m - \delta & (j \leq m \leq k-1), \\ Y_m & (\text{otherwise}). \end{cases}$$

In particular $Y'_n = Y_n = X_n$, which is (M2). For $m < j$ and for $m \geq k$ we have $Y'_m = Y_m \geq X_m$, which is (M1) there. For $j \leq m \leq k-1$, the identity (*) gives $y_i = x_i$ for $j < i \leq m$, hence

$$Y_m - X_m = Y_j - X_j.$$

Now $Y_j - X_j = (Y_{j-1} - X_{j-1}) + (y_j - x_j) \geq y_j - x_j \geq \delta$, using $Y_{j-1} \geq X_{j-1}$ (which is (M1), or $0 \geq 0$ when $j = 1$) and the definition of δ . Therefore

$$Y'_m - X_m = (Y_m - X_m) - \delta = (Y_j - X_j) - \delta \geq 0,$$

which is (M1) at m . So $\mathbf{x} \prec \mathbf{y}'$. This proves the lemma. ■

The choice of the two indices is the only delicate part, and it is worth saying why the obvious alternative fails, and where. Suppose j were taken as the *smallest* index with $y_j > x_j$, keeping k as the smallest index after j with $y_k < x_k$. Step 5 would survive: the minimality of k still gives $y_i \geq x_i$ for $j < i < k$, so the slack $Y_m - X_m$ is nondecreasing on $j \leq m \leq k-1$ instead of constant, and it is therefore at least its value $Y_j - X_j \geq y_j - x_j \geq \delta$ at $m = j$. What breaks is Step 4. The entries strictly between j and k need no longer agree with $\mathbf{x}$'s, and lowering y_j by δ can push it below y_{j+1} , so that $\mathbf{y}'$ comes out unsorted and Lemma 21.1.2 cannot be applied to it a second time. Take $\mathbf{x} = (5, 5, 0)$ and $\mathbf{y} = (6, 6, -2)$, both decreasing, with $\mathbf{x} \prec \mathbf{y}$ since $5 \leq 6$, $10 \leq 12$, and both totals are 10. Choosing j smallest gives $j = 1$ and $k = 3$, so $\delta = \min\{6 - 5, 0 - (-2)\} = 1$ and $\mathbf{y}' = (5, 6, -1)$, which is not decreasing; the rule of the lemma gives $j = 2$ instead, and $\mathbf{y}' = (6, 5, -1)$, which is. Taking j last and k first after it makes the interval between them a stretch where the two vectors already coincide, and that stretch is what protects the order.

21.1.4 The theorem

Theorem 21.1.3. Hardy–Littlewood–Pólya Theorem

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$\mathbf{x} \prec \mathbf{y} \iff \mathbf{x} = \mathbf{D}\mathbf{y} \text{ for some } \mathbf{D} \in \Omega_n.$$

Moreover, if $\mathbf{x}$ and $\mathbf{y}$ are both decreasing, $\mathbf{D}$ may be taken to be a product of at most $n - 1$ T-transforms.

Idea. ($\Leftarrow$) is Chapter 19’s Proposition 19.7.6 and needs nothing new.

For ($\Rightarrow$), sort both vectors first, since neither side of the equivalence sees the order. Then count agreements. Lemma 21.1.2 raises that count by at least one, keeps both hypotheses alive, and therefore may be applied again; the count is at most n , so the process stops, at $\mathbf{y}$ turned into $\mathbf{x}$. The bound $n - 1$ comes from noticing that the count can never be exactly $n - 1$: two vectors with equal totals that agree in all but one coordinate agree in that one too.

Proof. ($\Leftarrow$) If $\mathbf{x} = \mathbf{D}\mathbf{y}$ with $\mathbf{D} \in \Omega_n$, then $\mathbf{x} \prec \mathbf{y}$ by Proposition 19.7.6.

($\Rightarrow$) Suppose $\mathbf{x} \prec \mathbf{y}$.

Step 1: the decreasing case. Assume first that $\mathbf{x}$ and $\mathbf{y}$ are both decreasing. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ write $\text{dis}(\mathbf{u}, \mathbf{v})$ for the number of indices i with $u_i \neq v_i$. We claim:

Claim. If $\mathbf{x}$ and $\mathbf{y}$ are decreasing with $\mathbf{x} \prec \mathbf{y}$ and $d = \text{dis}(\mathbf{x}, \mathbf{y})$, then $\mathbf{x} = \mathbf{T}_r \cdots \mathbf{T}_1 \mathbf{y}$ for some T-transforms $\mathbf{T}_1, \dots, \mathbf{T}_r$ with $r = 0$ if $d = 0$, and $r \leq d - 1$ otherwise.

