CHAPTER I

BASIC DEFINITIONS AND RESULTS

§ 1 Regularisation of Functions

The convolution of two functions leads to a new function that inherits the stronger of the smoothness properties of the two original functions. This is the idea behind "regularisation" of functions.

Let $\phi$ be a real function of class $C^\infty$ with the following properties: $\phi \geq 0$, $\phi$ is even, the support $\text{supp} \phi = [-1, 1]$, and $\int \phi = 1$. For each $\varepsilon > 0$, let $\phi_\varepsilon(x) = \frac{1}{\varepsilon} \phi\left(\frac{x}{\varepsilon}\right)$. Then $\text{supp} \phi_\varepsilon = [-\varepsilon, \varepsilon]$ and $\phi_\varepsilon$ has all the other properties of $\phi$ listed above. The functions $\phi_\varepsilon$ are called mollifiers or smooth approximate identities, see Bhatia [16].

If $f$ is a locally integrable function, we define its regularisation of order $\varepsilon$ as the function

$$f_\varepsilon(x) = (f * \phi_\varepsilon)(x) = \int f(x - y) \phi_\varepsilon(y) dy = \int f(x - \varepsilon t) \phi(t) dt.$$ 

The family $f_\varepsilon$ has the following properties.

1. Each $f_\varepsilon$ is a $C^\infty$ function.

2. If the support of $f$ is contained in a compact set $K$, then the support of $f_\varepsilon$ is contained in an $\varepsilon$-neighbourhood of $K$.
3. If \( f \) is continuous at \( x_0 \), then \( \lim_{\varepsilon \to 0} f_\varepsilon(x_0) = f(x_0) \).

4. If \( f \) has a discontinuity of the first kind at \( x_0 \), then
   \[
   \lim_{\varepsilon \to 0} f_\varepsilon(x_0) = \frac{1}{2} [f(x_0^+) + f(x_0^-)].
   \]
   (A point \( x_0 \) is a point of discontinuity of the first kind if the left and right limits of \( f \) at \( x_0 \) exist; these limits are denoted as \( f(x_0^-) \) and \( f(x_0^+) \), respectively.)

5. If \( f \) is continuous, then \( f_\varepsilon(x) \) converges to \( f(x) \) as \( \varepsilon \to 0 \). The convergence is uniform on every compact set.

6. If \( f \) is differentiable, then, for every \( \varepsilon > 0 \), \( (f_\varepsilon)' = (f')_\varepsilon \).

7. If \( f \) is monotone, then, as \( \varepsilon \to 0 \), \( f_\varepsilon(x) \) converges to \( f'(x) \) at all points \( x \) where \( f'(x) \) exists. (Recall that a monotone function can have discontinuities of the first kind only and is differentiable almost everywhere.)

8. If \( f(x) \) belongs to \( L^p(\mathbb{R}) \) for \( 1 \leq p \leq \infty \), then this also holds for the regularisations and \( \|f_\varepsilon\|_p \leq \|f\|_p \); as \( \varepsilon \) approaches zero, \( f_\varepsilon \) converges to \( f \) in \( L^p \).

9. The regularisations \( f_\varepsilon(x) \) converges to \( f(x) \) at every Lebesgue point of \( f(x) \), and hence almost everywhere, and in particular at every point of continuity of \( f(x) \).

### § 2 Divided Differences

Let \( f \) be a continuously differentiable function on \((-1, 1)\). Then we denote by \( f^{[1]} \) the function on \((-1, 1) \times (-1, 1)\) defined as

\[
f^{[1]}(\lambda, \mu) = \begin{cases} 
   \frac{f(\lambda) - f(\mu)}{\lambda - \mu}, & \text{if } \lambda \neq \mu \\
   f'(\lambda), & \text{if } \lambda = \mu.
\end{cases}
\]
The expression $f^{[1]}(\lambda, \mu)$ is called the \textbf{first divided difference} of $f$ at $(\lambda, \mu)$. Clearly, $f^{[1]}(\lambda, \mu)$ is symmetric in $\lambda$ and $\mu$.

For higher order divided differences we require the continuous differentiability property of the corresponding order for the function $f$ and we may define the higher order divided differences inductively as follows. If $\lambda_1, \lambda_2, \ldots, \lambda_{k+1}$ are distinct

$$f^{[k]}(\lambda_1, \lambda_2, \ldots, \lambda_{k+1}) = \frac{f^{[k-1]}(\lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \lambda_k) - f^{[k-1]}(\lambda_2, \lambda_3, \ldots, \lambda_{k-1}, \lambda_k)}{\lambda_k - \lambda_{k+1}}.$$  

For other values of $\lambda_1, \lambda_2, \ldots, \lambda_{k+1}$, $f^{[k]}$ is defined by continuity; e.g.,

$$f^{[k]}(\lambda, \lambda, \ldots, \lambda) = \frac{1}{k!} f^{(k)}(\lambda),$$

where, $f^{(k)}(\lambda)$ denotes the $k$th-derivative of $f$.

The notion of divided differences can also be extended to the functions of two variables as follows.

Let $f(x, y)$ be a continuously differentiable function on $(-1, 1) \times (-1, 1)$. Then the first divided differences of $f(x, y)$ are functions on $(-1, 1) \times (-1, 1) \times (-1, 1)$ and defined as

$$f^{[1,0]}(\lambda_1, \lambda_2; \mu) = \begin{cases} \frac{f(\lambda_1, \mu) - f(\lambda_2, \mu)}{\lambda_1 - \lambda_2}, & \text{if } \lambda_1 \neq \lambda_2 \\ f^{[1,0]}(\lambda, \mu), & \text{if } \lambda_1 = \lambda_2 \end{cases}$$

$$f^{[0,1]}(\lambda; \mu_1, \mu_2) = \begin{cases} \frac{f(\lambda, \mu_1) - f(\lambda, \mu_2)}{\mu_1 - \mu_2}, & \text{if } \mu_1 \neq \mu_2 \\ f^{[0,1]}(\lambda, \mu), & \text{if } \mu_1 = \mu_2 \end{cases}$$
where, \( f^{(k,1)} \) denotes the \( k \)-th-partial derivative with respect to first variable and \( 1 \)-partial derivative with respect to second variable.

If \( f(x, y) \) is twice continuously differentiable on \((-1, 1) \times (-1, 1)\), then the second mixed divided difference
\[
f^{[1,1]}(x_1, x_2; y_1, y_2), \quad x_1 \neq x_2, \quad y_1 \neq y_2,
\]
is defined as
\[
f^{[1,1]}(x_1, x_2; y_1, y_2) = \frac{f(x_1, y_1) - f(x_1, y_2) - f(x_2, y_1) + f(x_2, y_2)}{(x_1 - x_2)(y_1 - y_2)}.
\]
The divided difference is to be interpreted appropriately when \( x_1 = x_2 \) or \( y_1 = y_2 \).

We may define the higher order divided differences inductively:
\[
f^{[k+1,m]}(x_1, \ldots, x_{k+2}; y_1, \ldots, y_{m+1}) = \frac{f^{[k,m]}(x_1, \ldots, x_k, x_{k+2}; y_1, \ldots, y_{m+1}) - f^{[k,m]}(x_1, \ldots, x_k, x_{k+1}; y_1, \ldots, y_{m+1})}{(x_{k+1} - x_{k+2})}.
\]
One may easily check that the divided differences are symmetric in variables \( x \) and \( y \) separately. For details see Donoghue [24] and Brown and Vasudeva [22].

