Greed and Majorization (ABSTRACT)

D.Stott Parker
Prasad Ram
Computer Science Department
University of California
Los Angeles, CA 90024-1596

November 29, 1994

Abstract

Modern analyses of greedy-solvable problems (in terms of matroids, greedoids, submodularity, etc.) have grown in sophistication to the point that uninitiated readers can come away with a sense of confusion and disbelief. Despite the sophistication of these analyses and the great importance of greedy algorithms, currently there appears to be no truly satisfactory resolution of *what* a greedy algorithm is, or *when* and *why* greedy algorithms work. The pervasiveness of greedy algorithms is due to their simplicity, so it is disturbing that an arcane theory would be required to explain them.

Our contribution here is a new theory, giving a simple linear algebraic framework of greed. First, we introduce a generalized **majorization** ordering on numeric sequences, and identify conditions for functions on sequences to preserve this ordering. Generalized majorization is not a total order, but a preorder. It is naturally viewed as an 'exchange' ordering where the exchanges are specific linear transformations. Second, we show that greedy-solvable problems can be formalized as optimization problems in which the objective is monotone with respect to this exchange ordering. We outline greedy algorithms for such problems, including those that exploit additional properties of the objective, such as convexity. Examples from the literature illustrate how diverse well-known applications of greed can be expressed.

1 Introduction

The greedy method is a well-known approach for solving optimization problems. The study of greedy methods is significant for three main reasons: greedy methods are space and time efficient, they define an important class of 'easy' optimization problems, and they usually serve as good heuristics in cases where they do not always yield the optimal solution. In the case of several known 'hard' problems, greedy algorithms result in solutions that are close to optimal. Space-time tradeoffs with optimality favor greedy algorithms as good heuristics. Characterizing the class of easy problems could aid in the discovery of more efficient algorithms for several problems that have not yet been proven to be hard.

Many studies of greed [7, 5, 13, 9, 8] have focused on greedy algorithms for specific classes of optimization problems. Although early studies emphasized the similarities between greediness and matroids or greedoids, these models are not general enough to model many known greedy-solvable problems. For example, matroids deal explicitly only with problems which seek to maximize a linear objective function, such as in the *Continuous Knapsack* problem

$$\text{knapsack value} \quad = \quad \sum_{i=1}^{n} w_i \; x_i$$

where w_i is the weight (value) of the i-th commodity, and $x_i \in [0,1]$ is the fraction of the knapsack volume devoted to it. Matroids do not support objective functions that reflect the ordering of choices made over the execution of the greedy algorithm. This limitation is perhaps the main reason that greedoids [6] were developed. For example,

$$\text{average file access time} \quad = \quad \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{i} \ell_j \right)$$

is the objective in the problem *Optimal Storage on Files on Tape*, where ℓ_j is the length of the j-th file on the tape. This problem can be represented easily with greedoids, but not matroids. However, neither matroids nor greedoids can cope with situations where the objective function uses different weights than the weights of the objects. Dechter & Dechter stress in [2] that the ordering used by the greedy algorithm in its selection need not necessarily correspond to the weights used in the optimal solution. For example, in *Single-Processor Average Flow Time*, the goal is to order jobs so as to minimize

$$\text{average flow time} \quad = \quad \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{i} t_j \right)$$

where w_i is the weight of the i-th job selected, and t_j is the processing time of the j-th job selected. When jobs are greedily selected in order of t/w, so $[t_1/w_1, \ldots, t_n/w_n]$ is nondecreasing, the resulting schedule is optimal. Thus the matroid and greedoid models are incomplete formal models of greed.

We show that a class of optimization problems, corresponding to maximization of a function that is monotone with respect to the majorization order on sequences, is greedy-solvable. The majorization order on sequences is a generalization of the majorization of Hardy-Littlewood-Pólya used in the study of inequalities. This class of optimization problems subsumes most problems commonly cited as greedy-solvable, including some problems that cannot be handled with e.g. matroids. This gives a new, easily understood framework for studying greed.

2 Majorization Preliminaries

Majorization was first defined by Muirhead [11], and was popularized by Hardy, Littlewood and Pólya in their study of symmetric means [3]. Classical majorization has found several applications in stochastic scheduling [10], optimal coding [14], etc. The book by Marshall and Olkin [10] provides a very good account of the classical theory and its applications. Our work is based on our treatment of majorization in [15], which relies only on linear algebra and convexity.

