A Linear Algebraic Reconstruction of Majorization

D. Stott Parker
stott@cs.ucla.edu
UCLA Computer Science Department
Los Angeles, CA 90095-1596

Prasad Ram
prasad@cp10.es.xerox.com
Xerox Corporation
El Segundo, CA 90245

September 15, 1997

Abstract

Majorization is an important partial order on multisets. Since its development at the turn of the twentieth century by economists formalizing the notion of 'exchange' among distributions, it has found applications ranging from stochastic scheduling to singular value theory.

In this paper we reconstruct majorization as a partial order on vectors (ordered sequences) using linear algebra. We do this in two different ways, one following a recent extension of majorization to permit partial ordering among the indices of the sequences, and the other following a model of majorization as 'exchangeability'. Majorization with respect to a partial order is shown to be a special case of majorization as exchangeability. In this special case, the resulting majorization order always defines a distributive lattice that is isomorphic to the standard real vector lattice.

In classical majorization theory, the semigroup of doubly-stochastic matrices plays a crucial role, both in the definition of majorization and in the modeling of exchange. We generalize majorization to permit any matrix semigroup of exchanges. Various semigroups of stochastic matrices make particularly interesting kinds of exchanges, and simultaneously define useful notions of majorization.

Finally, we investigate notions of convexity. When the functions in question are differentiable, we show that both Schur convexity and submodularity can be reexpressed naturally in this linear algebraic formulation.

Keywords: majorization, exchange, stochastic matrices, matrix semigroups, lattices, Möbius inversion, Schur convexity, submodularity

Copyright ©1997 D. Stott Parker, Prasad Ram

Contents

1 Introduction **3**

2 Classical Majorization **3**
2.1 Majorization and Exchange . . . 3
2.2 Definition of Classical Majorization . . . 4
2.3 Classical Majorization and Convexity . . . 4
2.4 Equivalent Formulations of Majorization . . . 5
2.5 Majorization relative to a Partial Order . . . 5

3 Background: Partial Orders, Lattices, and Möbius Inversion **6**
3.1 Partial Orders and Lattices . . . 6
3.2 Möbius Inversion . . . 6

4 Majorization relative to a Partial Order **8**
4.1 Möbius Inversion via Linear Algebra . . . 8
4.2 The Vector Lattice . . . 8
4.3 Definition of Majorization relative to a Partial Order . . . 9
4.4 The Majorization Lattice . . . 10

5 Majorization as Exchangeability **11**
5.1 Exchanges as Linear Transformations . . . 11
5.2 Exchangeability as a Matrix Semigroup . . . 11
5.3 Definition of Majorization as Exchangeability . . . 11
5.4 Equivalence of the various definitions for Majorization . . . 12

6 Majorization relative to a Partial Order vs. Exchangeability **13**
6.1 Pairwise Exchanges . . . 13
6.2 Semigroups generated by Pairwise Exchanges . . . 13
6.3 Majorization relative to a Partial Order is Exchangeability . . . 14

7 Majorization and Convexity **15**
7.1 Schur Convexity (Monotonicity with respect to Majorization) . . . 15
7.2 Schur Convexity as Monotonicity with respect to Exchange . . . 15
7.3 Submodularity on Lattice-Ordered Sequences . . . 16
7.4 Submodularity over the Vector Lattice and Majorization Lattice . . . 17

8 Conclusion **17**

1 Introduction

Classical majorization was first defined by Muirhead [15], and was developed by Hardy, Littlewood and Pólya in their study of symmetric means [7]. Their classical theory defines a majorization order on multisets, and although quite beautiful is also quite complex. The treatment in Marshall and Olkin's book [14] presents the theory with a variety of inequivalent formulations, each with its own notational conventions.[1]

A linear algebraic formalization of majorization is important for several reasons. First, replacing multisets with vectors permits many notational simplifications. Second, such a formalization allows us to model the basic 'exchange' notion of majorization with linear operators. When exchanges are defined by doubly-stochastic matrices we get the classical majorization of Hardy-Littlewood-Pólya. Other semigroups of exchange matrices define other notions of majorization. Third, it allows us to understand majorization better, in ways ranging from a deeper appreciation of the majorization ordering to clearer convexity results.

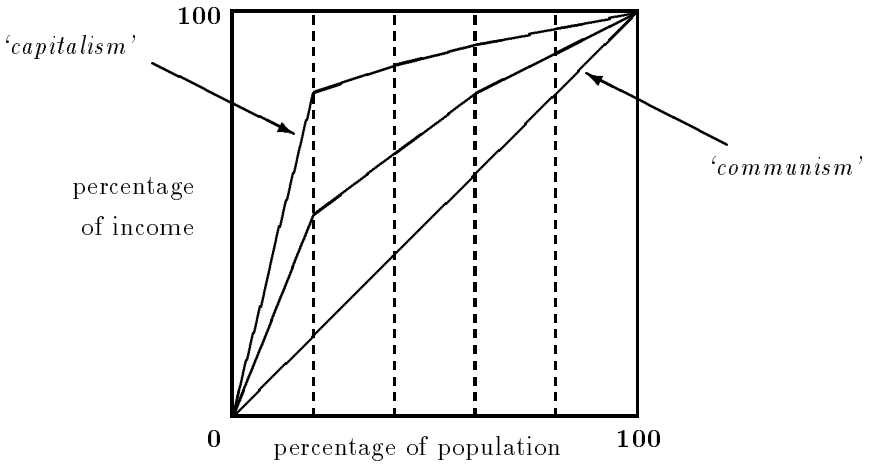
Defining majorization via linear algebra also generalizes neatly on the classical definition, and clarifies the significance of various previous generalizations of majorization listed in [14, ch.14]. All in all, the linear algebraic reconstruction of majorization appears very natural.

2 Classical Majorization

2.1 Majorization and Exchange

Majorization was studied early in the twentieth century by economists, in formalizing unevenness of distribution of income (and greed, in a sense). Lorenz noted in 1905 that one distribution could be said to more uneven than another precisely when it majorized the other. Dalton formalized his 1920 'principle of transfers', showing that a distribution is less uneven than another if it can be obtained from the other by transferring some income from a richer to a poorer income-receiver. A good summary is in chapter 1 of [14].

Of the three income distributions shown by the **Lorenz curves** in Figure 1, the bottommost (theoretical 'communism') is the least uneven: each member of the population receives the same income. The topmost (rampant 'capitalism') is the most uneven of the three: members of the population, sorted by descending income, receive increasingly smaller portions. Dalton's principle says that, by repeatedly transferring income from rich to poor (and only this), we can transform the 'capitalism' distribution to that of 'communism'.[2]



Income density	x_1	x_2	x_3	x_4	x_5
'capitalism'	80%	07%	05%	04%	04%
(in between)	50%	15%	15%	10%	10%
'communism'	20%	20%	20%	20%	20%

Figure 1: Three increasingly uneven income distributions for a population of five individuals

[1] Marshall and Olkin's book [14], the primary reference on the majorization literature, uses row vectors and matrix multiplication on the right, while we now use the more standard column vectors and matrix multiplication on the left.

[2] Former inhabitants of communist countries tell us that the curve for 'communism' in Figure 1 is highly idealized.

2.2 Definition of Classical Majorization

Classical majorization is the natural ordering on probability densities. In the Lorenz sense, one distribution is more uneven than another precisely when its probability density majorizes that of the other, i.e., when the integral of its density dominates the integral of the other density. As a result, classical majorization has found many applications in stochastic scheduling [14], optimal coding [17], etc. The book by Marshall and Olkin [14] provides a good survey of the classical theory and its applications.

Classical majorization is often defined as a binary relation '$\prec$' on multisets (or 'unordered sequences') [14]. In this paper, since we wish to both restate and extend the classical theory using linear algebra, we will model multisets with nonnegative n-vectors (or 'ordered sequences') ${\Re_+}^n$.

Definition 2.1 (Classical Majorization)
For $\mathbf{x}, \mathbf{y} \in {\Re_+}^n$,

$$\mathbf{x} \prec \mathbf{y} \quad \textbf{iff} \quad \begin{cases} \sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} & k = 1, \ldots, (n-1) \\ \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \end{cases}$$

where $[z_{[1]} \ldots z_{[n]}] = \mathrm{sort}{\downarrow}(\mathbf{z})$ *denotes the descendingly-sorted (nonincreasing) ordering of* $\mathbf{z} = [z_1 \ldots z_n]$.

Notice this definition is like that in Figure 1: one (sorted) density is majorized by another exactly when its graph lies below the other's:

$$[.20\ .20\ .20\ .20\ .20] \prec [.50\ .15\ .15\ .10\ .10] \prec [.80\ .07\ .05\ .04\ .04].$$

A consequence of this definition is that sequence ordering is immaterial. For example:

$$\begin{array}{ccccc} [.40\ .60] & \not\prec & [.50\ .50] & \prec & [.60\ .40] \\ [.60\ .40] & \not\prec & [.50\ .50] & \prec & [.40\ .60] \\ [.40\ .60] & \prec & [.60\ .40] & \prec & [.40\ .60]. \end{array}$$

As a result, classical majorization is only a **preorder** on sequences (it fails to be a partial order by lacking antisymmetry, since any sequence is majorized by any of its permutations).

However, classical majorization does define a partial order on *multisets* — multisets can be viewed as sequences that are equal modulo permutation. Using multisets instead of sequences can make some results more elegant, but on the other hand some results seem best expressed in terms of ordered sequences. Given the linear algebraic approach of this paper, classical majorization will be treated as a preorder on vectors.

2.3 Classical Majorization and Convexity

In 1923 Schur published the memorandum [22] that first connected majorization with convexity. Since that time there have been been a multitude of results in this vein, which make majorization an important tool in inequality theory [14].

Definition 2.2 *A* **Schur convex function** *is a function that is monotone with respect to* $\prec$.

