A variational proof of Birkhoff’s theorem

Q. J. Zhu
Department of Mathematics
Western Michigan University
Kalamazoo, MI 49008 USA

Supported by National Science Foundation under grant DMS 0102496.
**Doubly stochastic matrices**

An $N \times N$ matrix $A = (a_{nm})$ is *doubly stochastic* if

\[
a_{mn} \geq 0, \ m, n = 1, \ldots, N.
\]

\[
\sum_{n=1}^{N} a_{nm} = 1, \ m = 1, \ldots, N
\]

and

\[
\sum_{m=1}^{N} a_{nm} = 1, \ n = 1, \ldots, N.
\]

Denote the set of $(N \times N)$ doubly stochastic matrices by $\mathcal{A}$ and the set of permutation matrices by $\mathcal{P}$. Then

\[\mathcal{P} \subset \mathcal{A}.\]

Applications: Physics, stochastic process, economics...
Birkhoff Theorem

\[ \mathcal{A} = \text{conv} \ \mathcal{P}. \]
Approximate Fermat Principle

Let \( f : \mathbb{R}^N \rightarrow \mathbb{R} \) be a differentiable function bounded from below. Then, \( \forall \varepsilon > 0, \exists x \in \mathbb{R}^N \) such that
\[
\| f'(x) \| < \varepsilon.
\]

**Proof.** Let \( f(z) < \inf f + \varepsilon/2 \) and take \( x \) to be the minimizer of
\[
f(y) + \frac{\varepsilon}{2} \| y - z \|^2.
\]
Then
\[
f'(x) = -\varepsilon \| x - z \| \cdot \| \cdot \|'(x - z).
\]
The norm of the right hand side is \( \leq 1 \).
A variational proof of Birkhoff’s Theorem

Inclusion $\text{conv } P \subset A$. is easy to check. We show the opposite inclusion and for this we need a combinatorical lemma:

**Lemma 1.** For $A \in A$ there exists $P \in P$, the entries in $A$ corresponding to the 1’s in $P$ are all nonzero.

Let $\mathcal{E}$ be the Euclidean space of all $N \times N$ matrices with inner product

$$\langle A, B \rangle = \text{tr}(B^\top A) = \sum_{n,m=1}^{N} a_{nm}b_{nm}.$$

The key is

**Lemma 2.** Let $A \in A$. Then for any $B \in \mathcal{E}$ there exists $P \in P$ such that

$$\langle B, A - P \rangle \geq 0.$$
Proof. Induction on the number of nonzero elements of $A$. By Lemma 1 there exists $P \in \mathcal{P}$ such that the entries in $A$ corresponding to the 1’s in $P$ are all nonzero. Let $t \in (0, 1)$ be the minimum of these $N$ positive elements. Then we can verify that $A_1 = (A - tP)/(1 - t) \in \mathcal{A}$. Since $A_1$ has at least one less nonzero elements than $A$, by the induction hypothesis there exists $Q \in \mathcal{P}$ such that

$$\langle B, A_1 - Q \rangle \geq 0.$$ 

It follows that

$$\langle B, A - tP - (1 - t)Q \rangle \geq 0$$

and, therefore, at least one of $\langle B, A - P \rangle$ or $\langle B, A - Q \rangle$ is nonnegative. Q.E.D.
Now define \( f : \mathcal{E} \rightarrow \mathcal{R} \) by

\[
f(B) := \ln \left( \sum_{P \in \mathcal{P}} \exp \langle B, A - P \rangle \right).
\]

Then \( f \) is defined for all \( B \in \mathcal{E} \), is differentiable and is bounded from below by 0. By the approximate Fermat principle we can select a sequence \( B_i \in \mathcal{E} \) such that

\[
0 = \lim_{i \to \infty} f'(B_i) = \lim_{i \to \infty} \sum_{P \in \mathcal{P}} \lambda_i^P (A - P).
\]

where

\[
\lambda_i^P = \frac{\exp \langle B_i, A - P \rangle}{\sum_{P \in \mathcal{P}} \exp \langle B_i, A - P \rangle}.
\]

