A convexity result of Davis

Let $\mathcal{H}_n$ denote the set of all $n \times n$ Hermitian (self-adjoint) matrices. Suppose $f : \mathcal{H}_n \to \mathbb{R}$. The function f is said to be convex if

<1> $$f((1-t)A+tB) \le (1-t)f(A)+tf(B) \qquad \text{for all } 0 \le t \le 1 \text{ and } A, B \in \mathcal{H}_n.$$

The function is unitarily invariant if

<2> $$f(U^*AU) = f(A) \qquad \text{for each } A \in \mathcal{H}_n \text{ and unitary } U.$$

If $A = U\Lambda U^*$ with unitary U and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ then

$$f(A) = f(\Lambda) = \Psi(\lambda_1, \dots, \lambda_n).$$

We could also think of Ψ as a function defined for diagonal matrices with real elements. That is,

$$\Psi(\lambda_1, \dots, \lambda_n) = \Psi(\Lambda).$$

The function Ψ must be symmetric in its arguments, because permutation matrices are unitary. If f is also convex then clearly Ψ must also be convex as a function on $\mathbb{R}^n$: apply <1> to diagonal matrices A and B.

Davis (1957) proved that the implication also goes in the other direction. That is, convexity of Ψ for a unitarily invariant f implies convexity of f.

Remark. Actually Davis allowed f to take values in a partially ordered real vector space. The proof is essentially the same as the proof for real-valued f.

The proof starts from <1>, writing Λ for $(1-t)A+tB$. We may assume that the Hermitian matrix Λ is diagonal, $\text{diag}(\lambda_1, \dots, \lambda_n)$. The λ_j's are the eigenvalues of Λ with corresponding eigenvalues $e_1, \dots, e_n$, the usual orthonormal basis for $\mathbb{C}^n$. (The proof also works with real symmetric matrices and the $\{e_j\}$ interpreted as an onb for $\mathbb{R}^n$.)

For any $n \times n$ matrix M write $\text{diag}(M)$ for the $n \times n$ matrix with the same diagonal entries as M but zeros in the off-diagonal elements.

The proof works by first showing that

<3> $$\Psi(\text{diag}(M)) := f(\text{diag}(M)) \le f(M) \qquad \text{for } M \in \mathcal{H}_n.$$

From the equality

$$\Lambda = (1-t)\,\text{diag}(A) + t\,\text{diag}(B)$$

we then get

$$\begin{aligned} f(\Lambda) &= \Psi\left[(1-t)\text{diag}(A) + t\text{diag}(B)\right] \\ &\le (1-t)\Psi(\text{diag}(A)) + t\Psi(\text{diag}(B)) \qquad \text{by convexity of } \Psi \\ &\le (1-t)f(A) + tf(B) \qquad \text{by <3>.} \end{aligned}$$

To establish <3>, suppose M has eigenvalues μ_j with corresponding eigenvectors z_j, for $j = 1, \dots, n$. Write $\text{diag}(M)$ as $\text{diag}(m_1, \dots, m_n)$, where

$$\begin{aligned} m_j &= \langle e_j, Me_j \rangle \\ &= \langle e_j, \textstyle\sum_k \mu_k \langle e_j, z_k \rangle z_k \rangle \qquad \text{because } Me_j = M \textstyle\sum_k \langle e_j, z_k \rangle z_k \\ &= \textstyle\sum_k \mu_k S_{j,k} \qquad \text{where } S_{J,k} := \langle e_j, z_k \rangle^2. \end{aligned}$$

In vector form the equality becomes $m = S\mu$, where $m = (m_1, \dots, m_n)$ and $\mu = (\mu_1, \dots, \mu_n)$ are both $n \times 1$ vectors.

