CSCE 790S
Background Results

Stephen A. Fenner

January 17, 2020

Abstract

These results are background to the course CSCE 790S/CSCE 790B, Quantum Computation and Information (Spring 2007 and Fall 2011). Each result, or group of related results, is roughly one page long.

Contents

1 The Cauchy-Schwarz Inequality

This is one of the most versatile inequalities in all of mathematics.

Theorem 1.1 (Cauchy-Schwarz) *For any real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$,*

$$|a_1 b_1 + \cdots + a_n b_n| \leq \sqrt{(a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2)}\ , \tag{1}$$

with equality holding iff the two vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are linearly dependent.

Proof. There are many, many ways of proving this. Here is a direct calculation. We have,

$$\begin{aligned}
0 &\leq \sum_{1 \leq i < j \leq n} (a_i b_j - a_j b_i)^2 = \sum_{i<j} [a_i b_j (a_i b_j - a_j b_i) - a_j b_i (a_i b_j - a_j b_i)] \\
&= \sum_{i<j} [a_i b_j (a_i b_j - a_j b_i) + a_j b_i (a_j b_i - a_i b_j)] = \sum_{i<j} a_i b_j (a_i b_j - a_j b_i) + \sum_{i<j} a_j b_i (a_j b_i - a_i b_j) \\
&= \sum_{i<j} a_i b_j (a_i b_j - a_j b_i) + \sum_{j<i} a_i b_j (a_i b_j - a_j b_i) = \sum_{i \neq j} a_i b_j (a_i b_j - a_j b_i) = \sum_{i,j} a_i b_j (a_i b_j - a_j b_i) \\
&= \sum_{i,j} a_i^2 b_j^2 - \sum_{i,j} a_i b_i a_j b_j = \left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{j=1}^n b_j^2 \right) - \left(\sum_{i=1}^n a_i b_i \right)^2 .
\end{aligned}$$

Adding $\left(\sum_i a_i b_i\right)^2$ to both sides then taking the square root of both sides (noting that the square root function is strictly monotone increasing) yields the inequality (1). Clearly, equality holds above iff $a_i b_j - a_j b_i = 0$ for all $i < j$, or equivalently, $a_i b_j = a_j b_i$ for all $i < j$. It is not hard to check that this condition is equivalent to $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ being linearly dependent. □

Note that (1) still holds if we remove the absolute value delimiters from the left-hand side. In that case, equality holds iff there exists a $\lambda \geq 0$ such that either $(a_1, \ldots, a_n) = \lambda(b_1, \ldots, b_n)$ or $(b_1, \ldots, b_n) = \lambda(a_1, \ldots, a_n)$.

Corollary 1.2 (Triangle Inequality for Complex Numbers) *For any $z, w \in \mathbb{C}$, $|z + w| \leq |z| + |w|$.*

Proof. Writing $z = a_1 + a_2 i$ and $w = b_1 + b_2 i$ for real a_1, a_2, b_1, b_2, we have

$$\begin{aligned}
|z + w|^2 &= (a_1 + b_1)^2 + (a_2 + b_2)^2 = a_1^2 + a_2^2 + b_1^2 + b_2^2 + 2(a_1 b_1 + a_2 b_2) \\
&\leq a_1^2 + a_2^2 + b_1^2 + b_2^2 + 2\sqrt{(a_1^2 + a_2^2)(b_1^2 + b_2^2)} = \left(\sqrt{a_1^2 + a_2^2} + \sqrt{b_1^2 + b_2^2} \right)^2 = (|z| + |w|)^2 .
\end{aligned}$$

Taking the square root of both sides yields the corollary. □

Corollary 1.3 *For any complex numbers $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$,*

$$|z_1^* w_1 + \cdots + z_n^* w_n| \leq \sqrt{(|z_1|^2 + \cdots + |z_n|^2)(|w_1|^2 + \cdots + |w_n|^2)}\ . \tag{2}$$

Proof. We have

$$\begin{aligned}
|z_1^* w_1 + \cdots + z_n^* w_n| &\leq |z_1^* w_1| + \cdots + |z_n^* w_n| && \text{(by Corollary 1.2)} \\
&= |z_1||w_1| + \cdots + |z_n||w_n| \\
&\leq \sqrt{(|z_1|^2 + \cdots + |z_n|^2)(|w_1|^2 + \cdots + |w_n|^2)}\ . && \text{(by Theorem 1.1)}
\end{aligned}$$