Clearly, \( \lambda_i^P > 0 \) and \( \sum_{P \in \mathcal{P}} \lambda_i^P = 1 \). Thus, taking a subsequence if necessary
we may assume that, for each $P \in \mathcal{P}$, 
\[
\lim_{i \to \infty} \lambda^i_P = \lambda_P \geq 0
\]
and
\[
\sum_{P \in \mathcal{P}} \lambda_P = 1.
\]
Now taking limits as $i \to \infty$ in (1) we have
\[
\sum_{P \in \mathcal{P}} \lambda_P (A - P) = 0.
\]
It follows that $A = \sum_{P \in \mathcal{P}} \lambda_P P$, as was to be shown. Q.E.D.
Majorization

Let $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, we use $x^\downarrow$ to denote the vector by rearranging the components of $x$ in a decreasing order. Recall that $x < y$ ($x$ majorized by $y$), if

$$\sum_{n=1}^{N} x_n = \sum_{n=1}^{N} y_n$$

and, for $k = 1, \ldots, N$,

$$\sum_{n=1}^{k} x_n^\downarrow \leq \sum_{n=1}^{k} y_n^\downarrow.$$
Characterization of Majorization

\( x \prec y \) iff, for any \( z \in \mathbb{R}^N \),

\[ \langle z \downarrow, x \downarrow \rangle \leq \langle z \downarrow, y \downarrow \rangle. \]

**Proof.** Come out of Abel’s formula

\[
\begin{align*}
\langle z \downarrow, y \downarrow \rangle - \langle z \downarrow, x \downarrow \rangle &= \langle z \downarrow, y \downarrow - x \downarrow \rangle \\
&= \sum_{k=1}^{N-1} \left( z_k \downarrow - z_{k+1} \downarrow \right) \cdot \sum_{n=1}^{k} \left( y_n \downarrow - x_n \downarrow \right) \\
&\quad + z_N \downarrow \sum_{n=1}^{N} \left( y_n \downarrow - x_n \downarrow \right).
\end{align*}
\]
Level Sets of Majorization

The level set for \( y \in \mathbb{R}^N \) related to the majorization is \( l(y) := \{ x \in \mathbb{R}^N : x \prec y \} \). We have

\[
l(y) = \text{conv}\{Py : P \in \mathcal{P}\}.
\]

Proof. The inclusion

\[
\text{conv}\{Py : P \in \mathcal{P}\} \subset l(y)
\]

is straightforward. To proof the reversed inclusion, let \( x \prec y \). For any \( z \in \mathbb{R}^N \), choose \( P \in \mathcal{P} \) such that

\[
\langle z, Py \rangle = \langle z^{\downarrow}, y^{\downarrow} \rangle \geq \langle z^{\downarrow}, x^{\downarrow} \rangle \\
\geq \langle z, x \rangle.
\]

(2)
Then, the function
\[ g(z) := \ln \left( \sum_{P \in \mathcal{P}} \exp \langle z, Py - x \rangle \right) \]
is defined for all \( z \in \mathbb{R}^N \), differentiable and bounded from below (by 0). The rest of the proof is the same as that of Birkhoff’s theorem provided before.
Possible Alternative Variational Proof of Birkhoff’s Theorem

Let $C = \text{conv } \mathcal{P}$. Then $C$ is a convex compact set. For any $A \in \mathcal{A}$ let $P_C(A)$ be the projection of $A$ to $C$. Then $P_C(A)$ is characterized by

$$\langle B - P_C(A), A - P_C(A) \rangle \leq 0$$

for all $B \in C$. The proof will be completed if we can deduce $A = P_C(A)$ from the above necessary condition. Many examples verify this conclusion but no proof has been found yet.