Convexity and Linear Distortion

MASSEY UNIVERSITY

Seyed Mohsen Hashemi

Supervisor

Distinguished Professor Gaven J. Martin

Institute of Natural and Mathematical Science
Massey University of Albany
New Zealand

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Mathematics.

December 2017

ABSTRACT

This thesis is primarily concerned with the convexity properties of distortion functionals (particularly the linear distortion) defined on quasiconformal homeomorphisms of domains in Euclidean n-spaces, though we will mainly stick to three-dimensions. The principal application is in identifying the lower semi-continuity of distortion on uniformly convergent limits of sequences of quasiconformal mappings. For example, given the curve family or analytic definitions of quasiconformality - discussed in this thesis - it is known that if $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is a sequence of K-quasiconformal mappings (and here K depends on the particular distortion but is the same for every element of the sequence) which converges to a function $\mathbf{f}$, then the limit function is also K-quasiconformal.
Despite a widespread belief that this was also true for the geometric definition of quasiconformality (via the linear distortion $H(\mathbf{f})$ defined below) Tadeusz Iwaniec gave a specific surprising example to show that the linear distortion function is not lower semicontinuous. The main aim of this thesis is to show that this failure of lower semicontinuity is actually far more common, perhaps even generic in the sense that it might be true that under mild restrictions on a quasiconformal $\mathbf{f}$, there may be a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ with $\mathbf{f_n} \to \mathbf{f}$ uniformly and with $\limsup_{n\to\infty} H(\mathbf{f_n}) < H(\mathbf{f})$. The main result of this thesis is to show this is true for a wide class of linear mappings.

Acknowledgements

First and foremost I offer my sincerest gratitude and respect to my supervisor, **Distinguished Professor Gaven J. Martin** who has supported me throughout my education and thesis. He is one of the best mathematicians, not only for his knowledge in mathematics but also for his great character, humbleness and generosity. He is one of the most cheerful and smiling mathematicians that I have ever seen.
The greatest measure of appreciation goes to my dear wife **Leila** and my beautiful daughter **Haya** for their enduring support and patience. If I am in this position, it is because of my wife's unsparing supports and efforts. Furthermore, I would like to thank and acknowledge my parents for everything that they have given me in my life. They are the best and I have been fortunate enough to have such deserving parents. Finally, I am grateful to the all friendly staff of Institute Natural and Mathematical Science at Massey University of Albany and my best friends Valerie, Saghir, Ming and Faheem for their help.

Contents

List of Figures

List of Tables

Nomenclature

$\|\|A\|\|$	Norm of linear transformation A	19
$B^n(x,r)$	The ball with centre x and radius r	49
$m^*(S)$	Lebesgue outer measure	32
C^∞	Class on functions that they have continuous derivatives of all orders	22
$M_{m\times n}(\mathbb{R})$	Space of all $m \times n$ matrices over the field of real number $\mathbb{R}$	42
χ_S	Characteristic function for S	33
$\Delta\mathbf{f}$	The incremental change of $\mathbf{f}$	23
$\det A_{ij}$	(i,j)-minor of A	3
$\ell(T)$	Minimal stretching $\ell(T)$	35
$\ell(\alpha)$	Length of a path α	53
$\ell_\rho(\gamma)$	ρ-length of a path γ	54
$\ell_f(x,r)$	The minimal derivative of a function $\mathbf{f}$	50
Γ	Curve family	54
$\mathbb{S}^{n-1}$	The unit sphere in $\mathbb{R}^n$	53
$\mathcal{K}_{\alpha,\beta}(A)$	Sectional distortions of A	59
$\mathfrak{M}$	σ-algebra on a set X	27
$\operatorname{adj}(A)$	Adjoint or Adjugate of A	4
$\operatorname{diam}(A)$	The diameter of set A in $\mathbb{R}^n$	49
$\operatorname{dist}(A,B)$	The distance between the sets A and B in $\mathbb{R}^n$	49
$\operatorname{dom}\mathbf{f}$	Domain of $\mathbf{f}$	40
μ	Measure function	28
$\overline{A}$	Conjugate of matrix A with complex entries	9
∂E	The boundary of set E	39
$\upsilon_\rho(A)$	the ρ-volume of a Lebesgue measurable subset A of Ω	54

0.1 Introduction

The principal objects of study in this thesis are quasiconformal mappings, and in particular the convexity properties of the linear distortion are investigated. Quasiconformal mapping are generalizations of conformal mappings, the defining feature is that they are geometric homeomorphisms of "bounded distortion". The notion of distortion can be quantified in various ways and we will see some of them here, indeed the different distortions and their properties is of fundamental interest to us.
Quasiconformal mappings can be considered not only domains in the complex plane and in higher dimensional Euclidean spaces, but also on Riemann surfaces and on Riemannian manifolds in all dimensions, and even on arbitrary metric spaces. However, although our results have consequencese in this framework, we will not discuss these generalisations here.
The importance of quasiconformal mappings in complex analysis was realized by Ahlfors and Teichmüller in the 1930s. Ahlfors used quasiconformal mappings in his geometric approach to Nevanlinna's value distribution theory. He chose the term "quasiconformal" in his work in 1935. Teichmüller used quasiconformal mappings to find and measure a distance between two conformally inequivalent compact Riemann surfaces. For more details and historical progress, you can see "What is a quasiconformal mapping?" by Juha Heinonen or [25].
Let Ω be a domain in $\mathbb{R}^n$, let $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ be a continuous injection. The **linear distortion or dilatation** of $\mathbf{f}$ at the point x in A is the quantity $H_{\mathbf{f}}(x)$ or $H(x, \mathbf{f})$ defined by

$$H_{\mathbf{f}}(x) = H(x, \mathbf{f}) = \limsup_{r \to 0} \frac{L_f(x,r)}{\ell_f(x,r)}, \tag{1}$$

where for $0 < r < \text{dist}(x, \partial\Omega)$ we set

$$L_f(x,r) = \max_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|, \qquad \ell_f(x,r) = \min_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|.$$

In the other words,

$$H_{\mathbf{f}}(x) = H(x, \mathbf{f}) = \limsup_{r \to 0} \frac{\max_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|}{\min_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|}. \tag{2}$$

Let Ω and Ω' be domains in $\mathbb{R}^n$ with $n \geq 2$, and let $\mathbf{f}$ be a homeomorphism of Ω onto Ω'. The **linear distortion of function f** is defined

$$H(\mathbf{f}) = \sup\{H_{\mathbf{f}}(x) \; : \; x \in \Omega\}, \tag{3}$$

where $H_{\mathbf{f}}(x) \geq 1$ is the linear distortion of $\mathbf{f}$ at x as defined at (1). The homeomorphism $\mathbf{f}$ is a conformal mapping if and only if $H(\mathbf{f}) = 1$ (see [18] p.77). Quasiconformal mappings are principally mappings of bounded distortion. There are three ways to define quasiconformal mappings in Euclidean spaces that are called the modulus of curve family, analytic and geometric definition.
The central problem discussed in this thesis concerns the semicontinuity properties of distortion functionals for quasiconformal mappings. In particular, if $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is a sequence of quasiconformal mappings of $\hat{\mathbb{R}}^n$ onto itself and that

$$\mathbf{f_n} \longrightarrow \mathbf{f}, \tag{4}$$

uniformly on $\hat{\mathbb{R}}^n$, we ask for which distortion functionals $\mathcal{K} = \mathcal{K}(\mathbf{g})$, defined for a quasiconformal mapping $\mathbf{g}$ of $\hat{\mathbb{R}}^n$ do we have the lower semicontinuity property

$$\mathcal{K}(\mathbf{f}) \leq \liminf_{n\to\infty} \mathcal{K}(\mathbf{f_n}), \tag{5}$$

so that the limit $\mathbf{f}$ is itself a $\mathcal{K}$-quasiconformal mapping of $\hat{\mathbb{R}}^n$ onto itself. For instance, $\mathcal{K}$ could be the linear distortion, maximal distortion, outer and inner distortion etc. This lower semicontinuity property is related to issues of convexity of the functional $\mathcal{K}$ defined on the space of mappings or more precisely the pointwise differentials of these mappings. We will provide references which show that the functionals determined by the distortion of modulus of curve families and also that determined by the pointwise differential inequality do in fact have the lower semicontinuity property given at (5). For the linear distortion the question of lower semicontinuity has been answered negatively by Tadeusz Iwaniec (see [15]). In his paper [23] he proved one lemma and a theorem that there is a sequence of quasiconformal mappings as at (4) such that for the linear distortion

$$H(\mathbf{f}) > \lim_{n\to\infty} H(\mathbf{f_n}) \tag{6}$$

Notice that the limit here does exist and the limit mapping $\mathbf{f}$ above must be quasiconformal (and so $H(\mathbf{f}) < \infty$) since the other two functional distortions are controlled by the linear distortion functional.

Lemma 0.1.1. ***(Iwaniec's Lemma)*** *Given $n \geq 3$ and $H > 1$, there is a matrix A and a rank-one matrix B and numbers $t, s > 0$ such that*

$$H(A - sB) = H(A + tB) = H < H(A).$$

Theorem 0.1.2. ***(Iwaniec's Theorem)*** *For each $n \geq 3$ and $H > 1$, there exists a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of quasiconformal mappings $\mathbf{f_n} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ converging uniformly to a linear quasiconformal map $\mathbf{f_0} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that*

$$H(x, \mathbf{f_n}) \equiv H < H(x, \mathbf{f_0}), \qquad \textit{almost everywhere in } \mathbb{R}^n \quad n = 1, 2, \ldots$$

These lemma and theorem show that there exists at least one linear mapping T such that the linear distortion is not lower semicontinuous. The key idea in Iwaniec's work is that the linear distortion function fails to be rank-one convex in dimension $n \geq 3$. The linear distortion functional is defined pointwise from the differential matrix when it is nonsingular. At the points $x_0 \in \mathbb{R}^3$ of differentiability of a quasiconformal mapping a local analysis of the convexity properties of the linear distortion can be given if we only consider the nonsingular linear mappings $x \mapsto Tx$, $T = D\mathbf{f}(x_0) \in GL(3, \mathbb{R})$. To compute the linear distortion we study the singular values of

$$A = T^t T \in M_{3\times 3}(\mathbb{R}),$$

the space of symmetric positive definite 3×3 matrices. Given such an A the spectral theorem tells us A is orthogonally diagonalisable and so we may as well suppose it is diagonal since we can diagonalise it by orthogonal (conformal) transformations (at a point). In this way we reduce the problem of the convexity of the linear distortion functional to considering that functional defined on the space of 3×3 positive definite diagonal matrices $A = (a_{ij})$, $a_{ij} = 0$ if

$i \neq j$, all the diagonal entries a_{ii} positive and we can further suppose that $1 = a_{11} \leq a_{22} \leq a_{33}$, by scaling and a further conjugation by orthogonal matrices - neither of which affects the linear distortion.

Next, we know that if B is a rank-one matrix then it can be written as the tensor product of two vectors. In chapter 4 we will find conditions and relations between the entries of matrix A and entries of those two vectors to determine when the following is true:

The Main Question of This Thesis: Assume that

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

is diagonal and $B = \mathbf{u} \otimes \mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^3$ and $1 < a < b$. We want to show that for every matrix A (or for every a and b) there are vectors $\mathbf{u}$ and $\mathbf{v}$ and number t, such that

$$H(A + tB) = H(A + t\mathbf{u} \otimes \mathbf{v}) < H(A), \tag{7}$$

where H is the linear distortion functional.

Here we only consider the case of strict inequality. The method given for Iwaniec was based on an explicit choice of matrix A and rank-one B. Our aim was to find out how general this situation is. Again we can normalise B a little bit further since we can absorb the first entry of $\mathbf{u}$ and the first entry of $\mathbf{v}$ into the parameter t if neither of these entries are 0 as special case which is easier to deal with and gives nothing useful.

We can prove Iwaniec's theorem applies in more general conditions. In the following theorem, we will prove that for all diagonal nonsingular linear mappings T the above property is true.

Theorem 0.1.3. ***(New Result)*** *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a diagonal nonsingular linear mapping. Then there is a sequence of quasiconformal mappings (in fact piecewise linear) $\{T_n\}_{n=1}^{\infty}$ such that*

$$T_n \longrightarrow T \quad \text{uniformly in } \mathbb{R}^n,$$

and

$$H(T_n) = H < H(T).$$

For completeness, in the first two chapters we recall some linear algebra and non-linear analysis that we need to aid understanding our subject. We give proofs for some of these results and elementary consequences we need - such as the description of symmetric rank one matrices. However the expert reader should skip these. In chapter 3, we will present three different notions of convexity, that are named polyconvexity, quasiconvexity and rank-one convexity and their properties and relationships with each other. Again, those well versed in nonlinear analysis and the calculus of variations will not find a great deal that is new. In the next chapter, we define the linear distortion and quasiconformal mappings and their relations. The main new results of this thesis can be found at the end of this chapter and our solution for that. Chapter five is a conclusion and summary of our question and its solution.

Chapter 1

Linear Algebra

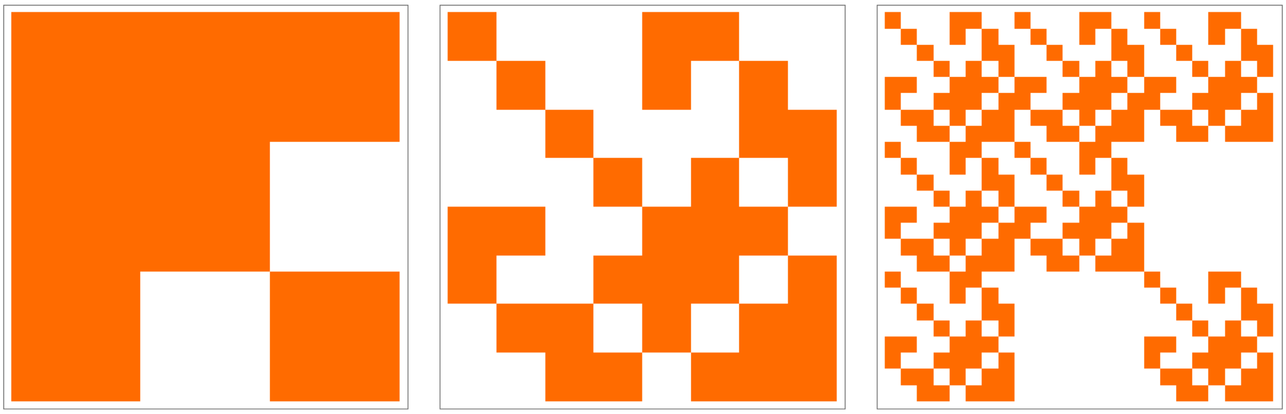

Figure 1.1: Tensor product
Source:www.quantumcalculus.org/tensor-products-everywhere/

Chapter 1

Linear Algebra

This chapter presents an introduction to linear algebra and some of its applications in geometry. It includes six sections that we need for the following chapters. In the first section we review the relations between matrices and determinants. The Second section is about rank of a matrix that is so important for studying matrices. In the third section we recall some definitions and theorems of eigenvalues and eigenvectors. We are interested to find these vectors, and their applications in geometry. Symmetric matrices have a major role in this thesis and we study these subjects in the fourth section. We will see that how we can define **linear distortion** by eigenvalues of matrices of non-linear maps. In the fifth section, we explain the definition of linear transformation and non-linear maps, but we are not interested in studying the relation between linear maps and matrices instead we would like to focus on non-linear maps and matrices. The Sixth section is about the tensor product that we use it in two last chapters. Tensor product has a lot of applications and intuitions in geometry and physics. In mathematics, tensors are geometric objects that describe linear relations between geometric vectors and scalars (see: https://en.wikipedia.org/wiki/Tensor).

1.1 Matrices and Determinants

In this section we try to define and recall determinant and its properties. Determinant is calculated by a function such that is called a **determinant function**. It is a real-valued function of a matrix variable.

Definition 1.1.1. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. We define the determinant of A (written $\det(A)$ or $|A|$) by

$$\det(A) = |A| = \sum \pm a_{1j_1} a_{2j_2} \cdots a_{nj_n} \tag{1.1}$$

where the summation ranges over all permutations $j_1 j_2 \cdots j_n$ of the set $S = \{1, 2, \cdots, n\}$ the sign is taken + or - according to whether the permutation $j_1 j_2 \cdots j_n$ is even or odd.[26]

This definition shows us that if A be a square matrix then, $\det(A)$ is a function that it associates a real number $\det(A)$ with a matrix A.

Example 1.1.1. *If* $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ *and* $B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, *then we can calculate their determinants with formula (1.1) and write*

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

and

$$\det(B) = |B| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}-a_{11}a_{23}a_{32}.$$

♣

Now, we recall some theorems and properties of the determinant that are useful for the following sections and chapters. We accept some of them without proof, for seeing their proofs you can see [26],[3] or [22].

Theorem 1.1.1. *Let A be a square matrix.*
(a) If A has a row of zeros or a column of zeros,then $\det(A) = 0$
(b) $\det(A) = \det(A^t)$

Theorem 1.1.2. *If A is an $n \times n$ triangular matrix (upper or lower triangular or diagonal),then* $\det(A)$ *is the product of the entries the main diagonal;that is,* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Theorem 1.1.3. *Let A be an $n \times n$ matrix.*
(a) If B is the matrix that results when a single row or single column of A is multiplied by a scalar k,then $\det(B) = k \cdot \det(A)$.
(b) If B is the matrix that results when two rows or two columns of A are interchanged, then $\det(B) = -\det(A)$.
(c) If B is the matrix that results when a multiple of one row of A is added to another row or when a multiple of one column is added to another column, then $\det(B) = \det(A)$.

Theorem 1.1.4. *If A is a square matrix with two proportional rows or two proportional columns, then* $\det(A) = 0$.

Theorem 1.1.5. *If A is an $n \times n$ matrix and k is a scalar, then* $\det(k.A) = k^n \cdot \det(A)$.

Theorem 1.1.6. *A square matrix A is invertible if and only if* $\det(A) \neq 0$.

Theorem 1.1.7. *If A and B are square matrices of the same size,then* $\det(A.B) = \det(A) \cdot \det(B)$.

Theorem 1.1.8. *If A is invertible, then*

$$\det(A^{-1}) = \frac{1}{\det(A)}. \tag{1.2}$$

The proofs of all theorems exist in every linear algebra books that they are suggested for Bachelor of Science. To see the proofs you can see [3]. If $A = [a_{ij}]$ is a square matrix, we denote A_{ij} as a submatrix A by deleting row i and column j. For any square matrix A, $\det A_{ij}$ is called the $(\mathbf{i}, \mathbf{j}) -$ **minor of A** . We define the $(\mathbf{i}, \mathbf{j}) -$ **cofactor of A** to be

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

The **cofactor matrix of** A is the matrix C whose (i, j) entry is the (i,j)-cofactor of A. The **adjoint** or **adjugate** of A is the transpose of the cofactor matrix C of A (see [31]). We denote adjugate of A by $\text{adj}(A)$.

$$\text{adj}(A) = C^t.$$

There are some properties of adj(A) that we can list them:

$$\begin{aligned} &A^{-1}\cdot\det(A)=\operatorname{adj}(A) && \operatorname{adj}(A.B)=\operatorname{adj}(B)\cdot\operatorname{adj}(A),\\ &\operatorname{adj}(I)=I, && \operatorname{adj}(cA)=c^{n-1}\operatorname{adj}(A),\\ &\operatorname{adj}(A^n)=\operatorname{adj}(A)^n, && \operatorname{adj}(A^t)=\operatorname{adj}(A)^t, \end{aligned} \tag{1.3}$$

Furthermore, if A is a $n\times n$ matrix with $n\geq 2$, then $\det(\operatorname{adj}(A))=\det(A)^{n-1}$, and if A is an invertible $n\times n$ matrix, then $\operatorname{adj}(\operatorname{adj}(A))=\det(A)^{n-2}A$.

1.2 Rank of Matrices

In this section we shall show three important vector spaces associated with matrices. The following definition defines those three vector spaces.

Definition 1.2.1. If A is $m\times n$ matrix, then the subspace of $\mathbb{R}^n$ spanned by the row vectors of A is called the **row space** of A, and the subspace of $\mathbb{R}^m$ spanned by the column vectors is called the **column space** of A. The solution space of the homogeneous system of equations $A\mathbf{x}=\mathbf{0}$, which is a subspace of $\mathbb{R}^n$, is called the **nullspace** of A.

In this section we want to answer to the following general questions:

- What relationships exist between the solutions of a linear system $A\mathbf{x}=\mathbf{b}$ and the row space, column space and nullspace of the coefficient matrix A?
- What relationships exist between the row space, column space and nullspace of a matrix?

Now,to answer these questions, we describe some theorems here. The reader can find the complete proofs in [3], [26] or [22].

Theorem 1.2.1. *For any matrix we can say:*
(a) Elementary row operations do not change the row space of a matrix.
(b) Elementary row operations do not change the nullspace of a matrix.

Remark. Although elementary row operations can change the column space of a matrix.

Theorem 1.2.2. *If A and B are row equivalent matrices, then:*
(a) A given set of column vectors of A is linearly independent if and only if the corresponding column vectors of B are linearly independent.
(b) A given set of column vectors of A forms a basis for the column space of A if and only if the corresponding column vectors of B forms a basis for the column space of B.

Theorem 1.2.3. *If a matrix A is in row-echelon form, then the row vectors with the leading $1'$s form a basis for the row space of A, and the column vectors with the leading $1'$s of the row vectors form a basis for the column space of R (R is a matrix that it is row equivalent with matrix A).*

Theorem 1.2.4. *If A is any matrix, then the row space and column space of A have the same dimension.*

Definition 1.2.2. The common dimension of the row space and column space of a matrix A is called the **rank** of A and is denoted by rank(A); the dimension of nullspace of A is called the **nullity** of A and is denoted by nullity(A).

We can also prove that $rank(A) = rank(A^t)$, because

$$rank(A) = dim(row\ space\ of\ A) = dim(column\ space\ of\ A^t) = rank(A^t).$$

Furthermore, we can explain the **Dimension Theorem** for matrices.
Theorem 1.2.5. *If A is a matrix with n column, then*

$$rank(A) + nullity(A) = n.$$

For seeing the complete proof you can see [3]. With this theorem we can say if A is an $m \times n$ matrix, then:
(1) rank(A)= the number of leading variables in the solution of $A\mathbf{x} = \mathbf{0}$.
(2) nullity(A)= the number of parameters in the solution of $A\mathbf{x} = \mathbf{0}$.

1.3 Eigenvalue and Diagonalization

We begin with a review of some definition and concepts. This section is useful for next chapters and we use these result for our problem.

Definition 1.3.1. If A is an $n \times n$ matrix, then a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ is called an **eigenvector** of A. If $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$; that is,

$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ. The scalar λ is called **eigenvalue** of A and $\mathbf{x}$ is said to be an eigenvector of A corresponding to λ.

To find the eigenvalues of an $n \times n$ matrix A, we can consider $A\mathbf{x} = \lambda\mathbf{x}$ as

$$A\mathbf{x} = \lambda I\mathbf{x}$$

or equivalently,

$$(A - \lambda I)\mathbf{x} = \mathbf{0}.$$

For λ to be an eigenvalue, there must be a nonzero solution of this equation. So, we have

$$det(A - \lambda I) = 0.$$

This is called the **characteristic equation** of A; the scalars satisfying this equation are eigenvalues of A. If A is an $n \times n$ matrix, then

$$det(A - \lambda I) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n$$

is called the **characteristic polynomial** of A.

It is obvious to see if A is an $n \times n$ triangular or diagonal matrix then the the eigenvalues of matrix A are the entries on the main diagonal of A. Also, by induction we can see if k is a positive integer, λ is an eigenvalue of a matrix A, and $\mathbf{x}$ is a corresponding eigenvector to λ, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector to λ^k.

Theorem 1.3.1. *A square matrix A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.*

To see proof of this theorem you can see [3] or [26].

Definition 1.3.2. A square matrix A is called **diagonalizable** if there is an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix; the matrix P is said to **diagonalize** A.

Now, we can describe three important theorems with their proof, because they are very useful for the rest of this thesis.

Theorem 1.3.2. *If A is an $n \times n$ matrix, then the following are equivalent.*
(a) A is diagonalizable.
(b) A has n linearly independent eigenvectors.

Proof. Assume A is diagonalizable,there is an invertible matrix P such that $P^{-1}AP$ is diagonal,say $P^{-1}AP = D$, where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

It follows from the formula $P^{-1}AP = D$ that $AP = PD$; that is,

$$AP = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}. \qquad (1.4)$$

If we now let $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_n}$ denote column vectors of P, then from (1.4) the successive columns of AP are $\lambda_1\mathbf{P_1}, \lambda_2\mathbf{P_2}, \ldots, \lambda_n\mathbf{P_n}$. However, we can say the successive columns are $A\mathbf{P_1}, A\mathbf{P_2}, \ldots, A\mathbf{P_n}$. Thus, we have

$$A\mathbf{P_1} = \lambda_1\mathbf{P_1}, \qquad A\mathbf{P_2} = \lambda_2\mathbf{P_2}, \quad \cdots \quad , A\mathbf{P_n} = \lambda_n\mathbf{P_n}. \qquad (1.5)$$

Since P is invertible, its column vectors are all nonzero; it follows from (1.5) that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A, and $\mathbf{P_1}, \mathbf{P_2}, \ldots, \mathbf{P_n}$ are corresponding eigenvectors. Since P is invertible then it has linearly independent columns, so, $\mathbf{P_1}, \mathbf{P_2}, \cdots, \mathbf{P_n}$ are linearly independent. Thus, A has n linearly independent eigenvectors.
To prove of converse of theorem assume A has n linearly independent eigenvectors $\mathbf{P_1}, \mathbf{P_2}, \cdots, \mathbf{P_n}$, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and let

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

be a matrix whose column vectors are $\mathbf{P_1}, \mathbf{P_2}, \cdots, \mathbf{P_n}$. The column vectors of the product AP are

$$A\mathbf{P_1},\ A\mathbf{P_2}, \ldots,\ A\mathbf{P_n}.$$

But

$$A\mathbf{P_1} = \lambda_1 \mathbf{P_1}, \qquad A\mathbf{P_2} = \lambda_2 \mathbf{P_2}, \quad \cdots \quad , A\mathbf{P_n} = \lambda_n \mathbf{P_n}, \tag{1.6}$$

so that

$$AP = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD \tag{1.7}$$

where D is the diagonal matrix having the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the main diagonal. Since the column vectors of P are linearly independent, P is invertible; thus, (1.7) can be rewritten as $P^{-1}AP = D$; that is, A is diagonalizable. ♠

Theorem 1.3.3. *If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, then $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}\}$ is a linearly independent set.*

To see the proof of this theorem you can see page 369 of [3]. As a consequence of theorem (1.3.3), we obtain the following important result.

Theorem 1.3.4. *If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.*

Proof. If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then by theorem (1.3.3), $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ are linearly independent. Thus, by theorem (1.3.2) A is diagonalizable (see [3]). ♠

1.4 Symmetric and Orthogonal Matrix

In this section we shall show consider certain classes of matrices that have special forms. The matrices that we explain in this section are among the most important in linear algebra.

Definition 1.4.1. A square matrix A is called **symmetric** if $A = A^t$.

It is easy to recognise symmetric matrices by inspection:
The entries on the main diagonal may be arbitrary, but mirror images of entries across the main diagonal must be equal [3]. We can say, a matrix $A = \left[a_{ij}\right]$ is symmetric if and only if $a_{ij} = a_{ji}$ for all values of i and j.

Theorem 1.4.1. *If A and B are symmetric matrices with the same size, and if k is any scalar, then:*
(a) A^t *is symmetric.*
(b) $A + B$ *and* $A - B$ *are symmetric.*
(c) kA *is symmetric.*

Proof of this theorem is easy and we don't try to prove it (you can see [26]).

Definition 1.4.2. If A and B are square matrices with same size such that $AB = BA$, then A and B are called **commutative**

It is not true, in general, that the product of symmetric matrices is symmetric. But it is easy to see, the product of two symmetric matrices is symmetric if and only if the matrices commute.

Theorem 1.4.2. *If A is an invertible symmetric matrix, then A^{-1} is symmetric.*

Proof. Assume that A is symmetric and invertible. Since A is symmetric, so, $A = A^t$. In linear algebra by some theorems we know that $(A^t)^{-1} = (A^{-1})^t$. Therefore

$$(A^{-1})^t = (A^t)^{-1} = A^{-1}$$

which proves that A^{-1} is symmetric. ♠

Theorem 1.4.3. *If A is an invertible matrix, then AA^t and A^tA are also invertible.*
Definition 1.4.3. A square matrix A with the property

$$A^{-1} = A^t$$

is said to be an **orthogonal matrix**

It follows from this definition that a square matrix A is orthogonal if and only if

$$AA^t = A^tA = I.$$

Example 1.4.1. *The matrix*

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

is orthogonal, since

$$A^tA = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

♣

Theorem 1.4.4. *For orthogonal matrices, we can prove the following statements:*
(a) The inverse of an orthogonal matrix is orthogonal.
(b) A product of orthogonal matrices is orthogonal.
(c) If A is orthogonal, then $\det(A) = 1$ *or* $\det(A) = -1$.

Definition 1.4.4. Let A is an $n \times n$ matrix. If there is orthogonal matrix P such that the matrix $P^{-1}AP$ is orthogonal, then A is called to be **orthogonally diagonalizable** and P is said to **orthogonally diagonalize** A.

Definition 1.4.5. A **quadratic form** in the n variables $x_1, x_2, \ldots, x_n$ is an expression that can be written as

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1.8}$$

where A is a symmetric $n \times n$ matrix. If we let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ then formula (1.8) can be written more compactly as

$$\mathbf{x^t} A\mathbf{x}. \tag{1.9}$$

If we use the fact that A is symmetric, that is, $A = A^t$, then the formula (1.9) can be expressed in terms of the Euclidean inner product by writing

$$\mathbf{x^t} A\mathbf{x} = \mathbf{x^t}(A\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A\mathbf{x}\rangle.$$

For more information you can see [3], [26] or [22].

Theorem 1.4.5. *Let A be a symmetric $n \times n$ matrix whose eigenvalues in decreasing size order are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. If $\mathbf{x}$ is constrained so that $\|\mathbf{x}\| = 1$ relative to the Euclidean inner product on $\mathbb{R}^{\ltimes}$, then :*
(a) $\lambda_1 \geq \mathbf{x^t} A\mathbf{x} \geq \lambda_n$.
(b) $\mathbf{x^t} A\mathbf{x} = \lambda_n$ *if $\mathbf{x}$ is an eigenvector of A corresponding to λ_n and $\mathbf{x^t} A\mathbf{x} = \lambda_1$ if $\mathbf{x}$ is an eigenvector of A corresponding to λ_1.*

We don't want to prove this theorem. Its proof exists in any linear algebra book. You can study page 480 of [3].

Definition 1.4.6. A quadratic form $\mathbf{x^t} A\mathbf{x}$ is called **positive definite** if $\mathbf{x^t} A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$, and the symmetric matrix A is called a **positive definite matrix**.

There is a property that a symmetric matrix A is positive definite if and only if all the eigenvalues of A are positive. Also we can prove a symmetric matrix A is positive definite if and only if the determinant of every principle submatrix is positive. For matrices with real entries, the orthogonal matrices and symmetric matrices play an important role in the orthogonal diagonalization problem. But we need something for matrices with complex entries. We review some definitions of matrices with complex entries. If A is a matrix with complex entries, then the **conjugate transpose** of A, denoted by A^*, is defined by

$$A^* = \overline{A}^t. \tag{1.10}$$

If A and B are matrices with complex entries and k is any complex number, then we can prove some properties like $(A^*)^* = A$, $(A+B)^* = A^* + B^*$, $(kA)^* = \overline{k}A^*$ or $(AB)^* = B^*A^*$. By definition (1.4.3) a matrix with real entries is called orthogonal if $A^{-1} = A^t$. For complex entries matrix, this property is called unitary matrix.

Definition 1.4.7. A square matrix A with complex entries is called **unitary** if

$$A^{-1} = A^*. \tag{1.11}$$

Recall that a square matrix A with real entries is called orthogonally diagonalizable if there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal. For complex matrices we have an analogous concept. A square matrix A with complex entries is called **unitarily diagonalizable** if there is a unitary P such that $P^{-1}AP = P^*AP$ is diagonal; the matrix P is said to **unitarily diagonalize** A.

Definition 1.4.8. A square matrix A with complex entries is called **Hermitian** or **self-adjoint** if

$$A = A^*.$$

It is easy to recognise Hermitian matrices by inspection: The entries on the main diagonal are real numbers, and the mirror of each entry across the main diagonal is its complex conjugate. Also, a matrix A with complex entries is called **normal** if

$$AA^* = A^*A.$$

We have two results with these definitions such that we will use them in the next chapters (you can see [3] or [26]).

Theorem 1.4.6. *The eigenvalues of a Hermitian matrix are real numbers.*

Proof. If λ is an eigenvalue and $\mathbf{v}$ a corresponding eigenvector of $n \times n$ Hermitian matrix A, then

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Multiplying each side of this equation on the left by the conjugate transpose of $\mathbf{v}$ yields

$$\mathbf{v}^*A\mathbf{v} = \mathbf{v}^*(\lambda\mathbf{v}) = \lambda\mathbf{v}^*\mathbf{v}. \tag{1.12}$$

We will show that the 1×1 matrices $\mathbf{v}^*A\mathbf{v}$ and $\mathbf{v}^*\mathbf{v}$ both have real entries, so it will follow from (1.12) that λ must be real number. Both $\mathbf{v}^*A\mathbf{v}$ and $\mathbf{v}^*\mathbf{v}$ are Hermitian, since

$$(\mathbf{v}^*A\mathbf{v})^* = \mathbf{v}^*A^*(\mathbf{v}^*)^* = \mathbf{v}^*A\mathbf{v}$$

and

$$(\mathbf{v}^*\mathbf{v})^* = \mathbf{v}^*(\mathbf{v}^*)^* = \mathbf{v}^*\mathbf{v}.$$

Since Hermitian matrices have real entries on the main diagonal, and since $\mathbf{v}^*A\mathbf{v}$ and $\mathbf{v}^*\mathbf{v}$ are 1×1, it follows that these matrices have real entries, which completes the proof. ♠

Theorem 1.4.7. *The eigenvalues of a symmetric matrix with real entries are real numbers.*

Proof. Let A be a symmetric matrix with real entries. Because the entries in A are real, it follows that

$$\overline{A} = A.$$

But this implies that A is Hermitian, since

$$A^* = (\overline{A})^t = A^t = A.$$

Thus, A has real eigenvalue by theorem (1.4.6). ♠

Now, we should review some properties of diagonalizable matrices. An $n \times n$ matrix A is called diagonalizable if it can be written as $A = P^{-1}DP$, where D is a diagonal matrix. This is possible if and only if there is a basis $B = \{\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_n}\}$ for $\mathbb{R}^n$ where $\mathbf{b_i}$ are eigenvectors of A. The corresponding eigenvalues sit along the diagonal of D, and the matrix $P = \begin{bmatrix}\mathbf{b_1} & \mathbf{b_2}, \ldots, \mathbf{b_n}\end{bmatrix}$. Thus $P = P_B$, the "change of coordinate" matrix $P_B[\mathbf{x}]_B = \mathbf{x}$ and $(P_B)^{-1}(\mathbf{x}) = [\mathbf{x}]_B$. An orthogonal matrix is a square matrix for which $A^{-1} = A^t$; equivalently, an orthogonal matrix is a square matrix with orthonormal columns.