The key insight is that the $n \times n$ matrix S is doubly stochastic because

$$1 = \|e_j\|^2 = \textstyle\sum_k \langle e_j, z_k \rangle^2 = \textstyle\sum_k S_{j,k}$$

with a similar equality for $\sum_j S_{j,k}$. By an elegant result of Birkoff, S can be written as a convex combination of permutation matrices,

$$S = \textstyle\sum_\sigma \gamma_\sigma P_\sigma.$$

That is, for each permutation σ of $[n]$,

$$P_\sigma[j,k] = \begin{cases} 1 & \text{if } j = \sigma(i) \\ 0 & \text{otherwise.} \end{cases}$$

Thus $m = \sum_\sigma \gamma_\sigma P_\sigma \mu$ and

$$\Psi(m) \le \sum\nolimits_\sigma \gamma_\sigma \Psi(P_\sigma \mu) = \Psi(\mu),$$

the last equality because Ψ is a symmetric function and P_σ merely permutes the components of μ. That is,

$$f(\text{diag}(M)) \le f(M),$$

as asserted by <3>.

References

Davis, C. (1957). All convex invariant functions of hermitian matrices. *Archiv der Mathematik 8*(4), 276–278.

Birkhoff's representation of doubly stochastic matrices

1 A Theorem

S:statement

The discussion at http://mathoverflow.net/questions/43569/ and the article at http://en.wikipedia.org/wiki/Birkhoff_polytope, suggest the following result was not originally due to Birkhoff.

A square matrix D is said to be doubly stochastic if it has nonnegative entries with all row and column sums equal to 1. A permutation matrix has exactly one 1 in each row and each column. More precisely, an $n \times n$ permutation matrix is specified by a permutation σ of $\{1, 2, \dots, n\}$:

$$P_\sigma[i, j] = \begin{cases} 1 & \text{if } j = \sigma(i) \\ 0 & \text{otherwise.} \end{cases}$$

I will write things like $P_{1,3,2}$ or P_σ when σ is the permutation for which $\sigma(1) = 1$, $\sigma(2) = 3$, and $\sigma(3) = 2$:

$$P_{1,3,2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

The subscripts give the location of the 1's in successive rows. In particular, $I_n = P_{1,2,\dots,n}$.

ds <*1*> **Theorem.** *Every doubly stochastic matrix can be written as a convex combination of permutation matrices.*

Geometrically, the set $\mathcal{D}_n$ of all $n \times n$ doubly stochastic matrices can be identified with a compact, convex subset of $\mathbb{R}^{n^2}$. It is the convex polytope defined by n^2 inequality constraints ($D[i,j] \ge 0$ for all (i,j)) and $2n$ linear equalities. The extreme points of that polytope correspond to the $n!$ possible permutation matrices.

2 An Example

S:example

The following Example previews the main idea behind the proof of Theorem <1>.

eg <2> **Example.** Here is one way to represent a doubly stochastic matrix as a convex combination of permutation matrices. Suppose

$$D_0 = \begin{bmatrix} .6 & .2 & .2 \\ .3 & .5 & .2 \\ .1 & .3 & .6 \end{bmatrix} = \tfrac{1}{2} I_3 + \begin{bmatrix} .1 & .2 & .2 \\ .3 & 0 & .2 \\ .1 & .3 & .1 \end{bmatrix} = \tfrac{1}{2} P_{1,2,3} + \tfrac{1}{2} D_1$$

$$D_1 = \begin{bmatrix} .2 & .4 & .4 \\ .6 & 0 & .4 \\ .2 & .6 & .2 \end{bmatrix} = \tfrac{4}{10} P_{3,1,2} + \begin{bmatrix} .2 & .4 & 0 \\ .2 & 0 & .4 \\ .2 & .2 & .2 \end{bmatrix} = \tfrac{4}{10} P_{3,1,2} + \tfrac{6}{10} D_2$$

$$D_2 = \begin{bmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} = \tfrac{1}{3} P_{1,3,2} + \tfrac{2}{3} D_3$$

$$D_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix} = \tfrac{1}{2} P_{2,1,3} + \tfrac{1}{2} P_{2,3,1}.$$

Thus $10D = 5P_{1,2,3} + 2P_{3,1,2} + P_{1,3,2} + P_{2,1,3} + P_{2,3,1}$.