§ 3 Convex and Log-convex Functions

Let \( I \subseteq \mathbb{R} \) be an open interval. A mapping \( f : I \to \mathbb{R} \) is said to be convex if
\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)
\]
for all \( x, y \in I \) and \( 0 \leq \lambda \leq 1 \).

A positive function \( f \) defined on an open interval \( I \subseteq \mathbb{R} \) is said to be log-convex or multiplicatively convex if for \( x, y \in I \) and \( 0 \leq \lambda \leq 1 \), the inequality
\[
\log f(\lambda x + (1 - \lambda)y) \leq \lambda \log f(x) + (1 - \lambda) \log f(y)
\]
or equivalently,
\[
f(\lambda x + (1 - \lambda)y) \leq [f(x)]^\lambda [f(y)]^{1-\lambda}
\]
holds. For properties of such functions, the reader may refer to Roberts and Varberg [55].

The convex functions involving two variables are defined as follows:

Suppose that \( D \) is a convex domain in the \( x,y \) plane, that is to say, a domain which contains with any two of its points the line segment between them. A real-valued function \( f(x,y) \) is said to be \textit{convex} in \( D \) if it is defined everywhere in \( D \) and

\[
f(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \leq \lambda f(x_1, y_1) + (1 - \lambda)f(x_2, y_2),
\]

for all \((x_1, y_1)\) and \((x_2, y_2)\) in \( D \) and \( 0 < \lambda < 1 \).

The above definition asserts more than convexity in \( x \) and \( y \) separately; thus \( xy \) is convex function of \( x \) for every \( y \) and is convex function of \( y \) for every \( x \), but is not a convex function of \( x \) and \( y \). Indeed, a function is convex in \( D \) if and only if its restriction to every line segment in \( D \) is convex on that segment.

**Theorem 1.3.1.** If \( f(x, y) \) is twice differentiable in an open domain \( D \), then a necessary and sufficient condition that it should be convex in \( D \) is that the quadratic form

\[
Q = f_{xx}u^2 + 2f_{xy}uv + f_{yy}v^2
\]

should be positive for all \( u, v \) and all \((x, y)\) in \( D \).

Proof. The condition is necessary: Let \((x,y)\) be in \( D \) and \((u,v)\) be arbitrary: when \( f(x,y) \) is convex, the function \( F(t) = f(x + ut, y + vt) \) is a convex function of \( t \) in some neighbourhood of the origin; since \( f(x,y) \) is \( C^2 \), so also is \( F(t) \), and \( F''(0) \geq 0 \). But
\[ F''(0) = f_{xx}u^2 + 2f_{xy}uv + f_{yy}v^2, \]
and because \((u,v)\) is arbitrary, the quadratic form above is positive.

The condition is sufficient: If the quadratic form \(Q\) is positive, functions of the type \(F(t)\) are convex in \(t\), hence \(f\) is convex, since its restriction to any line segment in \(D\) is convex.

The following Theorem is used in our subsequent chapters.

**Theorem 1.3.2.** Let \(f_n(x, y)\) be a sequence of non-negative convex functions defined on \((-1, 1) \times (-1, 1)\) which converges as a sequence of distributions to some distribution \(T\). Then there exists a subsequence \(f_{n_k}(x, y)\) which converges uniformly on compacts to some convex function \(f_0(x, y)\), and the distribution \(f_0(x, y)\) coincides with \(T\).

For a proof see Vasudeva (p.316, [64]).

### § 4 Subadditive Functions

The natural counterpart of Cauchy equation
\[ f(x + y) = f(x) + f(y), \]
is the class of functions
\[ f(x + y) \leq f(x) + f(y), \]
where \(f : I \to \mathbb{R}, I \subseteq \mathbb{R}\) and \(I\) being one of the intervals \((-\infty, \infty), (-\infty, 0)\) or \((0, \infty)\).

The functions satisfying the foregoing inequality are known as **subadditive functions**. These functions have been studied by several authors including Hille-Philips [34], Kuczma [43] and Rosenbaum [56].

It is well known that a measurable real valued subadditive function defined on \(\mathbb{R}\), fulfilling the
conditions, namely, \( \sup_{x > 0} \frac{f(x)}{x} \) and \( \inf_{x < 0} \frac{f(x)}{x} \) are finite, satisfies Lipschitz condition in \( \mathbb{R} \) see Kuczma (Theorem 5, p.412, [43]).

The Theorems 1.4.1 and 1.4.2 below will be used in the sequel.

**Theorem 1.4.1.** Let \( f : \mathbb{R} \to \mathbb{R} \) be subadditive. If \( f(0) = 0 \) has the Fréchet differential at zero, then there exist a \( c \in \mathbb{R} \) such that

\[ f(x) = cx. \]

For a proof see Kuczma, Theorem 2, p.409, [43].

**Theorem 1.4.2.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a measurable function, and let

\[ A = \inf_{x < 0} \frac{f(x)}{x}, \quad B = \sup_{x > 0} \frac{f(x)}{x}. \]

If \( A \) (respectively \( B \)) is finite, then

\[ A = \lim_{h \to 0^-} \frac{f(h)}{h}, \quad \text{(respectively } B = \lim_{h \to 0^+} \frac{f(h)}{h}). \]

§ 5 Monotone Functions of Two Variables

A real function \( f(x, y) \) of two variables defined in an open subset of the plane will be called **monotonic** if and only if

(i) It is locally integrable,

(ii) For every set of points \((x, y), (x + h, y), (x, y + k), (x + h, y + k)\) in the domain of \( f \) where \( h \) and \( k \) are positive, the quantity

\[ f(x + h, y + k) - f(x, y + k) - f(x + h, y) + f(x, y) \]

is non-negative.
A monotone function determines a distribution whose mixed partial derivative is the distribution limit of the quotients

\[ \frac{1}{t^2} \left[ f(x + t, y + t) - f(x, y + t) - f(x + t, y) + f(x, y) \right] \]

and since these are positive, the distribution \( \frac{\partial^2 f}{\partial x \partial y} \) is a positive measure.

Suppose, next, that \( \mu \) is a positive Radon measure with compact support in the plane and that \( H(x, y) \) is the characteristic function of the positive quadrant \( x > 0, y > 0 \). The convolution

\[ f(x, y) = (H \ast \mu)(x, y) = \int \int H(x - t, y - s) \, d\mu(t, s). \]

This is non-negative, bounded function, and in view of Fatou's Theorem, this is even a lower semi-continuous function. Since the mixed partial of \( H(x, y) \) is the delta distribution, we have \( \frac{\partial^2 f}{\partial x \partial y} = \mu \). Moreover, it is pretty easy to see that \( f(x, y) \) is monotone, since

\[ f(x + h, y + k) - f(x, y + k) - f(x + h, y) + f(x, y) \]

is exactly the \( \mu \)-measure of the rectangle determined by the inequalities \( x \leq t \leq x + h, y \leq s \leq y + k \).

More generally, if \( \mu \) is a positive Radon measure, we can write \( \mu = \sum \mu_i \) where the family \( \mu_i \) is determined by a suitable partition of unity. We then obtain \( f = \sum f_i \) in a natural way. \( f \) is then a monotone, non-negative, lower semi-continuous function whose mixed partial derivative is the measure \( \mu \).