Theorem 1.4.8. ***(The Spectral Theorem)*** *A (real) $n \times n$ matrix A is orthogonally diagonalizable if and only if A is symmetric.*

Proof. Assume A is orthogonally diagonalizable, then there are diagonal matrix D and orthogonal matrix P that $A = PDP^{-1} = PDP^t$. We can write

$$A^t = (PDP^t)^t = (P^t)^t D^t P^t = PDP^t = A,$$

so, A is symmetric.
Conversely, we need to show that a (real) $n \times n$ symmetric matrix A is orthogonally diagonalizable. We prove this part by mathematical induction. This is obviously true for every 1×1 matrix A: if $A = [a]$, then $A = [1][a][1] = PAP^t$. Assume that every $(n-1) \times (n-1)$ symmetric matrix is orthogonally diagonalizable that we show it with $(\star)$.
We will show that $(\star)$ forces it to be true that every $n \times n$ symmetric matrix must also be orthogonally diagonalizable. Consider an $n \times n$ symmetric matrix A where $n > 1$. By the preceding theorem, we can find a real eigenvalue λ_1 of A, together with a real eigenvector $\mathbf{v_1}$. By normalizing, we can assume $\mathbf{v_1}$ is a unit eigenvector. Add vector to extend $\{\mathbf{v_1}\}$ to a basis for $\mathbb{R}^n$ and the use the **Gram-Schmidt** process to get an orthonormal basis for $\mathbb{R}^n$: $B = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$.
Let $P = P_B = [\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}]$ = the change of coordinate matrix for B. Because P is orthogonal, $P^{-1} = P^t$. Now look at the matrix $P^{-1}AP$.
$P^{-1}AP$ is symmetric, because

$$(P^{-1}AP)^t = (P^tAP)^t = P^tA(P^t)^t = P^tAP = P^{-1}AP,$$

and its first column is

$$P^{-1}AP\mathbf{e_1} = P^{-1}A\mathbf{v_1} = P^{-1}\lambda_1\mathbf{v_1} = \lambda_1 P^{-1}\mathbf{v_1} = \lambda_1[\mathbf{v_1}]_B = \lambda_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Using the symmetry, partition $P^{-1}AP$ as a "block matrix"

$$\begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & F \end{bmatrix},$$

where $\mathbf{0}$ is a block with $(n-1)$ zeros, and F is a symmetric matrix. Then F has size $(n-1) \times (n-1)$, so our assumption $(\star)$ says that F is orthogonally diagonalizable. This means there is a diagonal matrix D' and an $(n-1) \times (n-1)$ orthogonal matrix Q for which

$$F = QD'Q^{-1} \quad \textit{or} \quad Q^{-1}FQ = D'.$$

Define a partitioned $n \times n$ matrix $R = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix}$. Since

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q^{-1} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix}.$$

We can see R is invertible: $R^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q^{-1} \end{bmatrix}$.

Let $U = PR$, since a product of orthogonal matrices is orthogonal (see theorem 1.4.4 part (b)), then U is orthogonal. Then

$$U^{-1}AU = (R^{-1}P^{-1})A(PR) = R^{-1}(P^{-1}AP)R$$

$$= R^{-1}\begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & F \end{bmatrix} R = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q^{-1} \end{bmatrix}\begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & F \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & Q^{-1}F \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & Q^{-1}FQ \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D' \end{bmatrix},$$

so $A = U\begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D' \end{bmatrix}U^{-1}$. Since $D = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D' \end{bmatrix}$ is orthogonal matrix, we have an orthogonal diagonalization of A:

$$A = UDU^{-1} = UDU^t.$$

This finishes the proof. ♠

For more information you can see (www.math.wustl.edu/~freiwald/309orthogdiag.pdf). We will finish this section with important theorem.

Theorem 1.4.9. *Let A be an $n \times n$ real symmetric matrix, and let*

$$Q(\mathbf{y}) = \mathbf{y}A\mathbf{y}^t = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}y_iy_j.$$

Then we have
(a) $Q(\mathbf{y}) > 0$ for all vectors $\mathbf{y} \neq 0$ (A is positive definite) if and only if all eigenvalues of A are positive real numbers.
(b) $Q(\mathbf{y}) < 0$ for all vectors $\mathbf{y} \neq 0$ (A is negative definite) if and only if all eigenvalues of A are negative real numbers.

Proof. According to the spectral theorem there is an orthogonal matrix C that reduces the quadratic form $\mathbf{y}A\mathbf{y}^t$ to a diagonal form. That is

$$Q(\mathbf{y}) = \mathbf{y}A\mathbf{y}^t = \sum_{i=1}^{n} \lambda_i x_i^2 \tag{1.13}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is the row matrix $\mathbf{x} = \mathbf{y}C$, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. The eigenvalues are real since A is symmetric. If all eigenvalues are positive, equation (1.13) shows that $Q(\mathbf{y}) > 0$ whenever $\mathbf{x} \neq 0$. But since $\mathbf{x} = \mathbf{y}C$ we have $\mathbf{y} = \mathbf{x}C^{-1}$, so $\mathbf{x} \neq 0$ if and only if $\mathbf{y} \neq 0$. Therefore $Q(\mathbf{y}) > 0$ for all vector $\mathbf{y} \neq 0$.
Conversely, if $Q(\mathbf{y}) > 0$ for all vector $\mathbf{y} \neq 0$ we can choose $\mathbf{y}$ so that $\mathbf{x} = \mathbf{y}C$ is the kth coordinate vector $\mathbf{e_k}$. For this $\mathbf{y}$ equation (1.13) gives us $Q(\mathbf{y}) = \lambda_k$, so each $\lambda_k > 0$. This proves part (a). The proof of (b) is entirely analogous. ♠

The last concept we will describe here is the singular values of a matrix that is used in chapters (4) and (5). If A be an $m \times n$ matrix, then $n \times n$ matrix A^tA is a symmetric matrix and hence can be orthogonally diagonalized, by the spectral theorem. Not only are the eigenvalues of A^tA all real (theorem (1.4.7)), they are all nonnegative.

Definition 1.4.9. If A is an $m \times n$ matrix, the **singular values** of A are the square roots of the eigenvalues of A^tA and are denoted by $\sigma_1, \sigma_2, \ldots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$.

We explain the following extremly important theorems and properties of singular values of a matrix A without proofs. For more details and proofs, you can see [31] pages 590 to 599.
Theorem 1.4.10. *If A is an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix*

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

such that

$$A = U\Sigma V^t,$$

where $1 \leq r \leq n$.

The factorisation of A as in theorem (1.4.10)is called a **singular value decomposition** or **SVD** of A. The columns of U are called **left singular vectors** of A, and the columns of V are called **right singular vectors** of A (see [31]). The SVD provides new geometric insight into the effect of transformations. We have said several times that an $m \times n$ matrix transforms the unit sphere in $\mathbb{R}^n$ into an ellipsoid in $\mathbb{R}^m$. For this propose, we can express the following theorem which is proved in [31] on page 598.

Theorem 1.4.11. *If A is an $m \times n$ matrix with rank r. Then the image of the unit sphere in $\mathbb{R}^n$ under the matrix transformation that maps $\mathbf{x}$ to $A\mathbf{x}$ is*

- *the surface of an ellipsoid in $\mathbb{R}^m$ if $r = n$.*
- *a solid ellipsoid in $\mathbb{R}^m$ if $r < n$.*

In general, we can describe the effect of an $m \times n$ matrix A on the unit sphere in $\mathbb{R}^n$ in terms of the effect of each factor in its SVD, $A = U\Sigma V^t$, from right to left. Since V^t is an orthogonal matrix, it maps the unit sphere to itself. The $m \times n$ matrix Σ does two things:

- The diagonal entries $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$ collapse n - r of the dimensions of the unit sphere, leaving an r-dimensional unit sphere, which the nonzero diagonal entries $\sigma_1, \sigma_2, \ldots, \sigma_r$ then distort into an ellipsoid.
- The orthogonal matrix U then aligns the axes of this ellipsoid with the orthonormal basis vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}$ in $\mathbb{R}^m$

For more information about theorems and applications of singular values of matrix A and its geometric approach, we refer you to [31] chapter 7 section 4.

1.5 Linear and Non-Linear Transformations

Recall that a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ was first defined as a function

$$T(v_1, v_2, \ldots, v_n) = (w_1, w_2, \ldots, w_m)$$

for which the equations relating $w_1, w_2, \ldots, w_m$ and $v_1, v_2, \ldots, v_n$ are linear. If these equations relating between $w_1, w_2, \ldots, w_m$ and $v_1, v_2, \ldots, v_n$ are not linear then this transformation is called **non-linear**. We review some properties of linear transformations.

Definition 1.5.1. If $T : V \longrightarrow W$ is a function from a vector space V into a vector space W, then T is called a **linear transformation** from V to W if for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and all scalars c

$$T(c\mathbf{u} + \mathbf{v}) = cT(\mathbf{u}) + T(\mathbf{v}).$$

In the special case $V = W$, the linear transformation $T : V \longrightarrow V$ is called a **linear operator** on V.

We shall call linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ **matrix transformations**, since they can be carried out by matrix multiplication. We can show that if $T : V \longrightarrow W$ is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$ or for all $\mathbf{u}$ and $\mathbf{v}$ in V, we can write

$$T(-\mathbf{v}) = -T(\mathbf{v})$$

and

$$T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}).$$

Definition 1.5.2. If $T : V \longrightarrow W$ is a linear transformation, the set of vectors in V that T maps into $\mathbf{0}$ is called the **kernel** of T; it is denoted by $ker(T)$. The set of all vectors in W that are images under T of at least one vector in V is called the **range** of T; it is denoted by $\mathrm{R}(T)$.

If $T : V \longrightarrow W$ is a linear transformation, then we can show that the kernel of T is a subspace of V and the range of T is a subspace of W. If $T : V \longrightarrow W$ is a linear transformation, then the dimension of the range of T is called the **rank of** T and is denoted by $\mathrm{rank}(T)$; the dimension of the kernel is called the **nullity of** T and is denoted by $\mathrm{nullity}(T)$. In section (1.2) we said some definitions and properties of rank and nullity of matrices. Now we can describe the relations between these definitions. If A an $m \times n$ matrix and $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is multiplication by A, then

$$nullity(T_A) = nullity(A)$$

and

$$rank(T_A) = rank(A).$$

We can explain **Dimension Theorem for Linear Transformation** like theorem (1.2.5).

Theorem 1.5.1. *If $T : V \longrightarrow W$ is a linear transformation from an n-dimensional vector space V to a vector space W, then*

$$rank(T) + nullity(T) = n$$

You see proof of this theorem in page 399 of [3].

1.6 Tensor Product and Rank-One Matrices

In this section we explain definitions and properties of new product of vectors. We had before inner product and cross product, but we need a new subject for another concepts.

Definition 1.6.1. (tensor product of vectors). If $\mathbf{u}$ and $\mathbf{v}$ are vectors on length m and n, respectively, their tensor product $\mathbf{u} \otimes \mathbf{v}$ is defined as an $m \times n$ matrix defined by $(\mathbf{u} \otimes \mathbf{v})_{ij} = u_i v_j$. In other word,

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathbf{t}}.$$

This is called the **outer product** of vectors $\mathbf{u}$ and $\mathbf{v}$. The outer product contrasts with the dot product, which takes as input a pair of coordinate vectors and products a scalar. For complex vectors, it is customary to use the conjugate transpose of $\mathbf{v}$, so in this case we can write

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^*$$

Example 1.6.1. *If* $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, *then the tenor product or outer product of vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *is*

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathbf{t}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \cdot \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \\ u_4v_1 & u_4v_2 & u_4v_3 \end{bmatrix}.$$

♣

Theorem 1.6.1. *If A is an $m \times n$ matrix then A is a rank-one matrix if and only if there two vectors $\mathbf{u}$ and $\mathbf{v}$ with m and n components, respectively such that*

$$A = \mathbf{u} \otimes \mathbf{v} \tag{1.14}$$

Proof. Assume A is a rank-one matrix. Having rank-one means that the columns of A are linearly dependent. So, if

$$A = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \dots & \mathbf{u_n} \end{bmatrix}$$

where $\mathbf{u_i}$ are columns of A with m components. Then, we can write

$$\mathbf{u_i} = v_i \mathbf{u}$$

where $i = 1, 2, \dots, n$ and $\mathbf{u}$ is a column matrix with m components. Therefore, the matrix A can be described as

$$A = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \dots & \mathbf{u_n} \end{bmatrix} = \begin{bmatrix} v_1\mathbf{u} & v_2\mathbf{u} & \dots & v_n\mathbf{u} \end{bmatrix} = \mathbf{u} \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} = \mathbf{u}\mathbf{v}^{\mathbf{t}} = \mathbf{u} \otimes \mathbf{v}.$$

For the other side we can write the relations conversely and this finishes the proof. ♠

It is easy to show that, if the rank of a real symmetric matrix is bigger than or equal to 1, then the diagonal elements of the matrix can not be all zero. We leave this proof for readers. Now, we want to explain another result for rank-one matrices. This theorem will be proved in 3-dimension, the general proof for n-dimension will be similar. It is used in the analysis of rank-one convexity for the linear distortion later.

Theorem 1.6.2. *Let A be a 3×3 rank-one symmetric matrix. Then A can be written as*

$$A = \lambda(1, x, y) \otimes (1, x, y). \tag{1.15}$$

In the other words, the matrix A can be expressed by

$$A = \lambda \mathbf{u} \otimes \mathbf{u} = \lambda \mathbf{u}\mathbf{u}^t,$$

where $\mathbf{u} = (1, x, y)$ is a vector in $\mathbb{R}^3$.

Proof. Since A is a rank-one matrix, by theorem (1.6.1), there are two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ that

$$A = \mathbf{v} \otimes \mathbf{w} = \mathbf{v} \cdot \mathbf{w}^t.$$

The vectors $\mathbf{v_1} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ and $\mathbf{w_1} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ are unit vectors. So, $\mathbf{v} = \|\mathbf{v}\|\mathbf{v_1}$ and $\mathbf{w} = \|\mathbf{w}\|\mathbf{w_1}$. Thus,

$$A = \mathbf{v} \cdot \mathbf{w}^t = \|\mathbf{v}\|\|\mathbf{w}\|\mathbf{v_1}\mathbf{w_1}^t.$$

If $\|\mathbf{v}\|\|\mathbf{w}\| = c \in \mathbb{R}$, then $A = c\mathbf{v_1}\mathbf{w_1}^t$. Since A is symmetric, we get

$$\mathbf{v_1}\mathbf{w_1}^t = \mathbf{w_1}\mathbf{v_1}^t, \tag{1.16}$$

where $\mathbf{v_1}$ and $\mathbf{w_1}$ are unit vectors. If the equation (1.16) is multiplied on the left by $\mathbf{w_1}^t$ and on the right by $\mathbf{v_1}$,

$$\begin{aligned}
\Rightarrow \quad & \mathbf{w_1}^t(\mathbf{v_1}\mathbf{w_1}^t)\mathbf{v_1} = \mathbf{w_1}^t(\mathbf{w_1}\mathbf{v_1}^t)\mathbf{v_1} \\
\Rightarrow \quad & (\mathbf{w_1}^t\mathbf{v_1})(\mathbf{w_1}^t\mathbf{v_1}) = (\mathbf{w_1}^t\mathbf{w_1})(\mathbf{v_1}^t\mathbf{v_1}) \\
\Rightarrow \quad & (\mathbf{w_1}^t\mathbf{v_1})^2 = \|\mathbf{v}\|^2\|\mathbf{w}\|^2 \\
\Rightarrow \quad & (\mathbf{w_1}^t\mathbf{v_1})^2 = 1 \times 1 = 1,
\end{aligned}$$

which implies

$$|\mathbf{v_1}.\mathbf{w_1}| = 1 = \|\mathbf{v_1}\|\|\mathbf{w_1}\|.$$

This is the equality case of Cauchy-Schwarz which holds if and only if $\mathbf{v_1}$ and $\mathbf{w_1}$ are linearly dependent. The two are both unit vectors so it follows that $\mathbf{v_1} = \pm\mathbf{w_1}$. Therefore, we have

$$A = \pm c\,(\mathbf{v_1} \otimes \mathbf{v_1}) = \pm c\,(\mathbf{v_1}\mathbf{v_1}^t).$$

Let $\mathbf{v_1} = (c_1, c_2, c_3)$, then $\mathbf{v_1} = c_1(1, \frac{c_2}{c_1}, \frac{c_3}{c_1})$. So,

$$A = \pm c\,(c_1c_2)(1, \frac{c_2}{c_1}, \frac{c_3}{c_1}) \otimes (1, \frac{c_2}{c_1}, \frac{c_3}{c_1}).$$

If $\pm c\,(c_1c_2) = \lambda$, $\frac{c_2}{c_1} = x$ and $\frac{c_3}{c_1} = y$, then

$$A = \lambda(1, x, y) \otimes (1, x, y),$$

as required. These arguments complete the proof of theorem (1.6.2). ♠

Chapter 2

Non-Linear Analysis

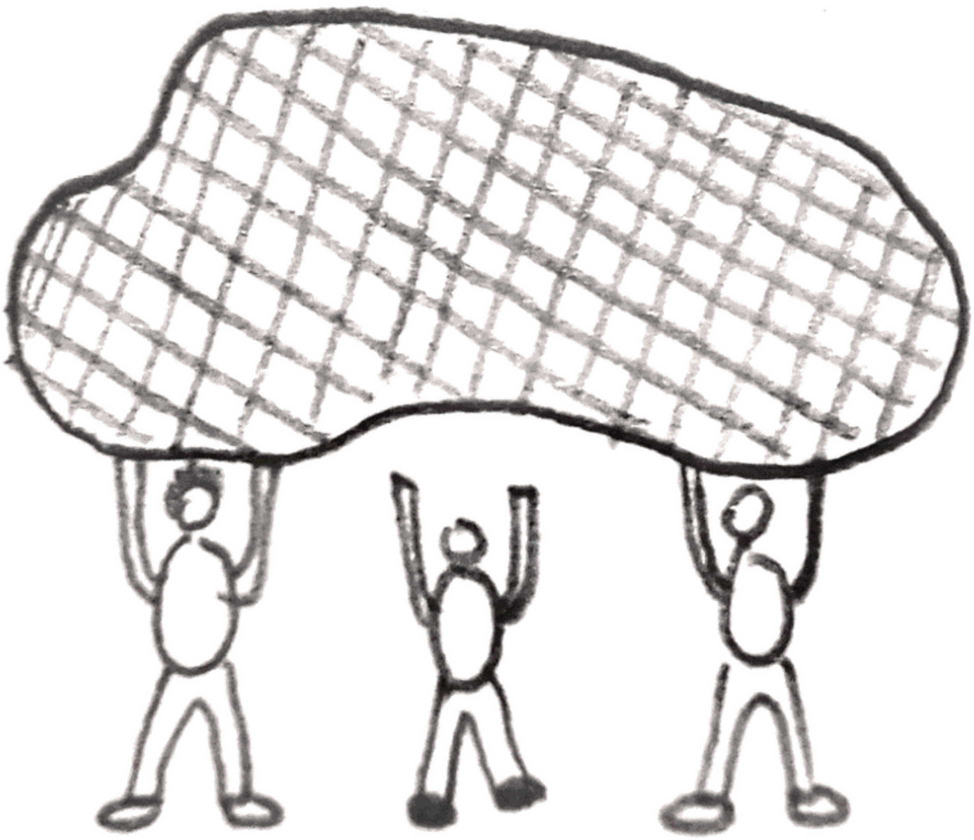

Figure 2.1:
Aha! Non-Linear Analysis
Source: https://enterfea.com/materially-nonlinear-analysis-explained/

Chapter 2

Non-Linear Analysis

This chapter presents functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and their properties. They are called vector fields. In section one, we introduce these kind of functions and their limit and continuity. In second section we will define first derivative and its properties of functions. This section will be very important to us. Finally, in last section, we will study about second derivative of functions, because we need this concept to investigate the convexity of functions.

2.1 Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ spaces

In this section, we shall consider functions from n-space $\mathbb{R}^n$ and with range in m-space $\mathbb{R}^m$. When both n and m are equal to 1, such a function is called a **real-valued function of a real variable**. When $n = 1$ and $m > 1$ it is called a **vector-valued function of a real variable**. In this chapter, we assume that $n > 1$ and $m \geq 1$. When $m = 1$, the function is called a **real-valued function of a vector variable** or, more briefly, a **scalar field**. When $m > 1$ it is called a **vector-valued function of a vector variable**, or simply a **vector field**. This section extends the concepts of limit, continuity, and in other sections we will define derivative and second derivative to scalar and vector fields (see [4]).
The concepts of limit and continuity are extended to scalar and vector field. We shall formulate the definitions for vector fields; they apply also to scalar fields.

Definition 2.1.1. We consider a function $\mathbf{f} : S \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$, where S is a subset of $\mathbb{R}^n$. If $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ we write

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{b} \tag{2.1}$$

to mean that

$$\lim_{\|\mathbf{x}\,-\,\mathbf{a}\|\to\mathbf{0}} \big\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\big\| = \mathbf{0} \tag{2.2}$$

The limit symbol in equation (2.2) is the usual limit of elementary calculus. In the definition it is not required that $\mathbf{f}$ be defined at the point a itself. If we write $\mathbf{h} = \mathbf{x} - \mathbf{a}$, equation (2.2) becomes

$$\lim_{\|\mathbf{h}\|\to\mathbf{0}} \big\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{b}\big\| = \mathbf{0}. \tag{2.3}$$

A function $\mathbf{f}$ is said to be **continuous** at $\mathbf{a}$ if $\mathbf{f}$ is defined at $\mathbf{a}$ and if

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}). \tag{2.4}$$

We say $\mathbf{f}$ is continuous on a set S if $\mathbf{f}$ is continuous at each point of S. Since these definitions are straightforward extensions of those in the one-dimensional case, it is not surprising to learn that many familiar properties of limits and continuity can also be extended. For example, the usual theorems concerning limits and continuity of sums, products, and quotients also hold for scalar fields. For vector field, quotients are not defined but for to prove of the theorems we can follow concerning sums, multiplication by scalars, inner products, and norms.

Theorem 2.1.1. *Let* $\mathbf{f}$ *and* $\mathbf{g}$ *be functions such that the composite function* $\mathbf{f} \circ \mathbf{g}$ *is defined at* $\boldsymbol{a}$*, where*

$$(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = \mathbf{f}(\mathbf{g}(\mathbf{x})).$$

If $\mathbf{g}$ *is continuous and if* $\mathbf{f}$ *is continuous at* $\mathbf{g}(\mathbf{a})$*, then the composition* $\mathbf{f} \circ \mathbf{g}$ *is continuous at* $\mathbf{a}$*.*

To see the proof of this theorem, you can study [4]. We want to recall some definition in linear transformation and the set of all linear transformations.

Definition 2.1.2. (a) Let $L(X, Y)$ be the set of all linear transformations of the vector space X into the vector space Y. Instead of $L(X, X)$, we shall simply write $L(X)$. If $A_1, A_2 \in L(X, Y)$ and if c_1, c_2 are scalar, define $c_1A_1 + c_2A_2$ by

$$(c_1A_1 + c_2A_2)(\mathbf{x}) = c_1A_1(\mathbf{x}) + c_2A_2(\mathbf{x}) \qquad (\mathbf{x} \in X).$$

It is then clear that $(c_1A_1 + c_2A_2) \in L(X, Y)$.
(b) If X, Y and Z are vector spaces, and if $A \in L(X, Y)$ and $B \in L(Y, Z)$, we define their product BA to be the composition of A and B:

$$(BA)(\mathbf{x}) = B(A(\mathbf{x})) \qquad (\mathbf{x} \in X).$$

Then $BA \in L(X, Z)$. Note that BA need not be same as AB, even if $X = Y = Z$.
(c) For $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, define the **norm** $\|A\|$ of A to be the supremum of all numbers $\|A\mathbf{x}\|$ where $\mathbf{x}$ ranges over all vectors in $\mathbb{R}^n$ with $\|\mathbf{x}\| \leq 1$. Observe that the inequality

$$\|A\mathbf{x}\| \leq \|A\| \, \|\mathbf{x}\| \tag{2.5}$$

holds for all $\mathbf{x} \in \mathbb{R}^n$. Also, if λ is such that $\|A\mathbf{x}\| \leq \lambda \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$, then $\|A\| \leq \lambda$.

We can prove that if $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|A\| < \infty$ and A is a uniformly continuous mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$. If $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ and c is a scalar, then

$$\|A + B\| \leq \|A\| + \|B\|, \qquad \|cA\| = |c| \, \|A\|.$$

with the distance between A and B defined as $\|A - B\|$, $L(\mathbb{R}^n, \mathbb{R}^m)$ is a **metric space**. You can see [32] for proofs of these statements.

2.2 Differentiation

Suppose $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a scalar field of n variables. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ denote a point of $\mathbb{R}^n$. A **partial function F with respect to the variable** x_i, is a one-variable function obtained from $\mathbf{f}$ by holding all variables constant except x_i. That is, we set x_j equal to a constant a_j for $j \neq i$. Then the partial function in x_i is defined by

$$F(x_i) = \mathbf{f}(a_1, a_2, \dots, x_i, \dots, a_n).$$

Now, we can defined the derivative for function F or $\mathbf{f}$ with respect to x_i.

Definition 2.2.1. The **partial derivative of f with respect to** x_i is the ordinary derivative of the partial function with respect to x_i. That is, the partial derivative with respect to x_i is $F'(x_i)$, in the notation of partial function. Standard notations for partial derivative of $\mathbf{f}$ with respect to x_i are

$$\frac{\partial \mathbf{f}}{\partial x_i}, \qquad D_{x_i}\mathbf{f}(x_1,\ldots,x_n), \qquad \mathbf{f}_{x_i}(x_1,\ldots,x_n).$$

Symbolically, we have

$$\frac{\partial \mathbf{f}}{\partial x_i} = \lim_{h\to 0} \frac{\mathbf{f}(x_1,\ldots,x_i+h,\ldots,x_n) - \mathbf{f}(x_1,\ldots,x_n)}{h}. \tag{2.6}$$

sometimes in some functions the partial derivatives exists but in other directions the answer of derivative are not equal to others. Because of that, we need to define the differentiability in the general case. Let X be an open set in $\mathbb{R}^n$ and $\mathbf{f} : X \longrightarrow \mathbb{R}$ be a scalar field; let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in X$. We say that $\mathbf{f}$ is **differentiable at a** if all the partial derivatives $\mathbf{f}_{x_i}(\mathbf{a})$, $i = 1, 2, \ldots, n$, exist and if the function $\mathbf{h} : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \mathbf{f}_{x_1}(\mathbf{a})(x_1 - a_1) + \ldots + \mathbf{f}_{x_n}(\mathbf{a})(x_n - a_n) \tag{2.7}$$

is a good linear approximation to $\mathbf{f}$ near $\mathbf{a}$, meaning that

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{h}(\mathbf{x})}{\|\mathbf{x}-\mathbf{a}\|} = \mathbf{0}. \tag{2.8}$$

We can use vector and matrix notation to rewrite things a bit. Define the **gradient** of a scalar field $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ to be a vector

$$\nabla \mathbf{f} = \left(\frac{\partial \mathbf{f}}{\partial x_1}, \ldots, \frac{\partial \mathbf{f}}{\partial x_n}\right).$$

Consequently,

$$\nabla \mathbf{f}(\mathbf{a}) = (\mathbf{f}_{x_1}(\mathbf{a}), \ldots, \mathbf{f}_{x_n}(\mathbf{a})).$$

Alternatively, we can use matrix notation and define the derivative of $\mathbf{f}$ at $\mathbf{a}$, denoted $D\mathbf{f}(\mathbf{a})$, to be a row matrix whose entries are the components of $\nabla\mathbf{f}(\mathbf{a})$; that is,

$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} \mathbf{f}_{x_1}(\mathbf{a}) & \mathbf{f}_{x_2}(\mathbf{a}) & \ldots & \mathbf{f}_{x_n}(\mathbf{a}) \end{bmatrix}.$$

Then, by identifying the vector $(\mathbf{x} - \mathbf{a})$ with the $n \times 1$ column matrix whose entries are the components of $(\mathbf{x} - \mathbf{a})$, we have

$$\nabla\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) = D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) = \begin{bmatrix} \mathbf{f}_{x_1}(\mathbf{a}) & \mathbf{f}_{x_2}(\mathbf{a}) & \ldots & \mathbf{f}_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix}$$

Hence, vector notation allows us to rewrite equation (2.7) quite compactly as

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \nabla\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

Thus, to say that $\mathbf{h}$ is a good linear approximation to $\mathbf{f}$ near $\mathbf{a}$ in equation (2.8) means that

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{a}) + \nabla\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})]}{\|\mathbf{x}-\mathbf{a}\|} = \mathbf{0}.$$

Let X be an open set in $\mathbb{R}^n$ and let $\mathbf{f}: X \longrightarrow \mathbb{R}^m$ be a vector field of n variables. We define the **matrix of partial derivatives of f**, denoted D$\mathbf{f}$, to be the $m \times n$ matrix whose ij-th entry is $\frac{\partial f_i}{\partial x_j}$, where $f_i : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ is the i-th component function of $\mathbf{f}$. That is,

$$D\mathbf{f}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{2.9}$$

The i-th row of D$\mathbf{f}$ is nothing more than Df_i, and the entries of Df_i are precisely the components of the gradient vector ∇f_i. (Indeed, in the case where $m = 1$, ∇f and $D\mathbf{f}$ mean exactly the same thing.) Now we can define derivative for vector field.

Definition 2.2.2. Let X be an open set in $\mathbb{R}^n$, let $\mathbf{f}: X \longrightarrow \mathbb{R}^m$, and let $\mathbf{a} \in X$. We say that $\mathbf{f}$ is **differentiable at a** if D$\mathbf{f}(\mathbf{a})$ exists and if the function $\mathbf{h}: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ defined by

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

is a good linear approximation to $\mathbf{f}$ near $\mathbf{a}$. That is, we require

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{h}(\mathbf{x})\|}{\|\mathbf{x}-\mathbf{a}\|} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})]\|}{\|\mathbf{x}-\mathbf{a}\|} = \mathbf{0} \tag{2.10}$$

In fact, we could have approached our discussion of differentiability much more abstractly right from the beginning. We could have defined a function $\mathbf{f}: X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ to be differentiable at a point $\mathbf{a} \in X$ to mean that there exists some linear mapping $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{a}) + A(\mathbf{x}-\mathbf{a})]\|}{\|\mathbf{x}-\mathbf{a}\|} = \mathbf{0}$$

We can say if $\mathbf{f}: X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$. Also, we can prove that if $\mathbf{f}: X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is such that, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ all $\frac{\partial f_i}{\partial x_j}$ exist and are continuous in a neighborhood of $\mathbf{a}$ in X, then $\mathbf{f}$ is differentiable at $\mathbf{a}$. In general we can show that a function $\mathbf{f}: X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable at $\mathbf{a} \in X$ if and only if each of its component functions $f_i : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, is differentiable at $\mathbf{a}$ for $i = 1, 2, \ldots, m$. If you want to see the proofs of these statements you can see [4], [32] or [11]. In general, if $\mathbf{f}: X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ is scalar field of n variables, the **kth-order partial derivative** with respect to the variables $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ (in that order), where $i_1, i_2, \ldots, i_k$ are integers in the set $\{1, 2, \ldots, n\}$ is the iterated derivative

$$\frac{\partial^k \mathbf{f}}{\partial x_{i_k} \ldots \partial x_{i_2} \partial x_{i_1}} = \frac{\partial}{\partial x_{i_k}} \cdots \frac{\partial}{\partial x_{i_2}} \frac{\partial}{\partial x_{i_1}} \mathbf{f}(x_1, x_2, \ldots, x_n).$$

Equivalent notation for this kth-order partial is

$$\mathbf{f}_{x_{i_1} x_{i_2} \cdots x_{i_k}}(x_1, x_2, \ldots, x_n).$$

Definition 2.2.3. Assume X is an open set in $\mathbb{R}^n$. A scalar field $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ whose partial derivatives up to (and including) order at least k exist and are continuous on X is said to be of **class** C^k. If $\mathbf{f}$ has continuous partial derivatives of all orders on X, then $\mathbf{f}$ is said of **class** C^∞ or **smooth**. A vector field $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is of class C^k (or of class C^∞) if and only if each of its component functions of class C^k (or of class C^∞).

Suppose $\mathbf{f}, \mathbf{g} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are two functions that are both differentiable at a point $a \in X$, and let $c \in \mathbb{R}$ be any scalar. Then the function $\mathbf{h} = \mathbf{f} + \mathbf{g}$ is also differentiable at $\mathbf{a}$, and we have

$$D\mathbf{h}(\mathbf{a}) = D(\mathbf{f} + \mathbf{g})(\mathbf{a}) = D(\mathbf{f})(\mathbf{a}) + D(\mathbf{g})(\mathbf{a}).$$

The function $\mathbf{k} = c \cdot \mathbf{f}$ is differentiable at $\mathbf{a}$ and

$$D\mathbf{k}(\mathbf{a}) = D(c \cdot \mathbf{f}))(\mathbf{a}) = c \cdot D\mathbf{f}(\mathbf{a})$$

Due to the nature of matrix multiplication, general versions of the product and quotient rules do not exist in any particularly simple form. However, for a scalar field, it is possible to prove the following:
Let $\mathbf{f}, \mathbf{g} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be differentiable at $\mathbf{a}$. Then
(a) The product function $\mathbf{fg}$ is also differentiable at $\mathbf{a}$, and

$$D(\mathbf{fg})(\mathbf{a}) = \mathbf{g}(\mathbf{a})D\mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{a})D\mathbf{g}(\mathbf{a}).$$

(b) If $\mathbf{g}(\mathbf{a}) \neq 0$, then the quotient function $\frac{\mathbf{f}}{\mathbf{g}}$ is differentiable at $\mathbf{a}$, and

$$D(\frac{\mathbf{f}}{\mathbf{g}})(\mathbf{a}) = \frac{\mathbf{g}(\mathbf{a})D\mathbf{f}(\mathbf{a}) - \mathbf{f}(\mathbf{a})D\mathbf{g}(\mathbf{a})}{(\mathbf{g}(\mathbf{a}))^2}.$$

We don't want to prove these theorems. To see proofs of these theorems you can see [11] or [32]

Theorem 2.2.1. *Let $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be a scalar field of class C^k. Then the order in which we calculate any kth-order partial derivative does not matter: If $(i_1, \ldots, i_k)$ are any k integers (not necessarily distinct) between 1 and n, and if $(j_1, \ldots, j_k)$ is any permutation of these integers, then*

$$\frac{\partial^k \mathbf{f}}{\partial x_{i_1} \ldots \partial x_{i_k}} = \frac{\partial^k \mathbf{f}}{\partial x_{j_1} \ldots \partial x_{j_k}}$$

We can also prove the chain rule is true for a scalar field and a vector field. This theorem has a long proof and you can see [11] or [32].

2.3 Higher-Order Derivative and Taylor's Theorem

Our goal in this section is to provide a means of approximating any scalar field by a polynomial of given degree, known as the **Taylor polynomial**.

Theorem 2.3.1. ***(First-order Taylor's formula in several variables)*** *Let X be an open set in $\mathbb{R}^n$ and suppose that $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at the point $\mathbf{a}$ in X. Let*

$$\mathbf{p_1}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}). \tag{2.11}$$

Then

$$\mathbf{f}(\mathbf{x}) = \mathbf{p_1}(\mathbf{x}) + R_1(\mathbf{x}, \mathbf{a}),$$

where $R_1(\mathbf{x}, \mathbf{a})/\left\|\mathbf{x} - \mathbf{a}\right\| \longrightarrow 0$ *as* $\mathbf{x} \longrightarrow \mathbf{a}$.

We may also express the first-order Taylor polynomial using the gradient in place of formula (2.11), we would have

$$\mathbf{p_1}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \nabla\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Before we explore higher-orders of Taylor's theorem in several variables, we consider the linear approximation in further detail.