Proof. First, $d \neq 1$: if $\mathbf{x}$ and $\mathbf{y}$ agreed in all indices but one, say i_0 , then subtracting those $n - 1$ equalities from the equal totals (M2) would give $x_{i_0} = y_{i_0}$ too, so $d = 0$. We induct on d .

If $d = 0$ then $\mathbf{x} = \mathbf{y}$, and the empty product, $r = 0$, works. Let $d \geq 2$, and assume the claim for every smaller value. By Lemma 21.1.2 there is a T-transform $\mathbf{T}_1$ with $\mathbf{y}' = \mathbf{T}_1 \mathbf{y}$ decreasing, $\mathbf{x} \prec \mathbf{y}'$, and $d' = \text{dis}(\mathbf{x}, \mathbf{y}') \leq d - 1$. If $d' = 0$, then $\mathbf{x} = \mathbf{T}_1 \mathbf{y}$ and $r = 1 \leq d - 1$. Otherwise $d' \geq 2$, since $d' \neq 1$ by the first paragraph applied to $\mathbf{x}$ and $\mathbf{y}'$; the inductive hypothesis gives $\mathbf{x} = \mathbf{T}_r \cdots \mathbf{T}_2 \mathbf{y}'$ with $r - 1 \leq d' - 1$, so $r \leq d' \leq d - 1$. This proves the claim. ■

Since $d \leq n$, the claim gives $r \leq d - 1 \leq n - 1$ in every case. This proves the “moreover” clause.

Step 2: products stay doubly stochastic. If $\mathbf{A}, \mathbf{B} \in \Omega_n$, then $\mathbf{AB}$ has nonnegative entries, and $\mathbf{AB}\mathbf{1} = \mathbf{A}\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top \mathbf{AB} = \mathbf{1}^\top \mathbf{B} = \mathbf{1}^\top$; so $\mathbf{AB} \in \Omega_n$ by Definition 18.8.4. By induction, any finite product of doubly stochastic matrices is doubly stochastic. With Proposition 21.1.1, the matrix $\mathbf{D}_0 = \mathbf{T}_r \cdots \mathbf{T}_1$ of Step 1 lies in Ω_n .

Step 3: the general case. Let $\mathbf{x} \prec \mathbf{y}$ be arbitrary. Majorization sees only rearrangements

(the remark after Definition 17.8.1), so $\mathbf{x}^\downarrow \prec \mathbf{y}^\downarrow$, and Step 1 gives $\mathbf{D}_0 \in \Omega_n$ with $\mathbf{x}^\downarrow = \mathbf{D}_0 \mathbf{y}^\downarrow$. Choose permutations $\sigma, \tau \in S_n$ with $\mathbf{y}^\downarrow = \mathbf{P}_\sigma \mathbf{y}$ and $\mathbf{x} = \mathbf{P}_\tau \mathbf{x}^\downarrow$; these exist because $\mathbf{y}^\downarrow$ is a rearrangement of $\mathbf{y}$ and $\mathbf{x}$ of $\mathbf{x}^\downarrow$. Each $\mathbf{P}_\sigma$ is doubly stochastic, having exactly one 1 in each row and each column and zeros elsewhere (the remark after Definition 3.4.2), so by Step 2

$$\mathbf{D} = \mathbf{P}_\tau \mathbf{D}_0 \mathbf{P}_\sigma \in \Omega_n, \quad \mathbf{D}\mathbf{y} = \mathbf{P}_\tau \mathbf{D}_0 \mathbf{y}^\downarrow = \mathbf{P}_\tau \mathbf{x}^\downarrow = \mathbf{x}.$$

This proves the theorem. ■

Two remarks on what the theorem does and does not say. It converts a family of $n-1$ inequalities and one equality into a single algebraic identity $\mathbf{x} = \mathbf{D}\mathbf{y}$, and an identity is much easier to feed into another argument, as the next section does, one row of $\mathbf{D}$ at a time. And the $n-1$ in the “moreover” clause is sharp; [Exercise \(C3: The count \$n-1\$ cannot be improved\)](#) exhibits a pair needing that many.