It follows that Schur convex functions are **symmetric** (invariant under permutation of their argument sequence). If the standard scalar product is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle \quad = \quad \mathbf{x}^\top \mathbf{y} \quad = \quad x_1 y_1 + \cdots + x_n y_n.$$

then a standard example of a Schur convex function is $F(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$ where $\mathbf{w}$ is a descendingly-sorted sequence in ${\Re_+}^n$, assuming also that $\mathbf{x}$ is drawn from the descendingly-sorted sequences in ${\Re_+}^n$. More generally, the function

$$F(\mathbf{z}) \quad = \quad \langle \mathrm{sort}{\downarrow}(\mathbf{w}), \mathrm{sort}{\downarrow}(\mathbf{z}) \rangle$$

is symmetric and Schur convex for $\mathbf{x} \in {\Re_+}^n$, given any sequence of nonnegative weights $\mathbf{w} = [w_1 \ldots w_n]$.

Theorem 2.1 *(Schur [22] and Ostrowski [16])*
For $\mathbf{z} = [z_1 \ldots z_n] \in {\Re_+}^n$, *a symmetric function* F *is Schur convex if and only if*

$$(z_i - z_j) \left(\frac{\partial F(\mathbf{z})}{\partial z_i} - \frac{\partial F(\mathbf{z})}{\partial z_j} \right) \quad \geq \quad 0 \qquad\qquad (1 \leq i, j \leq n).$$

Another connection between convexity and majorization with many facets is that, given an interval $I \subseteq \Re_+$ and continuous $\phi : I \to \Re_+$, the function $F(\mathbf{z}) = \sum_i \phi(z_i)$ is Schur convex if and only if ϕ is convex on I ([22], [14, p.68], [20, p.333]). This brings us to the following summary.

2.4 Equivalent Formulations of Majorization

One of the most provocative aspects of majorization is that it permits a number of equivalent formulations, and that these formulations are surprisingly diverse. The basic idea behind this paper is that an interesting formulation comes from generalizing the first theorem below, of Schur:

Theorem 2.2 *(Schur [22], Hardy-Littlewood-Pólya [7, p.45–49])*
For $\mathbf{x}, \mathbf{y} \in {\Re_+}^n$,

$$\mathbf{x} \prec \mathbf{y} \quad \textbf{iff} \quad \mathbf{x} = D\,\mathbf{y}$$

for some doubly stochastic matrix D (a nonnegative matrix each of whose rows and columns sum to 1).

Theorem 2.3 *(Hardy-Littlewood-Pólya [7, p.89]; also [14, p.11], [20, pp.320–321])*
If $\phi : I \to \Re_+$ *is any continuous convex function defined on an interval* $I \subseteq \Re_+$, *then for all* $\mathbf{x}, \mathbf{y} \in I^n$,

$$\mathbf{x} \prec \mathbf{y} \quad \textbf{iff} \quad \sum_{i=1}^{n} \phi(x_i) \quad \leq \quad \sum_{i=1}^{n} \phi(y_i).$$

Theorem 2.4 *(Pečarić-Proschan-Tong [20, pp.323–324])*
If $\phi : I \to \Re_+$ *is any continuous convex function defined on an interval* $I \subseteq \Re_+$, *then for all decreasing sequences* $\mathbf{x}, \mathbf{y} \in I^n$ *and sequences* $\mathbf{w}$ *such that* $\sum_{i=1}^{k} w_i x_i \leq \sum_{i=1}^{k} w_i y_i$ $(1 \leq k \leq (n-1))$ *and* $\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i$,

$$\mathbf{x} \prec \mathbf{y} \quad \textbf{iff} \quad \sum_{i=1}^{n} w_i\,\phi(x_i) \quad \leq \quad \sum_{i=1}^{n} w_i\,\phi(y_i).$$

In addition, we can formulate majorization in terms of a vector cone ordering [14, pp.424–426]. Define what we call the **zeta matrix** $\int = (\zeta_{ij})$ with $\zeta_{ij} =$ (if $i \geq j$ then 1 else 0). For example, when $n = 5$,

$$\int = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Theorem 2.5 *If* $\mathcal{S} \subseteq {\Re_+}^n$ *is a set of vectors* $\mathbf{x}$ *such that* $\langle \mathbf{1}, \mathbf{x} \rangle = \sum_i x_i = c$ *(for some constant* $c > 0$*), then for* $\mathbf{x}, \mathbf{y} \in \mathcal{S}$,

$$\mathbf{x} \prec \mathbf{y} \quad \textbf{iff} \quad \textstyle\int \mathrm{sort}{\downarrow}(\mathbf{x}) \;\leq_{\mathrm{vec}}\; \int \mathrm{sort}{\downarrow}(\mathbf{y}).$$

2.5 Majorization relative to a Partial Order

The classical notion of majorization was generalized by Hwang [9, 10, 11] and Lih [12] to a relation on what can be called 'partially ordered sequences'. The definition of $\mathbf{x} \prec \mathbf{y}$ in classical majorization (Definition 2.1) requires

$$\sum_{i \leq k} x_i \quad \leq \quad \sum_{i \leq k} y_i,$$

where k ranges over all index values in the finite total order $1 \leq k \leq n$ — $\mathbf{x}$ and $\mathbf{y}$ being descendingly-sorted — can be generalized so that k ranges over all index values in a partial order '$\sqsubseteq$':

$$\sum_{i \sqsubseteq k} x_i \quad \leq \quad \sum_{i \sqsubseteq k} y_i.$$

This might seem an artificial generalization, but it has an important application discussed in Section 7.3. We will develop an alternative generalization based on 'exchanges' that subsumes this one, and which we feel is more natural.

3 Background: Partial Orders, Lattices, and Möbius Inversion

Later we show the generalization of majorization just sketched, relative to partial orders, can be defined by the inverse of the Möbius function of the partial order, which is called the zeta function in combinatorial theory [21]. In this section we reproduce necessary background concepts, and in subsequent sections reconstruct them in terms of linear algebra.

3.1 Partial Orders and Lattices

We need elementary definitions for partial orders and lattices; for a comprehensive tutorial see [5].

A **partial order** is a pair $\langle \mathcal{S}, \sqsubseteq \rangle$ where $\mathcal{S}$ is a set and $\sqsubseteq$ is a binary relation on $\mathcal{S}$ satisfying

1. **reflexivity**: For all x in $\mathcal{S}$, $x \sqsubseteq x$.
2. **transitivity**: For all x, y, z in $\mathcal{S}$, if $x \sqsubseteq y$ and $y \sqsubseteq z$, then $x \sqsubseteq z$.
3. **antisymmetry**: For all x, y in $\mathcal{S}$, if $x \sqsubseteq y$ and $y \sqsubseteq x$, then $x = y$.

A **preorder** is a pair $\langle \mathcal{S}, \sqsubseteq \rangle$ satisfying the first two conditions.

If it exists, the **least element** of a partial order is written $\bot$, and $\bot \sqsubseteq x$ for all x in $\mathcal{S}$. If it exists, the **greatest element** is written $\top$, and $x \sqsubseteq \top$ for all x in $\mathcal{S}$.

A **total order** is a partial order $\langle S, \sqsubseteq \rangle$ in which for every pair of elements x, y either $x \sqsubseteq y$ or $y \sqsubseteq x$ holds. A partial order and total order are illustrated by the graphs in Figure 2. The greatest element of both partial orders is denoted by '1', and $4 \sqsubseteq 2$ and $2 \sqsubseteq 1$ in both orders.

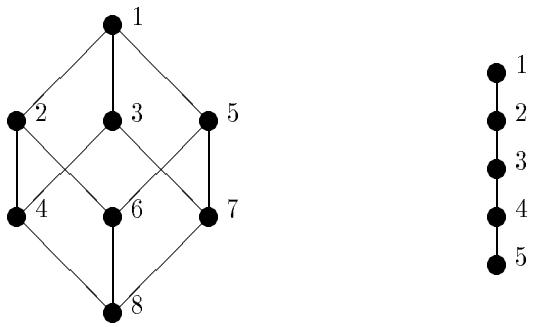


Figure 2: A partial order and a total order, each being a lattice and having a greatest and least element.

A **lattice** is an algebra $\langle S, \sqsubseteq, \sqcap, \sqcup \rangle$ in which $\langle S, \sqsubseteq \rangle$ is a partial order, and for all $a, b \in S$, there is a unique **greatest lower bound (glb)** $a \sqcap b$ and **least upper bound (lub)** $a \sqcup b$.

A lattice is called **distributive** if these operators satisfy the distributive law

$$\text{for all } a, b, c \text{ in } \mathcal{S}, \quad a \sqcap (b \sqcup c) \quad = \quad (a \sqcap b) \sqcup (a \sqcap c).$$

3.2 Möbius Inversion

The theory of Möbius inversion gives a generalized notion of differentiation and integration on discrete partially-ordered domains. Möbius developed the concept in 1832 in analyzing the distribution of primes. The theory gives an interesting and broad foundation on which to build a theory of majorization. We present here a classical formulation adapted from Rota [21].

Definition 3.1 *The* **zeta function** *of a partial order '$\sqsubseteq$' is defined by*

$$\zeta(x,y) \quad = \quad \begin{cases} 1 & \text{if } x \sqsubseteq y \\ 0 & \text{otherwise.} \end{cases}$$

The **Möbius function** *of a partial order is the multiplicative inverse of the zeta function:*

$$\mu(x,y) \quad = \quad \frac{-1}{\zeta(x,x)} \sum_{x \sqsubset z \sqsubseteq y} \zeta(x,z)\ \mu(z,y) \quad = \quad - \sum_{x \sqsubset z \sqsubseteq y} \mu(z,y).$$

The Möbius function is well-defined when the partial order $\langle S, \sqsubseteq \rangle$ is **locally finite**, *i.e., if for all $x, y \in \mathcal{S}$, the set $\{z \mid x \sqsubset z \sqsubseteq y\}$ is finite.*

Notice the zeta function is just the (transitive closure of the) incidence matrix of the graph of the partial order. Rota [21, pp.346–347] explains how the zeta function is in fact related to the Riemann zeta function of number theory.