Definition 2.3.1. Let $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ and let $\mathbf{a} \in X$. The **incremental change of f**, denoted $\Delta\mathbf{f}$, is

$$\Delta\mathbf{f} = \mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}),$$

where $\mathbf{h} = \mathbf{x} - \mathbf{a}$. The **total differential of f**, denoted $\mathrm{d}\mathbf{f}(\mathbf{a}, \mathbf{h})$, is

$$\mathrm{d}\mathbf{f}(\mathbf{a}, \mathbf{h}) = \frac{\partial\mathbf{f}}{\partial x_1}(\mathbf{a})h_1 + \frac{\partial\mathbf{f}}{\partial x_2}(\mathbf{a})h_2 + \ldots + \frac{\partial\mathbf{f}}{\partial x_n}(\mathbf{a})h_n.$$

The significance of the differential is that for $\mathbf{h} \approx \mathbf{0}$,

$$\Delta\mathbf{f} \approx \mathrm{d}\mathbf{f}.$$

We have abbreviated $\mathrm{d}\mathbf{f}(\mathbf{a}, \mathbf{h})$ by $\mathrm{d}\mathbf{f}$. The incremental change $\Delta\mathbf{f}$ equals the change in z-coordinate of the graph of $z = \mathbf{f}(x, y)$ as a point in $\mathbb{R}^2$ changes from $\mathbf{a} = (a, b)$ to $\mathbf{a} + \mathbf{h} = (a+h, b+k)$. The differential $\mathrm{d}\mathbf{f}$ equals the change in z-coordinate of the graph of the tangent plane at (a,b).

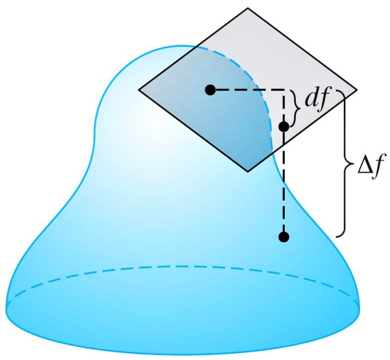

Figure 2.2: The total differential of $\mathbf{f}$
Source: Page 250 of book [11]

Now we state the second-order version of Taylor's theorem precisely. Taylor's theorem can be explained for every order but it requires that $\mathbf{f}$ be a C^k class function.

Theorem 2.3.2. ***(Second-order Taylor's formula)*** *Let* X *be an open set in* $\mathbb{R}^n$*, and suppose that* $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ *is of class* C^2*. Let*

$$\mathbf{p_2}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \mathbf{f}_{x_i}(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^{n} \mathbf{f}_{x_i x_j}(\mathbf{a})(x_i - a_i)(x_j - a_j). \tag{2.12}$$

Then

$$\mathbf{f}(\mathbf{x}) = \mathbf{p_2}(\mathbf{x}) + R_2(\mathbf{x}, \mathbf{a}),$$

where $\left|R_2(\mathbf{x}, \mathbf{a})\right| / \left\|\mathbf{x} - \mathbf{a}\right\|^2 \longrightarrow 0$ *as* $\mathbf{x} \longrightarrow \mathbf{a}$.

Proof of this theorem exists in [4] and [11]. Recall that the formula for the first-order Taylor polynomial $\mathbf{p_1}$ was written as

$$\mathbf{p_1}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \nabla\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

If we consider $\mathbf{h} = \mathbf{x} - \mathbf{a}$. Then the formula that was written in the above line becomes

$$\mathbf{p_1}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})\mathbf{h} = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{a})h_i. \tag{2.13}$$

We can show this formula by using vector and matrix notation. It turn out it is possible to do something similar for the second-order polynomial $\mathbf{p_2}$.

Definition 2.3.2. The **Hessian matrix** of a function $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ is the matrix whose ijth entry is $\partial\mathbf{f}/\partial x_j \partial x_i$. That is,

$$\mathrm{H}(\mathbf{f}) = \begin{bmatrix} \mathbf{f}_{x_1x_1} & \mathbf{f}_{x_1x_2} & \cdots & \mathbf{f}_{x_1x_n} \\ \mathbf{f}_{x_2x_1} & \mathbf{f}_{x_2x_2} & \cdots & \mathbf{f}_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_{x_nx_1} & \mathbf{f}_{x_nx_2} & \cdots & \mathbf{f}_{x_nx_n} \end{bmatrix}. \tag{2.14}$$

Now let's look again at the formula for $\mathbf{p_2}$ in theorem (2.3.2):

$$\mathbf{p_2}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \mathbf{f}_{x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^{n} \mathbf{f}_{x_ix_j}(\mathbf{a})(x_i - a_i)(x_j - a_j).$$

If we consider $h_i = (x_i - a_i)$ and $h_j = (x_j - a_j)$, then we can write

$$\mathbf{p_2}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \mathbf{f}_{x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^{n} \mathbf{f}_{x_ix_j}(\mathbf{a})h_ih_j.$$

We have let $\mathbf{h} = \mathbf{x} - \mathbf{a}$. This can be written as

$$\mathbf{p_2}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \begin{bmatrix} \mathbf{f}_{x_1}(\mathbf{a}) & \mathbf{f}_{x_2}(\mathbf{a}) & \cdots & \mathbf{f}_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}^t \begin{bmatrix} \mathbf{f}_{x_1x_1} & \mathbf{f}_{x_1x_2} & \cdots & \mathbf{f}_{x_1x_n} \\ \mathbf{f}_{x_2x_1} & \mathbf{f}_{x_2x_2} & \cdots & \mathbf{f}_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_{x_nx_1} & \mathbf{f}_{x_nx_2} & \cdots & \mathbf{f}_{x_nx_n} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$

Thus, we see that

$$\mathbf{p_2}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \mathbf{f}_{x_i}(\mathbf{a})(h_i) + \frac{1}{2}\mathbf{h}^t \mathrm{H}\mathbf{f}(\mathbf{a})\mathbf{h}. \tag{2.15}$$

For more informations you can see [11]. If we want to write the second-order Taylor's formula for function $\mathbf{f}$, then

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \sum_{i=1}^{n} \mathbf{f}_{x_i}(\mathbf{a})(h_i) + \frac{1}{2}\mathbf{h}^t \mathrm{H}\mathbf{f}(\mathbf{a})\mathbf{h} + R_2(\mathbf{x}, \mathbf{a}),$$

where $\left|R_2(\mathbf{x}, \mathbf{a})\right| / \left\|\mathbf{h}\right\|^2 \longrightarrow 0$ as $\mathbf{x} \longrightarrow \mathbf{a}$.

A scalar field $\mathbf{f}$ is said to have an **absolute maximum** at a point $\mathbf{a}$ of a set X in $\mathbb{R}^n$ if

$$\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{a})$$

for all $\mathbf{x} \in X$. The number $\mathbf{f}(\mathbf{a})$ is called the **absolute value** of $\mathbf{f}$ on X. The function $\mathbf{f}$ is said to have **relative maximum** at $\mathbf{a}$ if the top inequality is satisfied for every $\mathbf{x}$ in some n-ball B($\mathbf{a}$) lying in X. The terms **absolute minimum** and **relative minimum** are defined in an analogous fashion, using the opposite direction for top inequality. A number which is either a relative maximum or a relative minimum of $\mathbf{f}$ is called an **extremum of f.** Assume $\mathbf{f}$ is differentiable at $\mathbf{a}$. If $\nabla\mathbf{f}(\mathbf{a}) = D\mathbf{f}(\mathbf{a})$ is either zero or undefined, then the point $\mathbf{a}$ is called a **critical point of f**. A critical point is called a **saddle point** if every n-ball B($\mathbf{a}$) contains points $\mathbf{x}$ such that $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{a})$ and other points such that $\mathbf{f}(\mathbf{x}) > \mathbf{f}(\mathbf{a})$.

Theorem 2.3.3. *Let X be an open set in $\mathbb{R}^n$ and let $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be differentiable. If $\mathbf{f}$ has a relative extremum at $\mathbf{a} \in X$, then $D\mathbf{f}(\mathbf{a}) = \mathbf{0}$.*

We don't want to prove this theorem, if you want to see proof of this theorem, see [11]. The next theorem describes the kind and nature of the critical points in term of the algebraic sign of the quadratic form of the Hessian matrix.

Theorem 2.3.4. *Let X be an open set in $\mathbb{R}^n$ and $\mathbf{f} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function of class C^2. Suppose that $\mathbf{a} \in X$ is a critical point of $\mathbf{f}$*

1. *If the Hessian $H\mathbf{f}(\mathbf{a})$ is positive definite, then $\mathbf{f}$ has a relative minimum at $\mathbf{a}$.*
2. *If the Hessian $H\mathbf{f}(\mathbf{a})$ is negative definite, then $\mathbf{f}$ has a relative maximum at $\mathbf{a}$.*
3. *If $\det H\mathbf{f}(\mathbf{a}) \neq 0$ but $H\mathbf{f}(\mathbf{a})$ is neither positive nor negative definite, then $\mathbf{f}$ has a saddle point at $\mathbf{a}$.*

Chapter 3

Polyconvex, Quasiconvex and Rank-One Convex Functions

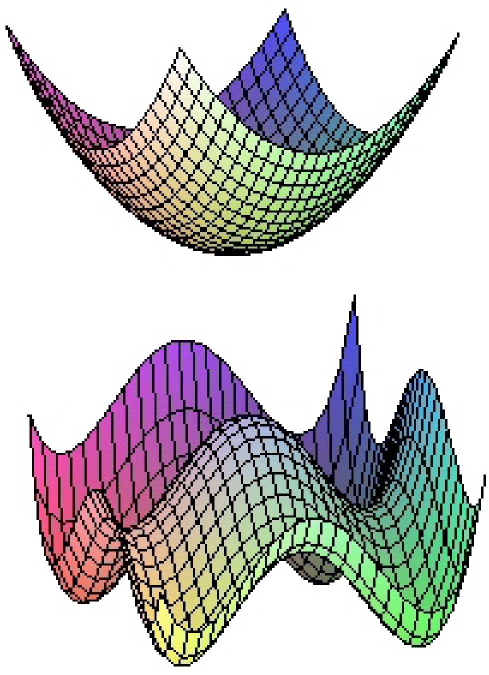

Figure 3.1: Convex and Non-convex Functions
Source: https://plus.maths.org/content/convexity

Chapter 3

Polyconvex, Quasiconvex and Rank-One Convex Functions

In this chapter we will explain the new concepts about norms, spaces, basic topology and convexity that are so important for our subjects. In the first section, we describe the norms and spaces that are useful for understanding the areas that we work in. In the second section we say some basic properties of topology and real analysis. In the third section we study the definitions of convexity of a function in one-dimension or higher dimensions. Finally in the last section, we describe the new definitions of polyconvex, quasiconvex, rank-one convex functions and the relations between them. This chapter has a major role in this thesis.

3.1 Norms and spaces

This section is about mathematical structures called norms and spaces. In mathematics a **space** is a set with some added structures. Mathematical spaces is a set of hierarchy, i.e., one space may inherit all the all the characteristics of a parent space.

Definition 3.1.1. A collection τ of subsets of a set X is said be a **topology** in X if τ has the following properties:

1. $\phi \in \tau$ and $X \in \tau$.
2. If $V_i \in \tau$ for $i = 1, 2, \ldots, n$, then $\bigcap\limits_{i=1}^{n} V_i \in \tau$.
3. If $\{V_\alpha\}$ is an arbitrary collection members of τ (finite, countable or uncountable), then $\bigcap\limits_{\alpha} V_\alpha \in \tau$.

Properly speaking, a **topological space** is an ordered pair (X, τ). The members of τ are called the open sets in X. If X and Y are topological spaces and if f is a mapping of X into Y, then f is said to be continuous provided that $f^{-1}(V)$ is an open set in X for every open set V in Y.
Definition 3.1.2. A collection $\mathfrak{M}$ of subsets of a set X is said to be a σ-algebra in X if $\mathfrak{M}$ has the following properties:

1. $X \in \mathfrak{M}$.
2. If $A \in \mathfrak{M}$, then $A^c \in \mathfrak{M}$ where A^c is the complement of A relative to X.
3. If $A = \bigcup_{n=1}^{\infty} A_n$ and if $A_n \in \mathfrak{M}$ for $n = 1, 2, \ldots$, then $A \in \mathfrak{M}$.

If $\mathfrak{M}$ is a σ-algebra in X, then X is called a **measurable space**, and the members of $\mathfrak{M}$ are called the **measurable sets** in X. The prefix σ refers to the fact that the property (3) is required to hold for all countable unions of members of $\mathfrak{M}$. If (3) is required for finite unions only, then $\mathfrak{M}$ is called the **algebra of sets**. If X is a measurable space, Y is a topological space and f is a mapping of X into Y, then f is said to be measurable provided that f^{-1} is a measurable set in X for every open set V in Y. Let X be a set and $\mathfrak{M}$ be a σ-algebra on X. A function μ from $\mathfrak{M}$ to the extended real number line is called the **measure** or **measure function** if it satisfies the following properties:

1. For all E in $\mathfrak{M}$, we have $\mu(E) \geq 0$.
2. $\mu(\phi) = 0$.
3. For all countable collections $\{E\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathfrak{M}$, we have

$$\mu(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu(E_k).$$

Property (3) is called **countably additive**. A triple $(X, \mathfrak{M}, \mu)$ is called a **measure space**. Measure space is a measurable space which has a positive measure μ defined on the σ-algebra of its measurable sets. Measure spaces are also topological spaces (see [33]). Let $(X, \mathfrak{M}_1)$ and $(Y, \mathfrak{M}_2)$ be two measurable spaces, then the function $f : X \longrightarrow Y$ is said to be a **measurable function** if the preimage of E under f is in $\mathfrak{M}_1$ for every $E \in \mathfrak{M}_2$.

Definition 3.1.3. A **metric space** is a set X in which a distance function or **metric** d is defined, with the following properties:

1. $0 \leq \mathrm{d}(x, y) \leq \infty$ for all x and y in X.
2. $\mathrm{d}(x, y) = 0$ if and only if $x = y$.
3. $\mathrm{d}(x, y) = \mathrm{d}(y, x)$ for all x and y in X.
4. $\mathrm{d}(x, y) \leq \mathrm{d}(x, z) + \mathrm{d}(z, y)$ for all x, y and z in X.

Let Y and Z be topological spaces, and let $g : Y \longrightarrow Z$ be continuous, if X is a topological space, if $f : X \longrightarrow Y$, and if $h = g \circ f$, then $h : X \longrightarrow Z$ is continuous. Also if X is a measurable space, if $f : X \longrightarrow Y$ is measurable, and if $h = g \circ f$, then $h : X \longrightarrow Z$ is measurable. Let u and v be real measurable functions on a measurable space X, let Φ be a continuous mapping of the plane into a topological space Y, and define

$$h(x) = \Phi(u(x), v(x)) \qquad \text{for } x \in X.$$

Then $h : X \longrightarrow Y$ is measurable. Proofs of these theorems exist in the amazing book [33]. Let X be a measurable space. If $f = u + iv$ where u and v are real measurable functions on X, then f is a complex measurable function on X. If $f = u + iv$ is a complex measurable function on X, then u, v, and $|f|$ are real measurable functions on X. If f and g are complex measurable functions on X, then so are $f + g$ and fg. If f is a complex measurable function on X, there is a complex measurable function α on X such that $|\alpha| = 1$ and $f = \alpha|f|$.

Theorem 3.1.1. *If $\mathcal{F}$ is any collection of subsets of X, there exists a smallest σ-algebra $\mathfrak{M}^*$ in X such that $\mathcal{F} \subset \mathfrak{M}^*$.*

This $\mathfrak{M}^*$ is sometimes called the σ-algebra generated by $\mathcal{F}$. We don't want to prove this theorem. There exists in [33] in page 12.

Definition 3.1.4. Let X be a topological space. By theorem (3.1.1), there exists a smallest σ-algebra $\mathcal{B}$ in X such that every open set in X belongs to $\mathcal{B}$. The members of $\mathcal{B}$ are called the **Borel sets of** X.

In particular, closed sets are Borel sets, and so are all countable unions of closed sets and all countable intersections of open sets. These last two are called $\mathbf{F}_\sigma$'s and $\mathbf{G}_\delta$'s, respectively, and play a considerable role. Since $\mathcal{B}$ is a σ-algebra, we may now regard X as a measurable space, with the Borel sets playing the role of the measurable sets; more concisely, we consider the measurable space $(X, \mathcal{B})$. If $f : X \longrightarrow Y$ is a continuous mapping of X, where Y is any topological space, then evident from the definitions that $f^{-1}(V) \in \mathcal{B}$ for every open set V in Y. In the other hand, every continuous mapping of X is Borel measurable. **Borel measurable mappings** are often called **Borel mappings** or **Borel functions**. We explain some other definitions that are useful for this thesis.

Definition 3.1.5. We need some spaces and their norm that are useful to understand the rest of this thesis.

1. The **closure** $\overline{E}$ of a set $E \subset X$ is the smallest closed set in X which contains E.
2. A set $K \subset X$ is **compact** if every open cover of K contains a finite subcover. In particular, if X is itself compact, then X is called a **compact space**.
3. X is a **Hausdorff space** if the following is true. If $p \in X$, $q \in X$ and $p \neq q$, then p has a neighbourhood U and q has a neighbourhood V such that $U \bigcap V = \phi$.
4. X is **locally compact space** if every point of X has a neighbourhood whose closure is compact.
5. **An inner product space** is a vector space V over the field F (F is either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$) with an inner product, i.e., with a map
$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow F$$
that satisfies the following axioms for all vectors $x, y, z \in V$ and all scalars $a \in F$:
(a) $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
(b) $\langle ax, y \rangle = a\langle x, y \rangle$.
(c) $\langle x + y, z \rangle = \langle x, z \rangle + \langle z, y \rangle$.
(d) $\langle x, x \rangle \geq 0$.
(e) $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.
The following norm that will be defined by
$$\|x\| = \langle x, x \rangle^{\frac{1}{2}} = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2},$$
where $x = (x_1, x_2, \ldots, x_n) \in V$, is called **inner product norm or Euclidean norm** in vector space V.
6. A **normed vector space** is a pair $(V, \|\cdot\|)$ where V is a vector space and $\|\cdot\|$ a inner product norm on V.
7. The vector space $\mathbb{R}^n$ with inner product norm is called a **Euclidean n-space.**
8. If we add one point to vector space $\mathbb{R}^n$ for compactification, then $\hat{\mathbb{R}}^n = \mathbb{R}^n \cup \{\infty\}$ is called **Möbius space**. Let $\pi : \hat{\mathbb{R}}^n \longrightarrow \mathbb{S}^n$, where the sphere $\mathbb{S}^n$ has its usual Euclidean metric topology, be a homeomorphism. The function π that is called **stereographic projection** of $\hat{\mathbb{R}}^n$ to $\mathbb{S}^n$, defined as follows : $\pi(\infty) = e_{n+1}$, and for $x \in \mathbb{R}^n$, $\pi(x)$ is the point of intersection with $\mathbb{S}^n$ of the Euclidean ray that issues from e_{n+1} and passes through x. We can find the equation of function $\pi(x)$ by using elementary analytic geometry as
$$\pi(x) = \Big(\frac{2x_1}{\|x_1\|^2 + 1}, \ldots, \frac{2x_n}{\|x_n\|^2 + 1}, \frac{\|x_n\|^2 - 1}{\|x_n\|^2 + 1} \Big).$$

It is easy to see that $\pi(x)$ is an injective function. So, its inverse is defined

$$\pi^{-1}(y) = \Big(\frac{y_1}{1-y_{n+1}}, \frac{y_2}{1-y_{n+1}} \dots, \frac{y_n}{1-y_{n+1}}\Big),$$

for $y \neq e_{n+1}$, while $\pi^{-1}(e_{n+1}) = \infty$. Stereographic projection provides a new metric structure into $\hat{\mathbb{R}}^n$ (see [18]). If x and y are two points in $\hat{\mathbb{R}}^n$, then the **chordal metric** on $\hat{\mathbb{R}}^n$ is defined by

$$q(x, y) = |\pi(x) - \pi(y)|.$$

9. A **Möbius transformation** is a mapping $\mathbf{f} : \hat{\mathbb{R}}^n \longrightarrow \hat{\mathbb{R}}^n$ such that $\mathbf{f}$ is a composition of a finite number of the following transformations (see [40] p.14)
   - Translation: $\mathbf{f}(x) = x + a$, for a fixed $a \in \mathbb{R}^n$.
   - Stretching: $\mathbf{f}(x) = rx$, where r is a positive real number.
   - Orthogonal mapping: $\mathbf{f}$ is linear and $\|\mathbf{f}(x)\| = \|x\|$ for all $x \in \mathbb{R}^n$.
   - Inversion in a sphere $\mathbb{S}^{n-1}(a, r)$: $\mathbf{f}(x) = a + \frac{r^2(x-a)}{\|x-a\|^2}$, where $r > 0$ is the radius and $a \in \mathbb{R}^n$ is the centre of sphere.
10. A metric space M is called **complete** or **Cauchy space** if every Cauchy sequence of points in M has a limit that is also in M. In the other words, if every Cauchy sequence in M converges in M.
11. A **Hilbert space** H is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product.
12. A **function space** is a set of functions of a given kind from a set X to a set Y. It can be a topological space, vector space or both of them.
13. A **Banach space** is a complete normed vector space. Thus, a Banach space is a vector space with a metric that allows the computation of vector length and distance between vectors and is complete in the sense that a Cauchy sequence of vectors always converges to the well defined limit that is within the space.
14. If $f : V \longrightarrow W$ is a function between two vector spaces V and W on field F, and k is an integer, then f is said to be homogeneous of degree k if

$$f(\alpha\mathbf{v}) = \alpha^k f(\mathbf{v}) \tag{3.1}$$

   for all nonzero $\alpha \in F$ and $\mathbf{v} \in V$. When the vector spaces involved are over the real numbers, a slightly less general form of homogeneity is often used, requiring only that the equation (3.1) holds for all $\alpha > 0$.
15. **Taxicab geometry** is a form of geometry in which the usual distance function or metric of Euclidean geometry is replaced by a new metric in which the distance between two points is the sum of the absolute of their cartesian coordinates. The taxicab metric is also known as **rectilinear distance**, $\mathbf{L_1}$ **distance** or $\boldsymbol{\ell_1}$ **norm**
16. For real number $p \geq 1$, the **p-norm** or $\mathbf{L^p}$ **– norm** of vector $\mathbf{x}$ in a real vector space is defined by

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{\frac{1}{p}}.$$

   The Euclidean norm from above falls into this class and is the 2-norm, and the 1-norm is the norm that corresponds to the rectilinear distance. The $\mathbf{L^\infty}$**-norm** is the limit of L^p-norm for $p \longrightarrow \infty$. It can describe as

$$\|\mathbf{x}\|_\infty = max\{|x_1|, |x_2|, \dots, |x_n|\}.$$

We can summarise all these definition in diagram below. An arrow from space A to space B implies that space A is also a kind of space B. That means for instance, that a normed vector space is also a metric space.

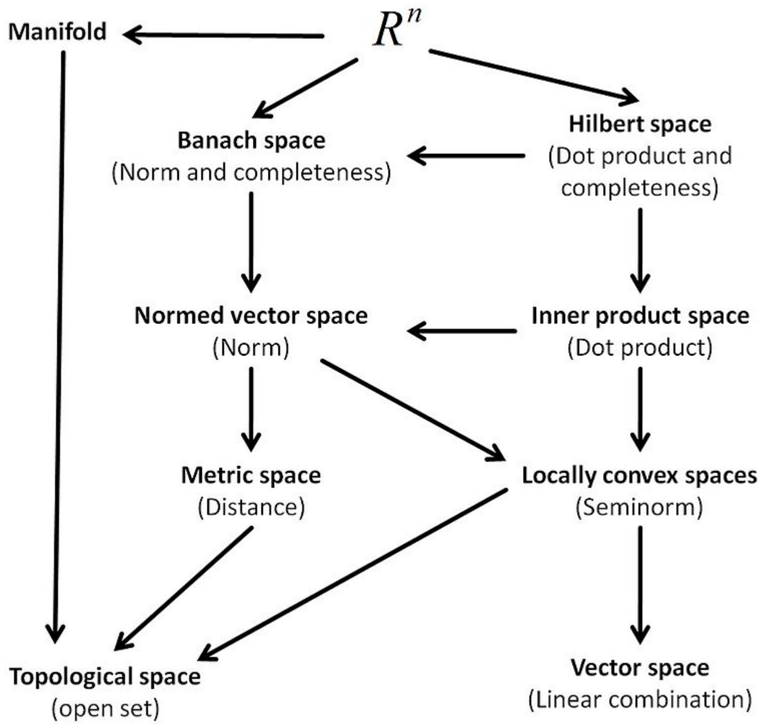

Figure 3.2: Abstract spaces
Source: https://en.wikipedia.org/wiki/Space_(mathematics)

All of these definitions exist in every advanced real and complex analysis book or measure theory book. If you want to see more information and theorems and examples you can see [32], [33], [5]. Also you can study a set of summarised definitions and properties in (https://en.wikipedia.org/wiki/Space_(mathematics)).

3.2 Lebesgue space and Conformal maps

In this section we introduce the **Lebesgue measure** and **conformal maps**. Let X be a measurable space and Y a topological space. A function $f : X \longrightarrow Y$ is measurable if $f^{-1}(U)$ is measurable in X for every open subset U of Y. We can say, if X be a measurable space and let $f : X \longrightarrow \mathbb{R}$ be a function, then f is measurable if and only if $f^{-1}((a, \infty))$ is a measurable subset of X for all $a \in \mathbb{R}$. If F is a symbol that denotes either $\mathbb{R}$ or $\mathbb{C}$ then let X be a measurable space then any constant function from X into F is measurable. The sum and product of two measurable functions from X into F are measurable. The complex conjugate of any measurable function from X into $\mathbb{C}$ is measurable. A function from X into $\mathbb{C}$ is measurable if and only if its real and imaginary parts are both measurable. All proofs of these statements exist in [41].

Definition 3.2.1. Let $(X, \mathfrak{M}, \mu)$ be a measure space. A measurable subset A of X is called a **null set** if $\mu(A) = 0$. A statement $P(x)$ is true for almost every x if there is a null set A such that $P(x)$ is true for every x not belonging to A. The measure μ is complete if every subset of every null set is measurable.

The **extended non-negative real axis** $[0, +\infty]$ is the non-negative real axis $[0, +\infty) = \{x \in \mathbb{R} : x \geq 0\}$with the additional element adjoined to it, which we label $+\infty$. We begin by defining the **Lebesgue outer measure** which assigns to each subset S of $\mathbb{R}$ an outer measure $m^*(S)$. Thus m^* will be a function

$$m^* : P(\mathbb{R}) \longrightarrow [0, +\infty]$$

where $P(\mathbb{R})$ denotes the power set of $\mathbb{R}$. m^* is not countably additive. Instead, it has the weaker property of countable subadditivity, meaning that,

$$m^*(\bigcup_{n\in\mathbf{N}} S_n) \leq \sum_{n\in\mathbf{N}} m^*(S_n)$$

for any sequence $\{S_n\}$ of subsets of $\mathbb{R}$.

Definition 3.2.2. If $S \subseteq \mathbb{R}$, the Lebesgue outer measure of S is defined by

$$m^*(S) = \inf \{ \sum_{k=1}^{\infty} \ell(I_k) \mid \quad S \subseteq \bigcup_{k=1}^{\infty} I_k \qquad \textit{where } I_k \textit{ are open intervals} \}.$$

The Lebesgue outer measure has these following properties:

1. m^* is defined for every set of real numbers.
2. $0 \leq m^*(S) \leq \infty$.
3. $m^*(A) \leq m^*(B)$ provided $A \subseteq B$ (monotonic).
4. $m^*(\phi) = 0$.
5. $m^*(\{a\}) = 0$.
6. $m^*(I) = \ell(I)$. I is an interval. The Lebesgue outer measure of an interval is its length.
7. $m^*(A + c) = m^*(A)$ (translation invariant).
8. $m^*(\bigcup_{n\in\mathbf{N}} S_n) \leq \sum_{n\in\mathbf{N}} m^*(S_n)$ for every sequence of sets of real numbers (countable subadditivity).

Proofs of these properties exist in pages (96) and (97) of [10] or in the book [38]. Also you can find good information about Lebesgue measure in (http://math.bard.edu/belk/math461/LebesgueMeasure.pdf).

Definition 3.2.3. A subset E of $\mathbb{R}$ is said to be **Lebesgue measurable** if

$$m^*(S \cap E) + m^*(S \cap E^c) = m^*(S)$$

for every S of $\mathbb{R}$. In this case, the outer measure $m^*(E)$ of E is called the **Lebesgue measure** of E, and is denoted $m(E)$.

Definition (3.2.3) is symmetric between E and E^c. Thus a set E is measurable if and only if its complement E^c is measurable. Lebesgue measure and Lebesgue measurable sets have the following properties:

1. If E and F are measurable subsets of $\mathbb{R}$, then $E \cup F$ is also measurable.
2. If E and F are measurable subsets of $\mathbb{R}$, then $E \cap F$ is also measurable.
3. Let $\{E_k\}$ be a sequence of pairwise disjoint measurable subsets of $\mathbb{R}$. Then the union $\bigcup_{k\in\mathbf{N}} E_k$ is measurable, and
$$m(\bigcup_{k\in\mathbf{N}} E_k) = \sum_{k\in\mathbf{N}} m(E_k).$$
4. Every interval I in $\mathbb{R}$ is Lebesgue measurable.
5. (**Heine-Borel Theorem**) Let $[a,b]$ be a closed interval in $\mathbb{R}$, and $\mathcal{C}$ be a family of open intervals that covers $[a,b]$. Then there exist a finite subcollection of $\mathcal{C}$ that covers $[a,b]$.
6. For every interval I in $\mathbb{R}$, $m(I) = \ell(I)$.

Proof. Here we only prove parts (5) and (6) the Heine-borel Theorem and measure of interval. Let S be the set of all points $s \in [a,b]$ for which the interval $[a,s]$ can be covered by some finite collection of $\mathcal{C}$. Note that $a \in S$, since the interval $[a,s]$ is just a single point. Our goal is to prove that $b \in S$.
Let $x = \sup(S)$. Since $S \subseteq [a,b]$, we know that $x \in [a,b]$. Therefore, there exists an interval $(c,d) \in \mathcal{C}$ that contains x. Since $c < x$, there is some point $s \in S$ that lies between c and x. Let $\{(c_1,d_1),(c_2,d_2),\ldots,(c_n,d_n)\}$ be finite subcollection of $\mathcal{C}$ that covers $[a,x]$. Then the collection $\{(c_1,d_1),(c_2,d_2),\ldots,(c_n,d_n),(c,d)\}$ covers $[a,x]$, which proves that $x \in S$.
Moreover, if $x < b$, then there exist an $\epsilon > 0$ such that $x+\epsilon \in [a,b]$ and $x+\epsilon \in (c,d)$. Then the collection $\{(c_1,d_1),(c_2,d_2),\ldots,(c_n,d_n),(c,d)\}$ covers $[a,x+\epsilon]$, which proves that $x+\epsilon \in S$, a contradiction since x is the supremum of S. We conclude that $x = b$
For part (6), we define the new function. If S is any subset of $\mathbb{R}$, the **characteristic function** for S in the function $\chi : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$\chi_S = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{if } x \notin S. \end{cases}$$

Note that if I is an interval then

$$\int_{-\infty}^{+\infty} \chi_I(x)\mathrm{dx} = \ell(I).$$

For every $\epsilon > 0$, there exists an interval $J^{'}$ containing J so that $\ell(J^{'}) \le \ell(J) + \epsilon$. Then the singleton collection $\{J^{'}\}$ of open intervals covers J, so

$$m(J) \le \ell(J^{'}) \le \ell(J) + \epsilon.$$

Since ϵ was arbitrary, it follows that $m(J) \le \ell(J)$.
Now let $\mathcal{C}$ be any collection of open intervals that covers J. Let $\epsilon > 0$, and let K be a closed subinterval of J such that $\ell(K) \ge \ell(J) - \epsilon$. By the Heine-Borel theorem, there exists a finite subcollection $\{I_1, I_2, \ldots, I_n\}$ of $\mathcal{C}$ that covers K. Then

$$\chi_{I_1} + \ldots + \chi_{I_n} \ge \chi_K$$

so

$$\sum_{I\in\mathcal{C}} \ell(I) \ge \ell(I_1) + \cdots + \ell(I_n) = \int_{-\infty}^{+\infty} \chi_{I_1}(x)\mathrm{dx} + \cdots + \int_{-\infty}^{+\infty} \chi_{I_n}(x)\mathrm{dx}$$

$$= \int_{-\infty}^{+\infty} (\chi_{I_1}(x) + \cdots + \chi_{I_n}(x))\mathrm{d}x \geq \int_{-\infty}^{+\infty} \chi_K(x)\mathrm{d}x = \ell(K) \geq \ell(J) - \epsilon.$$

Since ϵ was arbitrary, it follows that

$$\sum_{I \in \mathcal{C}} \ell(I) \geq \ell(J)$$

which proves that $m(J) \geq \ell(J)$. ♠

Proofs of all parts exist in (http://math.bard.edu/belk/math461/LebesgueMeasure.pdf) or you can see [10] and [38]. Assume $(X, \mathfrak{M})$ is a measurable space, then the function $f : X \longrightarrow \mathbb{R}$ is said to be a **Lebesgue measurable function** if and only if $\mathfrak{M}$ is the σ-algebra of Lebesgue measurable sets and for all $\alpha \in \mathbb{R}$ the set $\{ x \in X : f(x) < \alpha\}$ is Lebesgue measurable. This definition can also be described by $\{ x \in X : f(x) > \alpha\}$, $\{ x \in X : f(x) \geq \alpha\}$ and $\{ x \in X : f(x) \leq \alpha\}$. All definitions are equivalent. Lebesgue measurable functions are important in mathematical analysis because they can be integrated.
Let $(X, \mathfrak{M}, \mu)$ is a measure space. If $s : X \longrightarrow [0, \infty)$ is a measurable simple function, of the form

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i},$$

Where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the distinct values of s and A_i are disjoint members of $\mathfrak{M}$. If $E \in \mathfrak{M}$, we define

$$\int_E s\mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E).$$

The convention $0 \cdot \infty = 0$ is used here; it may happen that $\alpha_i = 0$ foe some i and that $\mu(A_i \cap E) = \infty$. If $f : X \longrightarrow [0, \infty)$ is measurable, and $E \in \mathfrak{M}$, we define

$$\int_E f\mathrm{d}\mu = \sup \int_E s\mathrm{d}\mu, \tag{3.2}$$

the supremum being taken over all simple measurable function s such that $0 \leq s \leq f$. The left side of equation (3.2) is called the **Lebesgue integral** of f over E, with respect to the measure μ. It is a number in $[0, \infty]$ (see [33] p.19).
Definition 3.2.4. Let $(X, \mathfrak{M}, \mu)$ be a measure space and $1 \leq p < \infty$. The $L^p(X)$**-space** is a set of equivalence classes of measurable functions $f : X \longrightarrow \mathbb{R}$ such that

$$\|f\|_p = \Big(\int_X |f|^p \mathrm{d}\mu \Big)^{\frac{1}{p}} < \infty.$$

The $\|f\|_p$ is called the L^p-norm of f. The notation $L^p(X)$ assumes that the measure μ is understood. For $1 \leq p < \infty$, the $L^p(X)$-space is a Banach space. So, we can say, $L^p(X)$ for $1 \leq p < \infty$ is the Banach space of measurable functions f with $|f|^p$ integrable in X.
Definition 3.2.5. Let Ω be an open set in the Euclidean space $\mathbb{R}^n$ and $f : \Omega \longrightarrow \mathbb{R}$ be a Lebesgue measurable function. If $1 \leq p \leq \infty$ the function f satisfies

$$\int_A |f|^p \mathrm{d}x < \infty,$$

This means $f \in L^p(A)$ for all compact subsets $A \in \Omega$. f is called locally integrable and the set of all functions is denoted by $L^p_{loc}(\Omega)$.

Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. We define the **maximal stretching L(T)** and **minimal stretching** $\ell(\mathbf{T})$ of T by

$$L(T) = \max_{|x|=1} |T(x)|, \qquad \ell(T) = \min_{|x|=1} |T(x)|.$$

The quantity $L(T)$ frequently goes by a different name, The **operator norm of T**, under the alternate notation $\|T\|$. For the composition ST of linear transformations $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $S : \mathbb{R}^m \longrightarrow \mathbb{R}^p$, it is true that

$$L(ST) \leq L(S)L(T), \qquad \ell(ST) \geq \ell(S)\ell(T). \tag{3.3}$$

A linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is nonsingular if and only if $\ell(T) > 0$ in which event

$$L(T^{-1}) = \ell(T)^{-1}, \qquad \ell(T^{-1}) = L(T)^{-1}. \tag{3.4}$$

Recall that the nonsingular linear transformations of $\mathbb{R}^n$ form a group under composition, the **general linear group GL(n)**. A linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is called an **orthogonal transformation** if $|T(x)| = |x|$ for every x in $\mathbb{R}^n$ or, equivalently, that $\langle T(x), T(y)\rangle = \langle x, y\rangle$ for all x and y in $\mathbb{R}^n$. The orthogonal transformations of $\mathbb{R}^n$ constitute a subgrup of GL(n), the **orthogonal group O(n)**. An element T of GL(n) belongs to $O(n)$ if and only if $T^{-1} = T^*$, where T^* denotes adjoint of T, the unique linear transformation $T^* : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ that satisfies

$$\langle T(x), y\rangle = \langle x, T^*(y)\rangle$$

for all x and y in $\mathbb{R}^n$. If U is an element of $O(n)$ (we represent orthogonal transformations by the letters U and V) then $L(U) = \ell(U) = 1$, while the determinant of U is either 1 or -1. If an orthogonal transformation U has $\det(U) = 1$, we call U a **rotation** of $\mathbb{R}^n$.
The **special orthogonal group SO(n)** is the subgroup of $O(n)$ consisting of all such rotations. It is not hard to see that

$$L(VTU) = L(T), \qquad \ell(VTU) = \ell(T), \tag{3.5}$$

for any linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ whenever $U \in O(n)$ and $V \in O(m)$. Let U be an open set in $\mathbb{R}^n$, and $x \in U$ we define the maximal stretching $L_{\mathbf{f}}(x)$ and the minimal stretching $\ell_{\mathbf{f}}(x)$ at x of a function $\mathbf{f} : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by :

$$L_{\mathbf{f}}(x) = \limsup_{h\to 0} \frac{|\mathbf{f}(x+h) - \mathbf{f}(x)|}{|h|},$$

$$\ell_{\mathbf{f}}(x) = \liminf_{h\to 0} \frac{|\mathbf{f}(x+h) - \mathbf{f}(x)|}{|h|}.$$

Theorem 3.2.1. *If* $\mathbf{f} : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *is differentiable at* $x \in U$*, then* $L_{\mathbf{f}}(x) = L(D\mathbf{f}(x)) = L(\mathbf{f}'(x))$ *and* $\ell_{\mathbf{f}}(x) = \ell(D\mathbf{f}(x)) = \ell(\mathbf{f}'(x))$*.*

Proof. We prove the first part. Recall that if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear transformation then

$$L(T) = \max_{|x|=1} |T(x)|, \qquad \ell(T) = \min_{|x|=1} |T(x)|.$$

We can chose a small h such that $x + h \in U$. By theorem (2.3.1), we can write

$$|\mathbf{f}(x+h) - \mathbf{f}(x)| = |\mathbf{f}'(x)h + h\epsilon(h)| \leq |h|L(\mathbf{f}'(x)) + |h||\epsilon(h)|.$$

So, we have

$$L_{\mathbf{f}}(x) \leq \limsup_{h\to 0}(L(\mathbf{f}'(x)) + |\epsilon(h)|) = L(\mathbf{f}'(x)).$$

Also, we can chose $h \in \mathbb{S}^{n-1}$ such that $L(\mathbf{f}'(x)) = |\mathbf{f}'(x)h|$. Then by definition of derivative we have

$$L_{\mathbf{f}}(x) \geq \lim_{t\to 0} \frac{|\mathbf{f}(x+th) - \mathbf{f}(x)|}{t} = |\mathbf{f}'(x)h| = L(\mathbf{f}'(x)),$$

and proof is completed. For second part proof is same as first part. ♠

Corollary 3.2.1.1. *If $\mathbf{f} : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is differentiable at $x \in U$, then $L_{\mathbf{f}}(x) = |\mathbf{f}'(x)|$.*

Let $n \geq 2$. A linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is **conformal** if T is nonsingular and preserves Euclidean angles, in the sense that

$$\theta[T(x), T(y)] = \theta(x, y)$$

for all nonzero vectors x and y in $\mathbb{R}^n$. For more information you can see [18].

Theorem 3.2.2. *Let $n \geq 0$. The following statements concerning a linear transformation $T \in GL(n)$ are equivalent :*

1. *T is conformal;*
2. *$T = \lambda U$, where λ is a positive number and U belongs to O(n);*
3. *$\|T\|^n = |\det(T)|$;*
4. *$T^*T = |\det(T)|^{2/n} I$.*

where I denotes the identity matrix.

Proof of this theorem exist in the pages (19) and (20) of book [18].

3.3 Convex Function in Higher Dimensions

In this section we defined the convex and concave function in $\mathbb{R}$ or higher dimensions. We start to define the convex function in $\mathbb{R}$ and extend this definition for higher dimension.

Definition 3.3.1. Let I be a nondegenerate interval of $\mathbb{R}$. A function $f : I \longrightarrow \mathbb{R}$ is called **convex** if

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{3.6}$$

for all x_1 and x_2 in I and $\lambda \in [0, 1]$. It is called **strictly convex** if the inequality (3.6) holds strictly whenever x_1 and x_2 are distinct points and $\lambda \in (0, 1)$. If the function $-f$ is convex or strictly convex then we say that f is **concave** or **strictly concave**, respectively. If f is both convex and concave in I, then f is said to be **affine**.

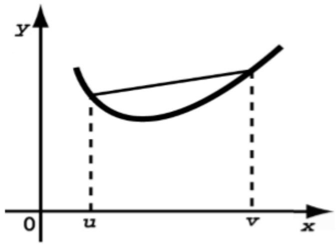

Figure 3.3: The graph of convex function is under the chord.
Source: Page 8 of book [30]

The convexity of a function $f : I \longrightarrow \mathbb{R}$ means geometrically that the points of the graph of f in the interval $[u, v]$ are under or on the chord joining the endpoints $(u, f(u))$ and $(v, f(v))$, for all $u,\ v \in I$ and $u < v$.

By the discrete case of **Jensen's inequality** we can say, a real-valued function f defined on an interval I is convex if and only if for all $x_1, \dots, x_2$ in I and all scalars $\lambda_1, \dots, \lambda_n$ in $[0, 1]$ with $\sum_{k=1}^{n} \lambda_k = 1$ we have

$$f(\sum_{k=1}^{n} \lambda_k x_k) \leq \sum_{k=1}^{n} \lambda_k f(x_k).$$

We can describe this inequality in other way. Next theorem is a general case of above inequality.

Theorem 3.3.1. *Let $f : I \longrightarrow \mathbb{R}$ be concave if and only if for any $a, b, c \in I$, with $a < b < c$,*

$$\frac{f(b) - f(a)}{b - a} \geq \frac{f(c) - f(b)}{c - b},$$

and,

$$\frac{f(b) - f(a)}{b - a} \geq \frac{f(c) - f(a)}{c - a}.$$

For strict concavity, the inequalities are strict. Let $f : I \longrightarrow \mathbb{R}$ be convex if and only if for any $a, b, c \in I$, with $a < b < c$,

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(b)}{c - b},$$

and,

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a}.$$

For strict convexity, the inequalities are strict.

Theorem 3.3.2. *(* ***J.Jensen Midpoint Convex****) Let $f : I \longrightarrow \mathbb{R}$ be a continuous function. Then f is convex if and only if it is midpoint convex, that is,*

$$f(\frac{x + y}{2}) \leq \frac{f(x) + f(y)}{2} \qquad for\ \ all\ \ x, y \in I.$$

Proof. We prove this theorem by contradiction. If f is not convex, then there exists a subinterval $[a, b]$ such that the graph of $f|_{[a,b]}$ is not under the chord joining $(a, f(a))$ and $(b, f(b))$;

that is, the function

$$\varphi(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a) - f(a), \qquad x \in [a,b]$$

verifies $\gamma = \{\varphi(x) | \, x \in [a,b]\} > 0$. Notice that φ is continuous and $\varphi(a) = \varphi(b) = 0$. Also, a direct computation shows that φ is also midpoint convex. Put $c = \inf\{\, x \in [a,b] | \varphi(x) = \gamma\}$; then necessarily $\varphi(c) = \gamma$ and $c \in (a,b)$. By definition of c, for every $h > 0$ for which $c \pm h \in (a,b)$ we have

$$\varphi(c-h) < \varphi(c) \qquad \varphi(c+h) < \varphi(c)$$

so that

$$\varphi(c) > \frac{\varphi(c-h) + \varphi(c+h)}{2}$$

in contradiction with the fact that φ is midpoint convex. ♠

Corollary 3.3.2.1. *Let $f : I \longrightarrow \mathbb{R}$ be a continuous function. Then f is convex if and only if*

$$f(x+h) + f(x-h) - 2f(x) \geq 0$$

for all $x \in I$ and all $h > 0$ such that $x+h$ and $x-h$ are in I.

Corollary 3.3.2.2. *(**The second derivative test for convexity**) Suppose that $f : I \longrightarrow \mathbb{R}$ is a twice differentiable function. Then:*

1. *f is convex if and only if $f''(x) \geq 0$ for all x in I;*
2. *f is strictly convex if and only if $f''(x) \geq 0$ and the set of points where f'' vanishes does not include intervals of positive length.*
3. *f is concave if and only if $f''(x) \leq 0$ for all x in I;*
4. *f is strictly concave if and only if $f''(x) \leq 0$ and the set of points where f'' vanishes does not include intervals of positive length.*

Theorem 3.3.3. ***(The operations with convex functions)***

1. *Adding two convex functions defined on the same interval we obtain a convex function; if one of them is strictly convex, then the sum is also strictly convex.*
2. *Multiplying a (strictly) convex function by a positive scalar we obtain also a (strictly) convex function.*
3. *The restriction of every (strictly) convex function to a subinterval of its domain is also a (strictly) convex function.*
4. *If $f : I \longrightarrow \mathbb{R}$ is a convex (strictly convex) function and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a nondecreasing (an increasing) convex function, then $g \circ f$ is a convex (strictly convex) function.*

For more information and seeing the proofs you can see chapter one of book [30]. Convex functions provide basic techniques in a series of domain like optimization theory, partial differential equation and geometric inequalities. The natural domain for a convex function is a convex set. That is why we shall start by recalling some basic facts on convex sets. A function f defined on an interval J is said to be **Lipschitz** if there exists a constant $L \geq 0$ such that

$$|f(x) - f(y)| \leq L|x-y| \qquad \text{for all } x, y \in J.$$

A subset C of a vector space (linear space are assumed to be real) E to be is said to be convex if it contains the line segment

$$[x, y] = \{(1-\lambda)x + \lambda y | \lambda \in [0,1]\}$$

connecting any of its points x and y.

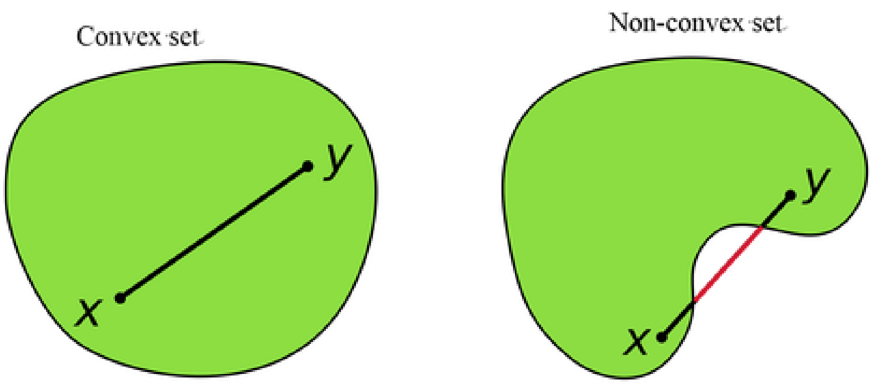

Figure 3.4: The convex and Non-convex (concave) sets
Source: convex function images

A subset A of E is said to be an **affine set** if it contains the whole line through any two of its points. Algebraically, this means

$$x, y \in A \; and \; \lambda \in \mathbb{R} \qquad (1-\lambda)x + \lambda y \in A.$$

For the given set $E \subset \mathbb{R}^n$, $\overline{E}$, ∂E, $intE$ and E^c respectively stand for the closure, the boundary, the interior and the complement of E. The **affine hull** of a set $E \subset \mathbb{R}^n$ is the smallest affine set containing E. It is denoted by *aff* E . A **hyperplane** $H \subset \mathbb{R}^n$ is a set of the form

$$H = \{x \in \mathbb{R}^n | \langle x, a \rangle = \alpha\}$$

where $a \in \mathbb{R}^n$, $a \neq 0$ and $\alpha \in \mathbb{R}$. Also, we can define the **convex hull** of a set $E \subset \mathbb{R}^n$, denoted by coE, is the smallest convex set containing E.

Definition 3.3.2. Let U be a convex subset of $\mathbb{R}^n$ (real linear space) and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be a function

1. The **subgraph of f**, denoted *sub* **f**, is the set

$$sub\, \mathbf{f} = \{(x, y) \in U \times \mathbb{R} \mid \mathbf{f}(x) \geq y\}.$$

2. The **epigraph of f**, denoted *epi* **f**, is the set

$$epi\, \mathbf{f} = \{(x, y) \in U \times \mathbb{R} \mid \mathbf{f}(x) \leq y\}.$$

3. The **domain** of **f** is defined as

$$\text{dom}\, \mathbf{f} = \{x \in \mathbb{R}^n \mid \mathbf{f}(x) < \infty\}.$$

4. The **lower contour set of f at** $\alpha \in \mathbb{R}$ **or lower-level set of f at** $\alpha \in \mathbb{R}$, is defined as
$$L_\alpha = \{x \in U \mid \mathbf{f}(x) \leq \alpha\}.$$

5. The **upper contour set of f at** $\alpha \in \mathbb{R}$ **or upper-level set of f at** $\alpha \in \mathbb{R}$, denoted U_α, is the set
$$U_\alpha = \{x \in U \mid \mathbf{f}(x) \leq \alpha\}.$$

The subgraph of a function is the area lying below the graph of the function, and the epigraph of a function is the area lying above the graph of the function. If U is a convex subset of $\mathbb{R}^n$ and $\mathbf{f} : U \longrightarrow \mathbb{R}$ is a function, we say that $\mathbf{f}$ is convex on U if *sub* $\mathbf{f}$ is a convex set. Also, we say that $\mathbf{f}$ is convex on U if *epi* $\mathbf{f}$ is a convex set. Note that concave and convex functions are required to have convex domains. For more information about these sets you can see [30] or [12]. Source of the figure below: (http://scottmccracken.weebly.com/uploads/9/0/6/6/9066859/convexity-print_version.pdf)

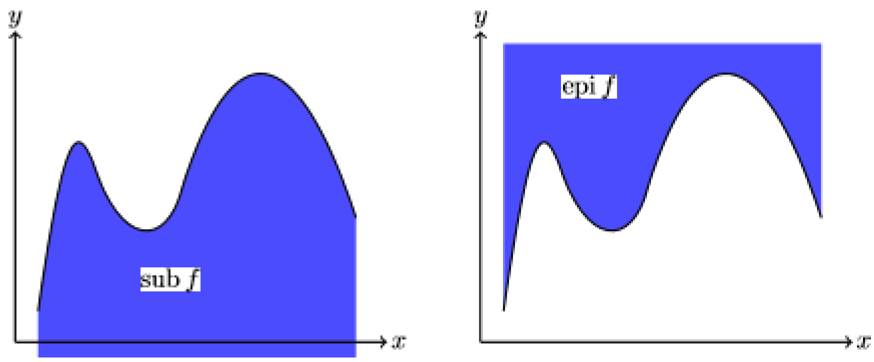

Figure 3.5: The subgraph and epigraph of f

Definition 3.3.3. Let U be a convex subset of $\mathbb{R}^n$ and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be a function. Then

1. $\mathbf{f}$ is convex if and only if for all $x, y \in U$ and $\lambda \in [0, 1]$, we have
$$\mathbf{f}(\lambda x + (1 - \lambda)y) \leq \lambda\mathbf{f}(x) + (1 - \lambda)\mathbf{f}(y).$$

2. $\mathbf{f}$ is concave if and only if for all $x, y \in U$ and $\lambda \in [0, 1]$, we have
$$\mathbf{f}(\lambda x + (1 - \lambda)y) \geq \lambda\mathbf{f}(x) + (1 - \lambda)\mathbf{f}(y).$$

3. $\mathbf{f}$ is strictly convex if and only if for all $x, y \in U$ and $\lambda \in (0, 1)$, we have
$$\mathbf{f}(\lambda x + (1 - \lambda)y) < \lambda\mathbf{f}(x) + (1 - \lambda)\mathbf{f}(y).$$

4. $\mathbf{f}$ is strictly concave if and only if for all $x, y \in U$ and $\lambda \in (0, 1)$, we have
$$\mathbf{f}(\lambda x + (1 - \lambda)y) > \lambda\mathbf{f}(x) + (1 - \lambda)\mathbf{f}(y).$$

They are easy to prove that if U be a convex subset of $\mathbb{R}^n$ and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be a function, then $\mathbf{f}$ is a convex function if and only if $-\mathbf{f}$ is concave and vice versa. If $\mathbf{f}_i : U \longrightarrow \mathbb{R}$ be convex (concave) functions and a_i be positive numbers $i = 1, \dots, k$, then

$$a_1\mathbf{f}_1 + \cdots a_k\mathbf{f}_k$$

is a convex (concave) function. Convexity in the case of several variables is equivalent with convexity on each line segment included in the domain of definition. So we can say, a function $\mathbf{f} : U \longrightarrow \mathbb{R}$ is convex if and only if for every two points x and y in U the function

$$\varphi : [0, 1] \longrightarrow \mathbb{R}, \qquad \varphi(t) = \mathbf{f}((1 - t)x + ty)$$

is convex. We can prove that each level set of height α of a convex function is a convex set. Convex functions exhibit a series of nice properties related to maximum and minimum of functions. We want to describe some theorems about convex and concave functions.

Theorem 3.3.4. *Let $U \subseteq \mathbb{R}^n$ be an open and convex set and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be either convex or concave function. Then $\mathbf{f}$ is a continuous function.*

Theorem 3.3.5. *Let $U \subseteq \mathbb{R}^n$ be an open and convex set and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be differentiable.*

1. *$\mathbf{f}$ is convex if and only if for any $x, y \in U$ we have*

$$\mathbf{f}(x) \geq \nabla\mathbf{f}(y)(x - y) + \mathbf{f}(y)$$

2. *$\mathbf{f}$ is concave if and only if for any $x, y \in U$ we have*

$$\mathbf{f}(x) \leq \nabla\mathbf{f}(y)(x - y) + \mathbf{f}(y)$$

The proofs of these theorems are long and we don't intend to describe them here. But you can see their proofs in books [30], [12] or article (https://pages.wustl.edu/files/pages/imce/nachbar/concavity.pdf). By theorem (2.2.1) we know that the order in which we calculate any kth-order partial derivative does not matter. The matrix of second order partial derivatives is Hessian denoted by $H\mathbf{f}(x) = D^2\mathbf{f}(x) = D(\nabla\mathbf{f})(x)$. By theorem (2.2.1) if f is C^2 then $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is symmetric.

Theorem 3.3.6. *Let $U \subseteq \mathbb{R}^n$ be open and convex set and let $\mathbf{f} : U \longrightarrow \mathbb{R}$ be C^2.*

1. *If $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is negative definite for every $x \in U$ then f is strictly concave.*
2. *If $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is negative semi-definite for every $x \in U$ then f is concave.*
3. *If f is concave and C^2 then $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is negative semi-definite for every $x \in U$.*
4. *If $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is positive definite for every $x \in U$ then f is strictly convex.*
5. *If $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is positive semi-definite for every $x \in U$ then f is convex.*
6. *If f is convex and C^2 then $H\mathbf{f}(x) = D^2\mathbf{f}(x)$ is positive semi-definite for every $x \in U$.*

Proof. Consider first the case $n = 1$, for proof part (1), in which case $D^2\mathbf{f}(x) \in \mathbb{R}$, hence $D^2\mathbf{f}(x)$ is negative definite if and only if $D^2\mathbf{f}(x) < 0$. Since $D^2\mathbf{f}(x) < 0$ for all x then $D\mathbf{f}(x)$ is strictly decreasing for all x, which implies that for any $a, b, c \in U$, $a < b < c$,

$$\frac{f(b) - f(a)}{b - a} > \frac{f(c) - f(b)}{c - b},$$

which, by theorem (3.3.1)implies that $\mathbf{f}$ is strictly concave. The proofs of parts (2), (4) and (5) are almost identical. It remains to prove parts (3) and (6). The proof of part (3) is by contradiction. Suppose that $D^2\mathbf{f}(y) > 0$ for some $y \in U$. Since $\mathbf{f}$ is C^2, $D^2\mathbf{f}(x) > 0$ for every x in some open ball containing y. Then $\mathbf{f}$ is not concave for x in this ball. The proof of part (6) is almost identical.
For $n > 1$, in which case $D^2\mathbf{f}(x)$ is a symmetric matrix. Consider any $a, b \in U$, $a \neq b$, and any $\lambda \in (0,1)$. Let $y = \lambda a + (1-\lambda)b$. To show part (1), we need to show that $\mathbf{f}(y) > \lambda\mathbf{f}(a) + (1-\lambda)\mathbf{f}(b)$. Let $g(\lambda) = b + \lambda(a-b)$ and $h(\lambda) = \mathbf{f}(g(\lambda)) = \mathbf{f}(b + \lambda(a-b))$. By the $n = 1$ step above, a sufficient condition for strict concavity of h is that $D^2h(\lambda) < 0$. By the chain rule, for any $\lambda \in (0,1)$, letting $v = a - b$, $Dh(\lambda) = D\mathbf{f}(g(\lambda))v$ and hence $D^2h(\lambda)$ is

$$D^2h(\lambda) = v'D^2\mathbf{f}(y)v.$$

Therefore, if $v'D^2\mathbf{f}(y)v < 0$ for every $\lambda \in (0,1)$ then h is strictly concave on $(0,1)$, which implies that $h(\lambda) > \lambda h(1) + (1-\lambda)h(0)$, which implies $\mathbf{f}(y) > \lambda\mathbf{f}(a) + (1-\lambda)\mathbf{f}(b)$, as was to be shown. The proof of the other parts of the theorem are similar. ♠

3.4 Polyconvex, quasiconvex and Rank-One Convex Functions

We start this section with three new definitions and their relationships to each others. One of these definitions is main thesis question and we will find some condition for it. One major application of convex and concave functions is optimization. As we will see, any relative maximum of a concave function is a absolute maximum.

Definition 3.4.1. Let $M_{m\times n}(\mathbb{R})$ denote the space of all $m \times n$ matrices over the field of real numbers $\mathbb{R}$ and U is an open subset of $M_{m\times n}(\mathbb{R})$. A function $\mathcal{F} : U \subset M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$ is said to be **rank-one convex** if for every $A \in U$ the function of real variable $t \mapsto \mathcal{F}(A + tB)$ is convex near $t = 0$.

Also we can say that, function $\mathcal{F} : U \subset M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$ is a rank-one convex function if $\mathcal{F}(tA + (1-t)B) \leq t\mathcal{F}(A) + (1-t)\mathcal{F}(B)$, whenever $t \in (0,1)$ and $A,\ B \in M_{m\times n}(\mathbb{R})$ with $\text{rank}(A-B) \leq 1$ (See [25], [34], [13] and [14]). Now we describe the others. At first we explain a easy definition for quasiconvex functions, and we extend the following definition to advance definition.

Definition 3.4.2. Let U be a convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : U \longrightarrow \mathbb{R}$ be a function. Then

1. A function $\mathcal{F}$ is **quasiconvex** if all its lower-level sets are convex.
2. A function $\mathcal{F}$ is **quasiconcave** if all its upper-level sets are concave.

If we note L_α the lower-level set of $\mathcal{F}$ at $\alpha \in \mathbb{R}$, then $x \in L_\alpha$ is equivalent to $\mathcal{F}(x) \leq \alpha$. So:

$$\begin{aligned}
\mathcal{F}\ \textit{is quasiconvex}\ &\Leftrightarrow \forall\alpha \in \mathbb{R},\ \forall x, y,\ \forall\lambda \in [0,1] : \Big[x, y \in L_\alpha \Rightarrow \lambda x + (1-\lambda)y \in L_\alpha\Big] \\
&\Leftrightarrow \forall\alpha \in \mathbb{R},\ \forall x, y,\ \forall\lambda \in [0,1] : \Big[\mathcal{F}(x) \leq \alpha, \mathcal{F}(y) \leq \alpha \Rightarrow \mathcal{F}(\lambda x + (1-\lambda)y) \leq \alpha\Big] \\
&\Leftrightarrow \forall\alpha \in \mathbb{R},\ \forall x, y,\ \forall\lambda \in [0,1] : \Big[\max\big\{\mathcal{F}(x), \mathcal{F}(y)\big\} \leq \alpha \Rightarrow \mathcal{F}(\lambda x + (1-\lambda)y) \leq \alpha\Big] \\
&\Leftrightarrow \forall x, y,\ \forall\lambda \in [0,1] : \Big[\mathcal{F}(\lambda x + (1-\lambda)y) \leq \max\big\{\mathcal{F}(x), \mathcal{F}(y)\big\}\Big]
\end{aligned}$$

Now, we can say that

1. A function $\mathcal{F}$ is quasiconvex if and only if

$$\forall x, y,\ \forall\lambda \in [0,1] : \ \Big[\mathcal{F}(\lambda x + (1-\lambda)y) \leq \max\big\{\mathcal{F}(x), \mathcal{F}(y)\big\}\Big].$$

2. A function $\mathcal{F}$ is quasiconcave if and only if

$$\forall x, y,\ \forall\lambda \in [0,1] : \ \Big[\mathcal{F}(\lambda x + (1-\lambda)y) \geq \min\big\{\mathcal{F}(x), \mathcal{F}(y)\big\}\Big].$$

Definition 3.4.3. Let U be a convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : U \longrightarrow \mathbb{R}$ be a function.

1. A function $\mathcal{F}$ is **strictly quasiconvex** if for all x, $y \in U$ with $x \neq y$ and all $\lambda \in (0,1)$, we have
$$\mathcal{F}(\lambda x + (1-\lambda)y) < \max\big\{\mathcal{F}(x), \mathcal{F}(y)\big\}.$$

2. A function $\mathcal{F}$ is **strictly quasiconcave** if for all x, $y \in U$ with $x \neq y$ and all $\lambda \in (0,1)$, we have
$$\mathcal{F}(\lambda x + (1-\lambda)y) > \min\big\{\mathcal{F}(x), \mathcal{F}(y)\big\}.$$

Let U be a convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : U \longrightarrow \mathbb{R}$ be a function. Then $\mathcal{F}$ is quasiconvex (or quasiconcave) if and only if $-\mathcal{F}$ is quasiconcave (or quasiconvex). Also, **f** is strictly quasiconvex (or strictly quasiconcave) if and only if the function $-\mathcal{F}$ is strictly quasiconcave (or strictly quasiconvex). In addition, let $\mathcal{F} : U \longrightarrow \mathbb{R}$ be a quasiconvex function. If $\mathcal{G} : \mathbb{R} \longrightarrow \mathbb{R}$ is increasing function, then $\mathcal{G} \circ \mathcal{F}$ is quasiconvex. If $\mathcal{F} : U \longrightarrow \mathbb{R}$ is a quasiconcave and $\mathcal{G} : \mathbb{R} \longrightarrow \mathbb{R}$ is increasing function, then $\mathcal{G} \circ \mathcal{F}$ is quasiconcave.

Theorem 3.4.1. *Let U be a convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : U \longrightarrow \mathbb{R}$ be a function. Then*

1. *$\mathcal{F}$ is quasiconvex if and only if*

$$\forall\, x, y \in U, \qquad \mathcal{F}(x) \geq \mathcal{F}(y) \Rightarrow D\mathcal{F}(x)(y-x) \leq 0.$$

2. *$\mathcal{F}$ is quasiconcave if and only if*

$$\forall\, x, y \in U, \qquad \mathcal{F}(y) \geq \mathcal{F}(x) \Rightarrow D\mathcal{F}(x)(y-x) \geq 0.$$

Proof. **(Proof of part 2)** Suppose first that $\mathcal{F}$ is quasiconcave on U. Fix $x, y \in U$ and assume $\mathcal{F}(y) \geq \mathcal{F}(x)$. For any $\lambda \in [0,1]$, we have

$$\mathcal{F}(x + \lambda(y-x)) = \mathcal{F}(\lambda y + (1-\lambda)x) \geq \min\big\{\mathcal{F}(x), \mathcal{F}(y)\big\} = \mathcal{F}(x).$$

So for all $\lambda \in [0,1]$ we can write

$$\frac{\mathcal{F}(x + \lambda(y-x)) - \mathcal{F}(x)}{\lambda} \geq 0.$$

We know that if $\lambda \longrightarrow 0^+$, then $\lambda(y-x) \longrightarrow 0^+$. Therefore,

$$\lim_{\lambda(y-x)\longrightarrow 0^+} \frac{\mathcal{F}(x + \lambda(y-x)) - \mathcal{F}(x)}{\lambda(y-x)}.(y-x) \geq 0.$$

This last limit implies $D\mathcal{F}(x)(y-x) \geq 0$. Conversely, assume that for all $x, y \in U$ such that $\mathcal{F}(y) \geq \mathcal{F}(x)$, we have $D\mathcal{F}(x)(y-x) \geq 0$. Choose any $x, y \in U$, and suppose without loss of generality that

$$\min\{\mathcal{F}(x), \mathcal{F}(y)\} = \mathcal{F}(y).$$

Now, we can show that for any $\lambda \in (0,1)$, we also have

$$\mathcal{F}(\lambda x + (1-\lambda)y) > \min\{\mathcal{F}(x), \mathcal{F}(y)\},$$

establishing the quasiconcavity of $\mathcal{F}$. For more information you can see (http://www.columbia.edu/~sd2702/convexity.pdf) and (http://scottmccracken.weebly.com/uploads/9/0/6/6/9066859/convexity-print_version.pdf). Also, you can study the books [12], [30]. ♠

This proof says that the angle between the gradient and the vector $(y-x)$ is acute or right. The part (2) is illustrated in the following figure (Source of figure : http://scottmccracken.weebly.com/uploads/9/0/6/6/9066859/convexity-print_version.pdf). We can explain another

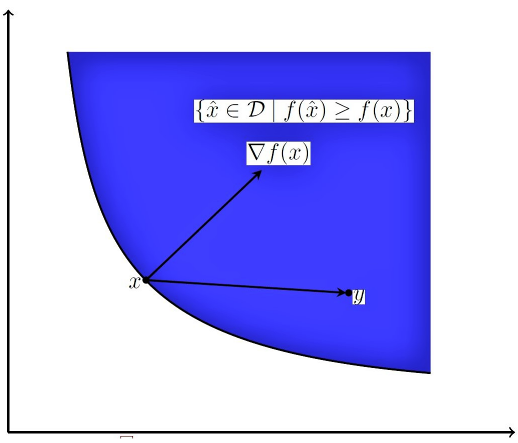

Figure 3.6: The angle between vectors $\nabla \mathbf{f}(x)$ and $(y-x)$ is acute.

definition for quasiconvex functions. This definition is formal and advance definition for quasiconvex functions.

Definition 3.4.4. Let $U \subset \mathbb{R}^n$ be a bounded open set, $\varphi : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is continuous, has uniformly bounded gradient and $\varphi = 0$ on ∂U. Let $M_{m\times n}(\mathbb{R})$ be the space of all $m \times n$ matrices over the field of real numbers $\mathbb{R}$. Then A continuous function $\mathcal{F} : M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$ is quasiconvex if

$$\int_U \big[\mho(A + \nabla\varphi(x)) - \mathcal{F}(A)\big]\mathrm{dx} \geq 0,$$

for every $A \in M_{m\times n}(\mathbb{R})$, for every bounded open set $U \subset \mathbb{R}^n$ and every function φ with those conditions.(See [36], [34] and [14])

Definition 3.4.5. A function $\mathcal{F} : M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$ is **polyconvex** if there exist a convex function $\mathcal{G} : M_{m\times n}(\mathbb{R}) \times \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$\mathcal{F}(A) = \mathcal{G}(A,\ minors\ \ of\ A),$$

for every $A \in M_{m\times n}(\mathbb{R})$. In other words, $\mathcal{F}$ is polyconvex if $\mathcal{F}(A)$ equals to convex function of subdeterminants of the matrix A or $\mathcal{F}$ is polyconvex if $\mathcal{F}(A)$ equals to convex function of minors of the matrix A (See [36], [37], [35] or [14]).

For example, $\mathcal{F} : M_{2\times 2}(\mathbb{R}) \longrightarrow \mathbb{R}$ is polyconvex if there exists a convex function $\mathcal{G} M_{2\times 2}(\mathbb{R}) \times \mathbb{R} \longrightarrow \mathbb{R}$ such that $\mathcal{F}(A) = \mathcal{G}(A, \det(A))$, for each $A \in M_{2\times 2}(\mathbb{R})$ (See [37], [35]). It is well-known that

$$\mathcal{F}\ \ convex \Rightarrow \mathcal{F}\ \ polyconvex \Rightarrow \mathcal{F}\ \ quasiconvex \Rightarrow \mathcal{F}\ \ rank-one\ \ convex.$$

Theorem 3.4.2. *Let $\mathcal{F} : M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$. Then*

1. $\mathcal{F}$ *convex* $\Rightarrow \mathcal{F}$ *polyconvex* $\Rightarrow \mathcal{F}$ *quasiconvex* $\Rightarrow \mathcal{F}$ *rank* $-$ *one convex.*
2. *If $m = 1$ or $n = 1$, then all these notions are equivalent.*
3. *If $\mathcal{F} : M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}$ is convex, polyconvex, quasiconvex or rank-one convex, then $\mathcal{F}$ is locally Lipschitz.*

There are long and complete proofs for this theorem in [7], [27] and [12]. The converse of all parts of result (1) above are false. For instance, when $m = n = 2$, the function $\mathcal{F}(A) = \det(A)$ is polyconvex but not convex.(See [14]). Also, we can find the quasiconvex functions such that they are not polyconvex. In particular, there are a few examples of quasiconvex functions which are not polyconvex. One such example has been given by Sverak (see [35]). You can find another given example by Alibert and Dacorogna in [2]. The existence of a rank-one convex function that is not quasiconvex when $m \geq 3$, $n \geq 2$ was proved by Vladimir Sverak [36]. A classical result in this direction is that a quadratic function $\mathbf{f}$ is quasiconvex if and only if it is rank-one convex (see [7], [27] and [37]). This can be proved by using the Fourier transformation.
In the case $m = 2$ and $n \geq 2$, however, it is an open question to determine whether

$$\mathcal{F}\ \ rank-one\ \ convex \Rightarrow \mathcal{F}\ \ quasiconvex.$$

If $m = n = 2$ the above implication would have far-reaching consequences. In article [9], the authors claimed to have found a counterexample for the case $m = n = 2$, that $\mathbf{f}$ is rank-one convex function such that is not quasiconvex. But in article [29], Patrizio Neff has shown that the example in [9] is not a counterexample and this case is still an open problem.