At each step I looked for a permutation σ_j with $D_j[i, \sigma_j(i)] > 0$, put $\delta_j = \min_i D_j[i, \sigma_j(i)]$, then subtracted off $\delta_j P_{\sigma_j}$ to leave a matrix R_j whose rows and columns all summed to $1 - \delta_j$. If $\delta_j < 1$ then $R_j = (1-\delta_j) D_{j+1}$, where D_{j+1} was a new doubly stochastic matrix that contained at least one more 0 than D_j. If $\delta_j = 1$ then $R_j = 0$ and we are done.

The calculations could also be written as

$$
\begin{aligned}
&D_0 = \delta_1 P_{\sigma_1} + (1-\delta_1)D_1 \qquad \text{with } \delta_1 = 1/2\\
&D_1 = \delta_2 P_{\sigma_2} + (1-\delta_2)D_2 \qquad \text{with } \delta_2 = 4/10\\
&D_2 = \delta_3 P_{\sigma_3} + (1-\delta_3)D_3 \qquad \text{with } \delta_3 = 1/3\\
&D_3 = \delta_4 P_{\sigma_4} + (1-\delta_4)D_4 \qquad \text{with } \delta_4 = 1/2\\
&D_4 = \delta_5 P_{\sigma_5} + 0 \qquad \text{with } \delta_5 = 1.
\end{aligned}
$$

Actually it is not necessary to standardize the remainder at each stage, Repeated substitution gives the representation of D as a convex combination of permutation matrices.

□

3 A Lemma

S:lemma

The proof of Theorem <1> consists of repeated appeals to the following Lemma, which formalizes the idea behind Example <2>.

diag <3> **Lemma.** *For each non-negative matrix D whose row- and column-sums all equal the same strictly positive number there exists a permutation σ for which $D[i, \sigma(i)] > 0$ for $i = 1, \dots, n$.*

Proof Without loss of generality suppose the row and column sums all equal 1, that is, D is doubly stochastic.

Write T for the index set $\{1, \dots, n\}$ for both the rows and the columns of the matrix D. For each $i \in T$ define $C_i = \{j \in T : D[i,j] > 0\}$. We seek a permutation σ for which $\sigma(i) \in C_i$ for all i. The Marriage Lemma (Pollard, 2001, Problem 10.5) states that this is possible iff

$$\# \cup_{i \in I} C_i \geq \# I \qquad \text{for each } I \subseteq T.$$

For a given I write J for $\cup_{i \in I} C_i$. We need to show that $k := \#I \leq \ell := \#J$. Without loss of generality suppose $I = \{1, \dots, k\}$ and $J = \{1, \dots, \ell\}$. (Equivalently, permute rows and columns to bring $I \times J$ to the top left corner.) Then D has the block form,

$$D = \begin{bmatrix} A_{k \times \ell} & 0_{k \times (n-\ell)} \\ B_{(n-k) \times \ell} & C_{(n-k) \times (n-\ell)} \end{bmatrix}$$

The block of zeros appears because $D[i,j] = 0$ if $i \in I$ and $j \notin J := \cup_{i \in I} C_i$.

Write $sum(matrix)$ for the sum of all the elements of a matrix. Each of the k rows of A sums to 1; each of the $(n-\ell)$ columns of C sums to 1; and each of the n rows of D sums to 1. Thus

$$n = sum(D) \geq sum(A) + sum(C) = k + (n - \ell),$$

which rearranges to $\ell \geq k$.

□

4 Proof of the Theorem

S:proof

Algorithm makeBirkhoff:

Input: An $n \times n$ doubly stochastic matrix D
Output: A representation of D as a convex combination of permutation matrices $\theta_1 P_{\sigma_1} + \cdots + \theta_k P_{\sigma_k}$
begin
  $j \leftarrow 1$; $R \leftarrow D$
  while $R \neq 0$ **do**
    find a permutation σ_j for which $\theta_j := \min R[i, \sigma_j(i)] > 0$
    $R \leftarrow R - \theta_j P_{\sigma_j}$
    $j \leftarrow j + 1$

References

PollardUGMTP

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.