We can think of ζ as the kernel of an integral operator (which transforms a function on the partial order to its bottom-to-top "integral"), with μ as the kernel of its inverse differential operator. Inversion formulas give a way to recover a function $g(x)$ from a (discrete approximation of an) integral $f(y) = \int g(x)\ \zeta(x,y)\, dx$:

Theorem 3.1 (Möbius inversion formula [21]):
Let $\langle \mathcal{S}, \sqsubseteq \rangle$ be a locally finite partially ordered set with least element $\perp$, and $g : \mathcal{S} \to \Re$ satisfy $g(\perp) = 0$. Define

$$f(y) \quad = \quad \sum_{x \sqsubseteq y} g(x) \quad = \quad \sum_{x} g(x)\ \ \zeta(x,y).$$

Then $f(\perp) = 0$ also, and

$$g(x) \quad = \quad \sum_{y \sqsubseteq x} f(y)\ \ \mu(y,x).$$

There are many important instances of this inversion formula [1, 21], including:

Finite differencing —— $\mathcal{S}$ is the set $\{0, \ldots, n-1\}$ with the '$\leq$' total order:

$$\begin{aligned} f(j) \quad &= \quad \sum_{i=0}^{j} g(i) \\ \mu(j,i) \quad &= \quad \begin{cases} 1 & \text{if } j = i \\ -1 & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases} \\ g(i) \quad &= \quad \sum_{j \leq i} f(j)\ \mu(j,i) \quad = \quad f(i)\ -\ f(i-1). \end{aligned}$$

Möbius inversion formula —— $\mathcal{S}$ is the set $\{0, \ldots, n-1\}$ with the partial order of divisibility:

$$\begin{aligned} f(n) \quad &= \quad \sum_{d \mid n} g(d) \\ \mu(d,n) \quad &= \quad \begin{cases} 1 & \text{if } n = d \\ (-1)^r & \text{if } n = d\, p_1\, p_2\, \cdots\, p_r \\ & (p_1, p_2, \ldots, p_r \text{ distinct primes}) \\ 0 & \text{otherwise} \end{cases} \\ g(n) \quad &= \quad \sum_{d \mid n} f(d)\ \mu(d,n). \end{aligned}$$

Principle of Inclusion and Exclusion —— $\mathcal{S}$ is the subsets of a finite set, ordered by inclusion:

$$\begin{aligned} f(t) \quad &= \quad \sum_{s \supseteq t} g(s) \\ \mu(t,s) \quad &= \quad (-1)^{\|s\| - \|t\|} \\ g(s) \quad &= \quad \sum_{t \supseteq s} f(t)\ (-1)^{\|s\| - \|t\|}. \end{aligned}$$

4 Majorization relative to a Partial Order

The notation and definitions in this section are ours, so the results vary from those in classical majorization. The linear algebraic perspective is successful in clarifying the essence of many results regarding majorization.

4.1 Möbius Inversion via Linear Algebra

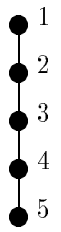
The zeta and Möbius functions of a partial order with n elements can be defined naturally as $n \times n$ matrices. Although such a formulation goes far to eliminate the magic surrounding the Möbius inversion theory, we have not seen it formulated directly with linear algebra before.

Definition 4.1 *For any partial order $\sqsubseteq$, the* **zeta matrix** $\int = (\zeta_{ij}) = (\zeta(i,j))$ *is the transitively closed incidence matrix of the partial order.*

The **Möbius matrix** $\partial = (\mu_{ij}) = (\mu(i,j))$ *is the inverse $\int^{-1}$ of the zeta matrix.*

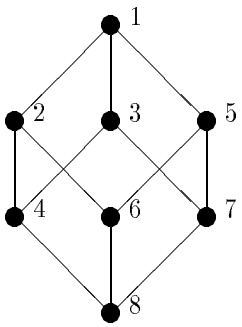
Since it can be determined from the context, we simply leave the partial order implicit, rather than complicate the notation by mentioning it explicitly.

The zeta and Möbius matrices for two orders are shown in Figures 3 and 4.



$$\int = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \qquad \partial = \int^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \end{array}$$

Figure 3: A total order, its zeta function (transitively closed incidence matrix), and its Möbius matrix



$$\int = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \qquad \partial = \int^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix} \end{array}$$

Figure 4: A partial order, its zeta function (transitively closed incidence matrix), and its Möbius matrix

This notation permits many results to be expressed more concisely. For example, Theorem 3.1 (the Möbius inversion theorem) can essentially be stated as

$$\mathbf{f} = \int \mathbf{g} \qquad \Leftrightarrow \qquad \mathbf{g} = \partial \mathbf{f}.$$

4.2 The Vector Lattice

Definition 4.2 *By a* **sequence x** *we mean a vector with nonnegative entries*[3] *in a set $\mathcal{S} \subseteq {\Re_+}^n$,*

The standard componentwise vector partial order is defined by

$$\mathbf{x} \leq_{\text{vec}} \mathbf{y} \quad \textbf{iff} \quad x_i \leq y_i \quad (1 \leq i \leq n).$$

[3] Often the results here can be generalized for arbitrary real sequences, without no insistence that the entries be nonnegative. For simplicity, though, we consider only the nonnegative sequences ${\Re_+}^n$.

Also define vector element-wise minima and maxima

$$\begin{array}{lllll} \mathbf{x} & \mathbf{min}_{\mathrm{vec}} & \mathbf{y} & = & [\ \min(x_1,y_1) \quad \ldots \quad \min(x_n,y_n)\ ] \\ \mathbf{x} & \mathbf{max}_{\mathrm{vec}} & \mathbf{y} & = & [\ \max(x_1,y_1) \quad \ldots \quad \max(x_n,y_n)\ ]. \end{array}$$

Theorem 4.1 *The nonnegative vectors* $\langle \Re_+{}^n, \leq_{\mathrm{vec}}, \mathbf{min}_{\mathrm{vec}}, \mathbf{max}_{\mathrm{vec}} \rangle$ *form a distributive lattice we call the* **vector lattice** *(sometimes also called a* **Riesz space***).*

Proof The one-dimensional algebra $\langle \Re_+, \leq, \min, \max \rangle$ is a distributive lattice. All vector properties we require follow from this. □

An important example is the vector lattice of **distribution sequences** of length n. The set $\mathcal{P}$ of ascending nonnegative vectors $\mathbf{v}$ of length n with $v_n = 1$, modeling discrete probability distributions, forms a vector lattice $\langle \mathcal{P}, \leq_{\mathrm{vec}}, \mathbf{min}_{\mathrm{vec}}, \mathbf{max}_{\mathrm{vec}} \rangle$ possessing least element $\bot = [0\,0 \cdots 0\,1]$ and greatest element $\top = [1\,1 \cdots 1\,1]$.

4.3 Definition of Majorization relative to a Partial Order

The zeta matrix $\int$ for any partial order '$\sqsubseteq$' can be used to obtain a generalized definition '$\preceq$' of majorization:

Definition 4.3 (Majorization relative to a Partial Order)
If $\mathcal{S} \subseteq \Re_+{}^n$ *is a set of vectors* $\mathbf{x}$*, then for* $\mathbf{x}, \mathbf{y} \in \mathcal{S}$*,*

$$\mathbf{x} \preceq \mathbf{y} \quad \textbf{iff} \quad \int \mathbf{x} \leq_{\mathrm{vec}} \int \mathbf{y}.$$

Vector majorization *is the important special case of majorization relative to the total order* $1 \sqsupseteq \cdots \sqsupseteq n$ *that was shown in Figure 3. Henceforth* '$\preceq$' *will denote vector majorization unless otherwise specified.*

This defines majorization relative to *any* finite partial order $\int$ on vector indices (i.e., sequence indices). It does *not* require $\mathbf{x}$ or $\mathbf{y}$ to be sorted, unlike some definitions.

Vector majorization is generally not equivalent to classical majorization. For example, when $\mathcal{S}$ is the set of nonnegative vectors $\mathbf{x}$ such that $\sum_{i=1}^{n} x_i = 1$, vector majorization and classical majorization differ:

$$\begin{array}{ccccccccc} [.60\ .40] & \prec & [.40\ .60] & \not\prec & [.50\ .50] & \prec & [.60\ .40] & \prec & [.40\ .60] \\ [.60\ .40] & \not\preceq & [.40\ .60] & \preceq & [.50\ .50] & \preceq & [.60\ .40] & \not\preceq & [.40\ .60] \\ [.04\ .04\ .05\ .07\ .80] & \not\prec & [.10\ .10\ .15\ .15\ .50] & \not\prec & [.20\ .20\ .20\ .20\ .20] & \prec & [.50\ .15\ .15\ .10\ .10] & \prec & [.80\ .07\ .05\ .04\ .04] \\ [.04\ .04\ .05\ .07\ .80] & \preceq & [.10\ .10\ .15\ .15\ .50] & \preceq & [.20\ .20\ .20\ .20\ .20] & \preceq & [.50\ .15\ .15\ .10\ .10] & \preceq & [.80\ .07\ .05\ .04\ .04]. \end{array}$$

By restricting the set $\mathcal{S}$, however, vector majorization and classical majorization can be made equivalent. Specifically, when $\mathcal{S}$ is the set of *descendingly-sorted* nonnegative sequences $\mathbf{x} = \mathrm{sort}{\downarrow}(\mathbf{x})$ of length n such that $\langle \mathbf{1}, \mathbf{x} \rangle = c$ for some constant c then as in Theorem 2.5, vector majorization is equivalent to classical majorization. When the sequences $\mathbf{x}$ and $\mathbf{y}$ are in descendingly-sorted order, vector majorization is called **weak majorization** ($\mathbf{x} \prec_w \mathbf{y}$) in the majorization literature [14]. When the sequences in $\mathcal{S}$ are not sorted, then sequence ordering matters, and the relation '$\preceq$' is *not symmetric* in the sequence elements.