□

Corollary 1.4 *For any column vectors $u, v \in \mathbb{C}^n$,*

$$|\langle u, v \rangle| \leq \|u\| \|v\|\ .$$

2 The Schur Triangular Form and the Spectral Theorem

Theorem 2.1 (Schur Triangular Form) *For every $n \times n$ matrix M, there exists a unitary U and an upper triangular T (both $n \times n$ matrices) such that $M = UTU^*$.*

Proof. We prove this by induction on n. The $n = 1$ case is trivial. Now supposing the theorem holds for $n \geq 1$, we prove it holds for $n + 1$. Let M be any $(n + 1) \times (n + 1)$ matrix. We let A be the linear operator on $\mathbb{C}^{n+1}$ whose matrix is M with respect to some orthonormal basis. A has some eigenvalue λ with corresponding unit eigenvector v. Using the Gram-Schmidt procedure, we can find an orthonormal basis $\{y_1, \ldots, y_{n+1}\}$ for $\mathbb{C}^{n+1}$ such that $y_1 = v$. With respect to this basis, the matrix for A looks like

$$N = \left[\begin{array}{c|c} \lambda & w^* \\ \hline 0 & N' \end{array}\right],$$

where w is some vector in $\mathbb{C}^n$ and N' is an $n \times n$ matrix. Since M and N represent the same operator with respect to different orthonormal bases, they must be unitarily conjugate, i.e., there is a unitary V such that $M = VNV^*$. N' is an $n \times n$ matrix, so we apply the inductive hypothesis to get a unitary W' and an upper triangular T' (both $n \times n$ matrices) such that $N' = W'T'W'^*$. Now we can factor N:

$$N = \left[\begin{array}{c|c} \lambda & w^* \\ \hline 0 & W'T'W'^* \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & W' \end{array}\right] \left[\begin{array}{c|c} \lambda & w^*W' \\ \hline 0 & T' \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & W'^* \end{array}\right] = WTW^*,$$

where

$$W = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & W' \end{array}\right] \qquad \text{and} \qquad T = \left[\begin{array}{c|c} \lambda & w^*W' \\ \hline 0 & T' \end{array}\right].$$

T is clearly upper triangular, and it's easily checked that $WW^* = I$, using the fact that W' is unitary. Thus W is unitary, and we get $M = VNV^* = VWTW^*V^* = UTU^*$, where $U = VW$ is unitary. □

A *Schur basis* for an operator A is an orthonormal basis that gives an upper triangular matrix for A.

Theorem 2.2 *If an $n \times n$ matrix A is both upper triangular and normal, then A is diagonal.*

Proof. Suppose that A is upper triangular and normal, but not diagonal. Then there is some $i < j$ such that $[A]_{ij} \neq 0$. Let j be *least* such that there exists $i < j$ such that $[A]_{ij} \neq 0$. For this i and j, we get

$$[AA^*]_{ii} = \sum_{k=1}^{n} [A]_{ik}[A^*]_{ki} = \sum_{k=1}^{n} [A]_{ik}[A]_{ik}^* = \sum_{k=1}^{n} |[A]_{ik}|^2 = \sum_{k=i}^{n} |[A]_{ik}|^2 \geq |[A]_{ii}|^2 + |[A]_{ij}|^2 > |[A]_{ii}|^2. \tag{3}$$

The last inequality follows from the fact that $[A]_{ij} \neq 0$. Similarly,

$$[A^*A]_{ii} = \sum_{k=1}^{n} [A]_{ki}^*[A]_{ki} = \sum_{k=1}^{n} |[A]_{ki}|^2 = \sum_{k=1}^{i} |[A]_{ki}|^2 = |[A]_{ii}|^2. \tag{4}$$

The next to last equation holds because A is upper triangular, and the last equation holds because of our minimum choice of j and the fact that $i < j$. From (3) and (4), we have $[AA^*]_{ii} > [A^*A]_{ii}$. But A is normal, so these two quantities must be equal. From this contradiction we get that A must be diagonal. □

Corollary 2.3 (Spectral Theorem for Normal Operators) *Every normal matrix is unitarily conjugate to a diagonal matrix. Equivalently, every normal operator has an orthonormal eigenbasis.*

3 The Polar and Singular Value Decompositions

Theorem 3.1 (Polar Decomposition) *For every $n \times n$ matrix A there are is an $n \times n$ unitary matrix U and a unique $n \times n$ matrix H such that $H \geq 0$ and $A = UH$. In fact, $H = |A|$.*