2.1 Majorization on Sequences

By a **sequence** x we mean a (row) vector with nonnegative entries in a set $\mathcal{C} \subseteq \mathcal{R}_+{}^n$,[1] and by '$\leq^n$' the standard componentwise partial ordering on n-sequences, so $x \leq^n y$ iff $x_i \leq y_i$ for $1 \leq i \leq n$. **Majorization** is also defined as a partial order on sequences:[2]

Definition 2.1 *Define the* **Zeta matrix** $Z = (\zeta_{ij})$ *with* $\zeta_{ij} \;=\; (\textit{if } i \leq j \textit{ then } 1 \textit{ else } 0)$. *For* $x, y \in \mathcal{R}_+{}^n$, *define* $x \preceq y$ *iff* $xZ \leq^n yZ$ *and* $\sum_i x_i = \sum_i y_i$, *i.e.,*

$$x \preceq y \quad \textbf{iff} \quad \left\{ \begin{array}{lcll} \sum_{i=1}^k x_i & \leq & \sum_{i=1}^k y_i & k = 1, \ldots, n-1 \\ \sum_{i=1}^n x_i & = & \sum_{i=1}^n y_i . & \end{array} \right.$$

For example, when $n = 4$,

$$Z = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad Z^{-1} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and for instance

$$\begin{array}{ccccc} [2,4] & \preceq & [3,3] & \preceq & [4,2] \\ [2,5,3,7] & \preceq & [3,6,2,6] & \preceq & [3,7,2,5] \\ [0,a,\ldots,a] & \preceq & [a,0,a,\ldots,a] & \preceq & [a,\ldots,a,0] \qquad (\forall a \geq 0). \end{array}$$

Definition 2.2 *A square matrix is said to be* **(row) stochastic** *if its elements are all non-negative and all rows sum to unity. An upper (resp. lower) triangular row stochastic matrix is an* **upper triangular stochastic matrix (UTSM)** *(resp.* **lower triangular stochastic matrix (LTSM)***).*

Theorem 2.1 *[15]*
A necessary and a sufficient condition that $x \preceq y$ *is that there exist a UTSM* U *such that* $x = yU$. *Equivalently, there must exist an LTSM* L *such that* $xL = y$.

A function G is called **majorization-order preserving**, or **monotone with respect to** $\preceq$, if:

$$x \preceq y \quad \text{implies} \quad G(x) \leq G(y).$$

Theorem 2.2 *[15]*
A differentiable function $G : \mathcal{R}_+^+ \rightarrow \mathcal{R}_+$ *is majorization-order preserving on* $\mathcal{C} \subseteq \mathcal{R}_+^n$ *iff*

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

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

---

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

[2] We use row vectors and matrix multiplication on the right for inessential reasons: it is the convention established in the majorization literature [10], it is convenient to compose operators on the right, and row vectors 'look like' sequences. Note: the majorization relation $\preceq$ here is *not* equivalent to the classical majorization in [10]; see below.

An important example of a majorization-order preserving function is

$$G([z_1, \ldots, z_n]) \;=\; (w_1\, z_1 \;+\; \ldots \;+\; w_n\, z_n) \;=\; z\, w^{\top}$$

for any descending (nonincreasing) nonnegative weight sequence $w = [w_1, \ldots, w_n]$, and any $\mathcal{C} \subseteq \mathcal{R}_+^n$.

2.2 Classical Majorization

Classical majorization is often defined as a binary relation '$\prec$' on 'unsorted sequences' [10]:

Definition 2.3 *For $x, y \in \mathcal{R}_+^n$,*

$$x \prec y \quad \textbf{iff} \quad \left\{ \begin{array}{lcll} \sum_{i=1}^k x_{[i]} & \leq & \sum_{i=1}^k y_{[i]} & k = 1, \ldots, n-1 \\ \sum_{i=1}^n x_{[i]} & = & \sum_{i=1}^n y_{[i]}, & \end{array} \right.$$

where $sort(z) = [z_{[1]}, \ldots, z_{[n]}]$ denotes the descendingly-sorted permutation of $z = [z_1, \ldots, z_n]$.