Definition 4.3 (for '$\preceq$') has several advantages over Definition 2.1 (for '$\prec$'):

1. It defines a bona-fide partial order, not just a preorder (as with classical majorization).
2. It actually induces a majorization lattice that is isomorphic to a natural vector lattice — see Section 4.4.
3. It defines majorization only over a set $\mathcal{S}$, significantly reducing notational complexity. With classical majorization, the page is cluttered with repeated statements that

$$\sum_{i=1}^{n} x_i \quad = \quad \sum_{i=1}^{n} y_i$$

but these can be replaced with the understanding that we require $\mathbf{x}$ and $\mathbf{y}$ to belong to a predefined set $\mathcal{S}$, which contains only vectors $\mathbf{z}$ such that $\langle \mathbf{1}, \mathbf{z} \rangle = c$ for a given constant c. Also, restricting $\mathcal{S}$

to consist only of descendingly-sorted sequences avoids the need for notation such as $x_{[i]}$ for the i-th largest element of $\mathbf{x}$.

4. It is a linear algebraic definition, permitting more concise statements, such as

$$\int \mathbf{x} \quad \leq_{\text{vec}} \quad \int \mathbf{y}$$

instead of

$$\sum_{i=1}^{k} x_i \quad \leq \quad \sum_{i=1}^{k} y_i \qquad (k = 1, \ldots, (n-1)).$$

Interestingly, sorted sequences can also be characterized naturally in terms of the zeta matrix of a total order. In particular, for the total order $1 \sqsubseteq \cdots \sqsubseteq n$, the space $\mathcal{S}$ of descendingly-sorted nonnegative sequences of length n that is used in classical majorization theory is just $\mathcal{S} = \int {\Re_+}^n$:

Theorem 4.2 *When $\int$ is the zeta matrix for the total order $1 \sqsubseteq \cdots \sqsubseteq n$, then for all nonnegative $\mathbf{x} \in {\Re_+}^n$, $\int \mathbf{x}$ is sorted in descending order. If we reverse the order and let $\int$ be the zeta matrix for $1 \sqsupseteq \cdots \sqsupseteq n$, then $\int \mathbf{x}$ is sorted in ascending order.*

Proof For the order $1 \sqsubseteq \cdots \sqsubseteq n$ ($\zeta_{ij} = 1$ iff $i \leq j$), $\int \mathbf{x} = [(x_1 + x_2 + \cdots + x_n)\ (x_2 + \cdots + x_n)\ \ldots\ x_n]$. Because $(x_1 + x_2 + \cdots + x_n) \geq (x_2 + \cdots + x_n) \geq \cdots \geq x_n$, $\int \mathbf{x} = \text{sort}{\downarrow}\left(\int \mathbf{x}\right)$ whenever $\mathbf{x} \in {\Re_+}^n$. (With this total order, vector majorization is called **weak supermajorization** ($\mathbf{y} \prec^w \mathbf{x}$) in the literature [14, p.10].)

Similarly, if the order is reversed (so $1 \sqsupseteq \cdots \sqsupseteq n$ and $\zeta_{ij} = 1$ iff $i \geq j$, as in Figure 3), then relative to this order we find the reverse characterization $\int \mathbf{x} = [x_1\ (x_1 + x_2)\ \ldots\ (x_1 + x_2 + \cdots + x_n)]$. Because $x_1 \leq (x_1 + x_2) \leq \ldots \leq (x_1 + x_2 + \cdots + x_n)$, $\int \mathbf{x} = \text{sort}{\uparrow}\left(\int \mathbf{x}\right)$, where $\text{sort}{\uparrow}(\mathbf{z})$ denotes the ascendingly-sorted (nondecreasing) ordering of $\mathbf{z} = [z_1 \ldots z_n]$. □

4.4 The Majorization Lattice

All partial orders $\sqsubseteq$ are 'invertible' in the sense that the zeta matrix $\int$ is invertible: the zeta matrix is always triangular, and its diagonal entries are never zero. Therefore, the Möbius matrix $\partial = \int^{-1}$ is defined. As the Möbius and zeta transformations are inverses of one another,

$$\mathbf{x} \leq_{\text{vec}} \mathbf{y} \quad \textbf{iff} \quad (\partial\,\mathbf{x}) \preceq (\partial\,\mathbf{y}).$$

This isomorphism between $\leq_{\text{vec}}$ and $\preceq$ implies the following result:

Theorem 4.3 *A convex subset $\mathcal{S}$ of ${\Re_+}^n$ ordered by majorization is a distributive lattice $\langle \mathcal{S}, \preceq, \sqcap, \sqcup \rangle$, which we call the* **majorization lattice***, when the majorization lub and glb operators are defined by*

$$\begin{aligned} \mathbf{x} \sqcup \mathbf{y} &= \partial((\textstyle\int \mathbf{x}) \ \mathbf{max}_{\text{vec}} \ (\int \mathbf{y})) \\ \mathbf{x} \sqcap \mathbf{y} &= \partial((\textstyle\int \mathbf{x}) \ \mathbf{min}_{\text{vec}} \ (\int \mathbf{y})). \end{aligned}$$

Proof Since the transformation $\mathbf{x} \mapsto \int \mathbf{x}$ defines a lattice isomorphism between the vector and majorization lattices, $\mathbf{x} \sqcap \mathbf{y}$ and $\mathbf{x} \sqcup \mathbf{y}$ are defined just so as to be the majorization glb and lub:

$$\begin{array}{l} \mathbf{z} \preceq \mathbf{x},\ \ \mathbf{z} \preceq \mathbf{y} \\ \Leftrightarrow \quad (\int \mathbf{z}) \ \leq_{\text{vec}} \ (\int \mathbf{x}),\ \ (\int \mathbf{z}) \ \leq_{\text{vec}} \ (\int \mathbf{y}) \\ \Leftrightarrow \quad (\int \mathbf{z}) \ \leq_{\text{vec}} \ ((\int \mathbf{x}) \ \mathbf{min}_{\text{vec}} \ (\int \mathbf{y})) \\ \Leftrightarrow \quad \mathbf{z} \preceq \partial((\int \mathbf{x}) \ \mathbf{min}_{\text{vec}} \ (\int \mathbf{y})) \\ \Leftrightarrow \quad \mathbf{z} \preceq \mathbf{x} \sqcap \mathbf{y}. \end{array} \qquad \begin{array}{l} \mathbf{x} \preceq \mathbf{z},\ \ \mathbf{y} \preceq \mathbf{z} \\ \Leftrightarrow \quad (\int \mathbf{x}) \ \leq_{\text{vec}} \ (\int \mathbf{z}),\ \ (\int \mathbf{y}) \ \leq_{\text{vec}} \ (\int \mathbf{z}) \\ \Leftrightarrow \quad ((\int \mathbf{x}) \ \mathbf{max}_{\text{vec}} \ (\int \mathbf{y})) \ \leq_{\text{vec}} \ (\int \mathbf{z}) \\ \Leftrightarrow \quad \partial((\int \mathbf{x}) \ \mathbf{max}_{\text{vec}} \ (\int \mathbf{y})) \preceq \mathbf{z} \\ \Leftrightarrow \quad \mathbf{x} \sqcup \mathbf{y} \preceq \mathbf{z}. \end{array}$$

Thus the majorization lub and glb operators determine a distributive lattice. □

An important example is the majorization lattice of **density sequences** of length n. The set $\mathcal{D}$ of density sequences of length n (nonnegative $\mathbf{x}$ with $\sum_{i=1}^{n} x_i = 1$), modeling discrete probability densities, is a majorization lattice $\langle \mathcal{D}, \preceq, \sqcap, \sqcup \rangle$ that possesses least element $\bot = [0\,0 \cdots 0\,1]$ and greatest element $\top = [1\,0 \cdots 0\,0]$.

5 Majorization as Exchangeability

A key benefit of using a linear algebraic formulation is that modeling densities as vectors, with distributions as their left-to-right 'integral', also permits modeling exchanges as linear transformations. Thus, if $\mathbf{x} = T\mathbf{y}$, the matrix T is precisely an **exchange**, or **transfer** — an operator transforming the density $\mathbf{y}$ to $\mathbf{x}$.

5.1 Exchanges as Linear Transformations

Definition 5.1 *A square matrix is said to be* **(column) stochastic** *if its elements are all non-negative and all columns sum to 1, and* **sub-stochastic** *if they sum to at most 1. We define the following kinds of matrix:*

- **lower triangular stochastic matrix (LTSM)**
  a column stochastic matrix whose nonzero entries are all in the lower triangle.
- **upper triangular stochastic matrix (UTSM)**
  a column stochastic matrix whose nonzero entries are all in the upper triangle.
- **doubly stochastic matrix (DSM)**
  a row and column stochastic matrix — a matrix whose rows and columns each sum to 1.
- **weak doubly stochastic matrix (wDSM)**
  a doubly sub-stochastic matrix — a matrix whose rows and columns each sum to at most 1.
- **permutation matrix (PM)**
  a $\{0,1\}$-matrix each of whose rows and columns contains one 1.