Proof. First uniqueness. If $A = UH$ with U unitary and $H \geq 0$, then

$$|A| = \sqrt{A^*A} = \sqrt{H^*U^*UH} = \sqrt{H^*H} = |H| = H.$$

Now existence. Let $\{e_1, \ldots, e_n\}$ be the standard orthonormal basis for $\mathbb{C}^n$. We first prove the special case where $|A|$ is the diagonal matrix $\operatorname{diag}(s_1, s_2, \ldots, s_n)$ for some real values $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$. Let $0 \leq k \leq n$ be largest such that $s_k > 0$ ($k = 0$ if $|A| = 0$). Thus we have

$$|A| = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right],$$

where D is the $k \times k$ nonsingular matrix $\operatorname{diag}(s_1, \ldots, s_k)$. If $j > k$, then $|A|e_j = 0$, and thus $0 = |A|e_j = |A|^2 e_j = A^*Ae_j$, whence $\|Ae_j\|^2 = \langle Ae_j, Ae_j\rangle = \langle e_j, A^*Ae_j\rangle = \langle e_j, 0\rangle = 0$, and so $Ae_j = 0$. This means that $A = \left[\begin{array}{c|c} B & 0 \end{array}\right]$, where B is some $n \times k$ matrix, and the last $n - k$ columns of A are 0. We have

$$\left[\begin{array}{c|c} B^*B & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c} B^* \\ \hline 0 \end{array}\right] \left[\begin{array}{c|c} B & 0 \end{array}\right] = A^*A = |A|^2 = \left[\begin{array}{c|c} D^2 & 0 \\ \hline 0 & 0 \end{array}\right],$$

and so $B^*B = D^2$. Let W be an $n \times (n-k)$ matrix whose columns are unit vectors orthogonal to all the columns of B and to each other. (There are many possibilities for W if $k < n$; the columns of W can be any orthonormal set in the orthogonal complement of the space spanned by the columns of B.) By our choice of W, we have $B^*W = 0$, $W^*B = 0$, and $W^*W = I$. Finally, define $U := \left[\begin{array}{c|c} BD^{-1} & W \end{array}\right]$. We claim that U is unitary and that $A = U|A|$. Noting that D^{-1} is Hermitean, we have

$$U^*U = \left[\begin{array}{c} D^{-1}B^* \\ \hline W^* \end{array}\right] \left[\begin{array}{c|c} BD^{-1} & W \end{array}\right] = \left[\begin{array}{c|c} D^{-1}B^*BD^{-1} & D^{-1}B^*W \\ \hline W^*BD^{-1} & W^*W \end{array}\right] = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array}\right] = I,$$

and therefore U is unitary. We also have

$$U|A| = \left[\begin{array}{c|c} BD^{-1} & W \end{array}\right] \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} B & 0 \end{array}\right] = A.$$

Now for the general case. Since $|A| \geq 0$ (and hence normal), there is a unitary V such that $V|A|V^* = \operatorname{diag}(s_1, \ldots, s_n)$ for some real values $s_1 \geq \cdots \geq s_n \geq 0$. Since

$$V|A|V^* = V\sqrt{A^*A}V^* = \sqrt{VA^*AV^*} = \sqrt{(VAV^*)^*(VAV^*)} = |VAV^*|,$$

we see that VAV^* satisfies the special case, above, and so there is a unitary U such that $VAV^* = U|VAV^*|$. It follows that

$$A = V^*VAV^*V = V^*U|VAV^*|V = V^*UV|A|V^*V = V^*UV|A|,$$

which proves the theorem because V^*UV is unitary. □

Theorem 3.2 (Singular Value Decomposition) *For any $n \times n$ matrix A there exist $n \times n$ unitary matrices V, W and unique real values $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ such that $A = VDW$, where $D = \operatorname{diag}(s_1, \ldots, s_n)$. Furthermore, $s_1, \ldots, s_n$ are the eigenvalues of $|A|$.*

The $s_1, \ldots, s_n$ are known as the *singular values* of A.

Proof. For uniqueness, if $A = VDW$ as above, then

$$|A| = \sqrt{A^*A} = \sqrt{W^*DV^*VDW} = \sqrt{W^*D^2W} = W^*\sqrt{D^2}W = W^*DW,$$

and so the diagonal entries of D must be the eigenvalues of $|A|$. For existence, the Polar Decomposition gives a unitary U such that $A = U|A|$. Since $|A| \geq 0$ (and hence is normal), there exists a unitary Y such that $|A| = YDY^*$, where $D = \operatorname{diag}(s_1, \ldots, s_n)$ for some $s_1 \geq \cdots \geq s_n \geq 0$. Then $A = U|A| = UYDY^*$. Setting $V := UY$ and $W := Y^*$ proves the theorem. $\square$

4 Sterling's Approximation

Theorem 4.1 (Sterling's Approximation) $n! \sim \sqrt{2\pi n}(n/e)^n$.