Thus, $x \prec y$ iff $sort(x)Z \leq^n sort(y)Z$ and $\sum_i x_i = \sum_i y_i$. Although this definition is similar to the earlier definition of majorization on sequences, it is clearly *not* equivalent. For example:

$$\begin{array}{ccccc} [2,4] & \not\prec & [3,3] & \prec & [4,2] \\ [4,2] & \not\prec & [3,3] & \prec & [2,4] \\ [2,4] & \prec & [4,2] & \prec & [2,4]. \end{array}$$

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).

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. However, we define classical majorization as a preorder on sequences (vectors) because this fits our linear algebraic framework.

Definition 2.4 *An $n \times n$ matrix whose entries are all nonnegative and each row and column sum to unity is called a* **doubly stochastic matrix (DSM)***.*

Corresponding to Theorem 2.1 for UTSMs and majorization on sequences is a direct analogue for DSMs and classical majorization:

Theorem 2.3 *(Hardy-Littlewood-Pólya [3], p. 45–49)*
For $x, y \in \mathcal{R}_+^n$, $x \prec y$ if and only if there is a doubly stochastic matrix D such that $x = yD$.

Definition 2.5 *[10] A* **Schur convex function** *is a function that is monotone with respect to $\prec$.*

Since all permutation matrices P are DSMs, it is always true that $xP \prec x$, and as mentioned above the permutations of x define an equivalence class under $\prec$. It follows that Schur convex functions are **symmetric** (invariant under permutation of their argument sequence), because if $P' = P^{\top} (= P^{-1})$ then $G(x) = G(xPP') \leq G(xP) \leq G(x)$, i.e., $G(x) = G(xP)$ for every permutation P.

Theorem 2.4 *(Schur [16] and Ostrowski [12])*
For $z = [z_1, \ldots, z_n] \in \mathcal{R}_+^n$, a symmetric function G is Schur convex if and only if

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

For example, the function $G(z) = \left(sort_w(z)\ w^\top\right)$ is symmetric and Schur convex on $\mathcal{R}_+^n$ for any sequence of nonnegative weights $w = [w_1, \ldots, w_n]$, given that $[z_{[1]}, \ldots, z_{[n]}] = sort_w(z)$ is some permutation of z *sorted in the same order as* w (i.e., $(z_{[i]} - z_{[j]})(w_i - w_j) \geq 0$ for all i and j).

2.3 Generalized Majorization

We can generalize on the definitions of majorization above. Marshall and Olkin ([10], ch.14) list many previous generalizations. Ours is certainly related to some of these, but is not equivalent.

Definition 2.6 *Let $\mathcal{T}$ be a semigroup of $n \times n$ matrices ($\mathcal{T}$ contains the identity and is closed under multiplication), and let x and y be sequences of length n (n-vectors). Then the preorder of $\mathcal{T}$-**majorization**, written $x \ \preceq_{\mathcal{T}}\ y$, holds whenever there is a matrix $T \in \mathcal{T}$ such that $x = yT$.*

That this is a preorder follows from $\mathcal{T}$ being a semigroup, since the identity matrix makes the relation reflexive, and closure under multipication makes it transitive.

Ordered majorization on sequences can thus be called **UTSM-majorization**, written '$\preceq_{UTSM}$'. Also $x \preceq_{UTSM} y$ iff $y \preceq_{LTSM} x$. Also **DSM-majorization** is definable by $x \preceq_{DSM} y$ iff $x \prec y$.

Let **wDSM** be the set of **doubly sub-stochastic matrices** (nonnegative matrices each of whose rows and columns totals *at most* 1). **wDSM-majorization** is then almost identical to the DSM-majorization but $x \preceq_{wDSM} y$ also holds for sequences x whose total is less than that of y, and so is equivalent to $x \leq^n yD$ for some DSM D. This is called **weak majorization** ($x \prec_w y$) in ch.2.C of [10].