The following matrices illustrate exchanges on a density sequence with an LTSM and a DSM:

$$\begin{pmatrix} .4 & 0 & 0 & 0 \\ .2 & .3 & 0 & 0 \\ 0 & .2 & .9 & 0 \\ .4 & .5 & .1 & 1 \end{pmatrix} \begin{pmatrix} .40 \\ .30 \\ .20 \\ .10 \end{pmatrix} = \begin{pmatrix} .16 \\ .17 \\ .24 \\ .43 \end{pmatrix} \qquad \begin{pmatrix} .4 & .1 & .2 & .3 \\ .1 & .2 & .3 & .4 \\ .3 & .4 & .1 & .2 \\ .2 & .3 & .4 & .1 \end{pmatrix} \begin{pmatrix} .40 \\ .30 \\ .20 \\ .10 \end{pmatrix} = \begin{pmatrix} .26 \\ .20 \\ .28 \\ .26 \end{pmatrix}.$$

5.2 Exchangeability as a Matrix Semigroup

Theorem 5.1 *Each of the sets UTSM, LTSM, DSM, wDSM, PM forms a semigroup — a set of matrices closed under multiplication that includes the identity.*

Proof The permutations actually form a group. Each of the other sets contains the identity matrix, and is closed under multiplication because every matrix can be expressed as a convex sum of elementary matrices, and products of these convex sums are convex sums. For example, the Birkhoff-von Neumann theorem [2, 14] states that, if P is the set of permutation matrices, DSM can be generated as

$$DSM \quad = \quad convex(P) \quad = \quad \{\ \lambda_1 P_1 + \cdots + \lambda_m P_m \ \mid\ P = \{P_1, \ldots, P_m\},\ \ 0 \le \lambda_i \le 1,\ \ \sum_{i=1}^{m} \lambda_i = 1\ \ \}.$$

The same statement also can be made for LTSMs and UTSMs if P is replaced by the lower-triangular or upper-triangular $\{0,1\}$-matrices with exactly one 1 in each column. For wDSM, one instead replaces $\sum_{i=1}^{m} \lambda_i = 1$ with $\sum_{i=1}^{m} \lambda_i \le 1$. □

5.3 Definition of Majorization as Exchangeability

Definition 5.2 **($\mathcal{T}$-Majorization, Exchangeability)**
Let $\mathcal{T}$ be a semigroup of $n \times n$ matrices, so that $\mathcal{T}$ contains the identity and is closed under multiplication. For $\mathbf{x}$, $\mathbf{y}$ *in* $\mathcal{S} \subseteq {\Re_+}^n$, *the preorder* $\mathbf{x} \preceq_{\mathcal{T}} \mathbf{y}$ *of* **$\mathcal{T}$-majorization** *is defined by*

$$\mathbf{x} \quad \preceq_{\mathcal{T}} \quad \mathbf{y} \qquad \textbf{iff} \qquad \text{there is a matrix } T \in \mathcal{T} \text{ such that } \mathbf{x} = T\mathbf{y}.$$

This is **exchangeability** *— the existence of an operator that takes* $\mathbf{y}$ *and gives* $\mathbf{x}$.

That this is a preorder follows from $\mathcal{T}$ being a semigroup. The relation is reflexive because the identity is in $\mathcal{T}$, and closure under multiplication makes it transitive. Notice that '$\preceq_{\mathcal{T}}$' is a partial order iff the semigroup lacks any multiplicative inverses (except for the identity itself). For example, '$\preceq_{LTSM}$' is a partial order.

This extremely general definition permits the following useful instances:

UTSM-majorization:	$\mathbf{x} \preceq_{UTSM} \mathbf{y}$	**iff**	$\mathbf{x} = U\mathbf{y}$ for some UTSM U.
LTSM-majorization:	$\mathbf{x} \preceq_{LTSM} \mathbf{y}$	**iff**	$\mathbf{x} = L\mathbf{y}$ for some LTSM L.
PM-majorization:	$\mathbf{x} \preceq_{PM} \mathbf{y}$	**iff**	$\mathbf{x} = P\mathbf{y}$ for some PM P.
DSM-majorization:	$\mathbf{x} \preceq_{DSM} \mathbf{y}$	**iff**	$\mathbf{x} = D\mathbf{y}$ for some DSM D.
wDSM-majorization:	$\mathbf{x} \preceq_{wDSM} \mathbf{y}$	**iff**	$\mathbf{x} = W\mathbf{y}$ for some wDSM W.

An equivalent definition is $\mathbf{x} \preceq_{wDSM} \mathbf{y}$ **iff** $\mathbf{x} \leq_{\text{vec}} D\mathbf{y}$ for some DSM D. This is called **weak majorization** ($\mathbf{x} \prec_w \mathbf{y}$) [14, ch.2.C] when $\mathbf{x}$ and $\mathbf{y}$ are in descendingly-sorted order.

There are *many* other semigroups. The $n \times n$ nonnegative matrices $M_n(\Re_+)$, any linear group (subgroup of $GL_n(\Re)$), or the convex sums $convex(\mathcal{T})$ of a semigroup $\mathcal{T}$ are natural examples. Still other semigroups can be constructed as a subset of a semigroup that happens to be closed under multiplication: see Section 6.3.

5.4 Equivalence of the various definitions for Majorization

We conclude this section by showing that LTSM-majorization is equivalent to vector majorization on nonnegative descendingly-sorted sequences with a fixed sum. This is also equivalent to classical majorization, as we have already noted (p.9).

Earlier Theorem 2.2 showed that doubly stochastic matrices (DSMs) define a notion of exchange that is equivalent to classical majorization. The relation of majorization as LTSM-exchangeability '$\preceq_{LTSM}$' satisfies a direct analogue of Theorem 2.2, but using LTSMs instead of DSMs:

Theorem 5.2 *Let $\mathcal{S} \subseteq {\Re_+}^n$ be a set of nonnegative sequences all having sum $c > 0$.*
For $\mathbf{x}, \mathbf{y} \in \mathcal{S}$, a necessary and a sufficient condition for the vector majorization $\mathbf{x} \preceq \mathbf{y}$ is that $\mathbf{x} \preceq_{LTSM} \mathbf{y}$.

Proof For all $\mathbf{x}, \mathbf{y} \in {\Re_+}^n$, assuming $\int$ is the zeta matrix for the vector majorization total order $1 \sqsupseteq \cdots \sqsupseteq n$,

$$\mathbf{x} \preceq_{LTSM} \mathbf{y} \quad \Rightarrow \quad \mathbf{x} = L\mathbf{y} \quad \Rightarrow \quad \int \mathbf{x} = \int L\mathbf{y} \quad \Rightarrow \quad \int \mathbf{x} \leq_{\text{vec}} \int \mathbf{y} \quad \Rightarrow \quad \mathbf{x} \preceq \mathbf{y}.$$

For the converse, we assume $\mathbf{x} \preceq \mathbf{y}$ and derive an LTSM L that 'reduces' $\mathbf{y}$ to $\mathbf{x}$ (i.e., so that $\mathbf{x} = L\,\mathbf{y}$) by repeatedly replacing y_p with x_p where they differ, reducing the distance between $\mathbf{x}$ and $\mathbf{y}$.

We do this by induction on the 'distance' $d(\mathbf{x}, \mathbf{y})$ — the number of indices in which the two vectors differ. The basis where the distance is 0 is trivial.

For the induction step, let p and q be the two lowest indices such that $x_p \neq y_p$ and $x_q \neq y_q$. (There must be at least two such indices: two sequences with equal sums cannot differ in only one index.) Thus $p < q$ and $x_p < y_p$ because $\mathbf{x} \preceq \mathbf{y}$. Define the LTSM differing from the identity only in the $\{p, q\} \times \{p, q\}$ submatrix

$$L_{pq}(\epsilon) \quad = \quad \begin{array}{c} \\ \vdots \\ p \\ \vdots \\ q \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} \cdots & & p & \cdots & q & & \cdots \end{array} \\ \begin{pmatrix} \ddots & & & & & & \\ & 1-\epsilon & & & & & \\ & & & \ddots & & & \\ & \epsilon & & & 1 & & \\ & & & & & \ddots & \end{pmatrix} \end{array}$$

where $(1-\epsilon)\, y_p = x_p$, so that $\epsilon = (1 - x_p/y_p) \in (0, 1]$. Then

$$L_{pq}(\epsilon)\ \mathbf{y} \quad = \quad [\, y_1 \ \cdots \ y_{p-1} \qquad x_p \qquad y_{p+1} \ \cdots \ y_{q-1} \qquad (y_q - y_p + x_p) \qquad y_{q+1} \ \cdots \ y_n \,].$$

Set $T_1 = L_{pq}(\epsilon)$. Since $(T_1\mathbf{y})_p = x_p$ and $x_p < y_p$ it follows that $\mathbf{x} \preceq T_1\mathbf{y} \preceq \mathbf{y}$ and $d(\mathbf{x}, T_1\mathbf{y}) < d(\mathbf{x}, \mathbf{y})$. Multiplication by T_1 therefore preserves majorization and reduces the distance by at least 1. Notice the distance is at most n, and it cannot be 1. By induction then, $\mathbf{x}$ can be derived from $T_1\mathbf{y}$ with at most $(n-2)$ LTSMs. Thus, $\mathbf{x}$ can be derived from $\mathbf{y}$ in at most $(n-1)$ LTSMs, whose product is again an LTSM. □

A similar proof shows that $\mathbf{x} \preceq \mathbf{y}$ iff $\mathbf{y} \preceq_{UTSM} \mathbf{x}$. For example, since $[.40\ .60] \preceq [.50\ .50] \preceq [.60\ .40]$,

$$\begin{pmatrix} .40 \\ .60 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & 0 \\ \frac{1}{5} & 1 \end{pmatrix} \begin{pmatrix} .50 \\ .50 \end{pmatrix} \qquad \begin{pmatrix} 1 & \frac{1}{6} \\ 0 & \frac{5}{6} \end{pmatrix} \begin{pmatrix} .40 \\ .60 \end{pmatrix} = \begin{pmatrix} .50 \\ .50 \end{pmatrix}$$

$$\begin{pmatrix} .50 \\ .50 \end{pmatrix} = \begin{pmatrix} \frac{5}{6} & 0 \\ \frac{1}{6} & 1 \end{pmatrix} \begin{pmatrix} .60 \\ .40 \end{pmatrix} \qquad \begin{pmatrix} 1 & \frac{1}{6} \\ 0 & \frac{5}{6} \end{pmatrix} \begin{pmatrix} .50 \\ .50 \end{pmatrix} = \begin{pmatrix} .60 \\ .40 \end{pmatrix}.$$

Corollary 5.1 *For any convex subset $\mathcal{S}$ of ${\Re_+}^n$, both $\langle \mathcal{S}, \preceq_{LTSM}, \sqcap, \sqcup \rangle$ and $\langle \mathcal{S}, \preceq_{UTSM}, \sqcup, \sqcap \rangle$ are distributive lattices, where '$\sqcap$' and '$\sqcup$' are defined as in Theorem 4.3.*

6 Majorization relative to a Partial Order vs. Exchangeability

Building on the previous section, we show the UTSM semigroup can be generated by pairwise exchanges — exchanges among a pair of sequence indices. Partial ordering on sequence indices can then be added by restricting generators to include only those pairwise exchanges whose index pair appears in the partial order. Thus majorization as exchangeability generalizes majorization relative to a partial order.