Remark. The word **decreasing** in the “moreover” clause is not decoration. For $\mathbf{y} = (6, 4, 2, 0)$ and the unsorted $\mathbf{x} = (\frac{3}{2}, 1, \frac{9}{2}, 5)$ we have $\mathbf{x} \prec \mathbf{y}$, since $\mathbf{x}^\downarrow = (5, \frac{9}{2}, \frac{3}{2}, 1)$ has running totals $5, \frac{19}{2}, 11$ against $6, 10, 12$ and both totals are 12; yet no product of three T-transforms equals $\mathbf{x}$, since for each of the finitely many choices of the three pairs the best the three parameters can do still misses it. Four transforms suffice: the transforms on $\{1, 2\}$, $\{3, 4\}$, $\{1, 3\}$, $\{2, 4\}$ with $t = \frac{1}{2}, \frac{1}{2}, \frac{7}{8}, 1$ carry $\mathbf{y}$ through $(5, 5, 2, 0)$, $(5, 5, 1, 1)$, and $(\frac{3}{2}, 5, \frac{9}{2}, 1)$ to $\mathbf{x}$. Step 3 of the proof handles an unsorted pair with permutations instead, and a permutation matrix is a T-transform only when it is a transposition.

Example. Four coordinates, two transforms

Let $\mathbf{y} = (7, 5, 3, 1)$ and $\mathbf{x} = (5, 5, 4, 2)$. Verify $\mathbf{x} \prec \mathbf{y}$, run the construction of [Theorem 21.1.3](#), and write out the resulting $\mathbf{D}$.

Solution. Both vectors are decreasing with total 16, and the running totals are 5, 10, 14 for $\mathbf{x}$ against 7, 12, 15 for $\mathbf{y}$. So (M1) and (M2) hold and $\mathbf{x} \prec \mathbf{y}$.

First step. Here $\mathbf{y} - \mathbf{x} = (2, 0, -1, -1)$, so the only index with $y_i > x_i$ is $i = 1$, giving $j = 1$; the smallest index above it with $y_i < x_i$ is $k = 3$. Then

$$\delta = \min\{7 - 5, 4 - 3\} = 1, \quad t = \frac{1}{7 - 3} = \frac{1}{4},$$

and $\mathbf{T}_1$ is the T-transform on $\{1, 3\}$ with $t = \frac{1}{4}$:

$$\mathbf{y}' = \mathbf{T}_1 \mathbf{y} = (\frac{3}{4} \cdot 7 + \frac{1}{4} \cdot 3, 5, \frac{1}{4} \cdot 7 + \frac{3}{4} \cdot 3, 1) = (6, 5, 4, 1).$$

Now $\mathbf{y}'$ agrees with $\mathbf{x}$ in positions 2 and 3, where before it agreed only in position 2.

Second step. $\mathbf{y}' - \mathbf{x} = (1, 0, 0, -1)$, so $j = 1$ and $k = 4$, with $\delta = \min\{1, 1\} = 1$ and $t = 1/(6 - 1) = \frac{1}{5}$. The T-transform $\mathbf{T}_2$ on $\{1, 4\}$ gives

$$\mathbf{T}_2 \mathbf{y}' = (\frac{4}{5} \cdot 6 + \frac{1}{5} \cdot 1, 5, 4, \frac{1}{5} \cdot 6 + \frac{4}{5} \cdot 1) = (5, 5, 4, 2) = \mathbf{x}.$$

Two transforms, and $2 \leq n - 1 = 3$.

The matrix. Multiplying $\mathbf{T}_2\mathbf{T}_1$ row by row,

$$\mathbf{D} = \mathbf{T}_2\mathbf{T}_1 = \begin{pmatrix} \frac{3}{5} & 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ \frac{3}{20} & 0 & \frac{1}{20} & \frac{4}{5} \end{pmatrix}.$$

Every row adds up to 1, and so does every column: the first is $\frac{3}{5} + \frac{1}{4} + \frac{3}{20} = \frac{12+5+3}{20} = 1$, the third is $\frac{1}{5} + \frac{3}{4} + \frac{1}{20} = \frac{4+15+1}{20} = 1$, and the second and fourth are visibly 1. Finally

$$\mathbf{D}\mathbf{y} = \left(\frac{21+3+1}{5}, 5, \frac{7+9}{4}, \frac{21+3+16}{20}\right) = (5, 5, 4, 2) = \mathbf{x}.$$

Note that $\mathbf{D}$ is not a T-transform: it moves three of the four coordinates.