Definition 2.7 *A function $G : \mathcal{R}_+^+ \to \mathcal{R}_+$ is* **monotone with respect to** $\preceq_{\mathcal{T}}$ *on $\mathcal{C}$ if $x \preceq_{\mathcal{T}} y$ implies $G(x) \leq G(y)$. Also, G is* **antitone** *if $x \preceq_{\mathcal{T}} y$ implies $G(x) \geq G(y)$.*

Definition 2.8 *The preorder of* $(\mathcal{T}, G)$-**majorization** *is defined by*

$$x \ \preceq_{(\mathcal{T},G)}\ y \qquad \textbf{iff} \qquad x \ \preceq_{\mathcal{T}}\ y \quad \underline{and} \quad G(x) \ \leq \ G(y).$$

G is guaranteed to be monotone with respect to $\preceq_{(\mathcal{T},G)}$, and antitone with respect to $\preceq_{(\mathcal{T},-G)}$.

One situation where $(\mathcal{T}, G)$-majorization is useful is where $\mathcal{T}$ is a set of permutations, in which case optimizing $G(z)$ amounts to sorting z (permuting z into optimal order) with respect to G.

Other ('partially ordered') generalizations of majorization also have direct connections with submodularity (and matroids), but we omit these because of the paper length limit.

2.4 Connections Between Majorization and Convexity

Connections with convexity can make a class of optimization problems efficiently solvable. With UTSM-majorization or DSM-majorization, convexity can be connected with majorization in several ways. Although Schur convexity is just monotonicity with respect to $\prec$, there are many convexity results for Schur convex functions [10, 4]. For example, any symmetric function G that is convex is necessarily Schur convex.

Another connection between convexity and majorization with many facets is that, given an interval $I \subseteq \mathcal{R}_+$ and continuous $\phi : I \to \mathcal{R}_+$, the function $G(z) = \sum_i \phi(z_i)$ is Schur convex if and only if ϕ is convex on I ([10] p.68, [4], p.333). Stated more generally:

Theorem 2.5 *(Hardy-Littlewood-Pólya [3], p. 89; also [10], p.11, [4], pp.320–321)* *If $\phi : I \to \mathcal{R}_+$ is a continuous convex function (resp. continuous increasing convex function) defined on an interval $I \subseteq \mathcal{R}_+$, then for all $x, y \in I^n$, $\sum_{i=1}^n \phi(x_i) \leq \sum_{i=1}^n \phi(y_i)$ iff $x \preceq_{DSM} y$ (resp. $x \preceq_{wDSM} y$).*

This is the **'majorization theorem'**. There are a variety of extensions, which we omit.

2.5 Generalized Majorization as an Exchange Ordering

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 [10], including Lorenz curves related to those in Figure 1. Of the three income distributions shown, 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. One distribution majorizes another exactly when its graph lies above the other's. Dalton's principle says that, by repeatedly transferring income from rich to poor (and only this), we can gradually deform the 'capitalism' distribution to that of 'communism'.

The generalized majorization ordering $x \preceq_T y$ holds when there is a matrix T such that $x = yT$. Thus T is exactly an **exchange**, or **transfer** — an operator transforming y to x.

Definition 2.9 *The* **pairwise exchange UTSMs** $U_{pq}(\epsilon)$ *and* **pairwise exchange DSMs** $D_{pq}(\epsilon)$ $1 \leq p < q \leq n$, $\epsilon \in [0,1]$, *are matrices differing from the identity only in the* $\{p,q\} \times \{p,q\}$ *submatrix:*

$$U_{pq}(\epsilon) = \begin{pmatrix} \ddots & & & & \\ & 1-\epsilon & & \epsilon & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \ddots \end{pmatrix}, \quad D_{pq}(\epsilon) = \begin{pmatrix} \ddots & & & & \\ & 1-\epsilon & & \epsilon & \\ & & \ddots & & \\ & \epsilon & & 1-\epsilon & \\ & & & & \ddots \end{pmatrix}.$$

