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 2
2 The Schur Triangular Form and the Spectral Theorem 3
3 The Polar and Singular Value Decompositions 4
4 Sterling’s Approximation 6
5 Inequalities of Markov and Chebyshev 7
6 Relative Entropy 8
7 A Standard Tail Inequality 9
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 \(b_1, \ldots, b_n\),

\[
|a_1b_1 + \cdots + a_nb_n| \leq \sqrt{(a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2)},
\]

(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,

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

Adding \((\sum_i a_i b_i)^2\) 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. \(\square\)

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 \(z = a_1 + a_2i\) and \(w = b_1 + b_2i\) for real \(a_1, a_2, b_1, b_2\), we have

\[
|z + w|^2 = (a_1 + b_1)^2 + (a_2 + b_2)^2 = a_1^2 + a_2^2 + b_1^2 + b_2^2 + 2(a_1 b_1 + a_2 b_2) \\
\leq a_1^2 + a_2^2 + b_1^2 + b_2^2 + 2(\sqrt{a_1^2 + a_2^2})(\sqrt{b_1^2 + b_2^2}) = (\sqrt{a_1^2 + a_2^2} + \sqrt{b_1^2 + b_2^2})^2 = (|z| + |w|)^2.
\]

Taking the square root of both sides yields the corollary. \(\square\)

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

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

(2)

**Proof.** We have

\[
|z_1^* w_1 + \cdots + z_n^* w_n| \leq |z_1^* w_1| + \cdots + |z_n^* w_n| \quad \text{(by Corollary 1.2)} \\
= |z_1||w_1| + \cdots + |z_n||w_n| \\
\leq \sqrt{(|z_1|^2 + \cdots + |z_n|^2)(|w_1|^2 + \cdots + |w_n|^2)}. \quad \text{(by Theorem 1.1)}
\]

\(\square\)

**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 \( \lambda \) 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

\[
N = \begin{bmatrix}
\lambda & w^* \\
0 & N'
\end{bmatrix},
\]

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 \( N' = W'TW'^* \). Now we can factor \( N \):

\[
N = \begin{bmatrix}
\lambda & w^* \\
0 & W'TW'^*
\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & W' \end{bmatrix} \begin{bmatrix}
\lambda & w*W' \\
0 & T'
\end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & W'^* \end{bmatrix} = WTW^*,
\]

where

\[
W = \begin{bmatrix} 1 & 0 \\ 0 & W' \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} \lambda & w*W' \\ 0 & T' \end{bmatrix}.
\]

\( T \) 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 \( U = VV^* \) is unitary. \( \square \)

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

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

The last inequality follows from the fact that \( |A|_{ij} \neq 0 \). Similarly,

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

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 \( [A^*A]_{ii} > [A^*A]_{ij} \). But \( A \) is normal, so these two quantities must be equal. From this contradiction we get that \( A \) must be diagonal. \( \square \)

**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 $\text{diag}(s_1, s_2, \ldots, s_n)$ 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

$$|A| = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

where $D$ is the $k \times k$ nonsingular matrix $\text{diag}(s_1, \ldots, s_k)$. If $j > k$, then $|A|e_j = 0$, and thus $0 = |A|e_j = |A|^2e_j = A^*Ae_j$, 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 $A = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$, where $B$ is some $n \times k$ matrix, and the last $n-k$ columns of $A$ are $0$. We have

$$\begin{bmatrix} B^*B & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} B^* \\ 0 \end{bmatrix} \begin{bmatrix} B & 0 \end{bmatrix} = A^*A = |A|^2 = \begin{bmatrix} D^2 & 0 \\ 0 & 0 \end{bmatrix},$$

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 $U := \begin{bmatrix} BD^{-1} & W \end{bmatrix}$. We claim that $U$ is unitary and that $A = U|A|$. Noting that $D^{-1}$ is Hermitian, we have

$$U^*U = \begin{bmatrix} D^{-1}B^* \\ W^* \end{bmatrix} \begin{bmatrix} BD^{-1} & W \end{bmatrix} = \begin{bmatrix} D^{-1}B^*BD^{-1} & D^{-1}B^*W \\ W^*BD^{-1} & W^*W \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I,$$

and therefore $U$ is unitary. We also have

$$U|A| = \begin{bmatrix} BD^{-1} & W \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} B & 0 \end{bmatrix} = A.$$

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

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

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

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

which proves the theorem because $V^*UV$ is unitary. $\square$

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 $D = \text{diag}(s_1, \ldots, s_n)$. Furthermore, $s_1, \ldots, s_n$ are the eigenvalues of $|A|$. 

