Majorization and Applications to Optimization

Zvika Ben-Haim[*]   Tsvika Dvorkind[†]

January 6, 2004

Abstract

Majorization is a partial ordering on vectors which determines the degree of similarity between the vector elements. Majorization and the related concept of Schur-convexity can sometimes be used to prove certain properties of the solution to an optimization problem. For example, these concepts form a convenient tool for showing that in certain symmetric problems, the optimum is obtained when all optimization variables are equal.

1 Majorization and Schur-Convexity

We begin by defining the concepts of majorization and Schur-convexity. In the following discussion, boldface letters indicate real n-vectors. We use the notation $x_{(1)}$ to indicate the largest element in $\mathbf{x}$, $x_{(2)}$ to indicate the second-largest element, and so on.

Definition 1. The vector $\mathbf{x}$ is said to *majorize* the vector $\mathbf{y}$ (denoted $\mathbf{x} \succ \mathbf{y}$) if

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)}, \quad k = 1, 2, \ldots n-1, \tag{1}$$

$$\text{and} \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{2}$$

Majorization is a partial ordering among vectors, which applies only to vectors having the same sum. It is a measure of the degree to which the vector elements differ. For example, it can be shown that all vectors of sum s majorize the uniform vector $\mathbf{u}_s = \left(\frac{s}{n}, \frac{s}{n}, \ldots \frac{s}{n}\right)$. Intuitively, the uniform

[*]zvikabh@technion.ac.il

[†]dvorkind@technion.ac.il

vector is the vector with minimal differences between elements, so all other vectors majorize it. Formally, this follows directly from the fact that for any vector $\mathbf{x}$ of sum s,

$$\sum_{i=1}^{k} x_{(i)} \geq \frac{k}{n}\, s, \tag{3}$$

a fact which can be shown by induction on k.

Definition 2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *Schur-convex* if

$$\mathbf{x} \succ \mathbf{y} \implies f(\mathbf{x}) \geq f(\mathbf{y}). \tag{4}$$

Schur-convex functions translate the ordering of vectors to a standard scalar ordering. An example of a Schur-convex function is the max function, $\max(\mathbf{x}) = x_{(1)}$. Clearly, if $\mathbf{x} \succ \mathbf{y}$, then $x_{(1)} \geq y_{(1)}$.

The max function is symmetric in that any two of its arguments can be switched without modifying the value of the function. Symmetry is a necessary condition for a function to be Schur-convex. Thus, for example, linear functions are not Schur-convex unless they are symmetric.

However, if a function is symmetric and convex, then it is Schur-convex [1, 3.C.2], [4].

There are several simple rules for Schur-convexity of combinations of Schur-convex functions. For instance, suppose $h : \mathbb{R}^k \to \mathbb{R}$ is non-decreasing in each argument, and $f_1, \ldots f_k$ are Schur-convex functions $f_i : \mathbb{R}^n \to \mathbb{R}$. Then, the function $h(f_1(\mathbf{x}), \ldots f_k(\mathbf{x}))$ is Schur-convex [1, 3.B.1]. Additional combination laws are given in [1, 3.B].

2 Application to Optimization

2.1 Proving Uniform Optimality

The concept of majorization can be used as a tool for proving that the solution to an optimization problem occurs when all variables are equal. For unconstrained problems, this is true if the objective function is Schur-convex.

To see this, first consider the constrained optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t. } \sum x_i = s, \tag{5}$$

and assume f is Schur-convex. Since the uniform vector $\mathbf{u}_s$ is majorized by any other vector of sum s, we have

$$f(\mathbf{u}_s) \leq f(\mathbf{x}) \tag{6}$$

for any $\mathbf{x}$ having sum s. Thus, $\mathbf{u}_s$ is the solution to (5).

The unconstrained problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{7}$$

is equivalent to

$$\min_{s} \left(\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \ \sum x_i = s \right), \tag{8}$$

so that the solution to (7) is also $\mathbf{u}_s$, for some s. This result can be summarized in the following lemma.

Lemma 1. *In an unconstrained minimization problem* (7)*, where the objective function is Schur-convex, the optimum is obtained when all variables are equal.*

This result can also be extended to constrained optimization problems, as long as the uniform solution is always feasible.

Lemma 2. *Consider the constrained optimization problem*

$$\min_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}), \tag{9}$$

where f is Schur-convex. Assume that $\mathcal{A}$ has the property that, for any value p in the set $f(\mathcal{A})$, the sum-p uniform vector $\mathbf{u}_p$ is a member of $\mathcal{A}$. Then, the optimum value is obtained when all variables are equal.