It is well known that any distribution \( T \), solution of the differential equation \( \frac{\partial^2 T}{\partial x \partial y} = 0 \), is necessarily of the form \( T = X + Y \), where the distribution \( X \) depends only on
the $x$-coordinate, and $Y$ only on the $y$-coordinate. It follows, therefore that the most general function $g(x, y)$ having the mixed second partial $\mu \geq 0$ is necessarily of the form
\[ g(x, y) = f(x, y) + X(x) + Y(y) \]
where $f(x, y)$ is lower semi-continuous and the functions $X(x)$ and $Y(y)$ are arbitrary, measurable functions of one variable. It now becomes clear that a canonical determination is possible for the function $f$ having a given positive $\mu$ as its mixed partial derivative: we simply require
\[ f(x, 0) = f(0, y) = 0 \]
that is, the function be lower semi-continuous and vanish on the coordinate axes. Here we are tacitly supposing that the coordinate axes intersect the domain of definition of the function.

In the case of monotone functions of one variables, the first derivative is a positive distribution; a whole class of monotone functions correspond to that same derivative, and certain normalisation conditions must be imposed to make a canonical choice of the function. These conditions are usually two folds: the requirement that the function be lower semi-continuous and the requirement that the function vanishes at the origin. For higher dimensions the rules are almost the same: we need the function to vanish on the coordinate axes and to be lower semi-continuous.

For dimensions higher than 2 the situation is completely parallel to the two dimensional case: the monotone functions are the locally integrable functions with mixed $n^{th}$ order partial a positive measure; a canonical choice is always possible if it be required that the function vanish on the coordinate axes and be lower semi-continuous.
It should be remarked that our definition of monotonicity differs somewhat from that initially taken by W. H. Young in his study of these functions. The reader may refer to Hobson [35]. However, we find that the use of the terminology of distribution theory considerably simplifies the presentation without departing in any essential way from the ideas of Young and others.

§ 6 Positive Matrices

In what follows, \( C^n \) denotes the vector space of complex \( n \)-vectors, \( M_{m,n} \) (\( m, n \in \mathbb{N} \)) the space of \( m \times n \) complex matrices on \( C \) and \( M_{n,n} = M_n \) the algebra of complex matrices of order \( n \) on \( C \). \( I_n \) will denote the \( n \times n \) identity matrix, and \( 0_n \) will be the \( n \times n \) zero matrix. We denote by \( 0_{m,n} \) the \( m \times n \) zero matrix. For \( A \in M_n \), \( A^T \) and \( A^* \) denote the transpose and the conjugate transpose of \( A \) respectively. If \( A = A^* \) we say that \( A \) is Hermitian, and \( H_n \) will denote the set of all such matrices in \( M_n \). A matrix \( P \in H_n \) is called projection if \( P^2 = P \). The projections \( P \) and \( Q \) are called orthogonal if \( PQ = 0_n = QP \). We shall denote the set of all projections by \( P_n \). For \( A \in M_n \), \( \lambda(A) \) denotes the set of eigenvalues of \( A \) or the spectrum of \( A \). The matrix \( A \in H_n \) is positive definite (respectively semidefinite) if \( \lambda(A) \subset (0, \infty) \) (respectively \( \lambda(A) \subset [0, \infty) \)), equivalently \( x^*Ax > 0 \) (respectively \( x^*Ax \geq 0 \)) for all \( x \in C^n \) and \( x \neq 0 \). We shall denote by \( P_n \) (respectively \( S_n \)), the set of all positive definite (respectively semidefinite) matrices. It is clear that \( A^*A \) is positive semidefinite for all \( A \in M_n \). If
$A \in M_n$, then $(A^*A)^{1/2}$ is a positive semidefinite matrix, and so all its $n$ eigenvalues (not necessarily distinct) are non-negative. Thus, to each $A \in M_n$, we can associate an element $s(A)$ of $\mathbb{R}^n$ whose co-ordinates are eigenvalues of $(A^*A)^{1/2}$ appearing in descending order. We shall denote the positive operator $(A^*A)^{1/2}$ by $|A|$. The co-ordinates of $s(A)$ are called the singular values of $A$. If $A, B \in \mathbb{H}_n$ then $A > B$ (respectively $A \geq B$) or $B < A$ (respectively $B \leq A$) implies that $A - B \in \mathbb{P}_n$ (respectively $S_n$). We observe that if $A \geq B$ then $T^*AT \geq T^*BT$ for all matrices $T$ of order $n \times m$. A matrix $U \in M_n$ is called unitary if $UU^* = I_n = U^*U$.

§ 7 Tensor product of finite dimensional Hilbert Spaces

Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be finite dimensional Hilbert spaces. The algebraic tensor product $\mathbb{H}_1 \otimes \mathbb{H}_2$ is a vector space of all formal expressions of the type:

$$\sum_{i=1}^{k} \lambda_i (\phi_i \otimes \psi_i),$$

where $k \geq 1$, $\phi_i \in \mathbb{H}_1$, $\psi_i \in \mathbb{H}_2$, along with the following rules:

1. $(\phi_1 + \phi_2) \otimes \psi = \phi_1 \otimes \psi + \phi_2 \otimes \psi$
2. $\phi \otimes (\psi_1 + \psi_2) = \phi \otimes \psi_1 + \phi \otimes \psi_2$
3. $(\lambda \phi) \otimes \psi = \lambda (\phi \otimes \psi) = \phi \otimes (\lambda \psi)$.

We define an inner product in $\mathbb{H}_1 \otimes \mathbb{H}_2$ by setting

$$\langle \phi_1 \otimes \psi_1, \phi_2 \otimes \psi_2 \rangle = \langle \phi_1, \psi_1 \rangle \langle \psi_2, \phi_2 \rangle$$

for elements of the type $\phi \otimes \psi$, and then extend this in a bilinear way to other elements of $\mathbb{H}_1 \otimes \mathbb{H}_2$. The vector space
$H_1 \otimes H_2$ along with the inner product defined above is called **tensor product** of $H_1$ and $H_2$, and is denoted by $H_1 \otimes H_2$. If \{$e_1, e_2, \ldots, e_m$\} and \{$f_1, f_2, \ldots, f_n$\} are the bases of $H_1$ and $H_2$ respectively, then the set \{${e_i \otimes f_j : i = 1, 2, \ldots, m; j = 1, 2, \ldots, n}$\} is a basis in $H_1 \otimes H_2$. If the bases \{${e_1, e_2, \ldots, e_m}$\} and \{${f_1, f_2, \ldots, f_n}$\} are orthonormal then so is the basis \{${e_i \otimes f_j : i = 1, 2, \ldots, m; j = 1, 2, \ldots, n}$\}.

Let $A_1$ and $A_2$ be operators on $H_1$ and $H_2$ respectively. The product of $A_1$ and $A_2$ is an operator $A_1 \otimes A_2$ on $H_1 \otimes H_2$ defined by:

$$(A_1 \otimes A_2) \left( \sum_{i=1}^{k} \lambda_i \phi_i \otimes \psi_i \right) = \left( \sum_{i=1}^{k} \lambda_i (A_1 \phi_i \otimes A_2 \psi_i) \right)$$

for every $\sum_{i=1}^{k} \lambda_i \phi_i \otimes \psi_i$ in $H_1 \otimes H_2$.

The following identities are easy to verify:

(i) $A \otimes 0 = 0 \otimes B = 0$
(ii) $I \otimes I = I$
(iii) $(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B)$
(iv) $A \otimes (B_1 + B_2) = (A \otimes B_1) + (A \otimes B_2)$
(v) $(\alpha A) \otimes (\beta B) = \alpha \beta (A \otimes B)$, $\alpha, \beta$ are scalars,
(vi) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, where $A$ and $B$ are invertible;
(vii) $(A \otimes B)^* = A^* \otimes B^*$
(viii) $(A_1 A_2) \otimes (B_1 B_2) = (A_1 \otimes B_1) (A_2 \otimes B_2)$.