6.1 Pairwise Exchanges

Definition 6.1 *The* **pairwise exchange UTSMs** $U_{pq}(\epsilon)$ *and* **pairwise exchange DSMs** $D_{pq}(\epsilon)$*, defined for $1 \le p < q \le n$, $\epsilon \in [0,1]$, are matrices differing from the identity only in the $\{p,q\} \times \{p,q\}$ submatrix:*

$$U_{pq}(\epsilon) = \begin{array}{c} \\ \vdots \\ p \\ \vdots \\ q \\ \vdots \end{array} \begin{array}{c} \cdots \quad p \quad \cdots \quad q \quad \cdots \\ \begin{pmatrix} \ddots & & & & & & \\ & 1 & & & \epsilon & & \\ & & \ddots & & & & \\ & & & 1-\epsilon & & & \\ & & & & \ddots & & \end{pmatrix} \end{array}, \qquad D_{pq}(\epsilon) = \begin{array}{c} \\ \vdots \\ p \\ \vdots \\ q \\ \vdots \end{array} \begin{array}{c} \cdots \quad p \quad \cdots \quad q \quad \cdots \\ \begin{pmatrix} \ddots & & & & & & \\ & 1-\epsilon & & & \epsilon & & \\ & & \ddots & & & & \\ & \epsilon & & & 1-\epsilon & & \\ & & & & & \ddots & \end{pmatrix} \end{array}.$$

The **pairwise exchange LTSMs** $L_{pq}(\epsilon)$ *were introduced in the previous section; they could be defined by $L_{pq}(\epsilon) = R\,U_{pq}(\epsilon)\,R$, where R is the* **reflection permutation** *that maps i to $(n-i+1)$, for $1 \le i \le n$.*

The **pairwise permutation** P_{pq} *exchanges the p-th and q-th elements: $P_{pq} = D_{pq}(1)$.*

6.2 Semigroups generated by Pairwise Exchanges

Theorem 6.1 *If $\mathcal{T}$ is any of UTSM, LTSM, DSM, wDSM, or PM, then any $\mathcal{T}$-exchange among two vectors can be decomposed into a sequence of $(n-1)$ pairwise $\mathcal{T}$-exchanges.*

More precisely, let $\mathcal{S} \subseteq {\Re_+}^n$ be a set of vectors $\mathbf{x}$ such that $\langle \mathbf{1}, \mathbf{x} \rangle = c$ (for some constant c). For $\mathbf{x}, \mathbf{y} \in \mathcal{S}$, if $\mathbf{x} = T\mathbf{y}$ for some $T \in \mathcal{T}$ then there exists a $T' \in \mathcal{T}$ such that $\mathbf{x} = T'\mathbf{y}$, and T' is a product of at most $(n-1)$ pairwise exchanges in $\mathcal{T}$.

Proof The proof for PMs is extremely well-known. The result for DSMs and classical majorization is Theorem 2.2; see [14, ch.2]. The proof for wDSMs is essentially identical. For LTSMs, this is proven by the construction in the proof for Theorem 5.2, and the proof for UTSMs is similar. □

For example, for the proportion (density) sequences in Figure 1 we can transform *'capitalism'* to *'communism'* with the product of $5-1=4$ pairwise exchanges, derived precisely as in the proof of Theorem 5.2:

$$\begin{aligned} L &= L_{45}(1-\tfrac{.20}{.36})\ L_{34}(1-\tfrac{.20}{.52})\ L_{23}(1-\tfrac{.20}{.67})\ L_{12}(1-\tfrac{.20}{.80}) \\ &= \tfrac{1}{31356} \begin{pmatrix} 7839 & 0 & 0 & 0 & 0 \\ 7020 & 9360 & 0 & 0 & 0 \\ 6345 & 8460 & 12060 & 0 & 0 \\ 5640 & 7520 & 10720 & 17420 & 0 \\ 4512 & 6016 & 8576 & 13936 & 31356 \end{pmatrix} \simeq \begin{pmatrix} .25 & 0 & 0 & 0 & 0 \\ .22 & .30 & 0 & 0 & 0 \\ .20 & .27 & .39 & 0 & 0 \\ .18 & .24 & .34 & .56 & 0 \\ .15 & .19 & .27 & .44 & 1 \end{pmatrix}. \end{aligned}$$

Theorem 6.2 *Every UTSM U is equal to a product of $n(n-1)/2$ pairwise exchange UTSMs. Specifically*

$$U \quad = \quad \prod_{k=n-1}^{1} \prod_{i=k+1}^{n} U_{ik}(\epsilon_{ik})$$

where $\epsilon_{ik} = u_{ki} / \left(\prod_{j<k} (1-\epsilon_{ij}) \right)$ if $u_{ki} > 0$, and 0 otherwise. Furthermore, since $L = RUR$ is an LTSM, and $L_{pq}(\epsilon) = R\,U_{pq}(\epsilon)\,R$, reversing this statement with 'R' gives an almost identical result for LTSMs.

Proof Notice first that $U_{pq}(\epsilon)^\top$ is a row-stochastic, lower triangular matrix that looks very much like the Gauss transforms used in LU-decomposition, except that it has a non-unit diagonal element.

Set $L = U^\top$. A process very like LU-decomposition (Gaussian elimination) will yield a factorization $L = M_1 M_2 \cdots M_{n-1}$, where

$$M_k \quad = \quad U_{n,k}(\epsilon_{n,k})^\top \quad \cdots \quad U_{k+1,k}(\epsilon_{k+1,k})^\top$$

is a lower-triangular matrix having nonzero entries below the diagonal only in the k-th column. Specifically, we construct the matrices M_k by choosing values of ϵ_{ik} to have the Gaussian elimination property that

$$M_1 \quad \cdots \quad M_k$$

is a lower triangular matrix with zeroes below the diagonal after the first k columns, and whose first k columns are identical to those of L. If $L = (\ell_{ij})$, then it turns out that $\epsilon_{ik} = \ell_{ik} / \left(\prod_{j<k} (1-\epsilon_{ij}) \right)$ when $\ell_{ik} > 0$, and 0 otherwise. Notice that if $\epsilon_{ij} = 1$ for any j, then necessarily $\ell_{ik} = 0$ for $k > j$, so ϵ_{ik} is well-defined. Transposing the final result gives the stated product for $U = L^\top$. □

Definition 6.2 *If $\mathcal{G}$ is a set of $n \times n$ matrices, the* **semigroup of matrices generated by $\mathcal{G}$**

$$\mathit{Semigroup}(\ \mathcal{G}\ )$$

is the set of all finite products of matrices from $\mathcal{G}$.

Theorem 6.3 *LTSMs and UTSMs are semigroups generated by pairwise exchanges:*

$$\begin{aligned} \mathit{LTSM} \quad &= \quad \mathit{Semigroup}(\ \{\ L_{pq}(\epsilon) \mid \epsilon \in [0,1],\ \ 1 \le p < q \le n\ \}\ ) \\ \mathit{UTSM} \quad &= \quad \mathit{Semigroup}(\ \{\ U_{pq}(\epsilon) \mid \epsilon \in [0,1],\ \ 1 \le p < q \le n\ \}\ ). \end{aligned}$$

Proof The '$\subseteq$' direction is immediate from Theorem 6.2. The '$\supseteq$' direction is trivial. □

Perhaps surprisingly, no such theorem exists for arbitrary DSMs. For example, the DSM

$$\begin{pmatrix} .5 & .5 & 0 \\ .5 & 0 & .5 \\ 0 & .5 & .5 \end{pmatrix}$$

cannot be generated by any product of pairwise exchange DSMs [14, p.23]. This lack of a decomposition theorem and lack of a simple lattice structure are drawbacks of the classical majorization definition.

6.3 Majorization relative to a Partial Order is Exchangeability

Finally, we can give the more general result that the definition of majorization relative to a partial order (Definition 4.3) is equivalent to a special case of $\mathcal{T}$-majorization (Definition 5.2).