Warning. The matrix $\mathbf{D}$ is far from unique, and “doubly stochastic” cannot be relaxed to “row-stochastic”. For $\mathbf{y} = (2, 1, 0)$ and $\mathbf{x} = (1, 1, 1)$, both

$$\mathbf{D}_1 = \frac{1}{3}\mathbf{J}, \quad \mathbf{D}_2 = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

are doubly stochastic with $\mathbf{D}_i\mathbf{y} = \mathbf{x}$. And if only the **rows** are required to add up to 1, the conclusion fails: with $\mathbf{y} = (1, 0)$ and the row-stochastic $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{S}\mathbf{y} = (1, 1)$, whose total is $2 \neq 1$, so $\mathbf{S}\mathbf{y} \not\prec \mathbf{y}$. The column condition is what enforces the equal totals.

21.1.5 Majorization as a polytope

Birkhoff’s theorem turns the doubly stochastic matrix of [Theorem 21.1.3](#) into an average of permutations, and the statement becomes geometric.

Corollary 21.1.4. Majorization Is Membership in a Permutation Polytope

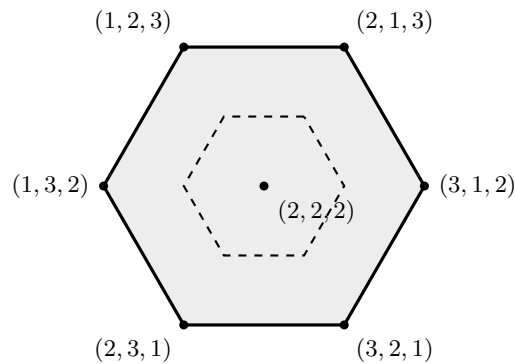
Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$\mathbf{x} \prec \mathbf{y} \iff \mathbf{x} \in \text{conv}\{\mathbf{P}_\sigma \mathbf{y} : \sigma \in S_n\}.$$

Proof. ($\Rightarrow$) By [Theorem 21.1.3](#), $\mathbf{x} = \mathbf{D}\mathbf{y}$ with $\mathbf{D} \in \Omega_n$. By Theorem 19.7.2, $\mathbf{D} \in \text{conv}\{\mathbf{P}_\sigma : \sigma \in S_n\}$, so by Theorem 18.1.3, $\mathbf{D} = \sum_{i=1}^m t_i \mathbf{P}_{\sigma_i}$ with $t_i \geq 0$ and $\sum_i t_i = 1$. Hence $\mathbf{x} = \sum_i t_i \mathbf{P}_{\sigma_i} \mathbf{y}$, a convex combination of the vectors $\mathbf{P}_{\sigma_i} \mathbf{y}$, which lies in their convex hull by Theorem 18.1.3.

($\Leftarrow$) By Theorem 18.1.3, $\mathbf{x} = \sum_i t_i \mathbf{P}_{\sigma_i} \mathbf{y}$ with $(t_1, \dots, t_m)$ a list of nonnegative weights adding up to 1. Put $\mathbf{D} = \sum_i t_i \mathbf{P}_{\sigma_i}$. Its entries are nonnegative, and $\mathbf{D}\mathbf{1} = \sum_i t_i \mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top \mathbf{D} = \mathbf{1}^\top$, so $\mathbf{D} \in \Omega_n$ and $\mathbf{x} = \mathbf{D}\mathbf{y}$. By Proposition 19.7.6, $\mathbf{x} \prec \mathbf{y}$. ■

The set $\text{conv}\{\mathbf{P}_\sigma \mathbf{y}\}$ is the **permutation polytope** of $\mathbf{y}$: the convex hull of the $n!$ rearrangements of $\mathbf{y}$. It lies in the affine subspace where the coordinate sum equals $\sum_i y_i$, so its dimension is at most $n - 1$. For $n = 3$ and distinct entries it is a hexagon.