Chapter 4

Convexity and Linear Distorsion

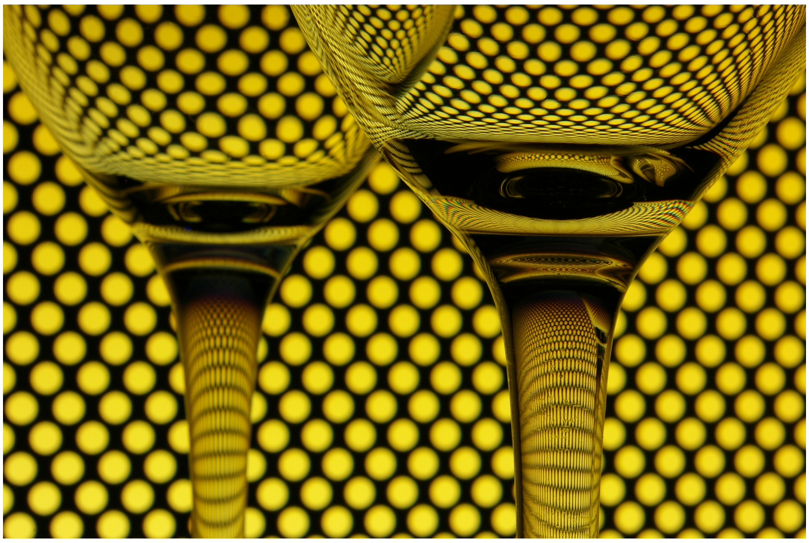

Figure 4.1: Distortion in Optics
Source: Images for linear distortion photography

Chapter 4

Convexity and Linear Distortion

In mathematics, the distortion is a measure of the amount by which a function from the Euclidean plane to itself distorts circles to ellipses. In the two first sections of this chapter we explain the concept and definitions of linear, outer and inner distortion and their properties. In the next two sections we focus on the main topic of quasiconformal mappings which define them from three perspectives that are called curve family,analytic and geometric definitions of quasiconformal mappings. In section five we introduce the main question in this thesis. Some properties like lower semicontinuity holds for curve family and analytic definitions but in section six we show that the linear distortion function is not lower semicontinuous. Also we illustrate some figures and examples.

4.1 Linear Distortion

In this section, we start to define some useful concepts such that they help to define a new concept that is called linear distortion. Let A be a subset of $\mathbb{R}^n$ and B be a subset of $\mathbb{R}^m$. Then we can prove that for a function $\mathbf{f} : A \longrightarrow B$, the following are equivalent:

- $\mathbf{f}$ is continuous.
- For every set U that is open in B, then $\mathbf{f^{-1}}(U)$ is open in A.
- For every set U that is closed in B, then $\mathbf{f^{-1}}(U)$ is closed in A.

Proof of this proposition exist in [28] or [39]. But you should note that $\mathbb{R}^n$ and $\mathbb{R}^m$ are Euclidean normed and also topological spaces (See figure (3.2)). We next wish to describe what it means for A and B to be topologically the same. There should be a bijection between them that pairs open sets with open sets (see [39]).

Definition 4.1.1. A function $\mathbf{f} : A \longrightarrow B$ is called a **homeomorphism** if $\mathbf{f}$ is bijective and continuous and $\mathbf{f^{-1}}$ is continuous. If such a function exists, then A and B are said to be **homeomorphic**.

So, another way to define a homeomorphism is to say that it is a bijective correspondence $\mathbf{f} : A \longrightarrow B$ such that $\mathbf{f}(U)$ is open in B if and only if U in open in A. This remark shows that a homeomorphism $\mathbf{f} : A \longrightarrow B$ gives us a bijective correspondence not only between A and B but also between the collections of open sets of A and of Y (see [28]). A subset $U \subset \mathbb{R}^n$ is called **path-connected** if for every pair $\mathbf{p}, \mathbf{q} \in U$, there exists a continuous function $\mathbf{g} : [0, 1] \longrightarrow U$ with $\mathbf{g}(0) = \mathbf{p}$ and $\mathbf{g}(1) = \mathbf{q}$. A set $U \subset \mathbb{R}^n$ is called **connected** if there is no subset of U that is both open in U and closed in U. We can prove that every path-connected set $U \subset \mathbb{R}^n$ is connected (see [39] and [28]). Let (X, d) be a metric space and $f : X \longrightarrow \mathbb{R}$ is a real valued function. We recall that U_y and L_y denote upper contour set of f and lower contour set of f at y. They are defined as

$$U_y = f^{-1}([y, \infty)) = \{x \in X \mid f(x) \geq y\}$$

$$L_y = f^{-1}((-\infty, y]) = \{x \in X \mid f(x) \leq y\}.$$

Let $f : X \longrightarrow \mathbb{R}$ be a function. Then we can prove that the following are equivalent.

- For any $y \in \mathbb{R}$, U_y is closed.
- For any $y \in \mathbb{R}$, $f^{-1}((-\infty, y)) = [U_y]^c$ is open.
- For any $x \in X$, if the sequence $\{x_n\}$ in X converges to x, then for any $\epsilon > 0$ there is a natural number N such that for all $n > N$, $f(x) > f(x_n) - \epsilon$.

Also, we can prove that the following are equivalent.

- For any $y \in \mathbb{R}$, L_y is closed.
- For any $y \in \mathbb{R}$, $f^{-1}((y, \infty)) = [L_y]^c$ is open.
- For any $x \in X$, if the sequence $\{x_n\}$ in X converges to x, then for any $\epsilon > 0$ there is a natural number N such that for all $n > N$, $f(x) < f(x_n) + \epsilon$.

We do not prove these theorems and skip on their proofs. You can find all of proofs in every topology and analysis books. Also you can see (https://pages.wustl.edu/files/pages/imce/nachbar/semi-continuity.pdf).

Definition 4.1.2. Let $f : X \longrightarrow \mathbb{R}$ be a function.

1. f is **upper semicontinuous** if and only if for any $y \in \mathbb{R}$, $f^{-1}((-\infty, y))$ is open.
2. f is **lower semicontinuous** if and only if for any $y \in \mathbb{R}$, $f^{-1}((y, \infty))$ is open.

Informally, a function is upper semicontinuous if it is continuous or, if not, it only jumps up; a function is lower semicontinuous if it is continuous or, if not, it only jumps down. We can show that f is continuous if and only if it is both upper and lower semicontinuous. This is obvious that f is lower semicontinuous if and only if $-f$ is upper continuous and vice versa. There is another definition for lower and upper semicontinuous. Let $f : X \longrightarrow \mathbb{R}$ be a function. f is called lower semicontinuous at $a \in X$ if

$$\liminf_{x \to a} f(x) \geq f(a),$$

and is called upper semicontinuous at $a \in X$ if

$$\limsup_{x \to a} f(x) \leq f(a).$$

Theorem 4.1.1. *Let $f : X \longrightarrow \mathbb{R}$ be a function. Then the following conditions are equivalent:*

1. *f is lower semicontinuous on X.*
2. *The lower contour set of f at $y \in \mathbb{R}$, L_y, is closed.*
3. *The epi f is closed in $X \times \mathbb{R}$.*

We can explain this theorem for upper semicontinuous function on X. For more details and seeing proof you can see [8] and [25]. An important property of lower semicontinuous functions is given by the following well-known Weierstrass theorem. A lower semicontinuous function f on a compact topological space X takes a minimum value on X (see [8]). We need to next theorem to define another definition.

Theorem 4.1.2. *Let A be a subset of $\mathbb{R}^n$, and let $\mathbf{f} : A \longrightarrow \mathbb{R}^n$ be an injective function. Suppose that $\mathbf{f}$ is continuous in some neighbourhood of a point x, that $\mathbf{f}$ is differentiable at x itself, and that $D\mathbf{f}(x)$ is nonsingular. Then $y = \mathbf{f}(x)$ is an interior point of $\mathbf{f}(A)$, the function $\mathbf{g} = \mathbf{f^{-1}} : \mathbf{f}(A) \longrightarrow A$ is differentiable at y, and*

$$D\mathbf{g}(y) = [D\mathbf{f}(x)]^{-1}$$

Consider an open set U in $\mathbb{R}^n$ and injective function $\mathbf{f}$ belonging to C^k for some $k \geq 1$. In case $k \geq 1$ assume additionally that $J_\mathbf{f}(x) \neq 0$ for every x in U, where $J_\mathbf{f}$ denotes the **Jacobian determinant** of $\mathbf{f}$, i.e., $J_\mathbf{f}(x) = \det[D\mathbf{f}(x)]$. (see [25], [18] and [6]).
Let A be a subset of $\mathbb{R}^n$ and B be a subset of $\mathbb{R}^m$, in this case, A and B are called **diffeomorphic**, if there exists a smooth bijective function $\mathbf{f} : A \longrightarrow B$ whose inverse is also smooth. $\mathbf{f}$ is called a **diffeomorphism**. In general, we can say, for k a positive integer or ∞, a C^k-diffeomorphism $\mathbf{f}$ between open sets A and B in $\mathbb{R}^n$ is a bijection $\mathbf{f} : A \longrightarrow B$ such that both $\mathbf{f}$ and $\mathbf{f^{-1}}$ are of class C^k. Unless otherwise stated, the word "diffeomorphism" will mean C^1-diffeomorphism. If $k \geq 1$ and the set A is connected, then either $J_\mathbf{f}(x) > 0$ for every x in A or $J_\mathbf{f}(x) < 0$ for all such x. Now we can explain a criterion that enables us to determine whether a function is, at least on a local level, a diffeomorphism.

Theorem 4.1.3. *Let U be an open set in $\mathbb{R}^n$, and let $\mathbf{f}$ be a member of the class C^k with $1 \leq k \leq \infty$. If x is a point of U for which $J_\mathbf{f}(x) \neq 0$, then there exists an $r > 0$ such that the restriction of $\mathbf{f}$ to the ball $B = B^n(x, r)$ is a C^k-diffeomorphism of B onto $\mathbf{f}(B)$.*

If an embedding $\mathbf{f} : U \longrightarrow \mathbb{R}^n$ is differentiable at $a \in U$, then the Jacobian determinant $J_\mathbf{f}(x) = \det[D\mathbf{f}(x)]$ represents the infinitesimal change of volume under the mapping $\mathbf{f}$. To be precise, we have

$$|J_\mathbf{f}(a)| = \lim_{r\to 0} \frac{|\mathbf{f}(B(a,r))|}{|B(a,r)|}, \tag{4.1}$$

where $|B|$ denotes the n-dimensional measure of B. For more details you can see [18] and [25].
If A and B are nonempty subset of $\mathbb{R}^n$, then $\operatorname{diam}(A)$ and $\operatorname{dist}(A, B)$ denote the **Euclidean diameter** of A and the **Euclidean distance** between A and B, respectively. Thus

$$\operatorname{diam}(A) = \sup\big\{\|x - y\| \,:\, x, y \in A\big\}, \qquad \operatorname{dist}(A, B) = \inf\big\{\|x - y\| \,:\, x \in A\,,\; y \in B\big\}.$$

We shorten $\operatorname{dist}(x, B)$ to the more compact $\operatorname{dist}(x, B)$ (see [18]). Now, we define the main subject of this thesis that is called linear distortion or dilatation. We know that the **domain in topology** means a open connected set. So, in the rest of this thesis, whenever we use the domain term, we mean a connected open set in $\mathbb{R}^n$.

Definition 4.1.3. Let Ω be a domain in $\mathbb{R}^n$, let $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ be a continuous injection. The **linear distortion or dilatation** of $\mathbf{f}$ at the point x in A is the quantity $H_\mathbf{f}(x)$ or $H(x, \mathbf{f})$ defined by

$$H_\mathbf{f}(x) = H(x, \mathbf{f}) = \limsup_{r\to 0} \frac{L_f(x,r)}{\ell_f(x,r)}, \tag{4.2}$$

where for $0 < r < \mathrm{dist}(x, \partial\Omega)$ we set

$$L_f(x,r) = \max_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|, \qquad \ell_f(x,r) = \min_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|.$$

In the other words,

$$H_{\mathbf{f}}(x) = H(x, \mathbf{f}) = \limsup_{r \to 0} \frac{\max_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|}{\min_{\|h\|=r} \|\mathbf{f}(x+h) - \mathbf{f}(x)\|}. \tag{4.3}$$

At points where the differential $D\mathbf{f}(a)$ exists and is nonsingular we clearly have the limit in (4.3) existing. We will show that the following equation is correct.

$$\max_{|h|=1} |D\mathbf{f}(x)h| \leq H(x, \mathbf{f}) \min_{|h|=1} |D\mathbf{f}(x)h| \tag{4.4}$$

Geometrically this means that the differential $D\mathbf{f}(a) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ maps the unit sphere to an ellipsoid for which ratio of the lengths of the largest and smallest semiaxes, the **eccentricity**, is equal to $H(a, \mathbf{f})$. If we define $C\mathbf{f}(x) = D^t\mathbf{f}(x).D\mathbf{f}(x)$, this matrix is referred to as the **right Cauchy-Green strain tensor**. The positive square roots of its eigenvalues, the singular values of $D\mathbf{f}(a)$, are the **principal stretchings**. The corresponding eigenvectors are called the **principal directions** of the deformation at $a \in \Omega$ (see [25]).
Our **distortion tensor** will be represented by the positive symmetric matrix defined by

$$G(x) = J_{\mathbf{f}}(x)^{-\frac{2}{n}} D^t\mathbf{f}(x) D\mathbf{f}(x). \tag{4.5}$$

Let

$$0 < \mu_1(x) \leq \mu_2(x) \leq \ldots \leq \mu_n(x) \tag{4.6}$$

denote the positive square roots of the eigenvalues of the direction tensor $G(x)$. Now returning to equation (4.3) one easily obtains the identity

$$H_{\mathbf{f}}(x) = H(x, \mathbf{f}) = \frac{\mu_n(x)}{\mu_1(x)} \tag{4.7}$$

Thus $H_{\mathbf{f}}(x) = H(x, \mathbf{f})$ measures the maximum possible relative distortion of linear objects (e.g. length of curves) at the infinitesimal level at $a \in \Omega$ (see [25]). Now we recall the following result of Rademacher-Stepanoff.

Theorem 4.1.4. ***(Rademacher-Stepanoff)*** *Let $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ be an arbitrary mapping defined on an open set $\Omega \subset \mathbb{R}^n$. Consider the set*

$$E = \Big\{ a \in \Omega \; : \; \limsup_{x \to a} \frac{|\mathbf{f}(x) - \mathbf{f}(a)|}{|x - a|} < \infty \Big\}.$$

Then E is a Lebesgue measurable set and $\mathbf{f}$ is differentiable at almost every point of E.

Proof. Proof of this theorem exist in Federer's book [16], page 218. ♠

In this thesis, we use the term **embedding** to mean a continuous injection. Also, we remark that in Rademacher-Stepanoff theorem the mapping $\mathbf{f}$ need not even be assumed measurable. Let Ω be an open subset of $\mathbb{R}^n$, $n \geq 2$, and let $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ be an embedding. Then the linear distortion function $H_{\mathbf{f}}(x) = H(x, \mathbf{f})$ defined in (4.3) is a Borel function (see [25] and [18]). The first step towards establishing the regularity properties of mapping of finite distortion is given by the following differentiability theorem.

Theorem 4.1.5. *Every embedding* $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ *is differentiable at almost every point of the set*

$$E = \{x \in \Omega \ : \ H(x, \mathbf{f}) < \infty\}.$$

Moreover, for almost every point $x \in E$ *we have*

$$|D\mathbf{f}(x)| \leq H(x, \mathbf{f}) |J_{\mathbf{f}}(x)|^{\frac{1}{n}} \tag{4.8}$$

and

$$\max_{|h|=1} |D\mathbf{f}(x)h| \leq H(x, \mathbf{f}) \min_{|h|=1} |D\mathbf{f}(x)h| \tag{4.9}$$

A proof can be found in [25], page 103 and 104. We can explain the following corollary that it is so useful.

Corollary 4.1.5.1. *Every continuous locally injective mapping* $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ *whose linear distortion function is finite almost everywhere is differentiable almost everywhere in* Ω*.*

4.2 Inner and Outer Distortions

In section (4.2) we defined the linear distortion. in this section we want to define its properties and two new definitions of inner and outer distortion. We recall some definitions and Linear algebra.

When a linear transformation $S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ enjoys the property that $S^* = S$, we say that S is symmetric or self-adjoint. For example, if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an arbitrary linear transformation, then both T^*T and TT^* are symmetric or self-adjoint. By spectral theorem (1.4.8) we can say, S is a symmetric linear transformation of $\mathbb{R}^n$, then there exists $U \in O(n)$ such that $D = U^{-1}SU$ has the form

$$D(x) = (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_n x_n).$$

In particular, by theorem (1.4.7) the eigenvalues of S (that is, $\lambda_1, \lambda_2, \ldots, \lambda_n$) are real. The previous transformation D is called **diagonal transformation**. We observe that

$$\min_{1 \leq i \leq n} |\lambda_i| \leq |D(x)| = \sqrt{\lambda_1^2 x_1^2 + \lambda_2^2 x_2^2 + \ldots + \lambda_n^2 x_n^2} \leq \max_{1 \leq i \leq n} |\lambda_i| \tag{4.10}$$

whenever $|x| = 1$, so $L(D) = \max_{1 \leq i \leq n} |\lambda_i|$ and $\ell(D) = \min_{1 \leq i \leq n} |\lambda_i|$. Also, we have $\det(D) = \lambda_1 \lambda_2 \ldots \lambda_n$. To say that a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is positive definite (respectively, positive semidefinite) means that $\langle x, T(x) \rangle > 0$ (respectively, $\langle x, T(x) \rangle \geq 0$) for every nonzero vector x in $\mathbb{R}^n$. For instance, if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an arbitrary linear transformation, then T^*T and TT^* are positive semidefinite. If, in addition, T is nonsingular, then T^*T and TT^* are actually positive definite. By theorem (1.4.9), any eigenvalue of a positive definite (respectively, positive semidefinite) linear transformation is positive (respectively, nonnegative). You can see [25] or [18].

Suppose that a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is both symmetric and positive semidefinite. We can list the eigenvalue of T in the manner $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$ and assert the existence of U in $O(n)$ for which $U^{-1}TU = D$, where $D(x) = (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_n x_n)$. According to formula (3.5), we can write that

$$L(T) = L(U^{-1}TU) = L(D) = \lambda_1.$$

Similarly, $\ell(T) = \lambda_n$ and $\det(T) = \lambda_1\lambda_2\ldots\lambda_n$. We can explain the next theorem.

Theorem 4.2.1. *Any linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ can be factored as $T = PU$, where U belongs to $O(n)$ and P is both symmetric and positive semidefinite.*

The linear transformation P in theorem (4.2.1) is uniquely determined by T: denoting by I the identity transformation of $\mathbb{R}^n$, we compute

$$TT^* = (PU)(PU)^* = PUU^*P^* = PIP = P^2,$$

and learn that P is the (known to be unique) symmetric, positive semidefinite square root of TT^* (see [18]). This observation entitle us to list the eigenvalues of P as $\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_n^{1/2}$, where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$ are the eigenvalues of TT^*. Therefore, there is an orthogonal transformation V such that $V^{-1}PV = D$, with $y = D(x)$ given as follows:

$$y_1 = \lambda_1^{1/2}x_1\,, \ldots,\; y_n = \lambda_n^{1/2}x_n.$$

If T is nonsingular then all eigenvalues of TT^* are positive, which makes it apparent that D maps the sphere $\mathbb{S}^{n-1}$ bijectively to the ellipsoid E whose equation is $(y_1^2/\lambda_1) + (y_2^2/\lambda_2) + \ldots + (y_n^2/\lambda_n) = 1$. Because $U(\mathbb{S}^{n-1}) = V(\mathbb{S}^{n-1}) = \mathbb{S}^{n-1}$, we can write

$$T(\mathbb{S}^{n-1}) = PU(\mathbb{S}^{n-1}) = P(\mathbb{S}^{n-1}) = PV(\mathbb{S}^{n-1}) = VD(\mathbb{S}^{n-1}) = V(E).$$

So, we can say

$$L(T) = L(PU) = L(P) = \sqrt{\lambda_1} = \sigma_1$$

and

$$\ell(T) = \ell(PU) = \ell(P) = \sqrt{\lambda_n} = \sigma_n,$$

where σ_1 and σ_n are singular values of T. A further result of theorem (4.2.1) is the useful inequality

$$\ell(T)^n \leq |\det(T)| \leq L(T)^n, \tag{4.11}$$

valid for any linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. We know that

$$|\det(T)| = \det(P) = \sqrt{\lambda_1\lambda_2\ldots\lambda_n},$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of TT^* (see [18]). Now, we can express a theorem such that it summarizes these above discussion.

Theorem 4.2.2. *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear transformation, and list the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of TT^*in nonincreasing order. Then there exist orthogonal transformations $U, V \in O(n)$ such that $VTU = D$, where*

$$D(x) = (\lambda_1^{1/2}x_1, \lambda_2^{1/2}x_2, \ldots, \lambda_n^{1/2}x_n).$$

Definition 4.2.1. We can define the **linear dilatation** $H(T)$, **inner dilatation** $H_I(T)$, and the **outer dilatation** $H_O(T)$ of a nonsingular linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ as

$$H(T) = \frac{L(T)}{\ell(T)}, \qquad H_I(T) = \frac{|\det(T)|}{\ell(T)^n}, \qquad H_O(T) = \frac{L(T)^n}{|\det(T)|}. \tag{4.12}$$

If T is singular it is customary to set $H(T) = H_I(T) = H_O(T) = \infty$ (see [18]).

In geometric terms, $H(T)$ measures the eccentricity of the ellipsoid $T(\mathbb{S}^{n-1})$ while $H_I(T)$ and $H_O(T)$ relate the volume of $T(\mathbf{B^n})$ to the volumes of the balls centered at the origin that are, respectively, inscribed in on circumscribed about $T(\mathbb{S}^{n-1})$. We remark that

$$H(T^{-1}) = H(T), \qquad H_I(T^{-1}) = H_O(T), \qquad H_O(T^{-1}) = H_I(T) \tag{4.13}$$

and that

$$H(VTU) = H(T), \qquad H_I(VTU) = H_I(T), \qquad H_O(VTU) = H_O(T) \tag{4.14}$$

if V and U are members of $O(n)$. Because of (4.14) and theorem (4.2.2) we can show that

$$1 \leq H_O(T) \leq H_I(T)^{n-1}, \qquad 1 \leq H_I(T) \leq H_O(T)^{n-1} \tag{4.15}$$

and also that

$$1 \leq \min\{H_I(T), H_O(T)\} \leq H(T)^{n/2} \leq \max\{H_I(T), H_O(T)\} \leq H(T)^{n-1} \tag{4.16}$$

We can see that as a result of (4.16) $H(T) = H_I(T) = H_O(T)$ when $n = 2$ (see [25] and [18]). Let Ω and Ω' be domains in $\mathbb{R}^n$ with $n \geq 2$, and let $\mathbf{f}$ be a homeomorphism of Ω onto Ω'. The **linear distortion of function f** is defined

$$H(\mathbf{f}) = \sup\{H_{\mathbf{f}}(x) \, : \, x \in \Omega\}, \tag{4.17}$$

where $H_{\mathbf{f}}(x) \geq 1$ is the linear distortion of $\mathbf{f}$ at x as defined at (4.2). $\mathbf{f}$ is a conformal mapping if and only if $H(\mathbf{f}) = 1$ (see [18] p.77).

4.3 The Moduli of A Curve Families

In this section we will explain the notion of the moduli of a curve families. This subject is so useful for describing the quasiconformal mappings in next section. We start with the some essential definitions.

Definition 4.3.1. A **path** in $\hat{\mathbb{R}}^n$ is a continuous mapping $\alpha : I \longrightarrow \hat{\mathbb{R}}^n$ where I is an interval in $\mathbb{R}$. The path is said to be closed or open if I is closed or open, respectively. The **locus** $|\alpha|$ of a path $\alpha : I \longrightarrow \hat{\mathbb{R}}^n$ is the point set $\alpha(I) \subset \hat{\mathbb{R}}^n$. The **subpath** of $\alpha : I \longrightarrow \hat{\mathbb{R}}^n$ is a restriction of α to a subinterval of I. Suppose $a = t_0 \leq t_1 \leq \ldots \leq t_n = b$ is a partition of interval $[a, b]$. Then we define the **length** of $\alpha : [a, b] \longrightarrow \hat{\mathbb{R}}^n$ and denote by $\ell(\alpha)$ as

$$\ell(\alpha) = \sup \Big\{ \sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})| \Big\}.$$

Thus $0 \leq \ell(\alpha) \leq \infty$ and it is obvious to see $\ell(\alpha) = 0$ if and only if α is a constant . If $\ell(\alpha) < \infty$ then α is called **rectifiable path**. Also a path in $\hat{\mathbb{R}}^n$ such that $\infty \in |\alpha|$ is non-rectifiable, except for the constant path $\alpha(t) = \infty$ (see [40] p.1).

In section (3.2) we defined the conformal mapping. Now we want to say another definition for it. Because we need to introduce the conformal maps with derivative and homeomorphism.

Definition 4.3.2. Let Ω and Ω' be domains in $\mathbb{R}^n$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Then $\mathbf{f}$ is conformal if $\mathbf{f} \in C^1$ and $|D\mathbf{f}(x)h| = |D\mathbf{f}(x)||h|$ for all $x \in \Omega$ and $h \in \mathbb{R}^n$.

We can also say a C^1 homeomorphism $\mathbf{f}$ is conformal if and only if $|\mathbf{f}'(x)|^n = J_{\mathbf{f}}(x)$ for all $x \in \Omega$. By **Liouville's theorem**, if Ω is a domain in $\mathbb{R}^n$ with $n \geq 3$, then any conformal mapping $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ is a restriction to Ω of a Möbius transformation (see [25] and [18]). Suppose that Γ is a **family of smooth paths in** Ω. Γ is called a **curve family** and it is simply a collection of curves. Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a diffeomorphism. Assume γ is a family of smooth paths in Ω. We denote $\mathbf{f}(\Gamma)$ for $\{\mathbf{f} \circ \gamma \ : \ \gamma \in \Gamma\}$. So, $\mathbf{f}(\Gamma)$ is also a family of smooth paths in Ω'.
Given a curve family Γ, an **admissible density** is a Borel function $\rho : \Omega \longrightarrow [0, \infty)$, if a ρ-length $\ell_\rho(\gamma)$ to any smooth path γ in Ω:

$$\ell_\rho(\gamma) = \int_\gamma \rho(x)|\mathrm{dx}| \geq 1 \qquad for \;\; every \;\; \gamma \in \Gamma$$

In the other words, the continuous function $\rho : \Omega \longrightarrow [0, \infty)$ is an admissible density if for all γ in Ω we have $\ell_\rho(\gamma) \geq 1$. There is an analogous notion of ρ-volume: the ρ-volume $\upsilon_\rho(A)$ of a Lebesgue measurable subset A of Ω is given by

$$\upsilon_\rho(A) = \int_A \rho^n(x)\mathrm{dx},$$

dx denoting with respect to Lebesgue measure in $\mathbb{R}^n$. The collection of all such admissible density is denoted by $\mathrm{Adm}(\Gamma)$ (see [18] p.78).
The condition $\ell_\rho(\gamma) \geq 1$ prevents ρ from being, on average, excessively small along the trajectory of any $\gamma \in \Gamma$. So, if ρ is an admissible density for Γ and if Γ contains sufficiently many paths and their collective trajectories fill up a substantial portion of Ω, one might expect that the associated volume $\upsilon_\rho(\Omega)$ could not itself be small. Now we can consider the new quantity to control the volume.

Definition 4.3.3. The **modulus of curve family** Γ is

$$M(\Gamma) = \inf\{\upsilon_\rho(\Omega) \ : \ \rho \in \mathrm{Adm}(\Gamma)\} = \inf \int_{\mathbb{R}^n} \rho^n(x)\mathrm{dx},$$

where the infimum is over all admissible densities for Γ. If $\mathrm{Adm}(\Gamma) = \phi$ then we set $M(\Gamma) = \infty$. From the definition of the modulus of curve family, we have $0 \leq M(\Gamma) \leq \infty$.

There are some properties of the modulus of a curve family and we explain and list them.

1. $\mathrm{Adm}(\Gamma) = \phi$ if and only if Γ contains a single constant path.
2. If $\Gamma_1 \subset \Gamma_2$, then $M(\Gamma_1) \leq M(\Gamma_2)$.
3. $M(\phi) = 0$.
4. $M(\bigcup_{i=1}^{\infty} \Gamma_i) \leq \sum_{i=1}^{\infty} M(\Gamma_i)$.
5. If Γ_1 and Γ_2 are curve families such that every curve in Γ_2 has a subcurve in Γ_1, then $M(\Gamma_1) \geq M(\Gamma_2)$.

All proofs exist in Väiälä's book [40] from page 16 to 20. One of the most important result is that modulus is a conformal invariant. It can be described by the following theorem.

Theorem 4.3.1. *Let Ω and Ω' be domains in $\mathbb{R}^n$, $\mathbf{f}:\Omega\longrightarrow\Omega'$ is a conformal map and $\tilde{\Gamma}=\mathbf{f}(\Gamma)$, then $M(\Gamma)=M(\tilde{\Gamma})$, for all $\Gamma\subset\Omega$.*

Proof. Let $\tilde{\rho}\in \mathrm{Adm}(\tilde{\Gamma})$, now we can define $\rho(x)=\tilde{\rho}(\mathbf{f}(x)).|D\mathbf{f}(x)|$. We have $\rho\in\mathrm{Adm}(\Gamma)$ since for all locally rectifiable $\gamma\in\Gamma$

$$\int_\gamma \rho \mathrm{ds}=\int_\gamma \tilde{\rho}(\mathbf{f}(x)).|D\mathbf{f}(x)|\mathrm{ds}=\int_{\mathbf{f}\circ\gamma}\tilde{\rho}(x)\mathrm{ds}\geq 1.$$

Therefore, we can write

$$\begin{aligned} M(\Gamma) &\leq \int_\Omega \rho^n \mathrm{dx} \\ &= \int_\Omega (\tilde{\rho}(\mathbf{f}(x)))^n.|D\mathbf{f}(x)|^n\mathrm{dx} \\ &= \int_\Omega (\tilde{\rho}(\mathbf{f}(x)))^n.J_{\mathbf{f}}(x)\mathrm{dx} \\ &= \int_\Omega (\tilde{\rho}(x))^n\mathrm{dx}. \end{aligned}$$

Now by taking the infimum over all $\tilde{\rho}$ we obtain $M(\Gamma)\leq M(\tilde{\Gamma})$. Since $\mathbf{f}$ is conformal then $\mathbf{f^{-1}}$ is conformal. So, with the same reason we have $M(\tilde{\Gamma})\leq M(\Gamma)$, and so, $M(\Gamma)=M(\tilde{\Gamma})$ (see [25] or [40]). ♠

Suppose that $\mathbf{f}:\Omega\longrightarrow\Omega'$ is a diffeomorphism between domains Ω and Ω' in $\mathbb{R}^n$, where $n\geq 2$. In the proof of theorem (4.3.1) we defined ρ with respect to $\tilde{\rho}$ that it is called **push forward**. Let $\rho:\Omega\longrightarrow[0,\infty)$ be defined by

$$\rho(x)=\tilde{\rho}\big(\mathbf{f}(x)\big).|\mathbf{f}'(x)|.$$

Let $\gamma:[a,b]\longrightarrow\Omega$ be a path in Γ. Then $\beta=\mathbf{f}\circ\gamma$ is a path and member of $\tilde{\Gamma}$, because it is easy to prove that

$$1\leq \ell_{\tilde{\rho}}(\beta)=\int_\beta \tilde{\rho}(x)|\mathrm{dx}|=\int_a^b \tilde{\rho}[\beta(t)]|\beta'(t)|\mathrm{dt}=\int_a^b \tilde{\rho}[\beta(t)]|\mathbf{f}'[\gamma(t)].\gamma'(t)|\mathrm{dt}$$

$$\leq \int_a^b \tilde{\rho}[\beta(t)]|\mathbf{f}'[\gamma(t)]|.|\gamma'(t)|\mathrm{dt}=\int_a^b \rho(\gamma(t))|\gamma'(t)|\mathrm{dt}=\int_\gamma \rho(x)|\mathrm{dx}|=\ell_\rho(\gamma).$$

In the definition (4.2.1) we defined linear, inner and outer dilatations for linear transformation T. We know that the derivative of $\mathbf{f}$ is a linear transformation (see [18] and [25]). We recall the definition of $H_O(T)$ for a linear transformation T given at (4.12), now we can define the **outer distortion**

$$K_O(\mathbf{f})=\sup_{x\in\Omega}H_O[\mathbf{f}'(x)]=\sup_{x\in\Omega}\frac{\|\mathbf{f}'(x)\|^n}{|J_{\mathbf{f}}(x)|} \tag{4.18}$$

Let $K_O(\mathbf{f})<\infty$. Then by standard change of variable formula, we can write

$$\nu_\rho(\Omega)=\int_\Omega \rho^n(x)\mathrm{dx}=\int_\Omega \tilde{\rho}^n[\mathbf{f}(x)]|\mathbf{f}'(x)|^n\mathrm{dx}$$

$$\leq K_O(\mathbf{f}) \int_{\Omega} \tilde{\rho}^n[\mathbf{f}(x)]|J_{\mathbf{f}}(x)|\mathrm{dx} = K_O(\mathbf{f}) \int_{\Omega} \tilde{\rho}^n(y)\mathrm{dy} = K_O(\mathbf{f})\nu_{\tilde{\rho}}(\Omega').$$

So, we have

$$\nu_\rho(\Omega) \leq K_O(\mathbf{f})\nu_{\tilde{\rho}}(\Omega') \tag{4.19}$$

If $\mathrm{Adm}(\tilde{\Gamma}) \neq \phi$, by (4.19), we have

$$M(\Gamma) \leq \nu_\rho(\Omega) \leq K_O(\mathbf{f})\nu_{\tilde{\rho}}(\Omega')$$

for every $\tilde{\rho}$ in $\mathrm{Adm}(\tilde{\Gamma})$, hence

$$M(\Gamma) \leq K_O(\mathbf{f})M(\tilde{\Gamma}). \tag{4.20}$$

Inequality (4.20) holds trivially when $\mathrm{Adm}(\tilde{\Gamma}) = \phi$, a case in which $M(\tilde{\Gamma}) = \infty$ by definition (see [18] p.80). Therefore, with the same process for $\mathbf{f^{-1}}$, the inequality (4.20) can be written as

$$M(\tilde{\Gamma}) \leq K_I(\mathbf{f})M(\Gamma) \tag{4.21}$$

when $K_I(\mathbf{f}) = K_O(\mathbf{f^{-1}}) < \infty$. Incidentally, it is a consequence of (4.13) that

$$K_I(\mathbf{f}) = \sup_{y\in D'} H_O[(\mathbf{f^{-1}})'(y)] = \sup_{x\in D} H_I[\mathbf{f}'(x)] = \sup_{x\in D} \frac{|J_{\mathbf{f}}(x)|}{\ell[\mathbf{f}'(x)]^n} \tag{4.22}$$

$K_I(\mathbf{f})$ is called the **inner distortion**. If the diffeomorphism $\mathbf{f}$ is a conformal mapping of D onto D', we have

$$H_O[\mathbf{f}'(x)] = H_I[\mathbf{f}'(x)]$$

for every x in D, in which event we obtain $K_O(\mathbf{f}) = K_I(\mathbf{f}) = 1$ (see [18] p.80). So, we can define inner and outer distortions by curve families.

Definition 4.3.4. Let Ω and Ω' be domains in $\mathbb{R}^n$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Let Γ be a curve family in Ω and its image $\tilde{\Gamma} = \{\mathbf{f} \circ \gamma \ : \ \gamma \in \Gamma\}$ is a curve family in Ω'. The outer distortion $K_O(\mathbf{f})$ and inner distortion $K_I(\mathbf{f})$ of a function $\mathbf{f}$ are defined as the following formulas

$$K_O(\mathbf{f}) = \sup_{\gamma\in\Gamma} \frac{M(\Gamma)}{M(\tilde{\Gamma})}, \qquad K_I(\mathbf{f}) = \sup_{\gamma\in\Gamma} \frac{M(\tilde{\Gamma})}{M(\Gamma)}, \tag{4.23}$$

where $M(\Gamma)$ and $M(\tilde{\Gamma})$ are not 0 or ∞. The maximal distortion of $\mathbf{f}$ is defined by

$$K(\mathbf{f}) = \max\{K_O(\mathbf{f}), K_I(\mathbf{f})\}.$$

We note that $K_I \geq 1$ or $K_O \geq 1$, hence $K \geq 1$. Since $M(\Gamma) = M(\tilde{\Gamma}) = \infty$, whenever one and hence the other of the families Γ and $\tilde{\Gamma}$ contain a constant path, and the constant paths have no influence on $K_I(\mathbf{f})$ and $K_O(\mathbf{f})$. In order to avoid technical difficulties, we shall assume from now on that every path family contains only non-constant paths (see [40] p.41 and 42).