4
The $s_1, \ldots, s_n$ are known as the singular values of $A$.

**Proof.** For uniqueness, if $A = VW$ as above, then
\[
|A| = \sqrt{A^*A} = \sqrt{W^*DV^*VDW} = \sqrt{W^*D^2W} = W^*\sqrt{D^2W} = W^*DW,
\]
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 $D = \text{diag}(s_1, \ldots, s_n)$ 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. \qed
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 \( n \),

\[
\frac{e}{\sqrt{2}} \sqrt{n} \left(\frac{n}{e}\right)^n \leq n! \leq e \sqrt{n} \left(\frac{n}{e}\right)^n.
\]

Proof. We start with an integral approximation. The theorem clearly holds for \( n = 1 \), so assume \( n \geq 2 \).

Since the log function is concave downward, we claim that for all \( i \) such that \( 2 \leq i \leq n \),

\[
\log i + \log(i - 1) \leq \int_{i-1}^{i} \log x \, dx \leq \log i - \frac{1}{2i}.
\]  
(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 \( e \), 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

\[
\log(n!) - \frac{\log n}{2} \leq \int_{1}^{n} \log x \, dx = n \log n - n + 1 \leq \log(n!) - \frac{1}{2} \sum_{i=2}^{n} \frac{1}{i},
\]  
(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:

\[
\sum_{i=2}^{n} \frac{1}{i} \geq \int_{2}^{n} \frac{dx}{x} = \log n - \log 2.
\]  
(7)

Equations (6) and (7) yield

\[
\log n! - \frac{\log n}{2} \leq n \log n - n + 1 \leq \log(n!) - \frac{\log n}{2} + \frac{\log 2}{2},
\]

and so

\[
n \log n - n + 1 + \frac{\log n}{2} - \frac{\log 2}{2} \leq \log n! \leq n \log n - n + 1 + \frac{\log n}{2}.
\]  
(8)

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

\[
\frac{e}{\sqrt{2}} \sqrt{n} \left(\frac{n}{e}\right)^n \leq n! \leq e \sqrt{n} \left(\frac{n}{e}\right)^n
\]

as desired. \( \square \)
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 \( \text{var}[X] \) 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 \( \Omega \) be the sample space for \( X \). We have

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

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.

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

**Proof.** We invoke Markov’s Inequality with the random variable \( Y = (X - E[X])^2 \), letting \( c = a^2 \). Note that \( Y \geq 0, E[Y] = \text{var}[X] \), 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

\[
H(q; p) = -\sum_i p_i \log \frac{q_i}{p_i},
\]

(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

\[
H(q; p) = -\sum_i p_i \log \frac{q_i}{p_i}
= -\frac{1}{\log 2} \sum_i p_i \log \frac{q_i}{p_i}
\geq -\frac{1}{\log 2} \sum_i p_i \left( \frac{q_i}{p_i} - 1 \right)
= \frac{1}{\log 2} \sum_i (p_i - q_i)
= \frac{1}{\log 2} \left( 1 - \sum_i q_i \right)
\geq 0.
\]

It is easy to see that equality holds above if and only if \( p = q \). \( \square \)

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

\[
H(q; p) = \lg n - H(p_1, \ldots, p_n).
\]

(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 $t \leq pn$. [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 $0 \leq t \leq pn$, then
\[ \sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} \leq 2^{-nh(p,t/n)}, \] (11)
where $h(\cdot; \cdot)$ is the binary relative entropy defined in Section 6.

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

Now suppose $0 < t \leq pn$. Set $\lambda = t/n$, and let $\mu = 1 - \lambda$. Note that $0 < \lambda \leq p < 1$ and $0 < q \leq \mu < 1$. Define
\[ C = \frac{p^t q^{n-t}}{\lambda^t \mu^{n-t}}. \]
For any $0 \leq i \leq t$, we have
\[ p^i q^{n-i} = C \left( \frac{q}{p} \right)^{t-i} \lambda^i \mu^{n-i} \leq C \left( \frac{\mu}{\lambda} \right)^{t-i} \lambda^t \mu^{n-t} = C \lambda^t \mu^{n-t}. \]
Therefore, starting with the left-hand side of (11), we get
\[ \sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} \leq C \sum_{i=0}^{t} \binom{n}{i} \lambda^i \mu^{n-i} \leq C \sum_{i=0}^{n} \binom{n}{i} \lambda^i \mu^{n-i} = C(\lambda + \mu)^n = C. \]

For the right-hand side of (11), we get
\[ 2^{-nh(t/n)} = 2^{-nh(p;\lambda)} = 2^{n[\lambda \log(p/\lambda) + \mu \log(q/\mu)]} = \left( \frac{p}{\lambda} \right)^{n\lambda} \left( \frac{q}{\mu} \right)^{n\mu} = \left( \frac{p}{\lambda} \right)^{t} \left( \frac{q}{\mu} \right)^{n-t} = C, \]
which proves the theorem. \qed