The solid hexagon is the permutation polytope of $\mathbf{y} = (3, 2, 1)$, drawn inside the plane $x_1 + x_2 + x_3 = 6$; the dashed one, at half the size, is the polytope of $(\frac{5}{2}, 2, \frac{3}{2})$. By [Corollary 21.1.4](#), $\mathbf{x} \prec \mathbf{y}$ says that $\mathbf{x}$ lies in the solid hexagon, and $\mathbf{x} \prec (\frac{5}{2}, 2, \frac{3}{2}) \prec \mathbf{y}$ says that the dashed hexagon sits inside the solid one. *The majorization order is containment of permutation polytopes.* The center $(2, 2, 2)$, the flat vector, lies in every one of them, which is Example (A chain, the two extremes, and size one) (b) again.

Quick check. Use [Corollary 21.1.4](#) to give a second proof that $\{\mathbf{x} : \mathbf{x} \prec \mathbf{y}\}$ is convex, a fact Chapter 18 proved by hand in Example (The vectors majorized by a fixed vector).

Solution. By the corollary, $\{\mathbf{x} : \mathbf{x} \prec \mathbf{y}\} = \text{conv}\{\mathbf{P}_\sigma \mathbf{y} : \sigma \in S_n\}$. A convex hull is convex (Theorem 18.1.3, or Definition 18.1.4 directly). So the set is convex.

21.1.6 Exercises

21.1.6.1 A. Check your understanding

Exercise. A1

- Define a T-transform, and say which coordinates of a vector it can change.
- State the Hardy–Littlewood–Pólya theorem in full.
- True or false: if $\mathbf{x} = \mathbf{S}\mathbf{y}$ with $\mathbf{S}$ row-stochastic (nonnegative entries, every row adding up to 1), then $\mathbf{x} \prec \mathbf{y}$. Justify your answer.
- True or false: if $\mathbf{x} \prec \mathbf{y}$, the doubly stochastic $\mathbf{D}$ with $\mathbf{x} = \mathbf{D}\mathbf{y}$ is unique. Justify your answer.
- In [Lemma 21.1.2](#), why is j chosen as the **largest** index with $y_j > x_j$ rather than the smallest?
- Why can the number of coordinates in which two decreasing vectors $\mathbf{x} \prec \mathbf{y}$ agree never be exactly $n - 1$?

Solution.

- (a) For $1 \leq j < k \leq n$ and $0 \leq t \leq 1$, it is $\mathbf{T} = (1-t)\mathbf{I} + t\mathbf{Q}$ with $\mathbf{Q}$ the permutation matrix of the transposition of j and k (Definition 21.1.1). It changes only the coordinates j and k , replacing each by an average of the two old values.
- (b) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} \prec \mathbf{y}$ if and only if $\mathbf{x} = \mathbf{D}\mathbf{y}$ for some doubly stochastic $\mathbf{D}$; and when $\mathbf{x}$ and $\mathbf{y}$ are both decreasing, $\mathbf{D}$ may be taken to be a product of at most $n-1$ T-transforms (Theorem 21.1.3).
- (c) False. With $\mathbf{y} = (1, 0)$ and $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{S}\mathbf{y} = (1, 1)$ has total 2 while $\mathbf{y}$ has total 1, so (M2) fails. Row sums alone control nothing but the individual entries; the column sums are what preserve the total.
- (d) False. For $\mathbf{y} = (2, 1, 0)$ and $\mathbf{x} = (1, 1, 1)$, both $\frac{1}{3}\mathbf{J}$ and the matrix $\mathbf{D}_2$ of the warning after Example (Four coordinates, two transforms) send $\mathbf{y}$ to $\mathbf{x}$.
- (e) Because it makes the entries strictly between j and k agree with $\mathbf{x}$'s, which is what Step 4 needs in order to keep $\mathbf{y}'$ decreasing. The majorization would survive the other choice: with j smallest, the slack $Y_m - X_m$ is nondecreasing on $j \leq m \leq k-1$ rather than constant, hence still at least δ throughout, and Step 5 goes through. The sorted order is what fails, and with it the possibility of applying the lemma again. For $\mathbf{x} = (5, 5, 0)$ and $\mathbf{y} = (6, 6, -2)$, taking j smallest gives $j = 1$, $k = 3$, $\delta = 1$, and $\mathbf{y}' = (5, 6, -1)$, which is not decreasing, while the rule of the lemma gives $j = 2$ and $\mathbf{y}' = (6, 5, -1)$, which is.
- (f) Suppose they agree in all indices but i_0 . Subtracting the $n-1$ equalities $x_i = y_i$ from the equal totals of (M2) leaves $x_{i_0} = y_{i_0}$, so in fact they agree everywhere.