Theorem 6.4 *Let '$\sqsubseteq$' be a partial order on the indices $\{1, \ldots, n\}$, with corresponding zeta matrix $\int$. Define*

$$\mathit{LTSM}_{\sqsubseteq} \quad = \quad \mathit{Semigroup}(\ \{\ L_{pq}(\epsilon) \mid \epsilon \in [0,1],\ \ p \sqsupset q\ \}\ ).$$

Then majorization relative to the partial order '$\sqsubseteq$' is equivalent to $\mathit{LTSM}_{\sqsubseteq}$-majorization: for $\mathbf{x}, \mathbf{y} \in \mathcal{S}$,

$$\int \mathbf{x} \ \le_{\mathrm{vec}} \ \int \mathbf{y} \qquad \textbf{iff} \qquad \mathbf{x} = T\mathbf{y} \quad \textit{for some} \quad T \in \mathit{LTSM}_{\sqsubseteq}.$$

Proof The 'only if' direction is like the proof of Theorem 5.2 for the total order $1 \sqsupseteq \cdots \sqsupseteq n$, but we allow as generators only those pairwise exchange LTSMs $L_{pq}(\epsilon)$ where $p \sqsupset q$. The 'if' direction is like Theorem 6.3. □

7 Majorization and Convexity

To conclude, we characterize the impact of our generalized majorization definition on two conventional notions of convexity: Schur convexity and submodularity. There are actually many connections between majorization and convexity. See for example [14] (in which submodular functions are called *$\mathcal{L}$-subadditive functions*), and [20]. In this section we illustrate how a linear algebraic approach again makes quite complex results straightforward.

7.1 Schur Convexity (Monotonicity with respect to Majorization)

Definition 7.1 *A function F is* **Schur convex** *(or:* **majorization-order preserving**, **monotone with respect to majorization***) if:*

$$\mathbf{x} \text{ is majorized by } \mathbf{y} \quad \text{implies} \quad F(\mathbf{x}) \leq F(\mathbf{y}).$$

We use this definition for all majorization relations (classical majorization '$\prec$', majorization relative to a partial order '$\preceq$', and $\mathcal{T}$-majorization '$\preceq_{\mathcal{T}}$'), if the majorization relation is understood from context.

For example, a function that is Schur convex with respect to UTSM-majorization is $F(\mathbf{z}) = \langle \mathbf{w}, \mathbf{z} \rangle$, for any descendingly-sorted $\mathbf{w} \in {\Re_+}^n$ and any $\mathbf{z} \in \mathcal{S} \subseteq {\Re_+}^n$. This function also happens to be Schur convex with respect to vector majorization.

Theorem 7.1 *With vector majorization:*
F is Schur convex on the majorization lattice iff $f(\mathbf{x}) = F(\boldsymbol{\partial}\,\mathbf{x})$ is monotone on the vector lattice.
Also, a differentiable function $F : {\Re_+}^n \to \Re_+$ is majorization-order preserving on $\mathcal{S} \subseteq {\Re_+}^n$ iff

$$\frac{\partial F}{\partial z_k}(\mathbf{z}) \quad \geq \quad \frac{\partial F}{\partial z_{k+1}}(\mathbf{z})$$

for all $\mathbf{z} = [z_1 \ \ldots \ z_n] \in \mathcal{S}$, and $1 \leq k < n$.

Proof A direct result of the bijection between the two lattices. If f is differentiable, f is monotone on the vector lattice iff $\nabla f(\mathbf{x}) = [\,\partial f/\partial x_1 \cdots \partial f/\partial x_n\,] \geq_{\text{vec}} \mathbf{0}$, which implies

$$\frac{\partial F}{\partial x_i}(\boldsymbol{\partial}\,\mathbf{x}) \ - \ \frac{\partial F}{\partial x_{i+1}}(\boldsymbol{\partial}\,\mathbf{x}) \quad = \quad \frac{\partial}{\partial x_i}(F(\boldsymbol{\partial}\,\mathbf{x})) \quad = \quad \frac{\partial}{\partial x_i} f(\mathbf{x}) \quad \geq \quad 0 \qquad (1 \leq i \leq n-1).$$

Thus $\partial F/\partial z_i \geq \partial F/\partial z_{i+1}$ when F is monotone with respect to vector majorization. □

7.2 Schur Convexity as Monotonicity with respect to Exchange

Theorem 7.2 *Let $\mathcal{T} = \text{Semigroup}(\,\mathcal{G}\,)$ be the semigroup generated by $\mathcal{G}$. If, for every family of generators $T(\epsilon) \in \mathcal{G}$, $T(0)$ is the identity matrix, and the* **Lie derivative with respect to** *$T(\epsilon)$ is defined and satisfies the inequality*

$$\frac{d}{d\epsilon}\, F(\ T(\epsilon)\ \mathbf{z}\ )\ \Big|_{\epsilon=0} \quad \leq \quad 0 \qquad\qquad \text{for all } \mathbf{z} \in \mathcal{S}$$

then F is Schur convex with respect to $\mathcal{T}$-majorization.

Proof Since $\mathbf{x} \preceq_{\mathcal{T}} \mathbf{y}$ iff $\mathbf{x} = T\mathbf{y}$ for some $T \in \mathcal{T}$, multiplication by T gives a reduction in the $\mathcal{T}$-majorization ordering, and $F(\,T\,\mathbf{z}\,) - F(\,\mathbf{z}\,) \leq 0$ when F is monotone (Schur convex) with respect to $\mathcal{T}$-majorization. □

For example, the $\preceq_{LTSM}$ majorization-preservation condition

$$\frac{\partial F}{\partial z_k}(\mathbf{z}) \quad \geq \quad \frac{\partial F}{\partial z_{k+1}}(\mathbf{z}) \qquad\qquad (1 \leq k \leq n)$$

in Theorem 7.1 is equivalent (for $\mathbf{z} \in {\Re_+}^n$) to the constraint on the Lie derivative with respect to $L_{ij}(\epsilon)$

$$\frac{d}{d\epsilon}\, F(\ L_{ij}(\epsilon)\ \mathbf{z}\ )\ \Big|_{\epsilon=0} \quad \leq \quad 0 \qquad\qquad (1 \leq i < j \leq n).$$

Similarly the Schur convexity condition

$$(z_i - z_j)\ \left(\frac{\partial F(\mathbf{z})}{\partial z_i}\ -\ \frac{\partial F(\mathbf{z})}{\partial z_j}\right)\quad \geq\quad 0 \qquad\qquad (1 \leq i, j \leq n)$$

mentioned in Theorem 2.1 is equivalent to the constraint on the Lie derivative with respect to $D_{ij}(\epsilon)$

$$\frac{d}{d\epsilon}\ F(\ D_{ij}(\epsilon)\ \mathbf{z}\ )\ \Big|_{\epsilon=0}\quad \leq\quad 0 \qquad\qquad (1 \leq i < j \leq n).$$

These easily-verified examples show that pairwise exchanges are sufficient to determine whether a function is monotone with respect to $\mathcal{T}$-majorization.

7.3 Submodularity on Lattice-Ordered Sequences

Submodularity is an interesting property of functions on lattices that combines aspects of both convexity and concavity. The survey [13] is particularly readable.

Definition 7.2 *A function $\mu : L \to \Re$ is* **submodular** *on a lattice $\langle L, \sqsubseteq, \sqcap, \sqcup\rangle$ if for all x, $y \in L$,*

$$\mu(x \sqcup y)\quad +\quad \mu(x \sqcap y)\quad \leq\quad \mu(x)\quad +\quad \mu(y)$$

where '$\sqcap$' and '$\sqcup$' are the lattice glb and lub operators.

Hwang [9] characterized majorization relative to a partial order in terms of finite sequences of 'flows' which are similar to UTSMs, using a result of Shapley concerning the existence of cores of convex (supermodular) characteristic function games. Subsequently, Lih [12] established the conditions for Schur convexity for Hwang's generalized form of majorization. Using the elegant presentation of Seymour and Welsh [23], Lih also reconciled Hwang's majorization with the Holley [8] and FKG [6] inequalities, two basic inequalities that underly results involving submodularity. All of these results rely in a central way on LTSMs.

Theorem 7.3 *(Holley [8, 23])*
Let L be a finite distributive lattice, and μ_1, $\mu_2 : L \to [0, 1]$ be probability densities such that

$$\mu_1(i)\quad \mu_2(j)\quad \leq\quad \mu_1(i \sqcup j)\quad \mu_2(i \sqcap j) \qquad \textit{for all } i, j \in L.$$

Then μ_1 majorizes μ_2 in the sense of Hwang. Furthermore, for any monotone[4] function f

$$\sum_{i \in L}\ f(i)\ \mu_1(i)\quad \geq\quad \sum_{i \in L}\ f(i)\ \mu_2(i).$$

Proof Holley showed that, on a distributive lattice, if

$$\mu_1(i)\ \ \mu_2(j)\quad \leq\quad \mu_1(i \sqcup j)\ \ \mu_2(i \sqcap j)\quad \text{for all } i, j \in L.$$

then there is a function $\nu : L \times L \to [0, 1]$ such that

$$\sum_j\ \nu(i,j)\quad =\quad \mu_1(i), \qquad \sum_i\ \nu(i,j)\quad =\quad \mu_2(j), \qquad \nu(i,j)\quad =\quad 0\quad \text{unless}\quad j \sqsubseteq i.$$

In other words, ν is an LTSM. Thus

$$\sum_{i \in L}\ f(i)\ \mu_1(i)\quad =\quad \sum_{i,\, j \in L}\ f(i)\ \nu(i,j)\quad =\quad \sum_{j\ \sqsubseteq\ i}\ f(i)\ \nu(i,j)\quad \geq\quad \sum_{j\ \sqsubseteq\ i}\ f(j)\ \nu(i,j)\quad =\quad \sum_{j \in L}\ f(j)\ \mu_2(j).$$

For each $i \in L$, $f(j) = \zeta(i,j)$ is a monotone function on L, so $\int \mu_1 \geq_{\text{vec}} \int \mu_2$, and μ_1 majorizes μ_2. □

[4] Seymour and Welsh, and also Lih, use the term 'increasing' for monotone.