If both $A$ and $B$ are invertible, then so is $A \otimes B$ and conversely, if $A \otimes B$ is invertible, then so are $A$ and $B$. If \{${\lambda_1, \lambda_2, \ldots, \lambda_n}$\} is the spectrum of a normal operator $A$ (An operator $A$ is called **normal** if $A^* A = A A^*$) and \{${e_1, e_2, \ldots, e_m}$\)
are the corresponding eigenvectors, and if \( \{ \mu_1, \mu_2, \ldots, \mu_n \} \) is the spectrum of another normal operator \( B \) and \( \{ f_1, f_2, \ldots, f_n \} \) are the corresponding eigenvectors, then the spectrum of \( A \otimes B \) is
\[
\{ \lambda_1 \mu_1, \lambda_2 \mu_2, \ldots, \lambda_1 \mu_n, \lambda_2 \mu_n, \ldots, \lambda_n \mu_1, \lambda_n \mu_2, \ldots, \lambda_n \mu_n \}
\]
with the corresponding eigenvectors:
\[
\{ e_1 \otimes f_1, e_1 \otimes f_2, \ldots, e_1 \otimes f_n, e_2 \otimes f_1, e_2 \otimes f_2, \ldots, e_2 \otimes f_n, \ldots, e_m \otimes f_1, e_m \otimes f_2, \ldots, e_m \otimes f_n \}.
\]

Let \( A \) be an operator on \( m \)-dimensional Hilbert space \( H_1 \) and \( S = (s_{ij}) \) be matrix representation of \( A \) with respect to an ordered orthonormal basis \( \{ e_1, e_2, \ldots, e_m \} \). The matrix \( S \) is of size \( m \times m \) and \( s_{ij} = \langle Se_j, e_i \rangle \). Also, let \( B \) be an operator on an \( n \)-dimensional Hilbert space \( H_2 \) and let \( T \) be the matrix representation of \( B \) with respect to an ordered orthonormal basis \( \{ f_1, f_2, \ldots, f_n \} \). Similarly, \( T = (t_{ij}) \) is an \( n \times n \) matrix where \( t_{ij} = \langle T f_j, f_i \rangle \). Then the matrix representation of the operator \( A \otimes B \) on an \( mn \) dimensional Hilbert space \( H_1 \otimes H_2 \) with respect to the orthonormal basis which is lexicographically ordered
\[
\{ e_1 \otimes f_1, e_1 \otimes f_2, \ldots, e_1 \otimes f_n, e_2 \otimes f_1, e_2 \otimes f_2, \ldots, e_m \otimes f_n \}
\]
is given by
\[
\begin{pmatrix}
S_{11} t_{11} & \cdots & S_{1n} t_{1n} & \cdots & S_{11} t_{11} & \cdots & S_{1n} t_{1n} \\
\vdots & & \vdots & & \vdots & & \vdots \\
S_{11} t_{n1} & \cdots & S_{1n} t_{nn} & \cdots & S_{11} t_{n1} & \cdots & S_{1n} t_{nn} \\
\vdots & & \vdots & & \vdots & & \vdots \\
S_{m1} t_{11} & \cdots & S_{m1} t_{1n} & \cdots & S_{m1} t_{11} & \cdots & S_{m1} t_{1n} \\
\vdots & & \vdots & & \vdots & & \vdots \\
S_{m1} t_{n1} & \cdots & S_{m1} t_{nn} & \cdots & S_{m1} t_{n1} & \cdots & S_{m1} t_{nn}
\end{pmatrix}
\]

Since the element in \((k,p)\)th row and \((j,q)\)th column is
\[
\langle (S \otimes T) (e_j \otimes f_q), (e_k \otimes f_p) \rangle = \langle Se_j \otimes Tf_q, e_k \otimes f_p \rangle \\
= \langle Se_j, e_k \rangle \langle Tf_q, f_p \rangle \\
= s_{kj} t_{pq}.
\]

This matrix is called the Tensor or Kronecker product of matrices \((s_{ij})\) and \((t_{ij})\).

Note that the above can be represented as block matrix in which each block is an \(n \times n\) matrix
\[
\begin{pmatrix}
 s_{11}^T & s_{12}^T & \cdots & s_{1m}^T \\
 s_{21}^T & s_{22}^T & \cdots & s_{2m}^T \\
 \vdots & \vdots & \ddots & \vdots \\
 s_{m1}^T & s_{m2}^T & \cdots & s_{mm}^T
\end{pmatrix}.
\]

It can be easily verified from the rules for operating with the tensor product of operators that if both \(A \in \mathcal{H}_m\) and \(B \in \mathcal{H}_n\) are positive semidefinite (resp. positive definite) then so is their tensor product. More generally, if \(A \geq B\) and \(C \geq D\) for \(A, B \in S_n\) and \(C, D \in S_n\) then \(A \otimes C \geq B \otimes D\).

\section*{§ 8 Hadamard Product}

Let \(A = (a_{ij})\) and \(B = (b_{ij})\) be in \(M_n\). Then a matrix \(C = (a_{ij} b_{ij})\) in \(M_n\) is called the Hadamard product of \(A\) and \(B\) and is denoted by \(A \circ B\). This product differs from the usual matrix product in many ways. The usual product of matrices is non-commutative while the Hadamard product of matrices is commutative.

There is an intimate relation between Hadamard product and tensor product of matrices in \(M_n\). Indeed, for \(A, B \in M_n\), we have
\[
A \circ B = U^* (A \otimes B) U,
\]
where, $U = (e_1 \otimes e_1 \; e_2 \otimes e_2 \; \ldots \; e_n \otimes e_n)$, a matrix of order $n^2 \times n^2$ and $e_i$, $i = 1, 2, \ldots, n$ is a standard orthonormal basis for $\mathbb{C}^n$. See, Zhang (p.191, [67]).

This gives immediately that if $A \otimes B$ is positive then so is $A \circ B$. See also Donoghue (p.9, [24]).

§ 9 Matrix Functions and Fréchet differential

Let $A \in \mathbb{H}_n$ and let $(u_1, u_2, \ldots, u_n)$ be an orthonormal set of eigenvectors corresponding to $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $A$. Then the matrix $U$ whose $i^{th}$ column is $u_i$, $i = 1, 2, \ldots, n$ is a unitary matrix and $U^* A U = D$, where, $D$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. Thus $A = U D U^*$, that is,

$$A = [u_1 \; \ldots \; u_n] \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix},$$

and so

$$A = \lambda_1 u_1 u_1^* + \ldots + \lambda_n u_n u_n^*.$$  \hspace{1cm} (1.1)

Suppose all $\lambda_i$'s are distinct. We denote by $P_{ij}$ the $n \times n$ matrix $u_i u_j^*$. Then

$$P_{ij} P_{kj} = u_i u_j^* u_j u_k^* = \delta_{jk} u_i u_k^* = \delta_{jk} P_{is}. $$

Hence \{$P_{ij} : i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$\} is a set of matrix units, and so is a basis of the space $M_n$. In this notation, (1.1) is

$$A = \lambda_1 P_{11} + \ldots + \lambda_n P_{nn}. $$  \hspace{1cm} (1.2)

Since $P_{ij} u_j = u_i u_j^* u_j = \delta_{ij} u_i$ and $P_{ij} P_{jj} = \delta_{ij} P_{jj}$, each $P_{ii}$ is the orthogonal projection matrix onto the subspace spanned
by the eigenvector \( u_i \). This expression of \( A \) is called the **spectral decomposition** of \( A \). See Mehta (p.48, [49]).