A more general version of this lemma is proved in the next subsection. An example which amounts to the use of this lemma in a particular optimization problem can be found in Lemma 2 of [3].

2.2 Minimizing the Maximum of k Functions

A slightly different application occurs in the optimization problem

$$\min_{\mathbf{x} \in \mathcal{A}} h(f_1(\mathbf{x}), \ldots f_k(\mathbf{x})), \tag{10}$$

where h is Schur-convex. For example, we may wish to minimize the worst-case (or maximum) among several different functions (as we have seen, the maximum is a Schur-convex function). It is sometimes useful to show that the optimum of (10) is obtained for $\mathbf{x}$ such that $f_1(\mathbf{x}) = \cdots = f_k(\mathbf{x})$. This can be shown using a generalization of Lemma 2.

Lemma 3. *Consider the optimization problem* (10)*, where* h *is a Schur-convex function. Let* $F(\mathbf{x}) = \sum f_i(\mathbf{x})$*. Suppose that for any value* s *in* $F(\mathcal{A})$*, there exists* $\mathbf{x} \in \mathcal{A}$ *such that* $f_1(\mathbf{x}) = \cdots = f_k(\mathbf{x}) = s/n$*. Then, the optimal solution to* (10) *satisfies* $f_1(\mathbf{x}) = \cdots = f_k(\mathbf{x})$*.*

Proof. Problem (10) is equivalent to

$$\min_s \left(\min_{\mathbf{x} \in \mathcal{A}} h(f_1(\mathbf{x}), \dots f_k(\mathbf{x})) \quad \text{s.t. } F(\mathbf{x}) = s \right). \tag{11}$$

For any s in $F(\mathcal{A})$, there exists $\mathbf{x}_s \in \mathcal{A}$ such that $(f_1(\mathbf{x}_s), \dots f_k(\mathbf{x}_s)) = \mathbf{u}_s$. Thus, for any $\mathbf{y}$ having sum s,

$$(f_1(\mathbf{y}), \dots f_k(\mathbf{y})) \succ (f_1(\mathbf{x}_s), \dots f_k(\mathbf{x}_s)), \tag{12}$$

so that

$$h(f_1(\mathbf{y}), \dots f_k(\mathbf{y})) \geq h(f_1(\mathbf{x}_s), \dots f_k(\mathbf{x}_s)). \tag{13}$$

Hence the solution to the problem

$$\min_{\mathbf{x}} h(f_1(\mathbf{x}), \dots f_k(\mathbf{x})) \quad \text{s.t. } F(\mathbf{x}) = s \tag{14}$$

is $\mathbf{x}_s$. Thus, the solution to (11) equals $\mathbf{x}_s$ for some value of s, so that for the optimal solution, $f_1(\mathbf{x}) = \cdots = f_k(\mathbf{x})$. □

3 Majorization and Linear Algebra

The following lemmas are useful in proving the the optimality of various matrix optimization problems in which a constraint on the trace of the matrix is given [2].

Lemma 4. *Let* $\mathbf{H}$ *be a Hermitian* $n \times n$ *matrix with eigenvalues* $\lambda = (\lambda_1, \dots \lambda_n)$*. Let* $\mathbf{h} = (H_{11}, H_{22}, \dots H_{nn})$*. Then,* $\lambda \succ \mathbf{h}$*.*

For a proof of this lemma, please see [1, 9.B.1].

Lemma 5. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two vectors in* $\mathbb{R}^n$ *such that* $\mathbf{x} \succ \mathbf{y}$*. Then, there exists a real symmetric matrix with diagonal elements given by* $\mathbf{y}$ *and eigenvalues given by* $\mathbf{x}$*.*

For a proof of this lemma, please see [1, 9.B.2]. An algorithm for finding the matrix satisfying these requirements is given in [2, p.68].

References

[1] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications,* Volume 143 in Mathematics in Science and Engineering Series, Academic Press, 1979.

[2] D.P. Palomar, *A Unified Framework for Communications through MIMO Channels,* Ph.D. Dissertation, Universitat Polytècnica di Catalunya, Barcelona, 2003.

[3] D.P. Palomar, J.M. Cioffi, and M.A. Lagunas, "Uniform power allocation in MIMO channels: a game-theoretic approach." IEEE Trans. Inf. Theory, 49: 7, pp. 1707-1727, 2003.

[4] I. Schur, "Über eine Klasse von Mittelbildungen mit Anwendungen die Determinanten." Theorie Sitzungsber. Berlin. Math. Gesellschaft, vol. 22, pp. 9-20, 1923.