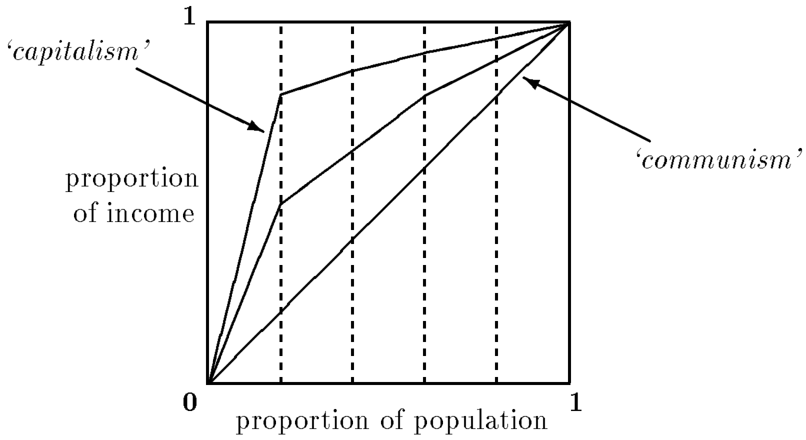

Figure 1: Three increasingly uneven income distributions for a population of five individuals

The **pairwise exchange LTSMs** $L_{pq}(\epsilon)$ *are defined by* $L_{pq}(\epsilon) = RU_{pq}(\epsilon)R$*, where* R *is the reflection permutation that maps* i *to* $n-i+1$*, for* $1 \leq i \leq n$*.*

Theorem 2.6 *[15]*
Every LTSM L *is equal to a product of* $n(n-1)/2$ *pairwise exchange LTSMs. Specifically*

$$L \quad = \quad (\ell_{ij}) \quad = \quad \prod_{k=1}^{n-1} \prod_{i=k+1}^{n} L_{ik}(\epsilon_{ik})$$

where $\epsilon_{ik} = \ell_{ik} / \left(\prod_{j<k}(1-\epsilon_{ij})\right)$ *if* $\ell_{ik} > 0$*, and 0 otherwise.*

If $\epsilon_{ij} = 1$ for any j, then necessarily $\ell_{ik} = 0$ for $k > j$, so this is well-defined. While a similar statement holds for UTSMs (via the reflection R), this is not so for DSMs ([10], p.23). However:

Theorem 2.7 *[10], p.21; [15]*
For $x, y \in \mathcal{R}_+^n$*, and* $\mathcal{T}$ *is any of UTSM, LTSM, DSM, or wDSM:*
If $x \preceq_{\mathcal{T}} y$ *then* $x = yT$ *where* $T \in \mathcal{T}$ *is a product of at most* $(n-1)$ *pairwise exchanges in* $\mathcal{T}$*.*

For example, the distribution sequences for Figure 1 are as in the table below:

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

We can transform *'capitalism'* to *'communism'* with the product of pairwise exchanges

$$U \quad = \quad U_{12}(\,1-\tfrac{.20}{.80}\,)\ U_{23}(\,1-\tfrac{.20}{.67}\,)\ U_{34}(\,1-\tfrac{.20}{.52}\,)\ U_{45}(\,1-\tfrac{.20}{.36}\,) \quad = \quad \begin{pmatrix} .25 & .23 & .20 & .18 & .15 \\ 0 & .30 & .27 & .24 & .19 \\ 0 & 0 & .38 & .35 & .27 \\ 0 & 0 & 0 & .56 & .44 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The denominators here reflect the total remaining amount that could be exchanged; for example $.52 = (.80 + .07 + .05) - (.20 + .20)$. The final matrix shows how much income each of the five individuals ultimately transfers to the others.

Theorem 2.8 *The $\preceq_{UTSM}$ majorization-preservation condition*

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

given in Theorem 2.2 is equivalent to the constraint on the **Lie derivative** *with respect to* U_{ij}

$$\frac{d}{d\epsilon}\, G(\ z\ U_{ij}(\epsilon)\ )\bigg|_{\epsilon=0} \quad \leq \quad 0 \qquad\qquad (1 \leq i < j \leq n).$$

The Schur convexity condition analogously constrains the Lie derivative with respect to D_{ij}.

This easily-verified theorem shows that pairwise exchanges are sufficient to determine whether a function is monotone with respect to majorization.

3 Greed

Greedy-solvable optimization problems are often defined with intuitive notions of greedy-solvability. Instances of problems such as *Continuous Knapsack*, *Storage of Files on Tape*, and *Job Sequencing* variants have been labeled as greedy problems. We show that concepts in greed are related to concepts in majorization in a direct way. This forms the basis for our motivation to understand greed in terms of majorization.