If an eigenvalue \( \lambda \) has multiplicity \( k \), that is, \( \lambda = \lambda_1 = \cdots = \lambda_k \) and the corresponding \( k \) orthonormal eigenvectors \( u_1, \ldots, u_k \), then it is an orthonormal basis of eigenspace

\[
E(\lambda) = \{ x \in \mathbb{C}^n : (A - \lambda I)x = 0 \}
\]

and \( P_k = P_{\lambda_1} \cdots P_{\lambda_k} \) is the orthogonal projection matrix onto \( E(\lambda) \). Therefore, counting the multiplicity of each eigenvalue, every Hermitian matrix \( A \) has a unique spectral decomposition

\[
A = \lambda_1 P_{\lambda_1} + \lambda_2 P_{\lambda_2} + \cdots + \lambda_m P_{\lambda_m},
\]

where, \( \lambda_1, \lambda_2, \ldots, \lambda_m \) are the distinct eigenvalues of \( A \).

Let \( H_n(I) \) denote the set of all elements in \( H_n \) whose spectrum is contained in the real open interval \( I \) and let \( f : I \to \mathbb{R} \) be a function. If \( A \) in \( H_n(I) \) has the spectral decomposition

\[
A = \sum_{i=1}^{m} \lambda_i P_{\lambda_i},
\]

then, we define

\[
f(A) = \sum_{i=1}^{m} f(\lambda_i) P_{\lambda_i}.
\]

The matrix \( f(A) \) is an element of \( H_n \). This defines the **matrix function** \( f : H_n(I) \to H_n \) corresponding to the function \( f : I \to \mathbb{R} \).

Observe that

\[
(f + g)(A) = f(A) + g(A),
\]

\[
(fg)(A) = f(A)g(A).
\]

Moreover, if \( f(x) \geq g(x) \) for all \( x \in I \), then \( f(A) \geq g(A) \) for all \( A \) in \( H_n(I) \).
Let $I$ denote an open interval of $\mathbb{R}$. A function $f$ defined on $H_n(I)$ is Fréchet differentiable at $A \in H_n(I)$ if there exists a bounded linear operator $df(A) \in \mathcal{L}(H_n, H_n)$ such that

$$\lim_{K \to 0} K^{-1} \left\| f(A + K) - f(A) - df(A)(K) \right\| = 0, \; K \in H_n.$$ 

Likewise $f$ is said to be Fréchet differentiable in $H_n(I)$, if $f$ is Fréchet differentiable at every point $A \in H_n(I)$. We say that $f$ is continuously Fréchet differentiable if the differential mapping $A \to df(A)$ from $H_n(I)$ into $\mathcal{L}(H_n, H_n)$ is continuous. It is easy to see that, if $f$ is differentiable at $A$, then for every $K \in H_n$,

$$df(A)(K) = \frac{d}{dt} \bigg|_{t=0} f(A + tK).$$

Let $C(I)$ denote the space of continuous functions $f : I \to \mathbb{R}$ and $C^L(I)$ the space of functions $f \in C(I)$ such that the derivative $f^{(1)}, \ldots, f^{(L)}$ exists and are continuous on $I$. Observe that $C^L(I)$ is a Fréchet space (Treves [63]). For each $L$, we denote the space of functions $F : H_n(I) \to H_n$ such that $F$ and its Fréchet derivatives $dF, \ldots, d^L F$ exist and are continuous on $H_n(I)$ by $C^L(H_n(I), H_n)$. The space $C^L(H_n(I), H_n)$ is Fréchet space with respect to the topology defined by seminorms

$$\left\| d^k F \right\|_{H_n(I)} = \sup_{A \in H_n(I)} \left\| d^k F(A) \right\|,$$

where $0 \leq k \leq L$ and $I'$ is a closed subinterval of $I$.

Our next theorem asserts that the matrix function $f$ inherits differentiability properties from $f$. 
Theorem 1.9.1. If \( f \in C^1(I) \) then the matrix function \( f \in C^1(H_n(I), H_n) \). If \( A \in H_n(I) \) and \( e_1, e_2, \ldots, e_m \) is an orthonormal basis of \( C^n \) consisting of eigenvectors of \( A \) corresponding to eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_m \) then the matrix \( df(A) (H) \) with respect to this basis is given by

\[
\left( df(A) (H) e_k, e_j \right) = f^{(1)}(\lambda_j, \lambda_k) \langle H e_k, e_j \rangle
\]

for all \( j, k \in \{1, 2, \ldots, n\} \).

This theorem is contained in Bhatia [16], and in general form in Brown and Vasudeva [22].

A function \( f : H_n(I) \to H_n \) is said to be **matrix monotone** of order \( n \) if \( A, B \in H_n(I), A \leq B \) implies \( f(A) \leq f(B) \).

A function \( f : H_n(I) \to H_n \) is called **matrix convex** on \( I \) of order \( n \) if

\[
f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda) f(B),
\]

for all \( A, B \in H_n(I) \) and for all \( 0 \leq \lambda \leq 1 \).

A function \( f : H_n(I) \to H_n \) is called **matrix concave** on \( I \) of order \( n \) if \(-f\) is matrix convex on \( I \) of order \( n \).

The function \( f \) is said to be **operator monotone** (respectively **convex**, **concave**) if \( f \) is matrix monotone (respectively convex, concave) of order \( n \) for each \( n \in \mathbb{N} \).

We shall denote by \( P_n(I) \) the set of all matrix monotone functions on \( I \) of order \( n \) and by \( C_n(I) \) the set of all matrix convex functions on \( I \) of order \( n \). Thus \( f \) is operator monotone (respectively convex) if and only if \( f \in \bigcap_{n=1}^{\infty} P_n(I) \).
(respectively $f \in \bigcap_{n=1}^{\infty} C_n(I)$). Also note that $P_{n+1}(I) \subset P_n(I)$, $C_{n+1}(I) \subset C_n(I)$, $n = 1, 2, \ldots$. Indeed, let $f \in P_{n+1}(I)$ and $A, B \in H_n(I)$ be such that $A \geq B$. Let
\[ A_1 = \begin{bmatrix} A & 0_{n,1} \\ 0_{1,n} & \alpha \end{bmatrix} \text{ and } B_1 = \begin{bmatrix} B & 0_{n,1} \\ 0_{1,n} & \alpha \end{bmatrix}, \]
where $\alpha \in \mathbb{I}$. Then $A_1, B_1 \in H_{n+1}(I)$ and $A_1 \geq B_1$. Consequently,
\[ f(A_1) = \begin{bmatrix} f(A) & 0_{n,1} \\ 0_{1,n} & f(\alpha) \end{bmatrix} \geq \begin{bmatrix} f(B) & 0_{n,1} \\ 0_{1,n} & f(\alpha) \end{bmatrix} = f(B_1). \]
Thus, it follows that $f(A) \geq f(B)$, that is $f \in P_n(I)$. Similarly, it can be shown that $C_{n+1}(I) \subset C_n(I)$.

The following differential characterisation of matrix monotonicity is due to Löwner (see Bhatia [16]).