21.1.6.2 B. Practice

Exercise. B1: Run the construction

Let $\mathbf{y} = (9, 5, 4)$ and $\mathbf{x} = (6, 6, 6)$. Verify $\mathbf{x} \prec \mathbf{y}$, produce T-transforms $\mathbf{T}_1, \mathbf{T}_2$ with $\mathbf{x} = \mathbf{T}_2\mathbf{T}_1\mathbf{y}$ by the recipe of Lemma 21.1.2, and write out $\mathbf{D} = \mathbf{T}_2\mathbf{T}_1$. Hence check that $\mathbf{D}$ is doubly stochastic.

Solution. Both vectors are decreasing with total 18, and the running totals are 6, 12 against 9, 14. So (M1) and (M2) hold and $\mathbf{x} \prec \mathbf{y}$.

First step. $\mathbf{y} - \mathbf{x} = (3, -1, -2)$, so $j = 1$ and $k = 2$. Then $\delta = \min\{3, 1\} = 1$ and $t = 1/(9-5) = \frac{1}{4}$, and $\mathbf{T}_1$ acts on $\{1, 2\}$:

$$\mathbf{y}' = \left(\frac{3}{4} \cdot 9 + \frac{1}{4} \cdot 5, \frac{1}{4} \cdot 9 + \frac{3}{4} \cdot 5, 4\right) = (8, 6, 4).$$

Second step. $\mathbf{y}' - \mathbf{x} = (2, 0, -2)$, so $j = 1$, $k = 3$, $\delta = \min\{2, 2\} = 2$, and $t = 2/(8-4) = \frac{1}{2}$. Then

$$\mathbf{T}_2\mathbf{y}' = \left(\frac{1}{2} \cdot 8 + \frac{1}{2} \cdot 4, 6, \frac{1}{2} \cdot 8 + \frac{1}{2} \cdot 4\right) = (6, 6, 6) = \mathbf{x}.$$

The matrix. $\mathbf{T}_1$ has rows $(\frac{3}{4}, \frac{1}{4}, 0)$, $(\frac{1}{4}, \frac{3}{4}, 0)$, $(0, 0, 1)$, and $\mathbf{T}_2$ has rows $(\frac{1}{2}, 0, \frac{1}{2})$, $(0, 1, 0)$,

$(\frac{1}{2}, 0, \frac{1}{2})$. Multiplying,

$$\mathbf{D} = \mathbf{T}_2 \mathbf{T}_1 = \begin{pmatrix} \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \end{pmatrix}.$$

Each row adds up to 1, and the columns add up to $\frac{3}{8} + \frac{1}{4} + \frac{3}{8} = 1$, $\frac{1}{8} + \frac{3}{4} + \frac{1}{8} = 1$, and $\frac{1}{2} + 0 + \frac{1}{2} = 1$. So $\mathbf{D} \in \Omega_3$, and $\mathbf{D}\mathbf{y} = (\frac{27}{8} + \frac{5}{8} + 2, \frac{9}{4} + \frac{15}{4}, \frac{27}{8} + \frac{5}{8} + 2) = (6, 6, 6)$.

Exercise. B2: Determine which are majorizations

For each pair, determine whether $\mathbf{x} \prec \mathbf{y}$. Where it holds, exhibit a doubly stochastic $\mathbf{D}$ with $\mathbf{x} = \mathbf{D}\mathbf{y}$; where it fails, name the clause of Definition 17.8.1 and the value of k at which it fails. Justify your answer.

- (a) $\mathbf{x} = (2, 2, 2)$, $\mathbf{y} = (4, 1, 1)$.
- (b) $\mathbf{x} = (3, 1, 0)$, $\mathbf{y} = (2, 2, 0)$.
- (c) $\mathbf{x} = (1, 3, 0)$, $\mathbf{y} = (2, 0, 2)$.
- (d) $\mathbf{x} = (1, 1)$, $\mathbf{y} = (2, 1)$.

Solution.

