Minkowski's Inequality

Minkowski Theorem. Suppose that \( r \) is finite, and \( \hat{a}_1, \ldots, \hat{a}_m \) are vectors in \( \mathbb{R}^{n^*} \) (as usual, they have non-negative entries).

\[
M_r (\hat{a}_1) + \ldots + M_r (\hat{a}_m) \geq M_r (\hat{a}_1 + \ldots + \hat{a}_m) \quad (r > 1)
\]
\[
M_r (\hat{a}_1) + \ldots + M_r (\hat{a}_m) = M_r (\hat{a}_1 + \ldots + \hat{a}_m) \quad (r = 1)
\]
\[
M_r (\hat{a}_1) + \ldots + M_r (\hat{a}_m) \leq M_r (\hat{a}_1 + \ldots + \hat{a}_m) \quad (r < 1)
\]

with equality (for \( r \neq 1 \)) only if (the \( \hat{a}_i \) are linearly dependent or \( r \leq 0 \) and for some \( j \) \( a_{2j} = a_{3j} = \ldots = a_{mj} = 0 \)).

Proof. Suppose wlog \( \Sigma p_i = 1 \), and let \( \hat{a}_1 + \ldots + \hat{a}_m = \hat{s} \), while \( M_r (\hat{s}) = S \).

Now
\[
S^r = \Sigma p_i s_i^r = \Sigma p_i s_i^{r-1} + \Sigma p_i a_{2i} s_i^{r-1} + \ldots + \Sigma p_i a_{mi} s_i^{r-1}
\]
\[
= \Sigma (p_i a_{2i}) (p_i s_i^{r-1}) + \ldots + \Sigma (p_i a_{mi}) (p_i s_i^{r-1})
\]
Each of these looks set up for Hölder. Assume \( r > 1 \). Then the conjugate \( r' = \frac{r}{r-1} \). Hölder says

\[
\sum a_i b_i < \left( \sum a_i^r \right)^\frac{1}{r} \left( \sum b_i^{r'} \right)^\frac{1}{r'}
\]

Applying this to the general term above,

\[
\sum (p_i^{\frac{1}{r}} a_{ki}) (p_i^{\frac{1}{r}} s_i)^{r-1} < \left( \sum p_i a_{ki}^r \right)^\frac{1}{r} \left( \sum p_i s_i^{r'} \right)^\frac{1}{r'}
\]

\[
= \left( \sum p_i a_{ki}^r \right)^\frac{1}{r} \left( \sum p_i s_i^{r'} \right)^\frac{1}{r'}
\]

\[
= M_r (\hat{a}_k) \left( \sum p_i s_i^r \right)^\frac{r-1}{r}
\]

\[
= M_r (\hat{a}_k) S^{r-1}
\]

So we have

\[
S^r < M_r (\hat{a}_1) S^{r-1} + \ldots + M_r (\hat{a}_m) S^{r-1}
\]

and dividing by \( S^{r-1} \) completes the proof.
The reversed sign case (r<1) is obtained from the r<1 case of Hölder:

$$\sum a_i b_i > (\sum a_i^r)^{\frac{1}{r}} (\sum b_i^r)^{\frac{1}{r}}$$

in a similar way. (see p.31 of H-L-P).

Here's a lovely form of Minkowaski's inequality! Suppose you have a matrix

$$
\begin{pmatrix}
    a_{11} & a_{1m} \\
    \vdots & \vdots \\
    a_{m1} & a_{mn}
\end{pmatrix}
$$

and you want to define a norm by taking the $M_r$-norm of each column and then the $M_s$-norm of the resulting $n$-vector. How would that compare to taking the $M_s$-norm of each row and then the $M_r$-norm of the resulting $m$-vector?
Theorem. (Minkowski, found by Ingraham, Jensen)

Let $M^{(i)}_r$ denote a mean taken by summing over columns in a matrix $A$ containing $a_{ij}$, and $M^{(j)}_r$ denote a mean taken by summing over rows.

If $A = (a_{ij})$ is $m \times n$, an $M^{(i)}_r$ mean has $m$ weights $p_1, \ldots, p_m$ and an $M^{(j)}_r$ mean has $n$ completely different weights $q_1, \ldots, q_n$.

(Minkowski)

Theorem. If $0 < r < s < \infty$, then

$$M^{(j)}_s M^{(i)}_r (A) \leq M^{(i)}_r M^{(j)}_s (A)$$

with equality only if $A$ is the rank-1 matrix formed by the outer product

$$A = BC$$

of a column vector $B$ and a row vector $C$. 

\[ m \times n \quad m \times 1 \quad 1 \times n \]
Proof. Let $s/r = K > 1$, and define $B_{ij} = p_i a_{ij}^r$. Then the statement is

$$
\left( \sum_j q_j \left( \sum_i p_i a_{ij}^r \right)^{s/r} \right)^{1/s} \leq \left( \sum_i p_i \left( \sum_j q_j a_{ij}^s \right)^{r/s} \right)^{1/r}
$$