**Theorem 1.9.2.** If $f \in C^1(I)$ then $f$ is operator monotone if and only if for each $n \in \mathbb{N}$, $0_n \leq df(A)(H)$, whenever $A \in H_n(I)$, $H \in S_n$.

Combining Theorem 1.9.1 and Theorem 1.9.2, we have the following characterisation. For a proof see Bhatia (p.126, [16]).

**Theorem 1.9.3.** If $f \in C^1(I)$, then $f$ is matrix monotone if and only if for $A \in H_n(I)$, the matrix $f^{[1]}(A)$ is positive semidefinite, where the matrix $f^{[1]}(A)$ with respect to the orthonormal basis $e_1, e_2, \ldots, e_n$ of eigenvectors of $A$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ is given by
\[
\left( f^{[1]}(A) \right)_{j,k} = \begin{cases} \frac{f(\lambda_j) - f(\lambda_k)}{\lambda_j - \lambda_k}, & j \neq k \\ f'(\lambda_j), & \text{otherwise.} \end{cases}
\]
Another characterisation due to Löwner [46] for operator monotone functions is via an integral representation. We state it below without proof.

**Theorem 1.9.4.** Let $f \in C^1(-1,1)$. Then $f$ is operator monotone if and only if

$$f(x) = f(0) + \int_{-1}^{1} \frac{x}{1 - tx} \, dv(t),$$

where $v$ is a positive bounded measure on $[-1,1]$.

If $f$ is an operator monotone function defined on $[0,\infty)$ then the above representation takes the form

$$f(x) = f(0) + bx + \int_{0}^{\infty} \frac{x}{\lambda + x} \, d\varpi(\lambda),$$

where $\varpi$ is a positive bounded measure. See Bhatia (p.144 [16]).

A. Korányi [40] introduced the notion of matrix functions of two variables, which is as follows:

Let $A \in H_m(I)$ and $B \in H_n(J)$ have spectral decompositions

$$A = \sum_{i=1}^{k} \lambda_i P_i$$

and

$$B = \sum_{j=1}^{r} \mu_j Q_j.$$

Then we define

$$f(A, B) = \sum_{i=1}^{k} \sum_{j=1}^{r} f(\lambda_i, \mu_j) P_i \otimes Q_j.$$

The matrix $f(A, B)$ is an element of $H_m \otimes H_n$. This defines the **matrix function** $f : H_m(I) \times H_n(J) \rightarrow H_m \otimes H_n$ of two variables corresponding to the function $f : I \times J \rightarrow \mathbb{R}$. 

Observe that if \( f(x, y) = g(x)h(y) \) then, \( f(A, B) = g(A) \otimes h(B) \).

Let \( f : H_m(I) \times H_n(J) \to H_m \otimes H_n \). Then the function \( f \) is said to have partial Fréchet differential at \( A \), where \( (A, B) \in H_m(I) \times H_n(J) \), if there exists \( T_1 \in \mathcal{L}(H_m(I), H_m \otimes H_n) \) such that for \( H \in H_m \) the following limit

\[
\lim_{H \to 0} \frac{1}{\|H\|} \left\| f(A + H, B) - f(A, B) - T_1(H) \right\| = 0
\]

holds and \( d_1 f(A, B) = T_1 \). In other words, the partial differential \( d_1 f(A, B) \) exists and equals \( T_1 \) if and only if the function \( H \to f(A + H, B) \) has a differential at \( H = 0 \) equal to \( T_1 \). The mixed partial differential of \( f \) with respect to the first and second variables denoted \( d_2 d_1 f \) is defined to be partial Fréchet differential of the function \( d_1 f \) with respect to the second coordinate so that \( d_2 d_1 f = d_2(d_1 f) \). The value \( d_2 d_1 f(A, B) \) of \( d_2 d_1 f \) at the point \( (A, B) \) of its domain is an element of \( \mathcal{L}(H_n, \mathcal{L}(H_m, H_m \otimes H_n)) \) which can be identified with a continuous bilinear map from \( H_m \times H_n \) into \( H_m \otimes H_n \). If \( f \) is twice differentiable at \( (A, B) \) then

\[
d_2 d_1 f(A, B) = d_1 d_2 f(A, B).
\]

The present definitions and various results from the theory of Fréchet differentiable functions that will be subsequently used are taken from Flett [26].

For each integer \( L \geq 0 \), we denote the space of functions \( f : H_m(I) \times H_n(J) \to H_m \otimes H_n \) such that \( f \) and its Fréchet derivatives \( df, d^2 f, \ldots, d^L f \) exist and are continuous on \( H_m(I) \times H_n(J) \), by \( C^L(H_m(I) \times H_n(J), H_m \otimes H_n) \). It is a Fréchet space with respect to the topology defined by the semi-norms

\[
\|d^L f\|_K = \sup_{(A, B) \in K} \left\| d^L f(A, B) \right\|
\]
where $0 < k \leq L$ and $K$ is a compact subset of $H_m(I) \times H_n(J)$.

The proof of this fact can be given by extending the arguments in the special case $C^1(I \times J)$.

For the following Theorem and its proof, the reader may refer to Brown and Vasudeva [22].

**Theorem 1.9.5.** If $f \in C^1(I \times J)$ then the operator function

$$f = O f \in C^1(H_m(I) \times H_n(J), H_m \otimes H_n).$$

The mapping

$$O : C^1(I \times J) \to C^1(H_m(I) \times H_n(J), H_m \otimes H_n)$$

is continuous.

If $f \in C^1(I \times J)$ and $I_i$ and $J_i$ are closed subintervals of $I$ and $J$ respectively then for $0 < k < L$

$$\|d^k f\|_{H_m(I_i) \times H_n(J_i)} \leq \sum_{r=0}^k C_{k,r}(m, n) \|f^{(r,k-r)}\|_{I_i \times J_i},$$

for some coefficients $C_{k,r}(m, n)$.

A function $f : H_m(I) \times H_n(J) \to H_m \otimes H_n$ is said to be **matrix monotone** of two variables of order $(m, n)$ if for all $A, B \in H_m(I)$ and $C, D \in H_n(J)$,

$$f(A, C) - f(A, D) - f(B, C) + f(B, D) \geq 0_{mn},$$

whenever $A \geq B$ and $C \geq D$.

A function $f$ is said to be **operator monotone** of two variables if $f$ is matrix monotone of two variables of order $(m, n)$ for each $m, n \in \mathbb{N}$.

Let $P_{m,n}(I,J)$ denote the set of all matrix monotone functions of two variables of order $(m,n)$ on $I \times J$. Then $P_{m+1,n}(I,J) \subseteq P_{m,n}(I,J), P_{m,n+1}(I,J) \subseteq P_{m,n}(I,J)$, $m, n = 1, 2, \ldots$. Indeed,
using the same technique as used in the case of matrix functions of one variable one can prove that, if \( f \in P_{n+1,n}(I, J) \) then \( f \in P_{m,n}(I, J) \).

The following is an analogue of Theorem 1.9.4 for the functions of two variables given by Korányi (Theorem 4, [40]).