Definition 4.3.5. Let Ω and Ω' be domains in $\mathbb{R}^n$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Let Γ be a curve family in Ω and its image $\tilde{\Gamma} = \{\mathbf{f} \circ \gamma \ : \ \gamma \in \Gamma\}$ is a curve family in Ω'. Then $\mathbf{f}$ is said to be **K-quasiconformal mapping** of Ω to Ω' provided $K(\mathbf{f}) < \infty$. If, in fact, $K(\mathbf{f}) \leq K < \infty$, then we call $\mathbf{f}$ a K-quasiconformal mapping. For this to be true it is necessary and sufficient that

$$\frac{1}{K}M(\Gamma) \leq M(\tilde{\Gamma}) \leq KM(\Gamma) \tag{4.24}$$

hold for every family Γ of curves in Ω (see [18] p.206 and [25] p.14).

From the definition of the inner and outer definition we will obtain the following relations :
Theorem 4.3.2. *Let Ω and Ω' be domains in $\mathbb{R}^n$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. The following properties hold for every $x \in \Omega$:*

$$\begin{array}{ll} 1)\ K_I(\mathbf{f}) = K_O(\mathbf{f^{-1}}) & 2)\ K_O(\mathbf{f}) = K_I(\mathbf{f^{-1}}) \\ 3)\ K(\mathbf{f^{-1}}) = K(\mathbf{f}) & 4)\ K_I(\mathbf{f}\circ\mathbf{g}) \leq K_I(\mathbf{f})K_I(\mathbf{g}) \\ 5)\ K_O(\mathbf{f}\circ\mathbf{g}) \leq K_O(\mathbf{f})K_O(\mathbf{g}) & 6)\ K(\mathbf{f}\circ\mathbf{g}) \leq K(\mathbf{f})K(\mathbf{g}) \end{array} \tag{4.25}$$

Proof. (Proof of part 1) Suppose that Γ is a family of paths in Ω and $\tilde{\Gamma} = \mathbf{f}(\Gamma)$. By definition (4.23), we have

$$\frac{M(\tilde{\Gamma})}{M(\Gamma)} = \frac{M(\tilde{\Gamma})}{M(\mathbf{f^{-1}}(\tilde{\Gamma}))} \leq K_O(\mathbf{f^{-1}})$$

and by taking the supremum over all Γ we obtain $K_I(\mathbf{f}) = K_O(\mathbf{f^{-1}})$. Proofs of part (2) and (3) are similar with (1).
(Proof of part 4) Consider $K_I(\mathbf{f}\circ\mathbf{g})$, by definition of inner distortion we have

$$\frac{M(\tilde{\Gamma})}{M(\Gamma)} = \frac{M((\mathbf{f}\circ\mathbf{g})\Gamma)}{M(\Gamma)} = \frac{M((\mathbf{f}\circ\mathbf{g})\Gamma)}{M(\mathbf{g}(\Gamma))}.\frac{M(\mathbf{g}(\Gamma))}{M(\Gamma)} \leq K_I(\mathbf{f})K_I(\mathbf{g}).$$

Taking the supremum over all Γ from both sides gives us $K_I(\mathbf{f}\circ\mathbf{g}) \leq K_I(\mathbf{f})K_I(\mathbf{g})$. For parts (5) and (6) we have same processes (see [40]). ♠

Corollary 4.3.2.1. *If $\mathbf{f}$ is a K-quasiconformal mapping, then $\mathbf{f^{-1}}$ is a K-quasiconformal mapping.*
Corollary 4.3.2.2. *If $\mathbf{h} = \mathbf{f}\circ\mathbf{g}$ and $\mathbf{f}$ is a K_1-quasiconformal mapping and $\mathbf{g}$ is a K_2-quasiconformal mapping, then $\mathbf{h}$ is a K_1K_2-quasiconformal mapping.*

Now, we can extend these functions for matrix. Suppose $A \in M_{n\times n}(\mathbb{R})$ is a nonsingular matrix, then we can define

- the linear distortion

$$H(A) = |A||A^{-1}| = \frac{\max\{|Ax| \ : \ |x| = 1\}}{\min\{|Ax| \ : \ |x| = 1\}} \tag{4.26}$$

- the outer distortion

$$K_O(A) = \frac{|A|^n}{|\det(A)|} \tag{4.27}$$

- the inner distortion

$$K_I(A) = K_I(A^{-1}) = \frac{|\text{adj}(A)|^n}{|\det(A)|^{n-1}} \tag{4.28}$$

- the maximal distortion

$$K(A) = \max\{K_O(A), K_I(A)\}. \tag{4.29}$$

In particular,

$$H(A) = (K_O(A)K_I(A))^{\frac{1}{n}} \tag{4.30}$$

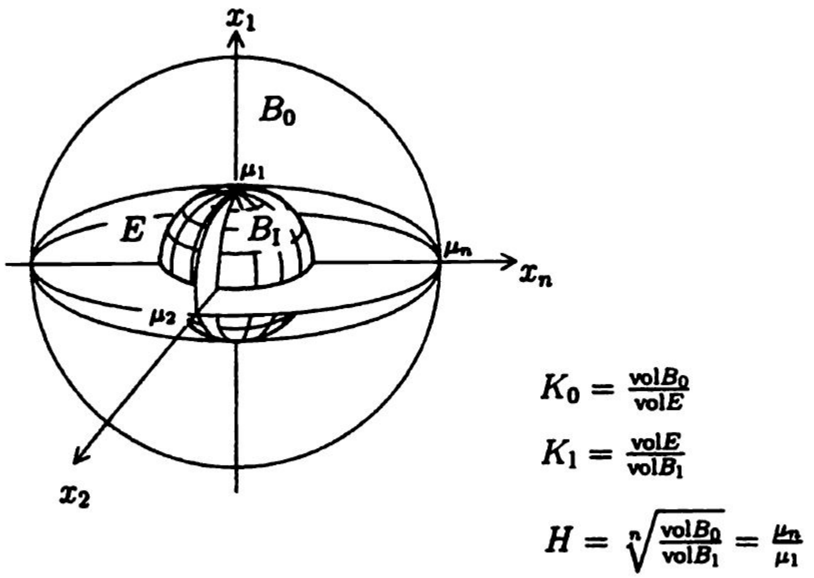

Figure 4.2: Eccentricity of the ellipsoid
Source: Page 110 of book [25]

If $0 < \mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ are eigenvalues of the matrix A^tA, then we can show the H, K_I and K_O in the following picture. This above figure shows the eccentricity of ellipsoid. Here and in what follows, all distortion functions of a singular matrix are assumed to be equal to ∞, except for the zero matrix where by convention they are all equal to 1. In two dimensions all these distortion functions coincide. This is not case for $n \geq 3$. However, these distortion functions are coupled by the inequalities (see [25])

$$\begin{array}{llll} H \leq K_O & K \leq K_O^{n-1} & K_O \leq K_I^{n-1} & K_I \leq K_O^{n-1} \\ H \leq K_I & K \leq K_I^{n-1} & K_O \leq H^{n-1} & K_I \leq H^{n-1} \\ H \leq K^{\frac{2}{n}} & K \leq H^{n-1} & K_O \leq K & K_I \leq K \end{array} \tag{4.31}$$

To prove of some of above inequalities, let $D\mathbf{f}(x)$ be a diagonal matrix with strictly positive entries

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0.$$

Then $|D\mathbf{f}(x)| = \lambda_1$, $|\text{adj}D\mathbf{f}(x)| = \lambda_1\lambda_2\ldots\lambda_{n-1}$ and $J_\mathbf{f}(x) = \lambda_1\lambda_2\ldots\lambda_n$. Hence it is easy to see

$$|\text{adj}D\mathbf{f}(x)| \leq |D\mathbf{f}(x)|^{n-1}$$

and therefore

$$K_I(x) = \frac{|\text{adj}D\mathbf{f}(x)|^n}{J_\mathbf{f}(x)^{n-1}} \leq \frac{|D\mathbf{f}(x)|^{n(n-1)}}{J_\mathbf{f}(x)^{n-1}} \leq K_O(x)^{n-1}.$$

Moreover, we have an equality if $\lambda_1 = \lambda_2 = \ldots = \lambda_{n-1}$. We can also prove that

$$|D\mathbf{f}(x)| = \lambda_1 \leq \lambda_1 \frac{\lambda_2}{\lambda_n}\frac{\lambda_3}{\lambda_n}\cdots\frac{\lambda_{n-1}}{\lambda_n} = \frac{(\lambda_1\lambda_2\ldots\lambda_{n-1})^{n-1}}{(\lambda_1\lambda_2\ldots\lambda_n)^{n-2}} = \frac{|\text{adj}D\mathbf{f}(x)|^{n-1}}{J_\mathbf{f}(x)^{n-2}}.$$

It follows that

$$K_O(x) = \frac{|D\mathbf{f}(x)|^n}{J_\mathbf{f}(x)} = \frac{|\mathrm{adj} D\mathbf{f}(x)|^{n(n-1)}}{J_\mathbf{f}(x)^{(n-1)(n-1)}} \leq K_I(x)^{n-1},$$

and for $\lambda_2 = \lambda_3 = \ldots = \lambda_n$ we have in fact equality (see [21] p.124). Let $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of matrix A. We call them the **sectional distortions** of A and they are defined for $1 \leq \alpha,\ \beta \leq n-1$ as

$$\mathcal{K}_{\alpha,\beta}(A) = \frac{(\lambda_n \lambda_{n-1} \cdots \lambda_{n-\alpha+1})^\beta}{(\lambda_1 \lambda_2 \cdots \lambda_\beta)^\alpha} \tag{4.32}$$

There is no repetition here. We can say $\mathcal{K}_{\alpha,\beta} = \mathcal{K}_{\alpha',\beta'}$ if and only if $\alpha = \alpha'$ and $\beta = \beta'$ (see [25]). Distortions can be described as particular cases of the $\mathcal{K}_{\alpha,\beta}$:

1. $\mathcal{K}_{1,n-1}(A) = (\lambda_1 \lambda_2 \cdots \lambda_n)^{-1} \lambda_n^n = K_O(A)$.
2. $\mathcal{K}_{n-1,1}(A) = (\lambda_1 \lambda_2 \cdots \lambda_n)^{1-n} (\lambda_2 \lambda_3 \cdots \lambda_n) = K_I(A)$.
3. $\mathcal{K}_{1,1}(A) = \lambda_n / \lambda_1 = H(A)$.

Also note the following symmetry relation:

$$\mathcal{K}_{\alpha,\beta}(A^{-1}) = \mathcal{K}_{\alpha,\beta}(A),$$

when $\det(A) \neq 0$ (see [25] p.110). Also we can write that

$$\mathcal{K}_{\alpha,\beta}(AB) \leq \mathcal{K}_{\alpha,\beta}(A) \mathcal{K}_{\alpha,\beta}(B).$$

The distortion functions $\mathcal{K}_{\alpha,\beta}$ with $\alpha + \beta = n$ have the very nice property of being polyconvex. This implies lower semicontinuity on the space ACL^n. In even dimensions, the distortion functions is expressed by $\mathcal{K} = (\mathcal{K}_{\ell,\ell})^{1/\ell}$, when $\ell = n/2$. This definition has all desired features mentioned above, which the inner, outer and linear distortion do not. So, for this reason we denote

$$\mathcal{K}_\ell(A) = \mathcal{K}_{\ell,n-\ell}(A) = \frac{(\lambda_n \lambda_{n-1} \cdots \lambda_{n-\ell+1})^n}{(\det(A))^\ell},$$

for $\ell = 1, 2, \ldots, n-1$, and in even dimensions we set

$$\mathcal{K}(A) = \frac{\lambda_n \cdots \lambda_{\ell+1}}{\lambda_\ell \cdots \lambda_1} \qquad \ell = \frac{n}{2}.$$

All the distortion functions $\mathcal{K}_{\alpha,\beta}$, $1 \leq \alpha,\ \beta \leq n-1$, can be expressed in terms of the distortion functions $\mathcal{K}_\ell$ by the formula

$$\mathcal{K}_{\alpha,\beta} = (\mathcal{K}_\alpha^\beta \mathcal{K}_{n-\beta}^\alpha)^{\frac{1}{n}}.$$

Having analysed these distortion functions for linear mappings, we define for a ACL^n mapping $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ the pointwise distortion functions

$$\mathcal{K}_{\alpha,\beta}(x, \mathbf{f}) = \mathcal{K}_{\alpha,\beta}(D\mathbf{f}(x)) \tag{4.33}$$

and in even dimensions

$$\mathcal{K}(x, \mathbf{f}) = \mathcal{K}(D\mathbf{f}(x)) \tag{4.34}$$

at the points $x \in \Omega$ where the differential exists (see [25] p.111). We will finish this section by recalling the definition of the distortion tensor

$$G(x) = \begin{cases} |J_\mathbf{f}(x)|^{-\frac{2}{n}} D^t\mathbf{f}(x) D\mathbf{f}(x) & if\, J_\mathbf{f}(x) \neq 0 \\ I & if\, J_\mathbf{f}(x) = 0. \end{cases} \tag{4.35}$$

4.4 Quasiconformal Mappings

Quasiconformal mappings are generalizations of conformal mappings. They can considered them not only in Riemann surfaces, but also on Riemannian manifolds in all domains, even on arbitrary metric spaces. Quasiconformal mappings occur naturally in various mathematical and often a priori unrelated contexts. There are three main definitions for quasiconformal mappings in Euclidean space: curve family, geometric and analytic. In the previous section we defined the quasiconformal maps with curve family. For more details, you can see (http://www.ams.org/notices/200611/whatis-heinonen.pdf) by Juha Heinonen or [40]. Now we explain these definitions. At first we recall the curve family definition.

Definition 4.4.1. (Curve family definition of quasiformality) Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Then $\mathbf{f}$ is **K-quasiconformal** if there exists a K, $1 \leq K \leq \infty$, such that

$$\frac{1}{K} M(\Gamma) \leq M(\tilde{\Gamma}) \leq K M(\Gamma) \tag{4.36}$$

for every curve family Γ in Ω and $\tilde{\Gamma} = \mathbf{f}(\Gamma)$ in Ω'.

We can express some theorem such that they lead us to other definitions. There exist complete processes in [25] and [40].

Theorem 4.4.1. *Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism and $\boldsymbol{f}$ is differentiable at the point $x \in \Omega$ and if $K_O(\boldsymbol{f}) < \infty$, then*

$$|D\mathbf{f}(x)|^n \leq K_O(\boldsymbol{f}) |J_{\mathbf{f}}(x)|.$$

There is a long and complete proof in Väiälä's book [40] page 47. In the all following theorems and corollaries we assume that Ω and Ω' are domains in $\mathbb{R}^n$, with $n \geq 2$.

Theorem 4.4.2. *Suppose $\mathbf{f} : \Omega \longrightarrow \Omega'$ is a diffeomorphism, then*

$$K_O(\mathbf{f}) = \sup_{x\in\Omega} H_O[\mathbf{f}'(x)], \qquad K_I(\mathbf{f}) = \sup_{x\in\Omega} H_I[\mathbf{f}'(x)].$$

Corollary 4.4.2.1. *Let $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a diffeomorphism, then $\mathbf{f}$ is a K-quasiconformal if and only if*

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K \, \ell(D\mathbf{f}(x))^n,$$

for all $x \in \Omega$.

Corollary 4.4.2.2. *If $\mathbf{f}$ is a differentiable quasiconformal mapping at point x in Ω, then either $D\mathbf{f}(x) = 0$ or $J_{\mathbf{f}}(x) \neq 0$.*

We don't want to prove these theorems and corollaries, they can be found in [40] in pages 46 to 48. If A is a bijective linear map, we denote $H(x, A) = H(A)$ for all $x \in \mathbb{R}^n$. If $\mathbf{f}$ is differentiable at x and $J_{\mathbf{f}}(x) \neq 0$, then by theorem (3.2.1) we have $H(x, \mathbf{f}) = H(\mathbf{f}'(x))$. We eliminate the very technical proof of the following theorem. It can be found in [40] pages 78 and 79.

Theorem 4.4.3. *Assume* $\mathbf{f} : \Omega \longrightarrow \Omega'$ *is a homeomorphism such that for some* $K < \infty$ *either* $K_O(\mathbf{f}) \leq K$ *or* $K_I(\mathbf{f}) \leq K$. *Then* $H(x, A) = H(A)$ *is bounded by a constant which depends only on* n *and* K.

Corollary 4.4.3.1. *If* $\mathbf{f} : \Omega \longrightarrow \Omega'$ *is a quasiconformal mapping, then* $H(x, \mathbf{f})$ *is bounded.*

Definition 4.4.2. A function $f : [a, b] \longrightarrow \mathbb{R}^m$ is said to be **absolutely continuous** if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\sum_{i=1}^{k} |f(b_i) - f(a_i)| < \epsilon$$

whenever $a = a_1 < b_1 \leq a_2 < b_2 \leq \ldots \leq a_k < b_k = b$ and

$$\sum_{i=1}^{k} |b_i - a_i| < \delta$$

.

It is to see that every absolutely continuous function is uniformly continuous and, therefore continuous. Also, we can prove that every Lipschitz-continuous function is absolutely continuous (see [25] and [33]).

Definition 4.4.3. For $j = 1, 2, \ldots, n$, let $I_j = [a_j, b_j]$ be a closed interval and $Q = I_1 \times I_2 \times \ldots \times I_n$ be a rectangular box lies in domain Ω in $\mathbb{R}^n$. Consider the jth face

$$Q_j = I_1 \times \ldots I_{j-1} \times \{a_j\} \times I_{j+1} \times \ldots I_n.$$

Then we can say a continuous function $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^m$ is $ACL(\Omega, \mathbb{R}^m)$, **absolutely continuous on lines** if for every $j = 1, 2, \ldots, n$ and almost every $a \in Q_j$, with respect to $(n-1)$-measure, the function

$$t \mapsto \mathbf{f}(a + te_j) \qquad a_j \leq t + \langle a, e_j \rangle \leq b_j$$

is absolutely continuous. e_j denotes the jth unit basis vector (see [25] p.106). For $p \geq 1$, an ACL mapping $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^m$ is said to be ACL^p if all partial derivatives of $\mathbf{f}$ are in $L^p_{loc}(\Omega)$. In other words,

$$\frac{\partial \mathbf{f}}{\partial x_i} \in L^p_{loc}(\Omega) \qquad 1 \leq i \leq n.$$

Sometimes, ACL^p is well known as $W^{1,p}$. There is a fundamental result of Gehring and Väiälä that its proof exists in Väiälä's book [40]. This result links geometric and analytic concepts in the theory of quasiconformal mappings.

Theorem 4.4.4. *(**Gehring-Väiälä**) Every quasiconformal mapping* $\mathbf{f} : \Omega \longrightarrow \mathbb{R}^n$ *lies in the space* $ACL(\Omega, \mathbb{R}^n)$.

There are some others good results in Väiälä's book [40] that we explain them.

Theorem 4.4.5. *Suppose that* Ω *and* Ω' *are domains in* $\mathbb{R}^n$*,with* $n \geq 2$*, and* $\mathbf{f} : \Omega \longrightarrow \Omega'$ *is a homeomorphism such that* $H(\mathbf{f})$ *is bounded. Then* $\mathbf{f}$ *is differentiable almost everywhere.*

Corollary 4.4.5.1. *Every quasiconformal mapping is differentiable almost everywhere.*

Theorem 4.4.6. *If* Ω *and* Ω' *are domains in* $\mathbb{R}^n$*, with* $n \geq 2$*, and* $\mathbf{f} : \Omega \longrightarrow \Omega'$ *is a homeomorphism and if* $1 \leq K \leq \infty$*, then the following are equivalent:*

1. *$K_O(\mathbf{f}) \leq K$*
2. *$\mathbf{f}$ is ACL, almost everywhere differentiable, and $|D\mathbf{f}(x)|^n \leq K|J_\mathbf{f}(x)|$ almost everywhere.*

Moreover, each of these conditions implies that $\mathbf{f} \in ACL^n$.

Corollary 4.4.6.1. *Every quasiconformal map is ACL^n (see [40] p.110 and 111).*

Theorem 4.4.7. *If Ω and Ω' are domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ is a homeomorphism, such that $H(x, \mathbf{f})$ is bounded by a constant C, then*

$$|D\mathbf{f}(x)|^n \leq C^{n-1}|J_\mathbf{f}(x)|.$$

Proof. By theorem (4.4.4) $\mathbf{f}$ is ACL, and by theorem (4.4.5) $\mathbf{f}$ is differentiable almost everywhere. Also, by corollary (4.4.2.2) and theorem (3.2.1) we can write that for $x \in \Omega$ either $D\mathbf{f}(x) = 0$ or

$$0 < H(D\mathbf{f}(x)) = H(x, \mathbf{f}) = C.$$

If $D\mathbf{f}(x) = 0$, then the result is obvious, therefore we choose $0 < H(D\mathbf{f}(x)) = H(x, \mathbf{f}) = C$. Accordingly,

$$\frac{|D\mathbf{f}(x)|^n}{|J_\mathbf{f}(x)|} = H_O(D\mathbf{f}(x)) \leq \max\{H_I(D\mathbf{f}), H_O(D\mathbf{f})\} \leq H(D\mathbf{f})^{n-1} = C^{n-1},$$

or

$$|D\mathbf{f}(x)|^n \leq C^{n-1}|J_\mathbf{f}(x)|.$$

♠

Theorem 4.4.8. *If Ω and Ω' are domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ is a homeomorphism, then $\mathbf{f}$ is a quasiconformal mapping if and only if $H(x, \mathbf{f})$ is bounded.*

Proof. Assume $\mathbf{f}$ is a quasiconformal map. Then by corollary (4.4.3.1) $H(x, \mathbf{f})$ is bounded. For another side, we suppose for all $x \in \Omega$, $H(x, \mathbf{f}) = C < \infty$. We know that by previous theorem $\mathbf{f}$ is ACL, and by theorem (4.4.5) $\mathbf{f}$ is differentiable almost everywhere. Therefore, for $x \in \Omega$, by corollary (4.4.2.2) and theorem (3.2.1), we can write $D\mathbf{f}(x) = 0$ or $0 < H(D\mathbf{f}(x)) = H(x, \mathbf{f}) = C$. By theorem (4.4.7), in both cases,

$$|D\mathbf{f}(x)|^n \leq C^{n-1}|J_\mathbf{f}(x)|.$$

By considering $C^{n-1} = K$ we have $K_O(\mathbf{f}) < K = C^{n-1}$. Since $K_O(\mathbf{f}) = K_I(\mathbf{f^{-1}})$, by theorem (4.4.3) we have $H(y, \mathbf{f^{-1}}) < \infty$ for all $y \in \Omega'$. With the same process, we can write $H(x, \mathbf{f}) < \infty$ for all $x \in \Omega$. So, $K_I(\mathbf{f}) = K_O(\mathbf{f^{-1}}) < \infty$, and hence $\mathbf{f}$ is quasiconformal (see [40] p.114). ♠

Theorem 4.4.9. *A homeomorphism $\mathbf{f} : \Omega \longrightarrow \Omega'$ is quasiconformal if and only if one of the distortions $K_O(\mathbf{f})$ or $K_I(\mathbf{f})$ is finite.*

Proof. Let $\mathbf{f}$ be quasiconformal, then by theorem (4.4.8) $H(x, \mathbf{f})$ is bounded and at least one of the inner or outer distortion is finite. Conversely, assume $K_O(\mathbf{f})$ or $K_I(\mathbf{f})$ is finite. By theorem (4.4.3) we can say, $H(x, \mathbf{f})$ is bounded and also by theorem (4.4.8) $\mathbf{f}$ is a quasiconformal map (see [40] p.114). ♠

Now we can explain another definition for quasiconformal mapping such that it has geometric perspective and intuition. For more details, you can see [25] or [18].

Definition 4.4.4. (Geometric definition of quasiconformality) Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Then $\mathbf{f}$ is called **H-quasiconformal**, if there is a constant $1 \leq H \leq \infty$ such that $H(\mathbf{f}) \leq H$.

This geometric definition is simple, direct, very general and quite appealing from a geometric-aesthetic vantage point. Further, it requires no regularity of differentiability properties of the mapping to formulate. Unfortunately, this geometric definition, while aesthetically pleasing, is difficult to work with. Nowadays, the following analytic definition of quasiconformality is more common (see [25] and [18]). But before the analytic definition we say some definitions that we need them to work with. At first, we represent a important theorem.

Theorem 4.4.10. *Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Then $\mathbf{f}$ is K-quasiconformal if and only if the following conditions are satisfied:*

1. *$\mathbf{f}$ is ACL,*
2. *$\mathbf{f}$ is differentiable almost everywhere,*
3. *For almost every $x \in \Omega$ we have*

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K\ \ell(D\mathbf{f}(x))^n.$$

Proof. Let $\mathbf{f}$ be quasiconformal, then by theorem (4.4.4) $\mathbf{f}$ is ACL, and by corollary (4.4.5.1) $\mathbf{f}$ is differentiable almost everywhere. So, properties (1) and (2) are true. Since $\mathbf{f}$ is K-quasiconformal, we can write

$$\frac{|D\mathbf{f}(x)|^n}{|J_{\mathbf{f}}(x)|} \leq K$$

for all $x \in \Omega$ that $\mathbf{f}$ is differentiable at x. We may also choose $|J_{\mathbf{f}}(x)| \neq 0$. We know that the $\mathbf{f^{-1}}$ is also K-quasiconformal and differentiable at $y = \mathbf{f}(x)$. Hence,

$$|J_{\mathbf{f}}(x)| = |J_{\mathbf{f^{-1}}}(y)| \leq K|D\mathbf{f^{-1}}(y)| = K\ \ell(D\mathbf{f}(x))^n.$$

The above relation shows us the condition (3) is also true. Conversely, assume that the three conditions hold. That means $\mathbf{f}$ is ACL, differentiable almost everywhere, and for almost $x \in \Omega$ the following inequality holds

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K\ \ell(D\mathbf{f}(x))^n.$$

Note that above inequality gives us $K_O(\mathbf{f}) \leq K$ and this relation by theorem (4.4.6) implies $K_O(\mathbf{f})$ is finite. Therefore, by theorem (4.4.9) $\mathbf{f}$ is a quasiconformal map. Since $|J_{\mathbf{f}}(x)| \neq 0$ almost everywhere, we have $K_I(\mathbf{f}) \leq K$. ♠

This theorem leads us to define a quasiconformal mapping in third way. We had curve family and geometric definition before, but next one is called analytic definition of quasiconformal mapping.

Definition 4.4.5. (Analytic definition of quasiconformality) Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. Then $\mathbf{f}$ is called **K-quasiconformal** if the first partial derivatives of $\mathbf{f}$ exist as Lebesgue functions, the Jacobian determinant of $\mathbf{f}$ is locally integrable and does not change sign in Ω and there is a measurable function $K_O = K_O(x) \geq 1$, finite almost everywhere, such that $\mathbf{f}$ satisfies the distortion inequality

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K\,\ell(D\mathbf{f}(x))^n$$

for almost everywhere in Ω (see [25] p.106). In simple way, we can say $\mathbf{f}$ is a quasiconformal if the following conditions are satisfied:

1. $\mathbf{f}$ is ACL,
2. $\mathbf{f}$ is differentiable almost everywhere,
3. For almost every $x \in \Omega$ we have

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K\,\ell(D\mathbf{f}(x))^n.$$

We finish this section by expressing a theorem that shows the equivalence of these three definitions.

Theorem 4.4.11. *Let Ω and Ω' be domains in $\mathbb{R}^n$, with $n \geq 2$, and $\mathbf{f} : \Omega \longrightarrow \Omega'$ be a homeomorphism. For all Γ in Ω the following are equivalent:*

1. $\frac{1}{K}M(\Gamma) \leq M(\tilde{\Gamma}) \leq KM(\Gamma)$.
2. $H(x, \mathbf{f})$ *is bounded.*
3. $\mathbf{f}$ *is ACL, $\mathbf{f}$ is differentiable almost everywhere and for almost every $x \in \Omega$ we have*

$$\frac{|D\mathbf{f}(x)|^n}{K} \leq |J_{\mathbf{f}}(x)| \leq K\,\ell(D\mathbf{f}(x))^n.$$

Proof. To get from (1) to (2) we can use the theorem (4.4.3) and corollary (4.4.3.1). Also we can prove by theorem (4.4.10), the condition (1) implies (3). By theorems (4.4.8), (4.4.9) and (4.4.10) we can show that (2) and (3) are equivalent. But the proof that (2) or (3) implies (1) is technical and we omit that. For seeing the complete proof you can see [40] pages 121 to 123. ♠

We should note that in 2-dimensions $H = K$. This means if $n = 2$ then the constant H and the constant K are equal. But if $n \geq 3$, H and K are different constants in geometric definition and analytic or curve family definitions. Because if $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, then in analytic definition we have

$$|D\mathbf{f}(x)|^n \leq K|J_{\mathbf{f}}(x)|.$$

Therefore,

$$(\lambda_n)^n \leq K\lambda_1\lambda_2\cdots\lambda_n.$$

On the other hand, we know that

$$\frac{\lambda_n}{\lambda_1} \leq H.$$

So, it is obvious to see that if $n = 2$, the constants H and K are equal.

4.5 New Question

In this section we will show that what is the main question in this thesis. In previous two sections we showed that there are three definitions of quasiconformality such that are equivalent. They were curve family, analytic and geometric definitions (see [20]). We recall some definitions and properties of convergence sequences and after that we start the convergence of quasiconformal mappings. In mathematical analysis, uniform convergence, locally uniform convergence and pointwise convergence have different meanings.
Let Ω and Ω' be domains in $\mathbb{R}^n$, and $\mathbf{f_n} : \Omega \longrightarrow \Omega'$ be a sequence of functions. We say that the sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is **uniformly convergent** with limit $\mathbf{f} : \Omega \longrightarrow \Omega'$ if and only if for every $\epsilon > 0$, there exists an integer N such that

$$\|\mathbf{f_n}(x) - \mathbf{f}(x)\| < \epsilon,$$

for all $n \geq N$ and all $x \in \Omega$. The sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is said to be **locally uniformly convergent** with limit $\mathbf{f}$, if Ω is a metric space and for every $x \in \Omega$, there exists an $r > 0$ such that the sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is uniformly convergent on $B(x, r) \cap \Omega$. The sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is **pointwise convergent** to $\mathbf{f}$ if and only if for every $\epsilon > 0$ and every $x \in \Omega$, there exists an integer N, depending on ϵ and on x, such that

$$\|\mathbf{f_n}(x) - \mathbf{f}(x)\| < \epsilon,$$

for all $n > N$. It is clear that uniformly convergence implies locally uniformly convergence and easy to see that local uniformly convergence results pointwise convergence (see [32]). Let (X, τ) is a topological space and (Y, d_Y) is a metric space. A sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of functions $\mathbf{f} : X \longrightarrow Y$ is said to converge of $\mathbf{f} : X \longrightarrow Y$ uniformly on compact subsets of X if, for every compact set $K \subseteq X$ and every $\epsilon > 0$, there is the positive integer $N_{K,\epsilon}$ such that

$$\|\mathbf{f_n}(x) - \mathbf{f}(x)\| < \epsilon,$$

for all $x \in X$ and $n \leq N$ (see [33] and [28]). Now we want to investigate the convergence of quasiconformal mappings. For this purpose, we explain some theorems such that their proofs are omitted.
Definition 4.5.1. Let $\{D_n\}$ be a sequence of sets in $\hat{\mathbb{R}}^n$ with $n \leq 2$. The kernel of $\{D_n\}$, denoted by $ker_{n\to\infty}D_n$ is the set defined as follows:

$$ker_{n\to\infty}D_n = \bigcup_{n=1}^{\infty} int\Big(\bigcap_{i=n}^{\infty} D_i\Big).$$

Hence, a point x belongs to $ker_{n\to\infty}D_n$ if and only if x has a neighbourhood U that is included in D_n for all sufficiently large n. This implies that each compact subset of $ker_{n\to\infty}D_n$ is contained in D_n once n suitably large. It is obvious that $ker_{n\to\infty}D_n$ is an open set and this set need not be connected, however, even if each of the sets D_n is a domain ([18] p.284).

Theorem 4.5.1. *Let $\{D_n\}$ be a sequence of domains in $\hat{\mathbb{R}}^n$ with $n \leq 2$, let $\mathbf{f_n}$ be a K-quasiconformal mapping of D_n onto a domain D'_n, and let D be a subdomain of $ker_{n\to\infty}D_n$. Assume that $\mathbf{f_n} \longrightarrow \mathbf{f}$ pointwise in D. There are the following three possibilities for the limit mapping $\mathbf{f}$:*

1. $\mathbf{f}$ *might take exactly two values in* D *(one of these at precisely one point of* D*) in which event the convergence of* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *is not locally uniform in* D*;*
2. $\mathbf{f}$ *might be a homeomorphism of* D *onto a subdomain of* $ker_{n\to\infty}D'_n$*, in which event the convergence of* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *is locally uniform in* D*;*
3. $\mathbf{f}$ *might be constant in* D*, in which event the convergence of* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *may or may not be locally uniform in* D*.*

Proof. There is long and complete proof in [18] in pages 284 to 286. Also you can see [40] pages 69 and 70 or [17]. ♠

Theorem 4.5.2. *Let* $\{D_n\}$ *be a sequence of domains in* $\hat{\mathbb{R}}^n$ *with* $n \leq 2$*, let* $\mathbf{f_n}$ *be a K-quasiconformal mapping of* D_n *into* $\hat{\mathbb{R}}^n$*, and let* D *be a subdomain of* $ker_{n\to\infty}D_n$*. Suppose that* $\mathbf{f_n} \longrightarrow \mathbf{f}$ *locally uniformly in* D*, where* $\mathbf{f}$ *is a homeomorphism. Then*

$$K_I(\mathbf{f}) \leq \liminf_{n\to\infty} K_I(\mathbf{f_n}), \qquad K_O(\mathbf{f}) \leq \liminf_{n\to\infty} K_O(\mathbf{f_n}).$$

In fact, if each of the mapping $\mathbf{f_n}$ *is K-quasiconformal mappings, then* $\mathbf{f}$ *is K-quasiconformal mappings as well.*

Corollary 4.5.2.1. *Let* D *be a domain in* $\mathbb{R}^n$*. The functions* $K_I(\mathbf{f})$*,* $K_O(\mathbf{f})$*, and* $K(\mathbf{f})$ *defined on the space of quasiconformal mappings of* D *into* $\mathbb{R}^n$ *with topology of local uniform convergence are lower semicontinuous (see [18] p.287).*

Theorem 4.5.3. *Assume that* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *is a sequence of K-quasiconformal mappings of* $\hat{\mathbb{R}}^n$ *onto itself and that* $\mathbf{f_n} \longrightarrow \mathbf{f}$ *uniformly on* $\hat{\mathbb{R}}^n$*. Then* $\mathbf{f}$ *is a K-quasiconformal mapping of* $\hat{\mathbb{R}}^n$ *onto itself. Moreover,* $\mathbf{f_n^{-1}} \longrightarrow \mathbf{f^{-1}}$ *uniformly on* $\hat{\mathbb{R}}^n$ *as well.*

Theorem 4.5.4. *If* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *is a sequence of homeomorphism of* $\hat{\mathbb{R}}^n$ *and if* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *converges uniformly on* $\hat{\mathbb{R}}^n$ *to a homeomorphism* $\mathbf{f}$*, then* $\mathbf{f_n^{-1}} \longrightarrow \mathbf{f^{-1}}$ *uniformly on* $\hat{\mathbb{R}}^n$*.*

There exist the technical proofs in [18] from page 283 to 293. More details can be founded in [1], [40] and [19]. These theorems show us if we have a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of K-quasiconformal mappings (with curve family definition) on $\hat{\mathbb{R}}^n$ and if $\{\mathbf{f_n}\}_{n=1}^{\infty}$ converges uniformly on $\hat{\mathbb{R}}^n$ to a $\mathbf{f}$, Then $\mathbf{f}$ is a K-quasiconformal mapping of $\hat{\mathbb{R}}^n$ onto itself and the distortion of $\mathbf{f}$ is equal or less than distortion of all functions $\mathbf{f_n}$. This fact is also true for for analytic definition for distortion. The following theorem expresses this subject.

