Minkowski's Inequality

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

$$M_r(\vec{a}_1) + \dots + M_r(\vec{a}_m) \geqslant M_r(\vec{a}_1 + \dots + \vec{a}_m) \quad (r > 1)$$

$$M_r(\vec{a}_1) + \dots + M_r(\vec{a}_m) = M_r(\vec{a}_1 + \dots + \vec{a}_m) \quad (r = 1)$$

$$M_r(\vec{a}_1) + \dots + M_r(\vec{a}_m) \leqslant M_r(\vec{a}_1 + \dots + \vec{a}_m) \quad (r < 1)$$

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

Proof. Suppose wlog $\sum p_i = 1$, and let

$$\vec{a}_1 + \dots + \vec{a}_m = \vec{s}, \quad \text{while } M_r(\vec{s}) = S.$$

Now

$$S^r = \sum p_i s_i^r = \sum p_i s_i s_i^{r-1} = \sum p_i a_{1i} s_i^{r-1} + \dots + \sum p_i a_{mi} s_i^{r-1}$$

$$= \sum (p_i^{1/r} a_{1i})(p_i^{1/r} s_i)^{r-1} + \dots + \sum (p_i^{1/r} a_{mi})(p_i^{1/r} s_i)^{r-1}$$

Each of these looks set up for Hölder.

Assume $r > 1$. Then the conjugate $r' = \frac{r}{r-1}$. Hölder says

$$\sum a_i b_i < \left(\sum a_i^r\right)^{1/r} \left(\sum b_i^{r'}\right)^{1/r'}$$

Applying this to the general term above,

$$\sum \left(p_i^{1/r} a_{ki}\right)\left(p_i^{1/r} s_i\right)^{r-1} < \left(\sum p_i a_{ki}^r\right)^{1/r} \left(\sum \left(p_i^{1/r} s_i\right)^{(r-1)r'}\right)^{1/r'}$$

$$= \left(\sum p_i a_{ki}^r\right)^{1/r} \left(\sum p_i s_i^r\right)^{1/r'}$$

$$= M_r(\vec{a}_k) \left(\sum p_i s_i^r\right)^{\frac{r-1}{r}}$$

$$= M_r(\vec{a}_k)\, S^{r-1}$$

So we have

$$S^r < M_r(\vec{a}_1) S^{r-1} + \dots + M_r(\vec{a}_m) S^{r-1}$$

and dividing by S^{r-1} completes the proof.

The reversed sign case ($r<1$) is obtained from the $r<1$ case of Hölder:

$$\sum a_i b_i > \left(\sum a_i^r\right)^{1/r} \left(\sum b_i^{r'}\right)^{1/r'}$$

in a similar way. (see p31 of H-L-P). $\square$

Here's a lovely form of Minkowski's inequality! Suppose you have a (≥ 0) matrix

$$\begin{pmatrix} a_{11} & & a_{1n} \\ & \ddots & \\ a_{m1} & & a_{mn} \end{pmatrix}$$

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 $M_r^{(i)}$ denote a mean taken by summing over <u>columns</u> in a matrix A containing a_{ij}, and $M_r^{(j)}$ denote a mean taken by summing over rows.

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

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

$$M_s^{(j)} M_r^{(i)}(A) \leq M_r^{(i)} M_s^{(j)}(A)$$

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

$$\underset{m\times n}{A} = \underset{m\times 1}{B}\,\underset{1\times n}{C}$$ of a column vector & a row vector.

Proof. Let $s/r = K > 1$, and define

$B_{ij} = p_i a_{ij}^r$. Then the statement is

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

Making the substitutions above, we have

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

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

$$\le \left(\sum_i \left(\sum_j q_j B_{ij}^{K=s/r}\right)^{1/K}\right)^{1/r}$$

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

$$\left(\sum_j q_j \left(\sum_i B_{ij}\right)^K\right)^{1/K} \le \sum_i \left(\sum_j q_j B_{ij}^K\right)^{1/K}$$

By Minkowski, ~~if we assume $\sum q_i = 1$~~ since $k>1$

$$M_k(\vec{B}_1) + \dots + M_k(\vec{B}_m) \geq M_k(\vec{B}_1 + \dots + \vec{B}_m)$$

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

The lhs above

$$\left(\sum_{j=1}^{n} q_j \Big(\underbrace{\sum_{i=1}^{m} B_{ij}}_{\text{column sum}} \Big)^{k} \right)^{1/k} = M_k(\vec{B}_1 + \dots + \vec{B}_m).$$

as desired. $\square$

We now specialize Minkowski to the case when all the weights $p_i = p$.