Making the substitutions above, we have

$$
\left( \sum_j q_j \left( \sum_i B_{ij}^k \right)^{1/k} \right)^{1/k} \leq \left( \sum_i \left( \sum_j q_j \left( p_i a_{ij}^r \right)^{s/r} \right)^{1/k} \right)^{1/K} \leq \left( \sum_i \left( \sum_j q_j B_{ij}^k \right)^{1/k} \right)^{1/K} \leq \left( \sum_i \left( \sum_j q_j B_{ij}^k \right)^{1/K} \right)^{1/K}
$$

Raising both sides to the $r$-th power ($r > 0$, so this doesn't reverse the inequality), we must show

$$
\left( \sum_j q_j \left( \sum_i B_{ij}^k \right)^{1/k} \right)^{1/k} \leq \left( \sum_i \left( \sum_j q_j B_{ij}^k \right)^{1/K} \right)^{1/K}
$$
By Minkowski, we assume $\sum q_j = 1$ since $k > 1$.

$$M_k(\hat{B}_1) + \ldots + M_k(\hat{B}_m) \geq M_k(\hat{B}_1 + \ldots + \hat{B}_m)$$

for any collection of $m$ vectors $\hat{B}_i$ in $\mathbb{R}^{n \times n}$, such as the rows $B_1, \ldots, B_m$ of $B$. The LHS here is the RHS above as long as we assume (as usual) that $\sum q_j = 1$.

The LHS above

$$\left( \sum_{j=1}^{n} q_j \left( \sum_{i=1}^{m} B_{ij} \right)^k \right)^{\frac{1}{k}} = M_k(\hat{B}_1 + \ldots + \hat{B}_m).$$

as desired. □

We now specialize Minkowski to the case when all the weights $p_i = p$. 
Then \( M_r(\hat{a}) = (\sum p_i \hat{a}_i)^{\frac{1}{r}} = p^{\frac{1}{r}} (\sum \hat{a}_i)^{\frac{1}{r}} \) and

**Proposition (Minkowski’s Inequality)**

\[
(\sum_{i=1} a_{1i}^r)^{\frac{1}{r}} + \ldots + (\sum_{i=1} a_{mi}^r)^{\frac{1}{r}} \geq \left( \sum_{i=1} (a_{1i} + \ldots + a_{mi})^r \right)^{\frac{1}{r}} \quad (r \geq 1)
\]

\[
(\sum_{i=1} a_{1i}^r)^{\frac{1}{r}} + \ldots + (\sum_{i=1} a_{mi}^r)^{\frac{1}{r}} = \left( \sum_{i=1} (a_{1i} + \ldots + a_{mi})^r \right)^{\frac{1}{r}} \quad (r = 1)
\]

\[
(\sum_{i=1} a_{1i}^r)^{\frac{1}{r}} + \ldots + (\sum_{i=1} a_{mi}^r)^{\frac{1}{r}} \leq \left( \sum_{i=1} (a_{1i} + \ldots + a_{mi})^r \right)^{\frac{1}{r}} \quad (r \leq 1)
\]

as long as the \( \hat{a}_1, \ldots, \hat{a}_m \) are non-negative vectors in \( IR^n \). There is equality only if (all \( \hat{a}_i \) are linearly dependent or (\( r < 0 \) and for some \( j^* \), all \( \hat{a}_{ij^*} = 0 \)).

We can give this a nice geometric interpretation. We are suppose we define distance in \( IR^n \) by

\[
|| \hat{x} - \hat{y} ||_r = (|x_1 - y_1|^r + \ldots + |x_n - y_n|^r)^{\frac{1}{r}} \quad (r \geq 1)
\]

Note the absolute values and the fact that \( r = 2 \) is the usual distance!
Then, 

\[
\| \hat{x} - \hat{y} \|_r + \| \hat{y} - \hat{z} \|_r \geq (\sum_i (|x_i - y_i| + |y_i - z_i|)^r)^{1/r}
\]

\[
\geq (\sum_i |x_i - z_i|^r)^{1/r}
\]

by the triangle inequality \( |x_i - y_i| + |y_i - z_i| \geq |x_i - y_i + y_i - z_i| \).

\[
= \| \hat{x} - \hat{z} \|_r
\]

so Minkowski's inequality proves a generalized triangle inequality for all these distances!

We're now going to work out some (more) consequences of our main theorems: the arithmetic-geometric mean inequality, Hölder, and Minkowski.
Suppose we have a parallelepiped in $\mathbb{R}^n$ with side lengths $a_1, \ldots, a_n$. It certainly seems reasonable to conjecture that the volume is maximized when the edges are orthogonal. But how to prove it?

Proposition (Hadamard) If $A$ is an $n \times n$ matrix, 
\[
\det(A)^2 \leq \left( \sum_{i=1}^{n} a_{i1}^2 \right) \cdots \left( \sum_{i=1}^{n} a_{in}^2 \right)
\]
with equality only if (the row vectors $\hat{a}_k$ are all orthogonal or some row vector $\hat{a}_k = \hat{0}$).