Theorem 4.5.5. *Suppose that the sequence* $\{\mathbf{f_n}\}_{n=1}^{\infty}$ *of mappings of finite distortion converges in* ACL^n *to a mapping* $\mathbf{f}$*. Suppose further that for some* $\ell \in \{1, 2, \ldots, n-1\}$*,*

$$\mathcal{K}_\ell(x, \mathbf{f_i}) \leq M_i(x) \qquad \textit{almost everywhere in } \Omega,\ i = 1, 2, \ldots \tag{4.37}$$

where $1 \leq M_i(x) < \infty$ *converge in a biting sense to a measurable function* $1 \leq M(x) < \infty$*. Then* $\mathbf{f}$ *has finite distortion and we have*

$$\mathcal{K}_\ell(x, \mathbf{f}) \leq M(x) \qquad \textit{almost everywhere in } \Omega. \tag{4.38}$$

The proof of the above theorem exists in [25] in pages 193 and 194. The theorem (4.5.5) can be explain in a slightly weaker form, dealing with the case $M_i(x) = M(x)$ for all i in positive integer numbers. For inner, outer and maximal distortion are proved in [19].

Theorem 4.5.6. *Suppose that the sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of mappings with finite distortion is bounded in ACL^n and satisfies*

$$\mathcal{K}_\ell(x, \mathbf{f_i}) \le M(x) < \infty \qquad \textit{almost everywhere in } \Omega,\ i = 1, 2, \ldots \tag{4.39}$$

Then $\{\mathbf{f_n}\}_{n=1}^{\infty}$ contains a subsequence converging in ACL^n locally uniformly on Ω to a mapping $\mathbf{f} \in ACL^n$ of finite distortion. The limit map $\mathbf{f}$ also satisfies the distortion inequality

$$\mathcal{K}_\ell(x, \mathbf{f}) \le M(x) < \infty \qquad \textit{almost everywhere in } \Omega. \tag{4.40}$$

This proof can be found in [25] page 195. These theorems show us that with curve family and analytic definitions of quasiconformal mappings, if $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is a sequence of K-quasiconformal mappings of $\hat{\mathbb{R}}^n$ onto itself and that $\mathbf{f_n} \longrightarrow \mathbf{f}$ uniformly on $\hat{\mathbb{R}}^n$, then $\mathbf{f}$ is a K-quasiconformal mapping of $\hat{\mathbf{R}}^{\mathbf{n}}$ onto itself. This theorem says the sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ and function $\mathbf{f}$ have the same constants and sometimes the constant of $\mathbf{f}$ is less than the constant of $\{\mathbf{f_n}\}_{n=1}^{\infty}$ (for more details and proofs you can see [25], [18], [40] and [24]). The theorems (4.5.5) and (4.5.6) show the lower semicontinuity holds for curve family and analytic definitions of quasiconformality. But for geometric definition of quasiconformality, we have not proved the linear distortion function to be lower semicontinuous. If $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is a sequence of H-quasiconformal mappings (geometric definition) of $\mathbb{R}^n$ onto itself and that $\mathbf{f_n} \longrightarrow \mathbf{f}$ uniformly on $\hat{\mathbb{R}}^n$, then $\mathbf{f}$ is a K-quasiconformal mapping of $\mathbb{R}^n$ but $\mathbf{f}$ is not a H-quasiconformal mapping. We can ask the following question regarding to lower semicontinuity of the linear distortion: In the above conditions for geometric definition, is

$$H(x, \mathbf{f}) \le \limsup_{n\to\infty} H(x, \mathbf{f_n})?$$

This question has been answered negatively by Tadeusz Iwaniec (see [15]). In article [23] He proved that $H(x, \mathbf{f}) > H(x, \mathbf{f_n})$ almost everywhere in $\mathbb{R}^n$. The key idea is that the linear distortion function fails to be rank-one convex in dimension $n \ge 3$. At first, we recall some definitions and after that explain the lemma such that it is useful for next theorem.
Recall that $M_{n\times n}(\mathbb{R})$ denotes the real vector space of all $n \times n$ matrices endowed with the norm

$$|A| = \max_{|h|=1} |Ah|.$$

By theorem (1.6.1), a matrix B is rank one if and only if it can be written as the tensor product of two vectors. Therefore, there are two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$ such that

$$B = \mathbf{u} \otimes \mathbf{v} = [b_{ij}] = [u_i v_j], \qquad i, j = 1, 2, \ldots, n.$$

By definition of linear distortion for matrix $A \in M_{n\times n}(\mathbb{R})$ (equation 4.26) we have

$$H(A) = \frac{\max_{|h|=1} |Ah|}{\min_{|h|=1} |Ah|}.$$

The linear distortion function $H(A)$ is not rank-one convex. This content is described in the next lemma.

Lemma 4.5.7. ***(Iwaniec's Lemma)*** *Given $n \ge 3$ and $H > 1$, there is a matrix A and a rank-one matrix B and numbers $t, s > 0$ such that*

$$H(A - sB) = H(A + tB) = H < H(A).$$

There is a long and technical proof with interesting computation in [23] in dimension 3. This proof can be extended to higher dimensions. Now we can explain the theorem that the linear distortion function fails to be rank-one convex in dimension higher than 2 (see [23]).

Theorem 4.5.8. ***(Iwaniec's Theorem)*** *For each $n \geq 3$ and $H > 1$, there exists a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of quasiconformal mappings $\mathbf{f_n} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ converging uniformly to a linear quasiconformal map $\mathbf{f_0} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that*

$$H(x, \mathbf{f_n}) \equiv H < H(x, \mathbf{f_0}), \qquad \textit{almost everywhere in } \mathbb{R}^n \quad n = 1, 2, \ldots$$

Proof. This follows with the aid of the lemma. We have $B = \mathbf{u} \otimes \mathbf{v}$ for vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$. Consider $\mathbf{f_0}(x) = Ax$, and for $n = 1, 2, \ldots$, we define a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ by equation

$$\mathbf{f_n}(x) = Ax + \frac{1}{n} h(n\mathbf{u} \cdot x)\mathbf{v}$$

where h is a periodic piecewise linear function on the real line as follows: Given s and t of the lemma

$$h(r) = \begin{cases} -rs & -s^{-1} \leq r \leq 0, \\ rt & 0 \leq r \leq t^{-1}. \end{cases}$$

Then we can extend h to the entire line (saw-tooth function). h is a bounded Lipschitz function whose derivative assumes only the two values $-s$ and t. The sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ converges uniformly to $\mathbf{f_0}$. The differential of $\mathbf{f_n}$ also assumes only two values, which are independent of n.

$$D\mathbf{f_n} = A + h'(n\mathbf{u} \cdot x)\mathbf{u} \otimes \mathbf{v} \in \{A - sB, A + tB\}.$$

In either case, the linear distortion of $D\mathbf{f_n}(x)$ is equal to H:

$$H(x, \mathbf{f_n}) = H(D\mathbf{f_n}(x)) \equiv H < H(A) \equiv H(x, \mathbf{f_0}).$$

This proves the theorem. ♠

Now, it is time to express that what is the main question in this thesis. We know that by theorems (1.4.7), (1.4.8) and (1.4.9), if matrix A is symmetric then its eigenvalues are real numbers and by spectral theorem the matrix A is orthogonally diagonalizable and finally A is positive definite if and only if all eigenvalues of A are positive real numbers. If $A = D\mathbf{f}(x)$ then all above results hold. So, we can consider A is diagonal matrix with positive entries.

The Main Question of This Thesis: Assume that

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

is diagonal and $B = \mathbf{u} \otimes \mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^3$ and $1 < a < b$. We want to show that for every matrix A (or for every a and b) there are vectors $\mathbf{u}$ and $\mathbf{v}$ and number t, such that

$$H(A + tB) = H(A + t\mathbf{u} \otimes \mathbf{v}) < H(A).$$

In next section we try to answer this question for different kinds of $\mathbf{u}$ and $\mathbf{v}$.

4.6 Solutions and Examples

In this section we want to answer the main question. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

be a diagonal matrix and B is a rank-one matrix such that $B = \mathbf{u} \otimes \mathbf{v} = \mathbf{u} \cdot \mathbf{v}^t$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^3$. It is easy to see, if $t = 0$, then $H(A + tB) = H(A)$. Let $t \neq 0$, then the matrix A has three eigenvalues $L_1 = 1$, $L_2 = a$ and $L_3 = b$ where $1 < a < b$. By definition $H(A)$ we have

$$H(A) = \frac{L_{\max}}{L_{\min}} = \frac{b}{1} = b.$$

But $H(A+tB)$ is a function with respect to t, x and y where they are real numbers and x and y are components of vectors $\mathbf{u}$ and $\mathbf{v}$. This function belongs to C^∞ with respect to t. Since A and B are symmetric then $A + tB$ is symmetric. So, all eigenvalues of $A + tB$ are positive real numbers. Assume that $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$ are eigenvalues of $A + tB$. The functions $H(A+tB)$, $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$ are continuous and differentiable infinite times with respect to t, because all of them are polynomial or rational functions. The equations of eigenvalues are so long and working with them is very difficult. Because of that, we should use the Taylor series for several variables functions to find the answers of limits and derivatives. Suppose that

$$\lim_{t\to 0} \lambda_1(t) = \lambda_1(0) = b \qquad \lim_{t\to 0} \lambda_2(t) = \lambda_2(0) = 1 \qquad \lim_{t\to 0} \lambda_3(t) = \lambda_3(0) = a$$

These limit show us for sufficiently small t in the neighbourhood of 0, we have

$$\lambda_2(t) < \lambda_3(t) < \lambda_1(t).$$

So, in this case

$$H(A + tB) = \frac{\lambda_1(t)}{\lambda_2(t)}.$$

If t tends to infinity then it implies

$$\lim_{t\to+\infty} H(A + tB) = \lim_{t\to+\infty} H(\frac{A}{t} + B) = H(B) = \infty,$$

because B is a rank-one matrix. As we know, the function $H(A + tB)$ belongs to C^∞ then H is derivative with respect to t infinite time. The derivative of H with respect to t is

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB) = \frac{\lambda_1'(t)\lambda_2(t) - \lambda_2'(t)\lambda_1(t)}{(\lambda_2(t))^2}. \tag{4.41}$$

To find the value of derivative at zero we can write

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB)\Big|_{t=0} = \frac{\lambda_1'(0)\lambda_2(0) - \lambda_2'(0)\lambda_1(0)}{(\lambda_2(0))^2}. \tag{4.42}$$

We know that $\lambda_1(0) = b$ and $\lambda_2(0) = 1$, so,

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB)\Big|_{t=0} = \lambda_1'(0) - b\lambda_2'(0) \tag{4.43}$$

The sign of equation (4.43) shows us the monotonicity of function $H(A+tB)$ in the neighbourhood of 0. But for convexity and concavity of function $H(A+tB)$ we need the second derivative of the function. For this purpose, we take the derivative from equation (4.41) with respect to t. Therefore,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB) &= \frac{(\lambda_1'(t)\lambda_2(t)-\lambda_2'(t)\lambda_1(t))'\lambda_2^2(t)-(\lambda_2^2(t))'(\lambda_1'(t)\lambda_2(t)-\lambda_2'(t)\lambda_1(t))}{(\lambda_2(t))^4}\\ &= \frac{\lambda_1''(t)\lambda_2^2(t)-\lambda_2''(t)\lambda_1(t)\lambda_2(t)-2\lambda_1'(t)\lambda_2'(t)\lambda_2(t)+2\lambda_1(t)(\lambda_2'(t))^2}{(\lambda_2(t))^3}.\end{aligned} \tag{4.44}$$

So,

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} = \frac{\lambda_1''(0)\lambda_2^2(0)-\lambda_2''(0)\lambda_1(0)\lambda_2(0)-2\lambda_1'(0)\lambda_2'(0)\lambda_2(0)+2\lambda_1(0)(\lambda_2'(0))^2}{(\lambda_2(0))^3}.$$

But the equations $\lambda_1(0)=b$ and $\lambda_2(0)=1$ imply that

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} = \lambda_1''(0)-b\lambda_2''(0)+2\lambda_2'(0)(b\lambda_2'(0)-\lambda_1'(0)). \tag{4.45}$$

By equations (4.43) and (4.45) and their signs the functions $H(A)$ and $H(A+tB)$ have the six following statuses in the neighbourhood of 0 in next page. In figures (4.3) and (4.8) for positive small t near to 0 $H(A+tB)$ is less than $H(A)$ and for negative t near to 0 is bigger than $H(A)$. In figures (4.4) and (4.7) the function $H(A+tB)$ for positive small t near to 0 is bigger and for negative t near to 0 is less than $H(A)$.
Also, in the figure (4.5) we have

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0}=0, \qquad \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0}>0.$$

These equations show us the function $H(A+tB)$ on a neighbourhood of 0 is bigger than $H(A)$ (see figure (4.5)).
But in figure (4.6), the first and second derivative of $H(A+tB)$ are

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0}=0, \qquad \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0}<0.$$

These equations show that there is a neighbourhood of 0 such that for every $t\neq 0$ on this neighbourhood we have

$$H(A+tB)<H(A).$$

It is very obvious that in order to answer our main question, the state of figure (4.6) is a desirable answer. In this case for every sufficiently small t near to 0 there are the vectors $\mathbf{u}$ and $\mathbf{v}$ such that

$$H(A+t\mathbf{u}\otimes\mathbf{v})<H(A).$$

In the rest of this section we try to find some vectors $\mathbf{u}$ and $\mathbf{v}$ and relations between their components and the entries of matrix A such that

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0}=0, \qquad \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0}<0.$$

In the figure (4.6), the function $H(A+tB)$ is concave on the neighbourhood of 0 and this property has a main rule in our answer.

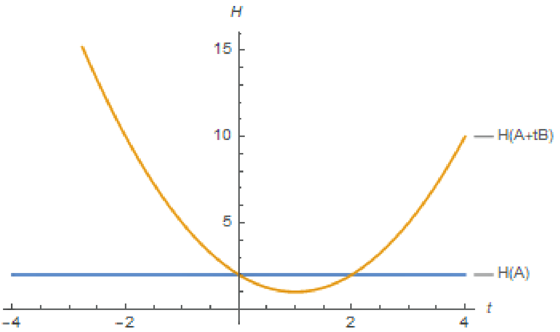

Figure 4.3: Linear distortion functions 1
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} < 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} > 0.$

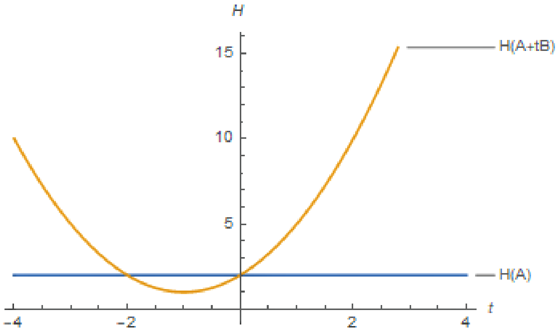

Figure 4.4: Linear distortion functions 2
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} > 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} > 0.$

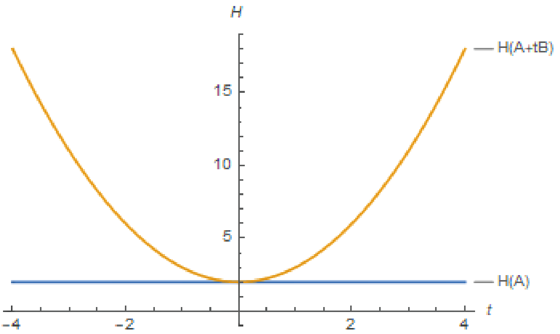

Figure 4.5: Linear distortion functions 3
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} = 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} > 0.$

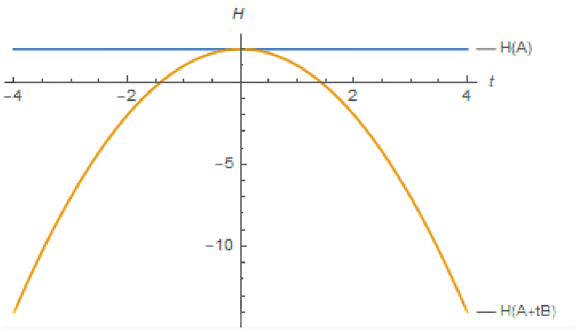

Figure 4.6: Linear distortion functions 4
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} = 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} < 0.$

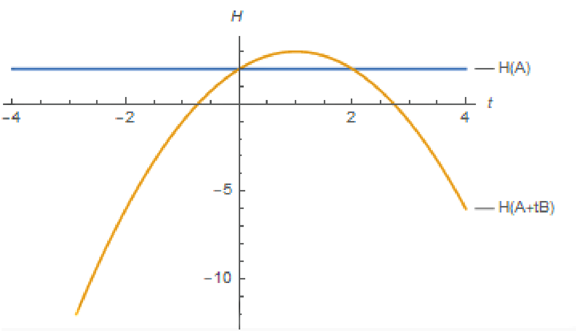

Figure 4.7: Linear distortion functions 5
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} > 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} < 0.$

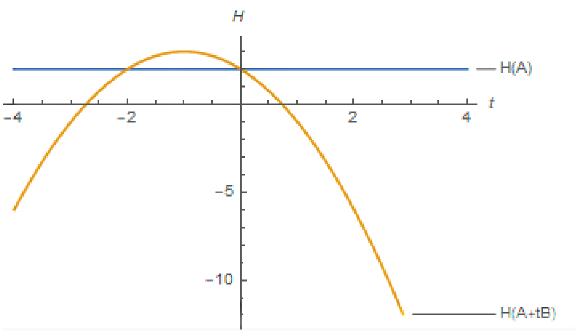

Figure 4.8: Linear distortion functions 6
$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} < 0,\ \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} < 0.$

General Case 1: Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

be a diagonal matrix and $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, y, x)$ are two vectors in $\mathbb{R}^3$ such that $B = \mathbf{u} \otimes \mathbf{v}$ is a rank-one matrix. The matrix B equals to

$$B = \begin{bmatrix} 1 & y & x \\ x & xy & x^2 \\ y & y^2 & xy \end{bmatrix}.$$

So,

$$A + tB = \begin{bmatrix} 1+t & ty & tx \\ tx & a + txy & tx^2 \\ ty & ty^2 & b + txy \end{bmatrix}.$$

We know that $H(A) = b$. Assume that the eigenvalues of matrix $A + tB$ are $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$. These eigenvalues are continuous and differentiable respect to t infinite times. By using the limit and Taylor series for multi-variables functions we can find

$$\lim_{t\to 0} \lambda_1(t) = \lambda_1(0) = b \qquad \lim_{t\to 0} \lambda_2(t) = \lambda_2(0) = 1 \qquad \lim_{t\to 0} \lambda_3(t) = \lambda_3(0) = a$$

These limits show us for sufficiently small t in the neighbourhood of 0, we have

$$\lambda_2(t) < \lambda_3(t) < \lambda_1(t).$$

So, in this case

$$H(A + tB) = \frac{\lambda_1(t)}{\lambda_2(t)}.$$

We note that these computations and formulas are very long and we do not write them down here in detail. We write down the results and only the final answers but the printed calculations that have been done in mathematica version 11.0 will be attached to this thesis. By these long calculation we have

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB)\Big|_{t=0} = xy - b, \tag{4.46}$$

and,

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} = 2x^2y^2 \left(\frac{(a-1)^5}{(b-1)^6} + \frac{(a-1)^4}{(b-1)^5} + \frac{(a-1)^3}{(b-1)^4} + \frac{(a-1)^2}{(b-1)^3} + \frac{a-1}{(b-1)^2} \right) + \frac{2(b-1)\mathrm{xy} + 2\mathrm{xy}}{a-1} + \frac{2x^2y^2 + 4\mathrm{xy}}{b-1} + 2b. \tag{4.47}$$

If x and y have same signs then we have three cases for first derivative and second derivative of $H(A+tB)$ is always positive. In this case, since $1 < a < b$, the second derivative is positive.

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} > 0.$$

We show all cases for function $H(A + tB)$ when x and y have same or different signs in the following table.

Table 4.1: Conditions for linear distortion on a neighbourhood of 0 in case one.

xy	Condition	H'	H''	monotonicity	Results
$xy > 0$	$xy > b$	positive	positive	increasing	For $t > 0$: $H(A+tB) > H(A)$
$xy > 0$	$xy > b$	positive	positive	increasing	For $t < 0$: $H(A+tB) < H(A)$
$xy > 0$	$xy = b$	zero	positive	minimum at 0	For all t : $H(A+tB) > H(A)$
$xy > 0$	$xy < b$	negative	positive	decreasing	For $t > 0$: $H(A+tB) < H(A)$
$xy > 0$	$xy < b$	negative	positive	decreasing	For $t < 0$: $H(A+tB) > H(A)$
$xy < 0$	$xy - b$	negative	unknown	decreasing	For $t > 0$: $H(A+tB) < H(A)$
$xy < 0$	$xy - b$	negative	unknown	decreasing	For $t < 0$: $H(A+tB) > H(A)$

If x and y have different signs then the first derivative of $H(A + tB)$ is negative and the function decreases on a neighbourhood of 0. In this situation, the sign second derivative is unknown. It may be positive or negative. The table (4.1) shows that for rank-one matrix $B = \mathbf{u} \otimes \mathbf{v}$ where $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, y, x)$ are two vectors in $\mathbb{R}^3$, there is not the neighbourhood of 0 such that for every $t \neq 0$ in this interval we have

$$H(A + tB) < H(A).$$

Example 4.6.1. *Let*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

be a diagonal matrix. In matrix A we have $a = 2$ and $b = 4$. We consider two vectors $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, y, x)$ such that $xy > b = 4$. For instance, assume that $\mathbf{u} = (1, -2, -5)$ and $\mathbf{v} = (1, -5, -2)$. The matrix $B = \mathbf{u} \otimes \mathbf{v}$. Therefore,

$$B = \begin{bmatrix} 1 & -5 & -2 \\ -2 & 10 & 4 \\ -5 & 25 & 10 \end{bmatrix}$$

. It is easy to see $H(A) = 4$ and

$$A + tB = \begin{bmatrix} 1+t & -5t & -2t \\ -2t & 2+10t & 4t \\ -5t & 25t & 4+10t \end{bmatrix}$$

The matrix $A + tB$ has three eigenvalues $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$ where

$$\lim_{t\to 0} \lambda_1(t) = \lambda_1(0) = b \qquad \lim_{t\to 0} \lambda_2(t) = \lambda_2(0) = 1 \qquad \lim_{t\to 0} \lambda_3(t) = \lambda_3(0) = a$$

Thus, for sufficiently small t on a neighbourhood the following inequality holds

$$\lambda_2(t) < \lambda_3(t) < \lambda_1(t).$$

So, in this example

$$H(A+tB)=\frac{\lambda_1(t)}{\lambda_2(t)}.$$

By rows one and two in table (4.1), for $t>0$ we have $H(A+tB)>H(A)$ and for $t<0$ we have $H(A+tB)<H(A)$. If $H1(t)=H(A+tB)$ we have $H1(0.01)=4.069467028518>4$ and $H1(0.0001)=4.0006010060034>4$, but $H1(-0.01)=3.950816325242<4$ and $H1(-0.001)=3.994101338150<4$. The following graph shows the functions $H(A)$ and $H(A+tB)$ around 0.

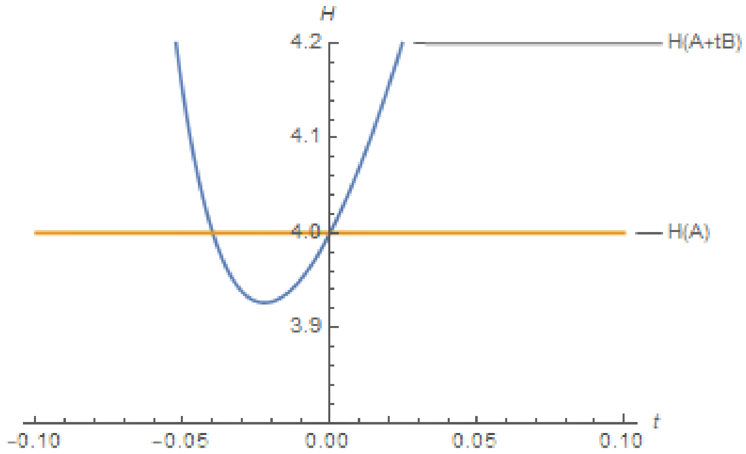

Figure 4.9: For $t>0$ we have $H(A+tB)>H(A)$ and for $t<0$ we have $H(A+tB)<H(A)$.

♣

Example 4.6.2. *Let*

$$A=\begin{bmatrix}1&0&0\\0&4&0\\0&0&6\end{bmatrix}$$

where $a=4$ and $b=6$ and $\mathbf{u}=(1,3,2)$ and $\mathbf{v}=(1,2,3)$ such that $xy=b=6$. Then

$$B=\begin{bmatrix}1&2&3\\3&6&9\\2&4&6\end{bmatrix},\qquad A+tB=\begin{bmatrix}1+t&2t&3t\\3t&4+6t&9t\\2t&4t&6+6t\end{bmatrix}.$$

By the third row of table (4.1), the function $H(A+tB)$ has a minimum at zero. Thus, for all $t\neq 0$ on a neighbourhood of 0, $H(A+tB)>H(A)$. It is clear that $H(A)=6$ and

$$H(A+tB)=\frac{\lambda_1(t)}{\lambda_2(t)},$$

where $\lambda_1(t)$ is the maximum eigenvalue and $\lambda_2(t)$ is the minimum eigenvalue of $A+tB$ on a neighbourhood of zero. If $t=0.01$ then $H(A+tB)=6.0037533314565>6$ and $t=0.001$ implies $H(A+tB)=6.000038311784>6$. For negative t, the values of function $H(A+tB)$ are also bigger than 6. For example, if $t=-0.01$ then $H(A+tB)=6.0039303910494>6$ and for $t=-0.001$ the value of function $H(A+tB)=6.00003848858>6$. In this function for all $t\neq 0$, the value of $H(A+tB)$ is bigger than $H(A)$. It shows us the main question doesn't hold in this condition. The following graph explain that what is happened to $H(a+tB)$ near to zero.

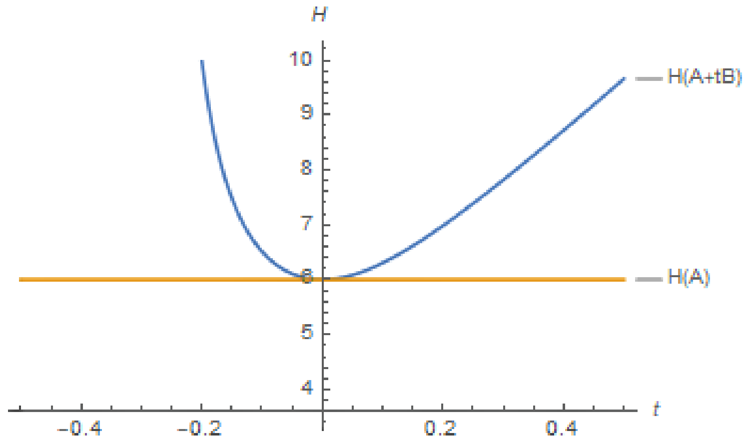

Figure 4.10: For all $t \neq 0$ on a neighbourhood of 0, we have $H(A+tB) > H(A)$.

♣

General Case 2: Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

be a diagonal matrix with $1 < a < b$ and $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, -x, y)$ are two vectors in $\mathbb{R}^3$ such that $B = \mathbf{u} \otimes \mathbf{v}$ is a rank-one matrix. The matrix B and $A + tB$ are equal

$$B = \begin{bmatrix} 1 & -x & y \\ x & -x^2 & xy \\ y & -xy & y^2 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & -tx & ty \\ tx & a - tx^2 & txy \\ ty & -txy & b + ty^2 \end{bmatrix}.$$

Suppose that the eigenvalues of matrix $A + tB$ are $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$. These eigenvalues belong to C^∞. By using the limit and Taylor series for multi-variabls functions we can find

$$\lim_{t\to 0} \lambda_1(t) = \lambda_1(0) = b \qquad \lim_{t\to 0} \lambda_2(t) = \lambda_2(0) = 1 \qquad \lim_{t\to 0} \lambda_3(t) = \lambda_3(0) = a$$

These limit show us for sufficiently small t in the neighbourhood of 0, we have

$$\lambda_2(t) < \lambda_3(t) < \lambda_1(t).$$

So, in this case

$$H(A + tB) = \frac{\lambda_1(t)}{\lambda_2(t)}, \qquad H(A) = b.$$

For this case the first and second derivative are found by the following equations:

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB)\Big|_{t=0} = y^2 - b, \tag{4.48}$$

and,

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} &= -2x^2y^2 \left(\frac{(a-1)^4}{(b-1)^5} + \frac{(a-1)^3}{(b-1)^4} + \frac{(a-1)^2}{(b-1)^3} + \frac{a-1}{(b-1)^2} \right) \\ &+ \frac{-2bx^2}{a-1} + \frac{2y^2(2-x^2)}{b-1} + 2b. \end{aligned} \tag{4.49}$$

We show all conditions of functions $H(A + tB)$ and $H(A)$ for this case in the following table.

Table 4.2: Conditions for linear distortion on a neighbourhood of 0 in case two.

Condition	H'	H''	monotonicity	Results
$y^2 > b$	positive	unknown	increasing	For $t > 0 :\ H(A+tB) > H(A)$
$y^2 > b$	positive	unknown	increasing	For $t < 0 :\ H(A+tB) < H(A)$
$y^2 = b$	zero	positive	minimum at 0	For all t $:\ H(A+tB) > H(A)$
$y^2 = b$	zero	negative	maximum at 0	For all t $:\ H(A+tB) < H(A)$
$y^2 < b$	negative	unknown	decreasing	For $t < 0 :\ H(A+tB) > H(A)$
$y^2 < b$	negative	unknown	decreasing	For $t > 0 :\ H(A+tB) < H(A)$

It is explicit that if a and b are big numbers and x and y are small then second derivative is positive and for opposite case is negative. For example for $a = 2$, $b = 4$, $t = 0$, $x = 1$ and $y = 2$ we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} H(A+tB)\Big|_{t=0} = 1.34979423868312.$$

Also for $a = 2$, $b = 4$, $t = 0$, $x = 5$ and $y = 6$ the second derivative is

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} H(A+tB)\Big|_{t=0} = -1040.2962962962963.$$

The following figure shows the situations of second derivative of $H(A+tB)$. The $H''(A+tB)$ is the yellow graph and the blue graph is the xy-plane. Fourth row in table (4.2) is the exact

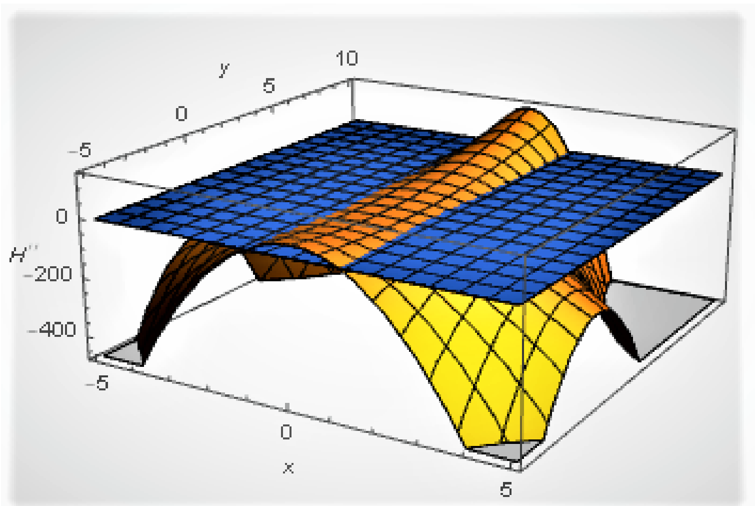

Figure 4.11: The figure of second derivative for case 2 with $a = 2$ and $b = 4$.

solution for the main question. Because there is a open interval of zero such that for every $t \neq 0$ in this interval $H(A+tB) < H(A)$. Here we explain some example.

Example 4.6.3. *Let*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

where $a = 3$ *and* $b = 10$ *and* $\mathbf{u} = (1, 2, -3)$ *and* $\mathbf{v} = (1, -2, -3)$ *such that* $y^2 = 9 < b = 10$. *Then*

$$B = \begin{bmatrix} 1 & -2 & -3 \\ 2 & -4 & -6 \\ -3 & 6 & 9 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & -2t & -3t \\ 2t & 3-4t & -6t \\ -3t & 6t & 10+9t \end{bmatrix}.$$

The linear distortion of A, $H(A)$*, is 10 and the function* $H(A + tB)$ *decreases on a neighbourhood of zero because*

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A + tB)\Big|_{t=0} = y^2 - b < 0.$$

So, for $t > 0$ *in this interval we have* $H(A+tB) < H(A)$ *and for* $t < 0$*,* $H(A+tB) > H(A)$*. Thus, this example is not desirable to main question. The following graph shows the situations of functions* $H(A + tB)$ *and* $H(A)$ *in an interval around 0.*

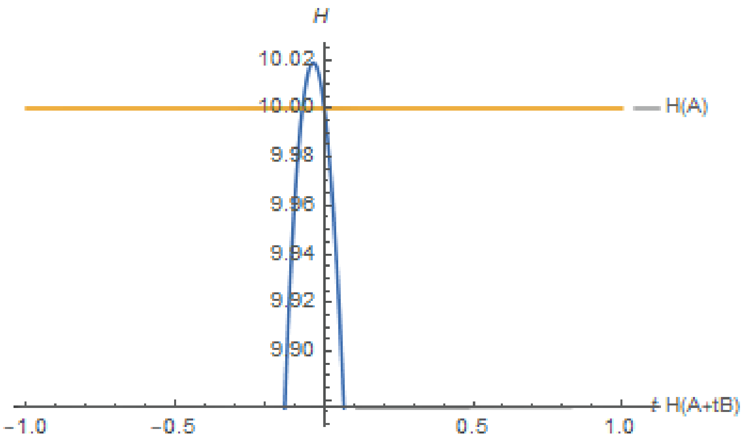

Figure 4.12: If $t > 0$ then $H(A + tB) < H(A)$ and for $t < 0$, $H(A + tB) > H(A)$.