7.4 Submodularity over the Vector Lattice and Majorization Lattice

We can also neatly characterize submodular functions on the vector and majorization lattice in terms of their derivatives, in a way reminiscent of Schur convex functions. Although Schur convexity and submodularity are consistent properties of a function, in the sense that functions can be both Schur convex and submodular, the following theorems show that they are inequivalent: Schur convexity imposes conditions on the first derivatives, while submodularity imposes conditions on second derivatives.

Here, we consider only vector majorization. Similar results are obtained for other kinds of majorization.

Theorem 7.4 *(Lorentz [14, p.150])*
When twice differentiable, f is submodular on the vector lattice $\langle {\Re_+}^n, \leq_{\text{vec}}, \mathbf{min}_{\text{vec}}, \mathbf{max}_{\text{vec}} \rangle$ iff

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \quad \leq \quad 0 \qquad\qquad (i \neq j,\ 1 \leq i, j \leq n).$$

Proof Essentially definition. To denote the expression $f([\, x_1 \ \ldots \ x_{i-1} \ \ u \ \ x_{i+1} \ \ldots \ x_{j-1} \ \ v \ \ x_{j+1} \ \ldots \ x_n \,])$ we use the shorthand $f[\, u \ v \,]$, giving the result

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \ = \ \lim_{\epsilon_i, \epsilon_j \to 0} \frac{f[\,(x_i{+}\epsilon_i)\ (x_j{+}\epsilon_j)\,] - f[\,(x_i{+}\epsilon_i)\ x_j\,] - f[\,x_i\ (x_j{+}\epsilon_j)\,] + f[\,x_i\ x_j\,]}{\epsilon_i \ \ \epsilon_j} \ \leq \ 0$$

where the inequality comes from the fact that f is submodular, because the points $\mathbf{x} = [\,(x_i{+}\epsilon_i)\ x_j\,]$ and $\mathbf{y} = [\,x_i\ (x_j{+}\epsilon_j)\,]$ have upper and lower bounds $\mathbf{x}\,\mathbf{max}_{\text{vec}}\,\mathbf{y} = [\,(x_i{+}\epsilon_i)\ (x_j{+}\epsilon_j)\,]$ and $\mathbf{x}\,\mathbf{min}_{\text{vec}}\,\mathbf{y} = [\,x_i\ x_j\,]$ with respect to $\leq_{\text{vec}}$. For the converse, if f is not submodular on a rectangle defined by $\mathbf{x} = [\,(x_i{+}a)\ x_j\,]$ and $\mathbf{y} = [\,x_i\ (x_j{+}b)\,]$ Lorentz pointed out that one can find a subrectangle on which $\partial^2 f / \partial x_i \partial x_j > 0$. □

Note: the derivatives $\frac{\partial^2 f}{\partial x_i^2}$ can still be positive. In fact, the Hessian $\nabla^2 f = (\frac{\partial^2 f}{\partial x_i \partial x_j})$ still can even be positive semidefinite (hence f can be convex), or be an **M-matrix** [14, p.508].

Theorem 7.5 *With vector majorization, a twice differentiable function F is submodular on the majorization lattice $\langle {\Re_+}^n, \preceq, \sqcap, \sqcup \rangle$ iff for all $i \neq j$ between 1 and $n-1$,*

$$\frac{\partial^2 F(\mathbf{z})}{\partial z_i \partial z_j} \quad - \quad \frac{\partial^2 F(\mathbf{z})}{\partial z_{i+1} \partial z_j} \quad - \quad \frac{\partial^2 F(\mathbf{z})}{\partial z_i \partial z_{j+1}} \quad + \quad \frac{\partial^2 F(\mathbf{z})}{\partial z_{i+1} \partial z_{j+1}} \quad \leq \quad 0.$$

Proof Theorem 4.3 shows that the Möbius transformation $\boldsymbol{\partial}$ gives a bijection between the majorization lattice $\langle {\Re_+}^n, \preceq, \sqcap, \sqcup \rangle$ and the vector lattice $\langle {\Re_+}^n, \leq_{\text{vec}}, \mathbf{min}_{\text{vec}}, \mathbf{max}_{\text{vec}} \rangle$. Thus $f(\mathbf{x}) = F(\boldsymbol{\partial}\,\mathbf{x})$ is submodular on the vector lattice when F is submodular on the majorization lattice. Expanding the inequality

$$\frac{\partial^2}{\partial x_i \partial x_j}(F(\boldsymbol{\partial}\,\mathbf{x})) \quad = \quad \frac{\partial^2}{\partial x_i \partial x_j}(f(\mathbf{x})) \quad \leq \quad 0$$

(which follows from the previous theorem) with the chain rule gives the stated result, because with vector majorization $\mathbf{z} = \boldsymbol{\partial}\,\mathbf{x} = [\,x_1\ (x_2{-}x_1)\ (x_3{-}x_2)\ \cdots\ (x_n{-}x_{n-1})\,]$. □

8 Conclusion

There are many perspectives on majorization, and sometimes it can be difficult to relate them. Classical majorization theory has not yet found a simple formulation, and research on understanding it better is still very lively. For example, very recently different authors have sought to derive precise characterizations of classical majorization in terms of restricted doubly stochastic matrices (e.g., [3] and [4]).

This paper is an attempt to improve our understanding of majorization theory by starting from scratch, and reconstructing it from linear algebra. Some benefits of the reconstruction include notational simplicity, decomposability of exchange matrices into pairwise exchanges, appreciation for a distributive lattice structure underlying majorization, and simple differential calculus statements for basic convexity results. We have used the theory developed here in obtaining results for Huffman Coding [19], and generally in formalizing Greedy Algorithms as convex optimization algorithms over a majorization lattice [18].

Our definition of $\mathcal{T}$-majorization generalizes on previous majorization definitions. It models 'exchanges' as linear transformations drawn from a given matrix semigroup, and equates majorization with exchangeability. The semigroup LTSM of lower triangular stochastic matrices is particularly useful: LTSM-majorization is equivalent to vector majorization, and also equivalent to classical majorization when the domain is restricted to descendingly-sorted sequences. At the same time, LTSM-majorization is a partial order on vectors (not a preorder) and the LTSM theory is simpler than the DSM theory, for example permitting pairwise decomposition.

We also pointed out that $\mathcal{T}$-majorization has intuitive properties for other semigroups $\mathcal{T}$ of stochastic matrices, some of which have already been studied extensively. It would be interesting to characterize other semigroups that constitute useful sets of exchanges, such as rotations or reflections. Our results for Schur convexity suggest also that it might be interesting to relate the theory of Lie groups and Lie algebras to the theory of exchangeability.

References

[1] C. Berge. *Principles of Combinatorics.* Academic Press, New York, 1971.

[2] G. Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tacuman Rev. Ser.*, 5:147–151, 1946.

[3] R.A. Brualdi, S.G. Hwang, and S.S. Pyo. Vector majorization via positive definite matrices. *Linear Algebra and its Applications*, 257:105–120, 1997.

[4] H. Cheng C.S. Wong. Vector majorization via nonnegative definite doubly stochastic matrices of maximum rank. *Linear Algebra and its Applications*, 261:187–194, 1997.

[5] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order.* Cambridge University Press, Cambridge, 1990.

[6] C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[7] G.H. Hardy, J.E. Littlewood, and G. Polya. *Inequalities.* Cambridge University Press, 1978.

[8] R. Holley. Remarks on the FKG inequalities. *Comm. Math. Phys.*, 36:227–237, 1974.

[9] F.K. Hwang. Majorization On A Partially Ordered Set. *Proc. American Math. Soc.*, 76:227–237, 1979.

[10] F.K. Hwang. Generalized Schur Functions. *Bulletin of the Institute of Mathematics – Academia Sinica*, 8(4):513–516, December 1980.

[11] F.K. Hwang and U.G. Rothblum. Majorization and Schur Convexity with Respect to Partial Orders. *Mathematics of Operations Research*, 18(4):928–944, November 1993.

[12] K. Lih. Majorization On Finite Partially Ordered Sets. *SIAM J. Alg. Disc. Meth.*, 3(4):495–503, December 1982.

[13] L. Lovász. Submodular functions and convexity. In B. Korte A. Bachem, M. Groetschel, editor, *Mathematical Programming: The State of the Art*, pages 235–257. Springer-Verlag, New York, 1983.

[14] A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press Inc., 1979.

[15] R.F. Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proc. Edinburgh Math. Soc.*, 21:144–157, 1903.

[16] A.M. Ostrowski. Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. *J. Math. Pures Appl.*, 31(9):253–292, 1952.

[17] D.S. Parker. Conditions for Optimality of the Huffman Algorithm. *SIAM J. Comput.*, 9(3):470–489, August 1980.

[18] D.S. Parker and Prasad Ram. Greed and Majorization. Technical report, UCLA Computer Science Dept., Los Angeles, CA 90095-1596, July 1996. Revised September 1997.

[19] D.S. Parker and Prasad Ram. The construction of Huffman codes is a submodular ('convex') optimization problem over a lattice of binary trees. Technical report, UCLA Computer Science Dept., Los Angeles, CA 90095-1596, October 1996. Revised September 1997; to appear, *SIAM J. Comput.*

[20] J.E. Pečarić, F. Proschan, and Y.L. Tong. *Convex Functions, Partial Orderings, and Statistical Applications.* Academic Press, San Diego, 1992.

[21] G.-C. Rota. On the Foundations of Combinatory Theory I. Theory of Möbius Functions. *Z. Wahrscheinlichkeitstheorie*, 2:340–368, 1964.

[22] I. Schur. Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. *Sitzungsber. Berl. Math. Ges.*, 22:9–20, 1923.

[23] P.D. Seymour and D.J.A. Welsh. Combinatorial applications of an inequality from statistical mechanics. *Math. Proc. Camb. Phil. Soc.*, 77:485–495, 1975.