Conceptually, a greedy problem is solvable by repeated selection of the 'best possible' exchange. As indicated above, we model exchanges as linear operators. We seek a sequence of exchanges that moves us from an initial feasible solution of the problem to a final, optimal solution. To this end we restrict the objective function G to be monotone with respect to $\preceq_{\mathcal{T}}$, the preorder defined by the semigroup of exchanges $\mathcal{T}$. The strategy of applying a sequence of exchanges that maximally improve the objective function G, it seems to us, can be viewed as the essence of greed. Any additional structure of G and $\mathcal{T}$, such as convexity, can however be exploited to give fast implementations of this greedy strategy.

3.1 Optimal Majorization Problems

Definition 3.1 *An* **optimal majorization problem** *has one of the two forms*

$$\boxed{\begin{array}{ll} \textit{maximize} & G(x) \\ \textit{subject to} & a \ \preceq\ x \ \preceq\ b \\ & x\ \in\ \mathcal{C} \end{array}} \qquad\qquad \boxed{\begin{array}{ll} \textit{minimize} & G(x) \\ \textit{subject to} & a \ \preceq\ x \ \preceq\ b \\ & x\ \in\ \mathcal{C} \end{array}}$$

where $\mathcal{C} \subseteq \mathcal{R}_+^n$, $\preceq$ is a generalized majorization ordering ($\preceq_{\mathcal{T}}$ or $\preceq_{(\mathcal{G},\mathcal{T})}$), $G : \mathcal{C} \to \mathcal{R}_+$ is either monotone or antitone with respect to $\preceq$ on $\mathcal{C}$, and a, b are constant sequences in $\mathcal{C}$.

A real sequence $y \in \mathcal{R}_+^n$ is called a **feasible solution** *to the optimal majorization problem if $y \in \mathcal{C}$ and $a \preceq_{\mathcal{T}} y \preceq_{\mathcal{T}} b$. $G(y)$ is called a* **feasible answer**. *A feasible solution $y \in \mathcal{C}$ is called the* **optimal solution** *if for all $x \in \mathcal{C}$, $G(x) \leq G(y)$. $G(y)$ is called the* **optimal answer**.

Example 3.1 (Continuous Knapsack)

$$\boxed{\begin{array}{lll} \textit{maximize} & G(x) \;=\; \sum_{i=1}^{n} w_i x_i \;=\; x\ w^\top \\ \textit{subject to} & [0,0,\ldots,0,1] \;\preceq_{UTSM}\; [x_1,\ldots,x_n] \\ & 0 \le x_i \le c_i \end{array}}$$

where commodities are ordered by their values w_i in descending order, so that $w_1 \geq \ldots \geq w_n$, and $c_i \in [0,1]$ is the maximum permissible proportion of the i-th commodity. The majorization constraint expresses the requirement that the proportions x_i total to 1.

The optimal solution is $x \;=\; [\ c_1,\ \ldots,\ c_r,\ (1-\sum_{i=1}^{r} c_i),\ 0,\ \ldots,\ 0\ ]$, where r is either the index such that $\sum_{i=1}^{r} c_i < 1$ but $\sum_{i=1}^{r+1} c_i \geq 1$, or n if no such index exists.

Example 3.2 (Storage of Files on Tape)

$$\boxed{\begin{array}{ll} \textit{minimize} & G(\ell) \;=\; \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{i} \ell_j\right) \;=\; \ell\ Z\ [\frac{1}{n},\ldots,\frac{1}{n}]^\top \\ \textit{subject to} & [L_1,\ldots,L_n] \;\preceq_{DSM}\; [\ell_1,\ldots,\ell_n] \;\preceq_{DSM}\; [L_1,\ldots,L_n] \end{array}}$$

where L_i is the length of the i-th file, and we seek a permutation $\ell \;=\; [\ell_1,\ldots,\ell_n]$ of $[L_1,\ldots,L_n]$ for storage of files on tape that minimizes the average file access time. The DSM-majorization constraint expresses the permutation requirement; extremal values of ℓ under any linear objective G will be defined by extremal DSMs, which are permutation matrices.

The optimal solution is the ascendingly sorted permutation ℓ of L.