- (a) Both decreasing, totals $6 = 6$; running totals 2, 4 against 4, 5. So $\mathbf{x} \prec \mathbf{y}$. Here $\mathbf{x}$ is the flat vector, and $\mathbf{D} = \frac{1}{3}\mathbf{J}$ works: every entry of $\frac{1}{3}\mathbf{J}\mathbf{y}$ is $\frac{1}{3}(4 + 1 + 1) = 2$.
- (b) Totals $4 = 4$, but at $k = 1$ the sorted first entries are $3 > 2$, so (M1) fails at $k = 1$ and $\mathbf{x} \not\prec \mathbf{y}$.
- (c) Sorted, $\mathbf{x}^\downarrow = (3, 1, 0)$ and $\mathbf{y}^\downarrow = (2, 2, 0)$; this is (b) again, so $\mathbf{x} \not\prec \mathbf{y}$, with (M1) failing at $k = 1$.
- (d) Totals $2 \neq 3$, so (M2) fails and $\mathbf{x} \not\prec \mathbf{y}$. (Both running totals do satisfy (M1), so $\mathbf{x} \prec_w \mathbf{y}$.)

Exercise. B3: Into the polytope

Let $\mathbf{y} = (4, 2, 0)$ and $\mathbf{x} = (2, 2, 2)$. Using Corollary 21.1.4, write $\mathbf{x}$ explicitly as a convex combination of rearrangements of $\mathbf{y}$, in two different ways.

Solution. The six rearrangements of $\mathbf{y}$ are the vectors with entries 4, 2, 0 in some order. Averaging the three cyclic shifts,

$$\frac{1}{3}[(4, 2, 0) + (0, 4, 2) + (2, 0, 4)] = \frac{1}{3}(6, 6, 6) = (2, 2, 2) = \mathbf{x}.$$

Alternatively, average a vector with its reversal:

$$\frac{1}{2}[(4, 2, 0) + (0, 2, 4)] = \frac{1}{2}(4, 4, 4) = (2, 2, 2) = \mathbf{x}.$$

Both are convex combinations, with weights $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{2})$. This is the nonuniqueness of the warning after Example (Four coordinates, two transforms), seen on the polytope: $\mathbf{x}$

is the center of the hexagon, lying strictly inside it rather than on an edge, and such a point has many representations. (The hexagon itself has empty interior in $\mathbb{R}^3$, since it lies in the plane $x_1 + x_2 + x_3 = 6$; “inside” here means inside the hexagon, within that plane.)

21.1.6.3 C. Going deeper

Exercise. C1: Row-stochastic is not enough, twice over

Let $\mathbf{S} \in M_n(\mathbb{R})$ be **row-stochastic**: every entry is nonnegative and every row adds up to 1.

- (a) Give such an $\mathbf{S}$ and a $\mathbf{y} \in \mathbb{R}^2$ with $\mathbf{S}\mathbf{y} \not\prec \mathbf{y}$.
- (b) Prove that if $\mathbf{S}\mathbf{y} \prec \mathbf{y}$ for **every** $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{S}$ is doubly stochastic.

Hint for (b): test the hypothesis on the standard basis vectors.

Solution.

- (a) Take $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{y} = (1, 0)$. Then $\mathbf{S}\mathbf{y} = (1, 1)$, whose total is 2, while $\mathbf{y}$ has total 1. So (M2) fails.
- (b) Apply the hypothesis with $\mathbf{y} = \mathbf{e}_j$. The vector $\mathbf{S}\mathbf{e}_j$ is the j -th column of $\mathbf{S}$, by Definition 0.9.9. From $\mathbf{S}\mathbf{e}_j \prec \mathbf{e}_j$, clause (M2) gives

$$\sum_{i=1}^n s_{ij} = \sum_{i=1}^n (\mathbf{e}_j)_i = 1.$$

So every column of $\mathbf{S}$ adds up to 1. The entries are nonnegative and the rows add up to 1 by hypothesis, so $\mathbf{S}$ is doubly stochastic by Definition 18.8.4.

Exercise. C2: Transitivity, a second time

- (a) Prove that if $\mathbf{A}, \mathbf{B} \in \Omega_n$ then $\mathbf{AB} \in \Omega_n$.
- (b) Hence deduce that $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{z}$ imply $\mathbf{x} \prec \mathbf{z}$, and compare this route with the proof of Proposition 17.8.1 (b).

Solution.