Proof. Suppose we have a positive-definite quadratic form $C$, with all diagonal $c_{ii} > 0$.

n.b. This is a symmetric matrix $C$ so that
\[
\langle C \hat{x}, \hat{x} \rangle \geq 0 \text{ for all } \hat{x} \text{ with equality only if } \hat{x} = \hat{0}.
\]
The eigenvalues of $C$ are all $\geq 0$, call them $\lambda_1, \ldots, \lambda_n$. Now

$$(\det C)^* = (\lambda_1 \cdots \lambda_n)^* = G(\lambda)^n \leq U(\lambda)^n = \left(\frac{\lambda_1 + \cdots + \lambda_n}{n}\right)^n$$

But $\lambda_1 + \cdots + \lambda_n$ is the trace of the matrix $C$, which is also the sum of the diagonal entries $c_{11}, \ldots, c_{nn}$, so

$$\det C \leq \left(\frac{c_{11} + \cdots + c_{nn}}{n}\right)^n$$

Now we're going to do something clever. Recall that if you multiply a row or column of a matrix by $K$, we scale the determinant by $K$. So we are going to build a new symmetric matrix $D$

$$d_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}$$
Notice that

$$\det D = \frac{1}{\sqrt{c_{11}} \cdot \sqrt{c_{nn}}} \cdot \frac{1}{\sqrt{c_{22}} \cdot \sqrt{c_{nn}}} \cdot \ldots \cdot \frac{1}{\sqrt{c_{mm}} \cdot \sqrt{c_{nn}}} \det C$$

rows columns

$$= \frac{1}{c_{11} \cdot \ldots \cdot c_{nn}} \det C$$

Notice also that $d_{ii} = \frac{c_{ii}}{\sqrt{c_{ii}} \cdot \sqrt{c_{ii}}} = 1$. Further,

$$\langle D \hat{x}, \hat{x} \rangle = \sum_{ij} \frac{c_{ij}}{\sqrt{c_{ii}} \cdot \sqrt{c_{jj}}} x_i x_j = \sum_{ij} c_{ij} \frac{x_i}{\sqrt{c_{ii}}} \frac{x_j}{\sqrt{c_{jj}}}$$

$$= \langle C \hat{y}, \hat{y} \rangle \quad \text{(where } y_i = \frac{x_i}{\sqrt{c_{ii}}})$$

$$> 0 \quad \text{(because } C \text{ was positive definite).}$$

So $D$ is positive-definite and (applying our last inequality),

$$\det D \leq \left( \frac{d_{11} + \ldots + d_{nn}}{n} \right)^n$$

or

$$\frac{\det C}{c_{11} \cdot \ldots \cdot c_{nn}} \leq 1 \implies \det C \leq c_{11} \cdot \ldots \cdot c_{nn}$$
(Pause to reflect: Isn't that awesome? We knew $\lambda_1 + \ldots + \lambda_n = c_1 + \ldots + c_n$. But did you know $\lambda_1 \ldots \lambda_n \leq c_1 \ldots c_n$?)

So now we return to our original matrix $A$. Consider the quadratic form

$$\sum_{i} (a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{ini} x_n)^2 = \langle A\hat{x}, A\hat{x} \rangle$$

$$\Leftrightarrow \sum_{i} = \langle A^* A \hat{x}, \hat{x} \rangle.$$ 

Certain facts about the symmetric matrix $A^* A^T$ are well known: the eigenvalues of $A^* A^T$ are the squares of the eigenvalues of $A$. So $A^* A^T$ is a positive-definite quadratic form or $A$ has a zero eigenvalue and $\det A = 0$. Also $\det A^* A^T = (\det A)^2$. 

\[ \boxed{\text{(Q.E.D.)}} \]
Now our previous theorem estimate applies to $A^*A^T$ and

$$(\det A^*A^T)^2 = \det A^*A^T \leq A^*A^T_{i1} \cdots A^*A^T_{in}$$

which is exactly the statement we wanted as

$$(A^*A^T)_{ii} = \sum_j a_{ij}a_{ji} = \sum_j a_{ij}^2 \geq \lambda_i$$

Now we have equality in $G(\lambda) \leq U(\lambda)$ (and hence in $\det AA^T \leq AA^T_{i1} \cdots AA^T_{in}$) only if all the $\lambda_i$ are equal. In this case the symmetric matrix $AA^T$ must be equal to $\lambda I$. (In the basis of eigenvectors of $AA^T$, this is obvious. But if you're a multiple of $I$ in one orthonormal basis, you're a multiple of $I$ in all of them.)
Now we backtrack. We have shown our theorem when $A$ has no zero eigenvalues. When $A$ has a zero eigenvalue,

$$(\det A)^2 = 0 \leq (\sum_i a_{ii}^2) \cdots (\sum_i a_{ni})$$

b/c the rhs is clearly $\geq 0$, and there is equality only if the rhs = 0, which can happen only if one of the $(\sum_i a_{ki})^2 = 0$. 