Example 3.3 (Single-Processor Average Flow Time)

$$\boxed{\begin{array}{ll} \textit{minimize} & G([(t_1,w_1),\ \ldots,\ (t_n,w_n)]) \;=\; \sum_{i=1}^{n}\ w_i\ \left(\sum_{j=1}^{i} t_j\right) \;=\; t\ Z\ w^\top \\ \textit{subject to} & [(t_1,w_1),\ \ldots,\ (t_n,w_n)] \;\preceq_{(t/w)}\; [(T_1,W_1),\ \ldots,\ (T_n,W_n)] \end{array}}$$

where T_i and W_i are respectively the processing time and (positive) weight of the i-th input job, and we seek a permutation $[(t_1,w_1),\ \ldots,\ (t_n,w_n)]$ of $[(T_1,W_1),\ \ldots,\ (T_n,W_n)]$ that sequences the jobs in a way which minimizes their average flow time. Here '$\preceq_{(t/w)}$' is a sort ordering of sequences by their t/w values; the optimal solution is the fully sorted sequence. Specifically, for $1 \leq p \leq n$,

$$[\ \ldots,\ (t_{p+1},w_{p+1}),\ (t_p,w_p),\ \ldots]\quad \preceq_{(t/w)} \quad [\ \ldots,\ (t_p,w_p),\ (t_{p+1},w_{p+1}),\ \ldots]$$

iff $(t_{p+1}/w_{p+1}) \leq (t_p/w_p)$, i.e., the exchange $(t_p,w_p) \leftrightarrow (t_{p+1},w_{p+1})$ improves t/w sort order, and generally $[(t_1,w_1),\ \ldots,\ (t_n,w_n)] \;\preceq_{(t/w)}\; [(T_1,W_1),\ \ldots,\ (T_n,W_n)]$ iff in the semigroup of permutation matrices there is some P, obtainable as a product of adjacent exchanges that each improve t/w sort order, such that $[(t_1,w_1),\ \ldots,\ (t_n,w_n)] \;=\; [(T_1,W_1),\ \ldots,\ (T_n,W_n)]\ P$.

The ordering $\preceq_{(t/w)}$ is precisely the $(\mathcal{P},G)$-majorization ordering $\preceq_{(\mathcal{P},G)}$, where $\mathcal{P}$ is the group of permutations. This is because $\preceq_{(t/w)}$ is the part of $\preceq_{\mathcal{P}}$ on which G is monotone:

$$\begin{aligned} G([\ \ldots,\ (t_{p+1},w_{p+1}),\ (t_p,w_p),\ \ldots]) \;&-\; G([\ \ldots,\ (t_p,w_p),\ (t_{p+1},w_{p+1}),\ \ldots]) \\ &= (\cdots\ w_{p+1}t_{p+1} + w_p t_{p+1} + w_p t_p\ \cdots) \;-\; (\cdots\ w_p t_p + w_{p+1} t_p + w_{p+1} t_{p+1}\ \cdots) \\ &= (w_p t_{p+1} \;-\; w_{p+1} t_p) \\ &= w_p\ w_{p+1} \left(\frac{t_{p+1}}{w_{p+1}} \;-\; \frac{t_p}{w_p} \right). \end{aligned}$$

3.2 Greedy-Solvability of Optimal Majorization Problems

As the objective function G to be maximized in an optimal majorization problem is restricted to be monotone with respect to the order $\preceq_{\mathcal{T}}$, we can improve any solution x by finding another feasible solution xT, for any $T \in \mathcal{T}$. Thus, all that is required for greedy-solvability of the problem is that there be a unique optimum x^*, i.e., a top (or bottom) to the order defined by $\preceq_{\mathcal{T}}$ and G. We identified the optimum for each of the examples above.

In practice, greedy algorithms take advantage of structure of the operator space $\mathcal{T}$. There seems to be a common structure amenable to greed: the semigroup $\mathcal{T}$ can be *convex*, and G can be a convex function. For example, the set of UTSMs is a convex space, since any convex combination of UTSMs is a UTSM, and any UTSM is a convex sum of upper triangular 0-1 matrices that have exactly one 1 in each row. Birkhoff [1] proved the set of DSMs is a convex space, whose extremal points are permutations. In other words, every DSM is a convex sum of permutation matrices.