**Theorem 1.9.6.** Let \( f : (-1, 1) \times (-1, 1) \to \mathbb{R} \) with \( f(x, 0) = 0 = f(0, y) \) for all \( x, y \in (-1, 1) \) be a function. Assume that the first partial derivatives and the mixed second partial derivative of \( f \) exist and are continuous. Then the following statements are equivalent:

(i) \( f \) is operator monotone on \((-1,1) \times (-1,1)\)

(ii) \( f \) admits the integral representation

\[
f(x, y) = \int_{(-1,1)^2} \frac{x}{1 - sx} \frac{y}{1 - sy} \, dv(s, t),
\]

where \( v \) is a positive bounded measure on \([-1,1] \times [-1,1]\)

(iii) \( f \) is analytic and can be continued analytically for all non-real values of two variables to a function \( g \), satisfying the following conditions:

(a) \( g(z_1, \bar{z}_2) = g(z_2, \bar{z}_1) \) for all \( \text{Im } z_1, \text{Im } z_2 \neq 0 \)

(b) \( g(z_1, z_2) - g(z_1, \bar{z}_2) - g(z_2, z_1) + g(\bar{z}_1, \bar{z}_2) = 2 \text{Re } [g(z_1, z_2) - g(z_1, \bar{z}_2)] \leq 0 \),

\[ \text{for all } \text{Im } z_1, \text{Im } z_2 > 0, \]

(c) for all \( \phi, \ 0 < \phi \leq \pi/2 \), there exists a constant \( C(\phi) \) such that \( |g(z_1, z_2)|/|z_1z_2| \leq C(\phi) \) for all \( z_1, z_2 \in \{ z : \phi \leq \arg z \leq \pi - \phi \} \).

It is remarked here that restriction to the interval \((-1,1)\) in Theorems 1.9.4 and 1.9.6 is unessential.
§ 10 Operator Means

Kubo and Ando [42] developed the axiomatic theory for connection and means of pair of positive operators. Let $A, B, C, \cdots$ be in $S_n$. A binary operation $\sigma$ defined on and with values in $S_n$ is called a connection if

(i) $A \leq C, B \leq D \Rightarrow A \sigma B \leq C \sigma D$

(ii) $C(A \sigma B) \leq (CAC) \sigma (CBC)$

(iii) $A_k \downarrow A$ and $B_k \downarrow B \Rightarrow (A_k \sigma B_k) \downarrow A \sigma B$.

A mean is a connection with normalisation condition,

(iv) $I_n \sigma I_n = I_n$.

Every connection possesses the property

(v) $(A \sigma B) + (C \sigma D) \leq (A + C) \sigma (B + D)$.

Examples. (i) The usual sum $A + B$ is a connection. The normalised sum is called arithmetic mean, and is denoted by $\sqrt{\frac{A}{2}} + \sqrt{\frac{B}{2}}$:

\[ A \sqrt{\frac{1}{2}} B = \frac{A + B}{2}. \]

(ii) The left trivial mean $\omega_l$ is defined to be $A \omega_l B = A$

and the right trivial mean $\omega_r$ is defined to be $A \omega_r B = B$.

(iii) For $A, B \in P_n$ the parallel sum is defined by $A : B = (A^{-1} + B^{-1})^{-1}$.

The normalised parallel sum $A \sqrt{\frac{1}{2}} B$ is given by

\[ A \sqrt{\frac{1}{2}} B = 2(A : B). \]

(iv) For $A, B \in P_n$, the geometric mean $A \#_{\sqrt{\frac{1}{2}}} B$ is defined by
If $A, B \in S_n$, the harmonic mean is defined by

$$A ^ {1/2} B = \lim_{\varepsilon \to 0} (A + \varepsilon I_n)^{1/2} (B + \varepsilon I_n)$$

and the geometric mean by

$$A ^ {\#} B = \lim_{\varepsilon \to 0} (A + \varepsilon I_n)^{\#} B.$$

The following theorem due to Kubo and Ando [42] plays an important role in the theory of operator means and leads to a fairly rich class of examples of operator means.

**Theorem 1.10.1.** For every connection $\sigma$ and $x > 0$, the operator $1\sigma x$ becomes a scalar. The map, $\sigma \mapsto f$, defined by

$$f(x) = 1\sigma x,$$

for $x > 0$ establishes an affine order-isomorphism from the class of connections onto the class of non-negative operator monotone functions on $(0, \infty)$.

Let $\sigma$ be a connection with $f(x)$ as the corresponding representing function. The **transpose** $\sigma'$ of a connection is defined by $A_0^* B = B_\sigma A$ with corresponding representing function $xf(x^{-1})$. The **adjoint** $\sigma'$ and **dual** $\sigma^\dagger$ of a non zero connection $\sigma$ are, respectively, defined by $A_0^* B = (A^{-1} \sigma B^{-1})^{-1}$ and $A_0^\dagger B = (B^{-1} \sigma A^{-1})^{-1}$ for $A, B \in P_n$ with corresponding representing functions $(f(x^{-1}))^{-1}$ and $xf(x)^{-1}$. A connection is called **symmetric** if $\sigma' = \sigma$. A connection is called **self-adjoint** if $\sigma' = \sigma$ and **self-dual** if $\sigma^\dagger = \sigma$. The arithmetic mean, geometric mean and harmonic mean are the examples of symmetric means. The arithmetic and harmonic means are adjoint of each other. The geometric mean is self-adjoint.
§ 11 Unitarily Invariant Norms

For a vector \( x = (x_1, x_2, \ldots, x_n) \) in \( \mathbf{C}^n \), a norm is defined by:

\[
\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} ; \quad 1 \leq p < \infty,
\]

\[
\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| = \lim_{p \to \infty} \|x\|_p.
\]

For each \( 1 \leq p \leq \infty \), \( \|x\|_p \), gives a norm on \( \mathbf{C}^n \) and these norms are called \( p \)-norms or the \( l_p \)-norms.

Note that the following properties are satisfied by \( p \)-norms.

1. \( \|x\|_p = \|y\|_p \) if \( x \leq y \).
2. \( \|x\|_p \leq \|y\|_p \) if \( |x| \leq |y| \).
3. \( \|x\|_p = \|Px\|_p \), for all \( x \in \mathbf{C}^n \).

Here \( |x| = (|x_1|, |x_2|, \ldots, |x_n|) \), \( |x| \leq |y| \) means that \( |x_i| \leq |y_i| \) for all \( i = 1, 2, \ldots, n \) and \( P \) is a permutation matrix, i.e., a matrix whose each column and each row has exactly one non zero entry which is equal to 1.

A norm on \( \mathbf{C}^n \) is called gauge invariant or absolute if it satisfies the first condition, monotone if it satisfies the second condition, and symmetric if it satisfies the third condition. The first two conditions are equivalent, for a proof see Bhatia [16].

Note that gauge invariant or absolute norms on \( \mathbf{C}^n \) are determined by the norms on \( \mathbf{R}^n \).

A map \( \Phi : \mathbf{R}^n \to \mathbf{R}_+ = \{ x \in \mathbf{R} : x \geq 0 \} \) is called a symmetric gauge function if

(i) \( \Phi \) is a norm,
(ii) $\Phi(Px) = \Phi(x)$ for all $x \in \mathbb{R}^n$ and $P$ a permutation matrix,
(iii) $\Phi(x) = \Phi(|x|)$.
We will always assume that $\Phi$ is normalised, so that 
(iv) $\Phi(1,0,\ldots,0) = 1$.

Here we notice that a symmetric gauge function is a convex function and completely determined by its values on 
$\mathbb{R}^+ = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0, \ i = 1, 2, \ldots, n}\).

If $A$ is a linear operator on $\mathbb{C}^n$, then we shall denote by $\|A\|$ the operator norm of $A$, defined by
$$\|A\| = \sup_{\|x\|=1} \|A(x)\| = s_1(A),$$
where $s_1(A)$ is the largest element of $s(A)$.