♣

Example 4.6.4. *Suppose that*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

where $a = 2$ *and* $b = 4$ *and* $\mathbf{u} = (1, 4, 2)$ *and* $\mathbf{v} = (1, -4, 2)$ *such that* $y^2 = b = 4$. *Then*

$$B = \begin{bmatrix} 1 & -4 & 2 \\ 4 & -16 & 8 \\ 2 & -8 & 4 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & -4t & 2t \\ 4t & 2-16t & 8t \\ 2t & -8t & 4+4t \end{bmatrix}.$$

The first derivative of $H(A+tB)$ *at* $t = 0$ *is equal zero, therefore,* $t = 0$ *is relative maximum or minimum of* $H(A+tB)$ *that it depends to sign of second derivative. The second derivative was*

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} = -2x^2y^2 \left(\frac{(a-1)^4}{(b-1)^5} + \frac{(a-1)^3}{(b-1)^4} + \frac{(a-1)^2}{(b-1)^3} + \frac{a-1}{(b-1)^2} \right) + \frac{-2bx^2}{a-1} + \frac{2y^2(2-x^2)}{b-1} + 2b.$$

If $a = 2$, $b = 4$, $x = 4$ *and* $y = 2$ *then*

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} = -178.40329218106996,$$

so, $H(A+tB)$ *at* $t = 0$, *has a maximum. Finally, we find the answer for our question. For instance,*

$$\begin{aligned}
&if\ \ t = 0.01 && \Rightarrow && H(A+tB) = 3.990337457010,\\
&if\ \ t = 0.001 && \Rightarrow && H(A+tB) = 3.99991011354399,\\
&if\ \ t = -0.01 && \Rightarrow && H(A+tB) = 3.99144763373212,\\
&if\ \ t = -0.001 && \Rightarrow && H(A+tB) = 3.9999111859119
\end{aligned}$$

We can see the graphs of functions $H(A+tB)$ *and* $H(A)$ *in below.*

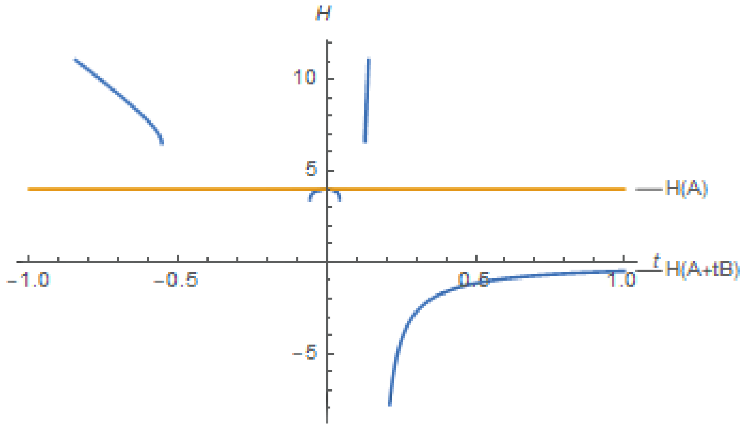

Figure 4.13: For all $t \neq 0$ on a neighbourhood of 0, we have $H(A+tB) < H(A)$.

♣

Example 4.6.5. *Now we want to present an another example such that it has a minimum at 0. Suppose that*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

where $a = 2$ *and* $b = 9$ *and* $\mathbf{u} = (1, 1, 3)$ *and* $\mathbf{v} = (1, -1, 3)$ *such that* $y^2 = b = 9$. *Then*

$$B = \begin{bmatrix} 1 & -1 & 3 \\ 1 & -1 & 3 \\ 3 & -3 & 9 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & -t & 3t \\ t & 2-t & 3t \\ 3t & -3t & 9+9t \end{bmatrix}.$$

The first derivative of $H(A+tB)$ *at* $t = 0$ *is equal zero, because* $y^2 = b = 9$. *Therefore,* $t = 0$ *is a relative maximum or minimum of* $H(A+tB)$ *that it depends to sign of second derivative. The sign of second derivative at* $t = 0$, *is*

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A + tB)\Big|_{t=0} = 1.92864990234375 > 0,$$

so, the function $H(A+tB)$ *has a minimum at* $t=0$. *Thus, the case*

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0} = y^2 - b = 0, \qquad \textit{and} \qquad \frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0} > 0$$

is not the answer for the main question, because in an open interval around 0 for all $t \neq 0$, $H(A+tB) > H(A)$.

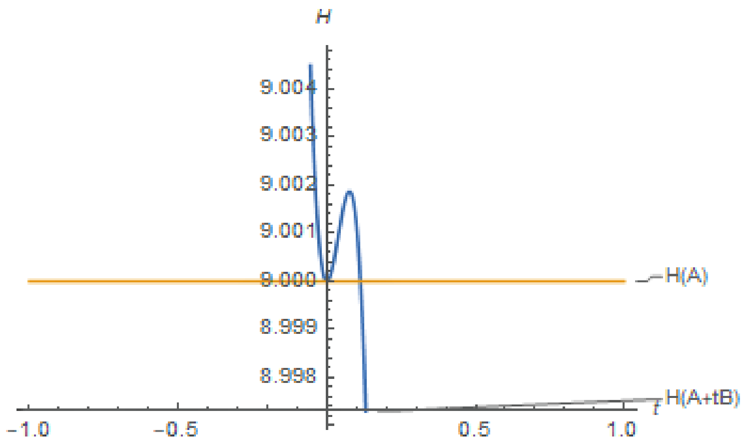

Figure 4.14: For all $t \neq 0$ on a neighbourhood of 0, we have $H(A+tB) > H(A)$.

♣

General Case 3: Finally, because of the earlier reductions reducing A to a diagonal matrix, it is of interest to consider B rank-one and symmetric. Then theorem (1.6.2) tells us that $B = \mathbf{u} \otimes \mathbf{u}$ where $\mathbf{u} = (1, x, y)$. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

be a diagonal matrix with $1 < a < b$ and $\mathbf{u} = (1, x, y)$ is a vector in $\mathbb{R}^3$ such that $B = \mathbf{u} \otimes \mathbf{u}$ is a rank-one matrix. The matrix B and $A + tB$ are equal

$$B = \begin{bmatrix} 1 & x & y \\ x & x^2 & xy \\ y & xy & y^2 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & tx & ty \\ tx & a+tx^2 & txy \\ ty & txy & b+ty^2 \end{bmatrix}.$$

Assume that the eigenvalues of matrix $A+tB$ are $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$. Like the previous two general cases these eigenvalues belong to C^∞. If we use the limit and Taylor series for multi-variabls functions same as before we will find

$$\lim_{t\to 0} \lambda_1(t) = \lambda_1(0) = b \qquad \lim_{t\to 0} \lambda_2(t) = \lambda_2(0) = 1 \qquad \lim_{t\to 0} \lambda_3(t) = \lambda_3(0) = a$$

These limit show us for sufficiently small t in the neighbourhood of 0, the following inequality holds

$$\lambda_2(t) < \lambda_3(t) < \lambda_1(t).$$

So, in this case

$$H(A+tB)=\frac{\lambda_1(t)}{\lambda_2(t)}, \qquad H(A)=b.$$

With long computations we can find that

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0}=y^2-b, \tag{4.50}$$

and,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{dt}^2}H(A+tB)\Big|_{t=0}&=2x^2y^2\left(\frac{(a-1)^5}{(b-1)^6}+\frac{(a-1)^4}{(b-1)^5}+\frac{(a-1)^3}{(b-1)^4}+\frac{(a-1)^2}{(b-1)^3}+\frac{a-1}{(b-1)^2}\right)\\&+\frac{2x^2b}{a-1}+\frac{4y^2+2x^2y^2}{b-1}+2b.\end{aligned} \tag{4.51}$$

Since $1<a<b$, it is clear that the second derivative is always positive in a neighbourhood of zero. Therefore, the function $H(A+tB)$ is convex on that interval. The following table displays the conditions for case 3. The table (4.3) shows that this case is not agreeable answer

Table 4.3: Conditions for linear distortion on a neighbourhood of 0 in case three.

Condition	H'	H''	monotonicity	Results
$y^2>b$	positive	positive	increasing	For $t>0:\ H(A+tB)>H(A)$
$y^2>b$	positive	positive	increasing	For $t<0:\ H(A+tB)<H(A)$
$y^2=b$	zero	positive	minimum at 0	For all t $:\ H(A+tB)>H(A)$
$y^2<b$	negative	positive	decreasing	For $t<0:\ H(A+tB)>H(A)$
$y^2<b$	negative	positive	decreasing	For $t>0:\ H(A+tB)<H(A)$

for main question. Thus, in this case there is no a neighbourhood of zero such that for all $t\neq 0$ in this interval $H(A+tB)<H(A)$. Now we give some examples.

Example 4.6.6. *Let*

$$A=\begin{bmatrix}1&0&0\\0&3&0\\0&0&10\end{bmatrix}$$

where $a=3$ *and* $b=10$ *and* $\mathbf{u}=(1,2,-3)$ *such that* $y^2=9<b=10$. *Then*

$$B=\begin{bmatrix}1&2&-3\\2&4&-6\\-3&-6&9\end{bmatrix}, \qquad A+tB=\begin{bmatrix}1+t&2t&-3t\\2t&3+4t&-6t\\-3t&-6t&10+9t\end{bmatrix}.$$

The linear distortion of A*,* $H(A)$*, is 10 and the function* $H(A+tB)$ *decreases on a neighbourhood of zero because*

$$\frac{\mathrm{d}}{\mathrm{dt}}H(A+tB)\Big|_{t=0}=y^2-b=9-10=-1<0.$$

So, for $t > 0$ in this interval we have $H(A+tB) < H(A)$ and for $t < 0$, $H(A+tB) > H(A)$. For example,

$$\begin{array}{lll} if\ \ t = 0.01 & \Rightarrow & H(A+tB) = 9.9935958377094, \\ if\ \ t = 0.001 & \Rightarrow & H(A+tB) = 9.9990370208181, \\ if\ \ t = -0.01 & \Rightarrow & H(A+tB) = 10.0138410160162, \\ if\ \ t = -0.001 & \Rightarrow & H(A+tB) = 10.0010372657234\,. \end{array}$$

Thus, this example is not desirable to main question. The following graph shows the situations of functions $H(A+tB)$ and $H(A)$ in an interval around 0.

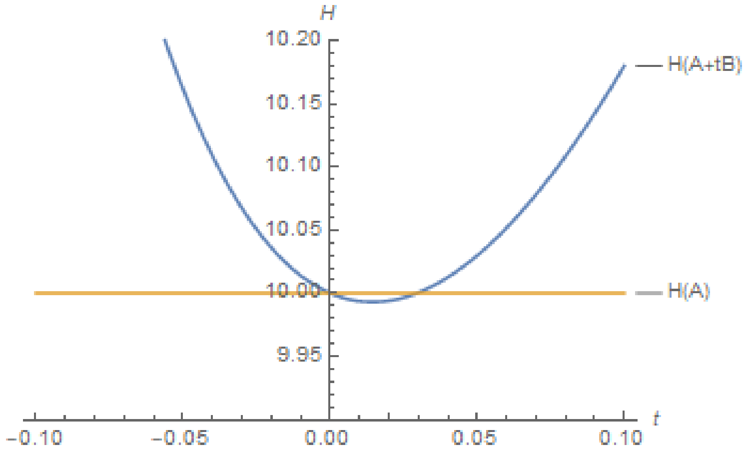

Figure 4.15: If $t > 0$ then $H(A+tB) < H(A)$ and for $t < 0$, $H(A+tB) > H(A)$.

♣

Example 4.6.7. *We will finish this chapter by presenting this example. Let*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

where $a = 5$ and $b = 9$ and $\mathbf{u} = (1, -2, 3)$ such that $y^2 = b = 9$. Then

$$B = \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{bmatrix}, \qquad A + tB = \begin{bmatrix} 1+t & -2t & 3t \\ -2t & 5+4t & -6t \\ 3t & -6t & 9+9t \end{bmatrix}.$$

By the third row of table (4.3), the function $H(A+tB)$ has a minimum at zero. Thus, for all $t \neq 0$ on a neighbourhood of 0, $H(A+tB) > H(A)$. It is clear that $H(A) = 9$. we have

$$\frac{\mathrm{d}}{\mathrm{dt}} H(A+tB)\Big|_{t=0} = y^2 - b = 9 - 9 = 0,$$

and,

$$\frac{\mathrm{d}^2}{\mathrm{dt}^2} H(A+tB)\Big|_{t=0} = 2x^2y^2 \left(\frac{(a-1)^5}{(b-1)^6} + \frac{(a-1)^4}{(b-1)^5} + \frac{(a-1)^3}{(b-1)^4} + \frac{(a-1)^2}{(b-1)^3} + \frac{a-1}{(b-1)^2} \right)$$

$$+ \frac{2x^2b}{a-1} + \frac{4y^2 + 2x^2y^2}{b-1} + 2b > 0,$$

for all x and y in $\boldsymbol{R}$ *(because $1 < a < b$). If we put $a = 5$ and $b = 9$ then the graph of second derivative of $H(A+tB)$, will be as follows. The yellow graph shows the function $H''(A+tB)$ and the blue one shows the xy-plane. In the figure (4.16) the graph of $H''(A+tB)$ is above of xy-plane that implies the function $H(A+tB)$ is a convex function in a neighbourhood of 0.*

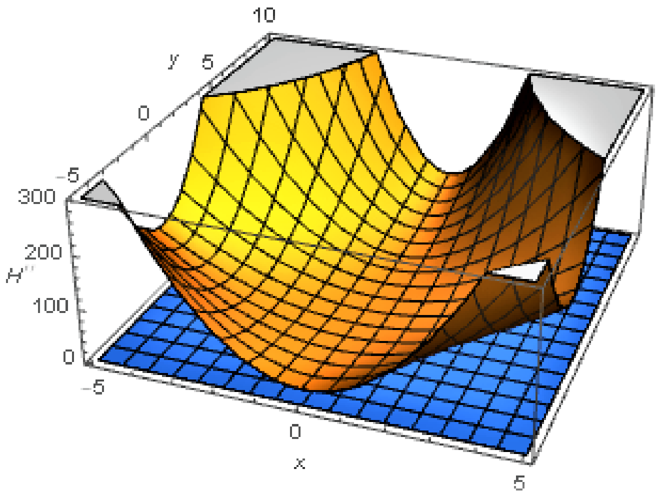

Figure 4.16: The figure of $H''(A+tB)$ for case 3 with $a = 5$ and $b = 9$.

As a result, there is an interval around 0 such that for all $t \neq 0$ we have $H(A+tB) > H(A)$. This conclusion shows us this example and in general case 3 is not answer for main question. For instance,

$$\begin{aligned} if\ \ t &= 0.001 && \Rightarrow && H(A+tB) = 9.00002917369, \\ if\ \ t &= -0.01 && \Rightarrow && H(A+tB) = 9.0030034576327, \\ if\ \ t &= -0.001 && \Rightarrow && H(A+tB) = 9.0000293266910\,. \end{aligned}$$

The following figure illustrates the situations of $H(A+tB)$ and $H(A)$.

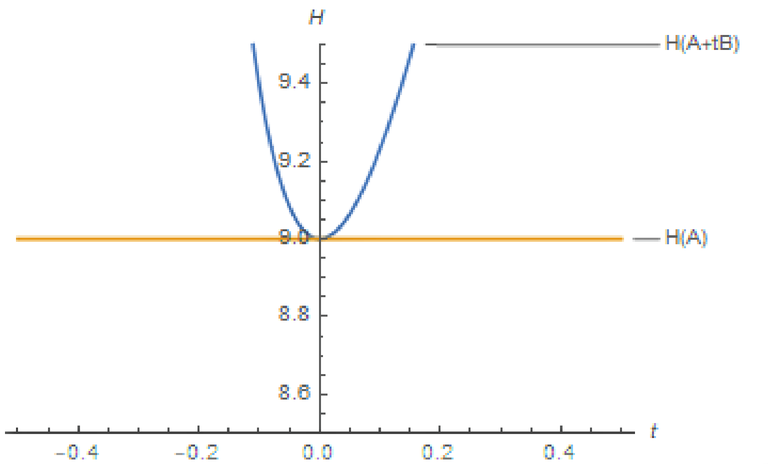

Figure 4.17: For all $t \neq 0$ on a neighbourhood of 0, we have $H(A+tB) > H(A)$.

♣

We now have the following improvement of theorem (4.5.8) showing that the situation of that theorem is much more common.

Theorem 4.6.1. ***(New Result)*** *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a diagonal nonsingular linear mapping. Then there is a sequence of quasiconformal mappings (in fact piecewise linear) $\{T_n\}_{n=1}^{\infty}$ such that*

$$T_n \longrightarrow T \quad \text{uniformly in } \mathbb{R}^n,$$

and

$$H(T_n) = H < H(T).$$

Proof. Since T is a diagonal nonsingular linear mapping we have

$$T(x) = (\alpha x_1, \beta x_2, \gamma x_3), \quad x = (x_1, x_2, x_3)$$

There corresponds a diagonal matrix A such that $T(x) = Ax$. The matrix A can be written as

$$A = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix}.$$

Then by using general case 2 above there exist two vectors $\mathbf{u}$ and $\mathbf{v}$ so that if we set

$$B = \mathbf{u} \otimes \mathbf{v}.$$

then the linear distortion function H is not rank-one convex in the direction B. There are $r, s > 0$ so that

$$H = H(A - sB) < H(A), \quad H = H(A + rB) < H(A)$$

For $n = 1, 2, \ldots$, we now define a sequence $\{T_n\}_{n=1}^{\infty}$ by equation

$$T_n(x) = Ax + \frac{1}{n} h(n\mathbf{u} \cdot x)\mathbf{v}$$

where h is a periodic piecewise linear function on real line as follow: Given s and t of the lemma

$$h(r) = \begin{cases} -rs & -s^{-1} \le r \le 0, \\ rt & 0 \le r \le t^{-1}. \end{cases}$$

Then we can extend h to the entire line (saw-tooth function). h is a bounded Lipschitz function whose derivative assumes only the two values $-s$ and t. The sequence $\{T_n\}_{n=1}^{\infty}$ converges uniformly to $T(x) = Ax$. The differential of T_n also assumes only two values, which are independent of n.

$$D\mathbf{T_n} = A + h'(n\mathbf{u} \cdot x)\mathbf{u} \otimes \mathbf{v} \in \{A - sB, A + tB\}.$$

In either case, the linear distortion of $DT_n(x)$ is equal to H:

$$H(x, T_n) = H(DT_n(x)) \equiv H < H(A) \equiv H(x, A) = H(T).$$

This proves the theorem. ♠

Iwaniec showed in [23] that there exists at least one linear mapping T with the above property. We have shown that this is true for all diagonal linear mappings T with distinct singular values. We believe this result will extend to all nonsingular linear mappings with distinct singular values, however those calculations are currently the subject of future research.

Chapter 5

Conclusion and Summary

Figure 5.1: Conclusion and Summary
Source: https://pixabay.com/en/photos/result/

Chapter 5

Conclusion and Summary

The central problem discussed in this thesis concerns the semicontinuity properties of distortion functionals for quasiconformal mappings. In particular, if $\{\mathbf{f_n}\}_{n=1}^{\infty}$ is a sequence of quasiconformal mappings of $\hat{\mathbb{R}}^n$ onto itself and that

$$\mathbf{f_n} \longrightarrow \mathbf{f}, \tag{5.1}$$

uniformly on $\hat{\mathbb{R}}^n$, we ask for which distortion functionals $\mathcal{K} = \mathcal{K}(\mathbf{g})$, defined for a quasiconformal mappings $\mathbf{g}$ of $\hat{\mathbb{R}}^n$ do we have the lower semicontinuity property

$$\mathcal{K}(\mathbf{f}) \le \liminf_{n\to\infty} \mathcal{K}(\mathbf{f_n}), \tag{5.2}$$

so that the limit $\mathbf{f}$ is itself a $\mathcal{K}$-quasiconformal mapping of $\hat{\mathbb{R}}^n$ onto itself. For instance, $\mathcal{K}$ could be the linear distortion, maximal distortion, outer and inner distortion and etc. This lower semicontinuity property is related to issues of convexity of the functional $\mathcal{K}$ defined on the space of mappings or more precisely the pointwise differentials of these mappings.

5.1 Summary of Thesis Problem

In this thesis we discussed three different notions of convexity, namely polyconvexity, quasiconvexity and rank-one convexity and their known (and unknown) relationships with one and other. We then discussed three different distortion functions, which give rise to the same class of mappings - namely the quasiconformal mappings - but with possibly different associated constants. Each such functional gave rise to a different definition of quasiconformality. These three different definitions were:

1. the definition via the distortion of the modulus of curve families
2. the geometric definition via the linear distortion, and
3. the analytic definition defined via a differential inequality.

We provided references which show that the functionals determined by the distortion of modulus of curve families and also that determined by the pointwise differential inequality do in fact have the lower semicontinuity property given at (5.2). For the linear distortion the question of lower semicontinuity has been answered negatively by Tadeusz Iwaniec (see [15]). In his paper [23] he proved in Theorem (4.5.8) that there is a sequence of quasiconformal mappings as at (5.1) such that for the linear distortion

$$H(\mathbf{f}) > \lim_{n\to\infty} H(\mathbf{f_n}) \tag{5.3}$$

Notice that the limit here does exist and the limit mapping $\mathbf{f}$ above must be quasiconformal (and so $H(\mathbf{f}) < \infty$) since the other two distortion functionals are controlled by the linear distortion functional.
The key idea in Iwaniec's work is that the linear distortion function fails to be rank-one convex in dimension $n \geq 3$. The linear distortion functional is defined pointwise from the differential matrix when it is nonsingular. At the points $x_0 \in \mathbb{R}^3$ of differentiability of a quasiconformal mapping a local analysis of the convexity properties of the linear distortion can be given if we only consider the nonsingular linear mappings $x \mapsto Tx$, $T = D\mathbf{f}(x_0) \in GL(3, \mathbb{R})$. To compute the linear distortion we study the singular values of

$$A = T^t T \in M_{3\times3}(\mathbb{R}),$$

the space of symmetric positive definite 3×3 matrices. Given such an A the Spectral Theorem tells us A is orthogonally diagonalisable and so we may as well suppose it is diagonal since we can diagonalise it by orthogonal (conformal) transformations (at a point). In this way we reduce the problem of the convexity of the linear distortion functional to considering that functional defined on the space of 3×3 positive definite diagonal matrices $A = (a_{ij})$, $a_{ij} = 0$ if $i \neq j$, all the diagonal entries a_{ii} are positive and we can further suppose that $1 = a_{11} \leq a_{22} \leq a_{33}$, by scaling and a further conjugation by orthogonal matrices - neither of which affects the linear distortion.
Next, we know that if B is rank-one matrix then it can be written as the tensor product of two vectors. In Chapter 4 we found conditions and relations between the entries of matrix A and entries of those two vectors to determine when the following is true:

The Main Question of This Thesis: Assume that

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

is diagonal and $B = \mathbf{u} \otimes \mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^3$ and $1 < a < b$. We want to show that for every matrix A (or for every a and b) there are vectors $\mathbf{u}$ and $\mathbf{v}$ and number t, such that

$$H(A + tB) = H(A + t\mathbf{u} \otimes \mathbf{v}) < H(A), \tag{5.4}$$

where H is the linear distortion functional.

Here we only consider the case of strict inequality. The method given for Iwaniec was based on an explicit choice of matrix A and rank-one B. Our aim was to find out how general this situation is. Again we can normalise B a little bit further since we can absorb the fist entry of $\mathbf{u}$ and the first entry of $\mathbf{v}$ into the parameter t is neither of these entries are 0 a special case which is easier to deal with and gives nothing useful.

5.2 Conclusion

Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix} \tag{5.5}$$

be the diagonal matrix in question with $1 < a < b$ and $B = \mathbf{u} \otimes \mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^3$,

$$u = (1, x, y), \quad \text{and} \quad v = (1, z, w) \tag{5.6}$$

We first investigated three particular cases that had additional symmetry.

Case 1: We supposed that $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, y, x)$. With long calculations we found the following result. The table (5.1) shows that choosing vectors $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, y, x)$ is not desirable for question. Because we can not find a neighbourhood of 0 such that for all $t \neq 0$ we have $H(A + tB) < H(A)$.

Case 2: We Assumed that $\mathbf{u} = (1, x, y)$ and $\mathbf{v} = (1, -x, y)$ are two vectors in $\mathbb{R}^3$ such that $B = \mathbf{u} \otimes \mathbf{v}$ is a rank-one matrix. Then, the table (5.2) shows for chosen vectors $\mathbf{u}$ and $\mathbf{v}$ we have the special conditions $y^2 = b$ and sufficient big x and y such that the function $H(A+tB)$ is concave in a neighbourhood of zero. Thus, in this case, there is a neighbourhood of 0 such that for all $t \neq 0$ we have $H(A + tB) < H(A)$.

Case 3: Let $\mathbf{u} = (1, x, y)$ be a vector in $\mathbb{R}^3$ such that $B = \mathbf{u} \otimes \mathbf{u}$ is a rank-one matrix. The table (5.3) shows in this case there is not a neighbourhood of zero that for all $t \neq 0$ we have $H(A + tB) < H(A)$ and choosing the vector $\mathbf{u} = (1, x, y)$ is not good.

Table 5.1: Conditions for linear distortion on a neighbourhood of zero in case one.

xy	Condition	H'	H''	monotonicity	Results
$xy > 0$	$xy > b$	positive	positive	increasing	For $t > 0$: $H(A + tB) > H(A)$
$xy > 0$	$xy > b$	positive	positive	increasing	For $t < 0$: $H(A + tB) < H(A)$
$xy > 0$	$xy = b$	zero	positive	minimum at 0	For all t : $H(A + tB) > H(A)$
$xy > 0$	$xy < b$	negative	positive	decreasing	For $t > 0$: $H(A + tB) < H(A)$
$xy > 0$	$xy < b$	negative	positive	decreasing	For $t < 0$: $H(A + tB) > H(A)$
$xy < 0$	$xy - b$	negative	unknown	decreasing	For $t > 0$: $H(A + tB) < H(A)$
$xy < 0$	$xy - b$	negative	unknown	decreasing	For $t < 0$: $H(A + tB) > H(A)$

Table 5.2: Conditions for linear distortion on a neighbourhood of zero in case two.

Condition	H'	H''	monotonicity	Results
$y^2 > b$	positive	unknown	increasing	For $t > 0$: $H(A + tB) > H(A)$
$y^2 > b$	positive	unknown	increasing	For $t < 0$: $H(A + tB) < H(A)$
$y^2 = b$	zero	positive	minimum at 0	For all t : $H(A + tB) > H(A)$
$y^2 = b$	zero	negative	maximum at 0	For all t : $H(A + tB) < H(A)$
$y^2 < b$	negative	unknown	decreasing	For $t < 0$: $H(A + tB) > H(A)$
$y^2 < b$	negative	unknown	decreasing	For $t > 0$: $H(A + tB) < H(A)$

Table 5.3: Conditions for linear distortion on a neighbourhood of 0 in case three.

Condition	H'	H''	monotonicity	Results
$y^2 > b$	positive	positive	increasing	For $t > 0:\ H(A+tB) > H(A)$
$y^2 > b$	positive	positive	increasing	For $t < 0:\ H(A+tB) < H(A)$
$y^2 = b$	zero	positive	minimum at 0	For all t $:\ H(A+tB) > H(A)$
$y^2 < b$	negative	positive	decreasing	For $t < 0:\ H(A+tB) > H(A)$
$y^2 < b$	negative	positive	decreasing	For $t > 0:\ H(A+tB) < H(A)$

Since the very beginning of the multidimensional theory of quasiconformal mappings, it has been widely believed that the class of K-quasiconformal mappings in $\mathbb{R}^n$ is closed with respect to uniform convergence, where K stands for the linear distortion. In article [23], T. Iwaniec proved one lemma and a theorem that refutes this belief. The key element of his construction is that the linear distortion function fails to be rank-one convex in dimensions $n \geq 3$. Those lemma and theorem are:

Lemma 5.2.1. ***(Iwaniec's Lemma)*** *Given $n \geq 3$ and $H > 1$, there is a matrix A and a rank-one matrix B and numbers $t, s > 0$ such that*

$$H(A - sB) = H(A + tB) = H < H(A).$$

Theorem 5.2.2. ***(Iwaniec's Theorem)*** *For each $n \geq 3$ and $H > 1$, there exists a sequence $\{\mathbf{f_n}\}_{n=1}^{\infty}$ of quasiconformal mappings $\mathbf{f_n} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ converging uniformly to a linear quasiconformal map $\mathbf{f_0} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that*

$$H(x, \mathbf{f_n}) \equiv H < H(x, \mathbf{f_0}), \qquad almost\ everywhere\ in\ \mathbb{R}^n \quad n = 1, 2, \ldots$$

These lemma and theorem show that there exists at least one linear mapping T that the linear distortion is not lower semicontinuous. But in the following theorem, we proved that for all linear mapping T with distinct singular values above property is true. For this purpose, we put $A = T^tT$. Also, we find t and s and rank-one matrix B with using case 2 such that

$$H(A + tB) < H(T^tT) = H(A) \qquad H(A - sB) < H(T^tT) = H(A),$$

and apply Iwaniec's construction.

Theorem 5.2.3. ***(New Result)*** *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a diagonal nonsingular linear mapping. Then there is a sequence of quasiconformal mappings (in fact piecewise linear) $\{T_n\}_{n=1}^{\infty}$ such that*

$$T_n \longrightarrow T \quad \text{uniformly in } \mathbb{R}^n,$$

and

$$H(T_n) = H < H(T).$$

5.3 Further Research

The results above show that it is always possible to find a rank-one matrix B so that (5.4) holds for diagonal matrix A with distinct positive eigenvalues. An interesting question - possibly connected with some aspects of materials science - is to determine the "rank-one direction" for which the linear distortional function at A is most concave. This might identify the structure of the laminations for the minimisers of certain stored energy functionals occuring in the calculus of variations. This problem is quite complicated since there are so many parameters, but we will address it in future research. An unanswered question, (partly posed by Gehring and Iwaniec) is to determine how big the difference between the left-hand side and the right-hand side of (5.3) can be? We hope to resolve this question, at least in the linear case. So, we can say:

The Extension Question: Assume that

$$A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

is diagonal and $B = \mathbf{u} \otimes \mathbf{v}$, where $\mathbf{u} = (x, y, z)$ and $\mathbf{v} = (r, s, w)$ are vectors in $\mathbb{R}^3$ and $1 < a < b < c$.

- We want to show that for every matrix A (or for every a, b and c) there are vectors $\mathbf{u}$ and $\mathbf{v}$ and number t, such that

$$H(A + tB) = H(A + t\mathbf{u} \otimes \mathbf{v}) < H(A).$$

- What is the best B such that if $H(A + tB) = H(A + t\mathbf{u} \otimes \mathbf{v}) < H(A)$, then the distance between $H(A + tB)$ and $H(A)$ will be the maximum. In the other words, what is the best B so that the minimum of $H(A + tB)$ has the smallest possible value. The second part is a question of the **optimization**.

Bibliography

[1] Lars Valerian Ahlfors, Clifford J Earle, Irwin Kra, Mitsuhiro Shishikura, and John Hamal Hubbard. *Lectures on quasiconformal mappings*, volume 3. van Nostrand Princeton, 1966.

[2] Jean-Jacques Alibert and Bernard Dacorogna. An example of a quasiconvex function that is not polyconvex in two dimensions. *Archive for rational mechanics and analysis*, 117(2):155–166, 1992.

[3] Howard Anton and Chris Rorres. Elementary linear algebra, john wiley & sons. *Inc., Newyork*, 1994.

[4] Tom M Apostol. Calculus, vol. ii. *Editora Reverté SA, Barcelona, Buenos Aires, Caracas, México, MCMLXXII*, 1969.

[5] Tom M Apostol. *Mathematical analysis.* Addison-Wesley Reading, MA, 1974.

[6] Kari Astala, Tadeusz Iwaniec, and Gaven Martin. *Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane (PMS-48).* Princeton University Press, 2008.

[7] John M Ball. Does rank-one convexity imply quasiconvexity? In *Metastability and incompletely posed problems*, pages 17–32. Springer, 1987.

[8] Viorel Barbu and Teodor Precupanu. *Convexity and optimization in Banach spaces.* Springer Science & Business Media, 2012.

[9] MK Benaouda and JJ Telega. Two-dimensional examples of rank-one convex functions that are not quasiconvex. In *Annales Polonici Mathematici*, volume 73, pages 291–295. L'INSTITUT DE MATHEMATIQUES DE L'ACADEMIE POLONAISE, 2000.

[10] Frank Burk. *Lebesgue measure and integration: an introduction*, volume 32. John Wiley & Sons, 2011.

[11] Susan Jane Colley. *Vector calculus.* Pearson,, 2012.

[12] Bernard Dacorogna. *Direct methods in the calculus of variations*, volume 78. Springer Science & Business Media, 2007.

[13] Bernard Dacorogna and Jean-Pierre Haeberly. Remarks on a numerical study of convexity, quasi-convexity, and rank-one convexity. *Progress in Nonlinear Differential Equations and Their Applications*, 25:143–154, 1996.

[14] Bernard Dacorogna and Jean-Pierre Haeberly. Some numerical methods for the study of the convexity notions arising in the calculus of variations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 32(2):153–175, 1998.

[15] Peter Duren, Juha Heinonen, Brad Osgood, and Bruce Palka. *Quasiconformal mappings and analysis: a collection of papers honoring FW Gehring.* Springer Science & Business Media, 2012.

[16] Herbert Federer. *Geometric measure theory.* Springer, 2014.

[17] Frederick W Gehring. *The Carathéodory convergence theorem for quasiconformal mappings in space.* Suomalainen Tiedeakatemia, 1963.

[18] Frederick W Gehring, Gaven J Martin, and Bruce P Palka. *An Introduction to the Theory of Higher-Dimensional Quasiconformal Mappings*, volume 216. American Mathematical Soc., 2017.

[19] FW Gehring and T Iwaniec. The limit of mappings with finite distortion. In *ANNALES-ACADEMIAE SCIENTIARUM FENNICAE SERIES A1 MATHEMATICA*, volume 24, pages 253–264. ACADEMIA SCIENTIARUM FENNICA, 1999.

[20] Juha Heinonen and Pekka Koskela. Definitions of quasiconformality. *Inventiones mathematicae*, 120(1):61–79, 1995.

[21] Stanislav Hencl and Pekka Koskela. *Lectures on mappings of finite distortion*, volume 2096. Springer, 2014.

[22] Kenneth Hoffman and Ray Kunze. Linear algebra. 1971. *Englewood Cliffs, New Jersey.*

[23] Tadeusz Iwaniec. The failure of lower semicontinuity for the linear dilatation. *Bulletin of the London Mathematical Society*, 30(1):55–61, 1998.

[24] Tadeusz Iwaniec. Nonlinear analysis and quasiconformal mappings from the perspective of pdes. *Banach Center Publications*, 48(1):119–140, 1999.

[25] Tadeusz Iwaniec and Gaven Martin. *Geometric function theory and non-linear analysis.* Clarendon Press, 2001.

[26] Bernard Kolman and D Hill. *Introductory linear algebra.* Pearson Prentice Hall, 2005.

[27] Charles Bradfield Morrey Jr. *Multiple integrals in the calculus of variations.* Springer Science & Business Media, 2009.

[28] James R Munkres. *Topology.* Prentice Hall, 2000.

[29] Patrizio Neff. Critique of" two-dimensional examples of rank-one convex functions that are not quasiconvex" by mk benaouda and jj telega. In *Annales Polonici Mathematici*, volume 86, page 193. L'INSTITUT DE MATHEMATIQUES DE L'ACADEMIE POLONAISE, 2005.

[30] Constantin Niculescu and Lars-Erik Persson. *Convex functions and their applications: a contemporary approach.* Springer Science & Business Media, 2006.

[31] David Poole. *Linear algebra: A modern introduction.* Cengage Learning, 2014.

[32] Walter Rudin. Principles of mathematical analysis (international series in pure & applied mathematics). 1976.

[33] Walter Rudin. *Real and complex analysis.* Tata McGraw-Hill Education, 1987.

[34] Vladimír Šverák. Examples of rank-one convex functions. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 114(3-4):237–242, 1990.

[35] Vladimir Sverak. Quasiconvex functions with subquadratic growth. *Proceedings: Mathematical and Physical Sciences*, pages 723–725, 1991.

[36] Vladimír Šverák. Rank-one convexity does not imply quasiconvexity. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 120(1-2):185–189, 1992.

[37] Vladimír Šverák. On lower-semicontinuity of variational integrals. *Equadiff 8*, pages 217–220, 1994.

[38] Terence Tao. *An introduction to measure theory*, volume 126. American Mathematical Soc., 2011.

[39] Kristopher Tapp. *Differential Geometry of Curves and Surfaces.* Springer, 2016.

[40] Jussi Väisälä. *Lectures on n-dimensional quasiconformal mappings*, volume 229. Springer, 2006.

[41] Nik Weaver. *Measure theory and functional analysis.* World Scientific Publishing Co Inc, 2013.

Index