Then $M_r(\vec{a}) = \left(\sum_i p a_i^r\right)^{1/r} = p^{1/r}\left(\sum_i a_i^r\right)^{1/r}$, and

Proposition (Minkowski's Inequality)

$$\left(\sum_i a_{1i}^r\right)^{1/r} + \dots + \left(\sum_i a_{mi}^r\right)^{1/r} \geq \left(\sum_i (a_{1i}+\dots+a_{mi})^r\right)^{1/r} \quad (r>1)$$

$$\left(\sum_i a_{1i}^r\right)^{1/r} + \dots + \left(\sum_i a_{mi}^r\right)^{1/r} = \left(\sum_i (a_{1i}+\dots+a_{mi})^r\right)^{1/r} \quad (r=1)$$

$$\left(\sum_i a_{1i}^r\right)^{1/r} + \dots + \left(\sum_i a_{mi}^r\right)^{1/r} \leq \left(\sum_i (a_{1i}+\dots+a_{mi})^r\right)^{1/r} \quad (r<1)$$

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

We can give this a nice geometric interpretation. ~~We are~~ Suppose we define r-distance in $\mathbb{R}^n$ by

$$\|\vec{x}-\vec{y}\|_r = \left(|x_1-y_1|^r + \dots + |x_n-y_n|^r\right)^{1/r}, \quad (r>1)$$

Note the absolute values and the fact that $r=2$ is the usual distance!

Then

Minkowski applied to $|\vec{x}-\vec{y}|, |\vec{y}-\vec{z}|$.

$$\|\vec{x}-\vec{y}\|_r + \|\vec{y}-\vec{z}\|_r \geq \left(\sum_i \left(|x_i-y_i|+|y_i-z_i|\right)^r\right)^{1/r}$$

$$\geq \left(\sum_i |x_i-z_i|^r\right)^{1/r}$$

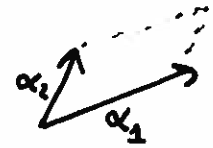
triangle inequality $|x_i-y_i|+|y_i-z_i| \geq |x_i-y_i+y_i-z_i|$.

$$= \|\vec{x}-\vec{z}\|_r$$

So Minkowski's inequality ~~is a~~ 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.

Ex.

Suppose we have a parallelepiped 

in $\mathbb{R}^n$ with side lengths $\alpha_1, \ldots, \alpha_n$. 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,

$$\det(A)^2 \leq \left(\sum_i a_{1i}^2\right) \cdots \left(\sum_i a_{ni}^2\right)$$

with equality only if (the row vectors $\vec{a}_k$ are all orthogonal <u>or</u> some row vector $\vec{a}_k = \vec{0}$).

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

$\langle C\vec{x}, \vec{x} \rangle \geq 0$ for all $\vec{x}$ with equality only if $\vec{x} = \vec{0}$.

The eigenvalues of C are all ≥ 0, call them $\lambda_1, \ldots, \lambda_n$. Now

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

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

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

Now we're going to do something clever. Recall that if the 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

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

Notice that

$$\det D = \underbrace{\frac{1}{\sqrt{c_{11}}} \cdots \frac{1}{\sqrt{c_{nn}}}}_{\text{rows}} \cdot \underbrace{\frac{1}{\sqrt{c_{11}}} \cdots \frac{1}{\sqrt{c_{nn}}}}_{\text{columns}} \det C$$

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

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

$$\langle D\vec{x}, \vec{x} \rangle = \sum_{i,j} \frac{c_{ij}}{\sqrt{c_{ii}}\sqrt{c_{jj}}} x_i x_j = \sum_{i,j} c_{ij} \frac{x_i}{\sqrt{c_{ii}}} \frac{x_j}{\sqrt{c_{jj}}}$$

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

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

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

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

or

$$\frac{\det C}{c_{11} \cdots c_{nn}} \leq 1 \implies \det C \leq c_{11} \cdots c_{nn}$$

(Pause to reflect: Isn't that awesome? We knew $\lambda_1 + \ldots + \lambda_n = c_1 + \ldots + c_n$. But did you know $\lambda_1 \cdots \lambda_n \leq c_1 \cdots c_n$?)

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

$$\sum_i (a_{1i} x_1 + a_{2i} x_2 + \ldots + a_{ni} x_n)^2 = \langle A^T \vec{x}, A^T \vec{x} \rangle$$

$$= \langle A^* A^T \vec{x}, \vec{x} \rangle.$$

Certain facts about the symmetric matrix $A^* A^T$ are well known: the eigenvalues of

~~The eigenvectors are the eigen~~

$A^* A^T$ are the squares of the eigenvalues of A. So $A^* A^T$ is a positive-definite quadratic form <u>or</u> A has a zero eigenvalue and $\det A = 0$. Also $\det A^* A^T = (\det A)^2$.

Now our previous ~~theorem~~ estimate applies to AA^T and

$$(\det A)^2 = \det AA^T \le AA^T_{11} \cdot \ldots \cdot AA^T_{nn}$$

which is exactly the statement we wanted as

$$(AA^T)_{ii} = \sum_j a_{ij} a^T_{ji} = \sum_j a_{ij}^2 .$$

Now we have equality in $G(\vec{\lambda}) \le U(\vec{\lambda})$ (and hence in $\det AA^T \le AA^T_{11} \cdot \ldots \cdot AA^T_{nn}$) only if all the λ_i are equal. In this case the symmetric matrix AA^T ~~has all~~ must be equal to λI. (In the orthonomal 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,

$$(\det A)^2 = 0 \leq \left(\sum_i a_{1i}^2\right) \cdots \left(\sum_i a_{ni}^2\right)$$

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