Matrix Scaling

Avi Wigderson

February 7, 2012

Matrix scaling is a technique invented by Sinkhorn (1964). It is an analytic technique, rather than a Fourier-analytic one. It arose from numerical analysis.

Uses:
- Linear systems
- MRI’s
- Approximating the Permanent
- LDC’s over the reals
- Discrete Geometry
- Arithmetic Combinatorics
- Quantum computation

We focus on the $L^1$ form of matrix scaling, and restrict to non-negative matrices.

Let $A$ be a non-negative real $n \times n$ matrix.

**Definition.** A matrix $A$ is scalable if there exist real numbers $\alpha_j$, $\beta_i$ such that the matrix $B$ defined by

$$B_{i,j} = \alpha_j \beta_i A_{i,j}$$

is doubly stochastic.

Clearly, the zero matrix isn’t scalable. Neither is

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

A more interesting example: neither is

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

**Theorem 1.** A matrix $A$ is scalable if and only if $\text{per}(A) > 0$. 
Proof. If \( \text{per}(A) = 0 \), then \( A \) cannot be scalable. (Exercise. Hint: Hall’s condition.)

If \( \text{per}(A) > 0 \), we may perform the following process. If \( s_i \) is the sum of column \( i \), then multiply each entry in column \( i \) by \( 1/s_i \). The column-sums of the new matrix are now all 1, but the row sums will not be (in general). Now do the same with the rows: if \( r_i \) is the sum of row \( i \), then multiply each entry in row \( i \) by \( 1/r_i \). But this alters the column-sums, so repeat the above two steps. Repeat. We claim that this process converges.

Why? It turns out that the permanent strictly increases in any two consecutive steps, unless it was doubly stochastic beforehand. Note that after the first step, either the row sums or the column sums are all 1; since \( \text{per}(A) \leq 1 \) if all the column sums of \( A \) are 1, or if the row-sums of \( A \) are all 1 (trivial upper bound), after the first step we have \( \text{per}(A) \leq 1 \). Hence, the permanent is always between 0 and 1 after the first step. In order to prove convergence, we must show that the permanent increases strictly at each step. The permanent is our ‘potential function.’

Suppose \( \text{per}(A) > 0 \). Let \( s_i \) be the sum of column \( i \). Multiply column \( i \) by \( 1/s_i \) to get \( A' \). Suppose \( A' \) has row sums \( r'_i \); multiply row \( i \) by \( 1/r'_i \) to get \( A'' \). Using the AM/GM inequality, we have

\[
\text{per}(A') = \prod_{i=1}^{k} s_i^{-1} \text{per}(A) \geq \text{per}(A)(n/\sum_{i=1}^{n} s_i)^n \geq \text{per}(A);
\]

we have equality above if and only if each \( s_i \) is 1.

Similarly, \( \text{per}(A'') \geq \text{per}(A') \), with equality if and only if each \( r'_i \) is 1. Hence, \( \text{per}(A'') \geq \text{per}(A) \), with equality if and only if \( A \) is doubly stochastic. \( \square \)