If $U, V \in M_n$ are unitary then $\|UAV\| = V^*\|A\|V$ and hence $\|A\| = \|UAV\|$ for all unitary $U, V \in M_n$. This property is known as unitary invariance. Several norms satisfy this property and are known as unitarily invariant norms. We shall use the symbol $\|\|$ to mean a unitarily invariant norm on $M_n$. We will normalise these norms by taking the value 1 on the matrix $\text{diag}(1,0,\ldots,0)$ in $M_n$.

The following theorem establishes a connection between these norms and symmetric gauge functions on $\mathbb{R}^n$.

**Theorem 1.11.1.** Given a symmetric gauge function $\Phi$ on $\mathbb{R}^n$, define a function on $M_n$ as 
$$\|A\|_{\Phi} = \Phi(s(A)),$$
Then this defines a unitarily invariant norm on $M_n$. Conversely, given any unitarily invariant norm $\|\cdot\|$ on $M_n$, define $\Phi_{\|\cdot\|}(x)$:

$$
\Phi_{\|\cdot\|}(x) = \|\text{diag}(x)\|,
$$

where $\text{diag}(x)$ is the diagonal matrix with entries $x_1, x_2, \ldots, x_n$ on its diagonal. Then this defines a symmetric gauge function on $\mathbb{R}^n$.

For a proof of Theorem 1.11.1 see Bhatia (p. 91, [16]).

Thus symmetric gauge functions on $\mathbb{R}^n$ lead to several examples of unitarily invariant norms on $M_n$. Two classes of such norms are specially important. The first is the class of Schatten $p$-norms defined as

$$
\|A\|_p = \Phi_p(s(A)) = \left[ \sum_{j=1}^n (s_j(A))^p \right]^{1/p}, \quad 1 \leq p < \infty,
$$

$$
\|A\|_\infty = \Phi_\infty(s(A)) = s_1(A) = \|A\|.
$$

The second is the class of Ky Fan $k$-norms defined as

$$
\|A\|_{(k)} = \sum_{j=1}^k s_j(A), \quad 1 \leq k \leq n.
$$

$\|A\|_{(n)}$ is also known as trace norm.

The following theorem describes the importance of Ky Fan norms.

**Theorem 1.11.2.** Let $A, B \in M_n$. If

$$
\|A\|_k \leq \|B\|_k, \quad 1 \leq k \leq n;
$$

then

$$
\|A\| \leq \|B\|.
$$
Proof. We shall prove this result for all symmetric gauge functions, which will in turn imply the result for all unitarily invariant norms.

Now
\[ \|A\|_k \leq \|B\|_k, \quad 1 \leq k \leq n \]
i.e.,
\[ \sum_{j=1}^{k} S_j(A) \leq \sum_{j=1}^{k} S_j(B), \quad 1 \leq k \leq n. \]
This implies that there exists a doubly stochastic matrix \( Q \) such that
\[ S(A) = Q(s(B)), \]
see Bhatia (p.39, [16]). Using Birkhoff’s theorem for doubly stochastic matrices and convexity, symmetric nature of the symmetric gauge functions, we obtain
\[ \Phi(S(A)) \leq \Phi(S(B)). \]
This completes the proof.

We also give below some examples of norms, which are not unitarily invariant see Horn and Johnson (p.292, [37]) and Mathias (p.17, [47]).

Let \( A \in \mathbb{M}_n \). The norm \( \|A\| \) defined by:

1. \[ \|A\| = \sum_{i,j=1}^{n} |a_{ij}|. \]
2. \[ \|A\| = n \max_{i=1, j=1}^{n} |a_{ij}|. \]
3. \[ \|A\|_c = \max_{u^*u = u'u = I_n} \{ \text{tr}(C^*AU) : C, U \in \mathbb{M}_n, \text{tr}C \neq 0 \text{ and } C \neq \lambda I_n \}. \]
Here \( \text{tr}X \) denotes the trace of \( X \).

The following results shall be used in the sequel:

**Lemma 1.11.3.** Let \( A, B \) be in \( \mathbb{M}_n \) such that the product \( AB \) is Hermitian. Then for every unitarily invariant norm, we have
\[ \|AB\| \leq \|\text{Re}(BA)\| \]

(Here \( \text{Re}(X) = \frac{X + X^*}{2} \)).

Proof: See Bhatia p.254, [16].

**Theorem 1.11.4.** Let \( A, B, X \) in \( M_n \). Then

\[ \|A^*XB\| \leq \frac{1}{2} \|AA^*X + XBB^*\| \]

for every unitarily invariant norm.

Proof. Case (i) If \( A, B, X \) are Hermitian and \( A = B \), then \( AXA \) is Hermitian, and so by the above lemma, we have

\[ \|AXA\| \leq \|\text{Re}(XA^2)\| = \frac{1}{2} \|A^2X + XA^2\|, \]

which is the required inequality in the special case.

Case (ii) Now let \( A \) and \( B \) be Hermitian and \( X \) be any matrix. Consider the following block matrices

\[ T = \begin{pmatrix} A & 0_n \\ 0_n & B \end{pmatrix}, \quad Y = \begin{pmatrix} 0_n & X \\ X^* & 0_n \end{pmatrix}. \]

By Case (i) for the Hermitian Matrices \( T \) and \( Y \), we have

\[ \|TYT\| \leq \frac{1}{2} \|T^2Y + YT^2\|. \]

Now on using that \( \|S\| = \|S\| \), we obtain

\[ \|AXB\| \leq \frac{1}{2} \|AA^*X + XBB^*\|. \]

Case (iii) Finally, let \( A, B \) and \( X \) be any matrices. Let \( A = A,U \), \( B = B,V \) be polar decompositions of \( A \) and \( B \) respectively. Then

\[ AA^*X + XBB^* = A^2X + XB^2 \]

\[ A^*XB = UA^*XB_{1}V. \]
On using the case (ii) for $A_1$ and $B_1$, we obtain
\[ \left\| A^* B \right\| = \left\| A_1 X B_1 \right\| \leq \frac{1}{2} \left\| A_1^2 X + X B_1^2 \right\| = \frac{1}{2} \left\| A A^* X + X B B^* \right\|. \]

The Theorem 1.11.5 below, which is stated without proof (For a proof see Horn and Johnson [37]), will be used in Theorem 1.11.6.

**Theorem 1.11.5.** Let $A, B \in M_n$ and let $A = X^* Y$. Then
\[ \sum_{i=1}^{k} s_i(A \circ B) \leq \sum_{i=1}^{k} C_i(X) C_i(Y) s_i(B), \]
where $C_i(S)$ denotes the length of $i$th column of the matrix $S$.

**Theorem 1.11.6.** If $A = (a_{ij})$ is positive semidefinite then
\[ \left\| A \circ B \right\| \leq \max_{i,j} |a_{ij}| \left\| B \right\|. \]

**Proof.** Since $A$ is positive semidefinite, we have
\[ A = A^{1/2} A^{1/2}. \]
Now using Theorem 1.11.5 we obtain,
\[ \sum_{i=1}^{k} s_i(A \circ B) \leq \sum_{i=1}^{k} \left( C_i(A^{1/2}) \right)^2 s_i(B). \]
If $A^{1/2} = (t_{ij})$ say, then $\left( C_i(A^{1/2}) \right)^2 = \sum_{j=1}^{n} |t_{ij}|^2$, which is exactly $a_{ii}$. This proves the desired inequality.