Here, $f(n) \sim g(n)$ means that $\lim_{n\to\infty} f(n)/g(n) = 1$.

We'll prove a slightly weaker version of Theorem 4.1 that nevertheless suffices for all our purposes, namely,

Theorem 4.2 (Weak Sterling) *For all positive integers n,*

$$\frac{e}{\sqrt{2}}\sqrt{n}\left(\frac{n}{e}\right)^n \le n! \le e\sqrt{n}\left(\frac{n}{e}\right)^n.$$

Proof. We start with an integral approximation. The theorem clearly holds for $n = 1$, so assume $n \ge 2$. Since the log function is concave downward, we claim that for all i such that $2 \le i \le n$,

$$\frac{\log i + \log(i-1)}{2} \le \int_{i-1}^{i} \log x\, dx \le \log i - \frac{1}{2i}. \tag{5}$$

The left-hand side is the area of the trapezoid T_1 formed by the points $(i-1, 0), (i, 0), (i, \log i), (i-1, \log(i-1))$, and the right-hand side is the area of the trapezoid T_2 formed by the points $(i-1, 0), (i, 0), (i, \log i), (i-1, \log i - 1/i)$. Note that T_2's upper edge is the tangent line to the curve $y = \log x$ at the point $(i, \log i)$. By concavity of log, the region under the curve $y = \log x$ in the interval $[i-1, i]$ contains T_1 and is contained in T_2, hence the inequalities (5).

Now note that $\log(n!) = \sum_{i=1}^{n} \log i = \sum_{i=2}^{n} \log i$. Summing (5) from $i = 2$ to n and simplifying, we get

$$\log(n!) - \frac{\log n}{2} \le \int_1^n \log x\, dx = n\log n - n + 1 \le \log(n!) - \frac{1}{2}\sum_{i=2}^{n}\frac{1}{i}, \tag{6}$$

using the closed form $\int \log x\, dx = x\log x - x + C$. The sum on the right-hand side of (6) is the Harmonic series, which satisfies another integral approximation:

$$\sum_{i=2}^{n}\frac{1}{i} \ge \int_2^n \frac{dx}{x} = \log n - \log 2. \tag{7}$$

Equations (6) and (7) yield

$$\log n! - \frac{\log n}{2} \le n\log n - n + 1 \le \log(n!) - \frac{\log n}{2} + \frac{\log 2}{2},$$

and so

$$n\log n - n + 1 + \frac{\log n}{2} - \frac{\log 2}{2} \le \log n! \le n\log n - n + 1 + \frac{\log n}{2}. \tag{8}$$

Taking e to the power of all three quantities in (8) and simplifying, we have

$$\frac{e}{\sqrt{2}}\sqrt{n}\left(\frac{n}{e}\right)^n \le n! \le e\sqrt{n}\left(\frac{n}{e}\right)^n$$

as desired. □

5 Inequalities of Markov and Chebyshev

We only consider random variables that are real-valued and over discrete sample spaces. If X is such a random variable, then we let $E[X]$ and $\mathrm{var}[X]$ respectively denote the expected value (mean) of X and the variance of X.

Theorem 5.1 (Markov's Inequality) *Let X be a random variable with finite mean, and suppose $X \geq 0$. For every real $c > 0$,*

$$\Pr[X \geq c] \leq \frac{E[X]}{c}.$$

Proof. Let Ω be the sample space for X. We have

$$\begin{aligned} E[X] &= \sum_{a \in \Omega} X(a) \Pr[a] = \sum_{a : X(a) \geq c} X(a) \Pr[a] + \sum_{a : X(a) < c} X(a) \Pr[a] \\ &\geq \sum_{a : X(a) \geq c} X(a) \Pr[a] \geq \sum_{a : X(a) \geq c} c \Pr[a] = c \Pr[X \geq c]. \end{aligned}$$

Dividing both sides by c proves the theorem. □

Theorem 5.2 (Chebyshev's Inequality) *Let X be a random variable with finite mean and variance, and let $a > 0$ be real.*

$$\Pr[\, |X - E[X]| \geq a \,] \leq \frac{\mathrm{var}[X]}{a^2}.$$

Proof. We invoke Markov's Inequality with the random variable $Y = (X - E[X])^2$, letting $c = a^2$. Note that $Y \geq 0$, $E[Y] = \mathrm{var}[X]$, and $\Pr[\, |X - E[X]| \geq a \,] = \Pr[Y \geq a^2]$. □