- (a) The entries of $\mathbf{AB}$ are $\sum_k a_{ik}b_{kj}$, sums of products of nonnegative numbers, hence nonnegative. Since $\mathbf{A}\mathbf{1} = \mathbf{1}$ and $\mathbf{B}\mathbf{1} = \mathbf{1}$, we get $(\mathbf{AB})\mathbf{1} = \mathbf{A}(\mathbf{B}\mathbf{1}) = \mathbf{A}\mathbf{1} = \mathbf{1}$, so every row of $\mathbf{AB}$ adds up to 1. Since $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$ and $\mathbf{1}^\top \mathbf{B} = \mathbf{1}^\top$, we get $\mathbf{1}^\top (\mathbf{AB}) = (\mathbf{1}^\top \mathbf{A})\mathbf{B} = \mathbf{1}^\top \mathbf{B} = \mathbf{1}^\top$, so every column does too. By Definition 18.8.4, $\mathbf{AB} \in \Omega_n$.
- (b) By Theorem 21.1.3 there are $\mathbf{D}, \mathbf{E} \in \Omega_n$ with $\mathbf{x} = \mathbf{D}\mathbf{y}$ and $\mathbf{y} = \mathbf{E}\mathbf{z}$. Then $\mathbf{x} = \mathbf{DEz}$, and $\mathbf{DE} \in \Omega_n$ by (a). By Theorem 21.1.3 again, $\mathbf{x} \prec \mathbf{z}$.

Proposition 17.8.1 (b) proved the same thing in one line, by chaining the inequalities $X_k \leq Y_k \leq Z_k$ between running totals. That proof is shorter and uses nothing. The route here is longer, but it explains transitivity rather than verifying it: a majorization is an averaging, and averaging twice is averaging.

Exercise. C3: The count $n - 1$ cannot be improved

Let $n \geq 2$, $\mathbf{y} = (n, 0, \dots, 0)$, and $\mathbf{x} = (1, 1, \dots, 1)$ in $\mathbb{R}^n$. For $\mathbf{z} \in \mathbb{R}^n$, write $\text{supp}(\mathbf{z}) = \{i : z_i \neq 0\}$.

- (a) Show that $\mathbf{x} \prec \mathbf{y}$.
- (b) Let $\mathbf{T}$ be a T-transform. Prove that $|\text{supp}(\mathbf{Tz})| \leq |\text{supp}(\mathbf{z})| + 1$ for every $\mathbf{z} \in \mathbb{R}^n$.
- (c) Deduce that no product of fewer than $n - 1$ T-transforms sends $\mathbf{y}$ to $\mathbf{x}$.

Solution.

- (a) Both vectors are decreasing with total n . Every running total of $\mathbf{y}$ is n , and the k -th running total of $\mathbf{x}$ is $k \leq n$. So (M1) and (M2) hold.
- (b) Let $\mathbf{T}$ act on $\{j, k\}$ with parameter t . By Eq. 21.1.1, $(\mathbf{Tz})_i = z_i$ for $i \notin \{j, k\}$, so

$$\text{supp}(\mathbf{Tz}) \subseteq \text{supp}(\mathbf{z}) \cup \{j, k\}.$$

If both $j, k \in \text{supp}(\mathbf{z})$, the right side is $\text{supp}(\mathbf{z})$ and there is nothing to prove. If neither is, then $z_j = z_k = 0$, so $(\mathbf{Tz})_j = (\mathbf{Tz})_k = 0$ and again $\text{supp}(\mathbf{Tz}) \subseteq \text{supp}(\mathbf{z})$. In the remaining case exactly one of them, say j , lies in $\text{supp}(\mathbf{z})$, and then $\text{supp}(\mathbf{Tz}) \subseteq \text{supp}(\mathbf{z}) \cup \{k\}$, a set with one more element. In all cases the bound holds.

- (c) Suppose $\mathbf{x} = \mathbf{T}_r \cdots \mathbf{T}_1 \mathbf{y}$. Put $\mathbf{z}^{(0)} = \mathbf{y}$ and $\mathbf{z}^{(i)} = \mathbf{T}_i \mathbf{z}^{(i-1)}$. Then $|\text{supp}(\mathbf{z}^{(0)})| = 1$ and $|\text{supp}(\mathbf{z}^{(r)})| = |\text{supp}(\mathbf{x})| = n$. By (b), the size of the support grows by at most 1 at each step, so

$$n \leq 1 + r, \quad \text{that is,} \quad r \geq n - 1.$$

Together with Theorem 21.1.3, which gives such a product with $r \leq n - 1$, exactly $n - 1$ transforms are needed for this pair.