Theorem 3.1 *When $\mathcal{T}$ is a convex set, G is convex on $\mathcal{C}$, the optimal majorization problem*

$$\boxed{\begin{array}{ll} \textit{maximize} & G(x) \\ \textit{subject to} & a \ \preceq_{\mathcal{T}} \ x \ \preceq_{\mathcal{T}} \ b \\ & x \ \in \ \mathcal{C} \end{array}}$$

is **greedy-solvable** *(assuming G is monotone with respect to $\preceq_{\mathcal{T}}$): repeated maximal improvement of G under exchanges ultimately yields an optimal solution.*

This can be established by noting that any iteration that initially assigns $x \leftarrow a$ or $x \leftarrow b$ and repeatedly replaces x by x', where $x = x'T$ for some exchange $T \in \mathcal{T}$ such that $a \preceq_{\mathcal{T}} x' \preceq_{\mathcal{T}} b$ and $x' \in \mathcal{C}$, must monotonically increase the value of $G(x)$, and ultimately converge to the unique optimum that is guaranteed by the convexity of G. Notice b is not guaranteed to be the optimal solution since $\preceq_{\mathcal{T}}$ can be a preorder (although b would be optimal if $\preceq_{\mathcal{T}}$ were a partial order).

Decomposability of elements of $\mathcal{T}$ into products of pairwise exchanges can be exploited in the knapsack problem. There the knapsack constraint $a \preceq_{UTSM} x$ is equivalent to $x \preceq_{LTSM} a$, and there is an LTSM L such that $x = aL$. By Theorem 2.6, L is a product of pairwise exchange LTSMs $L = \prod_{k=1}^{n-1} \prod_{i=k+1}^{n} L_{ik}(\epsilon_{ik})$, so we can step from a to x:

$$a \quad\rightarrow\quad aL_{21}(\epsilon_{21}) \quad\rightarrow\quad aL_{21}(\epsilon_{21})L_{31}(\epsilon_{31}) \quad\rightarrow\quad \cdots \quad\rightarrow\quad aL_{21}(\epsilon_{21})\ldots L_{n,n-1}(\epsilon_{n,n-1}) \quad=\quad x.$$

This yields a '*bubblesort*' scheme for greedy optimization: for the values of i and k indicated, repeatedly select the exchange $L_{ik}(\epsilon_{ik})$ that maximizes $G(\,aL_{21}(\epsilon_{21})\ldots L_{ik}(\epsilon_{ik})\,)$ under $\preceq_{UTSM}$.

4 Conclusion

The contribution of this paper is the identification of an important class of optimization problems for which the objective is a monotone function over a generalized majorization ordering. The significance of this class is that it apparently includes most problems commonly cited as greedy-solvable, including some problems that cannot be handled with matroids or greedoids. Furthermore, this class gives a clear formal notion of 'exchange' that is based on linear algebra and convexity, is reconcilable with submodularity, and is easily understood.

References

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

[2] A. Dechter and R. Dechter. On the Greedy Solution of Ordering Problems. *ORSA Journal on Computing*, 1(3):181–189, Summer 1989.

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

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

[5] B. Korte and L. Lovász. Mathematical Structures Underlying Greedy Algorithms. In ed. F. Gécseg, editor, *Fundamentals of Computation Theory (LNCS # 117)*, pages 205–209. Springer-Verlag, 1981.

[6] B. Korte, L. Lovász, and R. Schrader. *Greedoids.* Springer-Verlag, Berlin, 1991.

[7] E.L. Lawler. *Combinatorial Optimization: Networks and Matroids.* Holt, Rinehart & Winston, New York, 1976.

[8] E.L. Lawler. Submodular Functions and Polymatroid Optimization. In A.H.G. Rinnooy Kan (eds.) M. O'hEigeartaigh, J.K. Lenstra, editor, *Combinatorial Optimization: Annotated Bibliographies*, pages 32–38. J. Wiley & Sons, 1985.

[9] L. Lovasz. 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.

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

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

[12] 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.

[13] C.H. Papadimitriou and K.S. Stieglitz. *Combinatorial Optimization.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

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

[15] D.S. Parker and P. Ram. Majorization on Sequences. Technical report, UCLA Computer Science Dept., Los Angeles, CA 90024-1596, July 1994.

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