6 Relative Entropy

Let $p = (p_1, p_2, \ldots)$ and $q = (q_1, q_2, \ldots)$ be two probability distributions over some (finite or infinite) discrete sample space $\{1, 2, \ldots\}$. The *relative entropy of q with respect to p* is defined as

$$H(q;p) = -\sum_i p_i \lg \frac{q_i}{p_i}, \tag{9}$$

Where the sum is taken over all i such that $p_i > 0$. If $q_i = 0$ and $p_i > 0$ for some i, then $H(q;p) = \infty$. Otherwise, the sum in (9) may or may not converge, but we always have the following regardless:

Theorem 6.1 *$H(q;p) \geq 0$, with equality holding if and only if $p = q$.*

Proof. We use that fact that $\log x \leq x - 1$ for all $x > 0$, with equality holding iff $x = 1$. We have

$$\begin{aligned}
H(q;p) &= -\sum_i p_i \lg \frac{q_i}{p_i} \\
&= -\frac{1}{\log 2} \sum_i p_i \log \frac{q_i}{p_i} \\
&\geq -\frac{1}{\log 2} \sum_i p_i \left(\frac{q_i}{p_i} - 1 \right) \\
&= \frac{1}{\log 2} \sum_i (p_i - q_i) \\
&= \frac{1}{\log 2} \left(1 - \sum_i q_i \right) \\
&\geq 0.
\end{aligned}$$

It is easy to see that equality holds above if and only if $p = q$. □

An important special case is when $q = (q_1, \ldots, q_n) = (1/n, \ldots, 1/n)$ is the uniform distribution on a sample space of size n (and $p = (p_1, \ldots, p_n)$ is arbitrary). In this case, we have

$$H(q;p) = \lg n - H(p_1, \ldots, p_n). \tag{10}$$

If $(p, 1-p)$ and $(q, 1-q)$ are binary distributions, then we abbreviate $H((q, 1-q); (p, 1-p))$ by $h(q;p)$. Note that by (10), $h(1/2, p) = 1 - h(p)$.

7 A Standard Tail Inequality

It might be necessary to read Section 6 before this one.

Let $0 < p < 1$ and let $n > 0$ be an integer. In this section, we give an upper bound for the sum $\sum_{i=0}^{t} \binom{n}{i} p^i (1-p)^{n-i}$, where $t \le pn$. [For example, this sum is the probability of getting at most t heads among n flips of a p-biased coin (i.e., n identical Bernoulli trials with bias p). The expected number of heads among n flips is pn, and we want to show that the probability of getting significantly fewer than pn heads diminishes exponentially with n.]

Theorem 7.1 *Let n be a positive integer. Let $0 < p < 1$ be arbitrary, and set $q = 1 - p$. If t is an integer such that $0 \le t \le pn$, then*

$$\sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} \le 2^{-nh(p;t/n)}, \tag{11}$$

where $h(\cdot;\cdot)$ is the binary relative entropy defined in Section 6.

Proof. If $t = 0$, then $h(p;t/n) = h(p;0) = -\lg q$, and so both sides of (11) equal q^n and so the inequality is satisfied.

Now suppose $0 < t \le pn$. Set $\lambda = t/n$, and let $\mu = 1 - \lambda$. Note that $0 < \lambda \le p < 1$ and $0 < q \le \mu < 1$. Define

$$C = \frac{p^t q^{n-t}}{\lambda^t \mu^{n-t}}.$$

For any $0 \le i \le t$, we have

$$p^i q^{n-i} = C \left(\frac{q}{p}\right)^{t-i} \lambda^t \mu^{n-t} \le C \left(\frac{\mu}{\lambda}\right)^{t-i} \lambda^t \mu^{n-t} = C \lambda^i \mu^{n-i}.$$

Therefore, starting with the left-hand side of (11), we get

$$\sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} \le C \sum_{i=0}^{t} \binom{n}{i} \lambda^i \mu^{n-i} \le C \sum_{i=0}^{n} \binom{n}{i} \lambda^i \mu^{n-i} = C(\lambda + \mu)^n = C.$$

For the right-hand side of (11), we get

$$2^{-nh(p;t/n)} = 2^{-nh(p;\lambda)} = 2^{n[\lambda \lg(p/\lambda) + \mu \lg(q/\mu)]} = \left(\frac{p}{\lambda}\right)^{n\lambda} \left(\frac{q}{\mu}\right)^{n\mu} = \left(\frac{p}{\lambda}\right)^{t} \left(\frac{q}{\mu}\right)^{n-t} = C,$$

which proves the theorem. □