Matrix monotone and convex functions

Otte Heinävaara

November 30, 2017

Contents

1 Introduction **4**
- 1.1 Foreword . . . 4
- 1.2 Themes and structure . . . 6
- 1.3 Plan of attack . . . 8
- 1.4 How to rewrite this thesis . . . 11
- 1.5 Some random ideas . . . 11
- 1.6 Main TODO -list . . . 12
  - 1.6.1 Missing proofs . . . 12
  - 1.6.2 Sections to write . . . 12
  - 1.6.3 Figures to make . . . 13

2 Positive maps **14**
- 2.1 Motivation . . . 14
  - 2.1.1 The right definition . . . 14
  - 2.1.2 Real maps and adjoint . . . 16
  - 2.1.3 More convincing . . . 18
- 2.2 Spectral theorem . . . 19
  - 2.2.1 Commuting real maps . . . 22
  - 2.2.2 Symmetric product . . . 25
- 2.3 Congruence . . . 26
  - 2.3.1 $*$-conjugation . . . 26
  - 2.3.2 Block decomposition . . . 28
- 2.4 Loewner order . . . 31
- 2.5 Eigenvalue inequalities . . . 34
- 2.6 Notes and references . . . 37
- 2.7 Ideas . . . 37

3 k-tone functions **40**
- 3.1 Motivation . . . 40

3.2 Divided differences . . . 43
3.3 Cauchy's integral formula . . . 49
3.4 k-tone functions . . . 52
3.5 Smoothness . . . 53
3.6 Analyticity and Bernstein's theorems . . . 58

4 Matrix functions **62**
4.1 Functional calculus . . . 62
4.2 Holomorphic functional calculus . . . 63
4.3 Derivative of a matrix function . . . 64
4.4 Higher derivatives . . . 65

5 Pick-Nevanlinna functions **66**
5.1 Basic properties and examples . . . 66
5.2 Rigidity . . . 67
5.2.1 Schur functions . . . 67
5.2.2 Pick matrices . . . 69
5.3 Weak characterization . . . 71
5.4 Pick-Nevanlinna Interpolation theorem . . . 81
5.5 Schur transform . . . 85
5.6 Compactness . . . 87
5.7 Pick-Nevanlinna-Herglotz representation theorem . . . 87
5.8 Examples of representing measures . . . 88
5.9 Notes and references . . . 88

6 Monotone matrix functions **89**
6.1 Basic properties . . . 89
6.2 Successes . . . 91
6.3 Failures . . . 92
6.4 Pick functions are matrix monotone . . . 95
6.5 Derivative and Loewner's characterization . . . 96
6.6 Local characterization . . . 97
6.7 Regularity . . . 101
6.8 Main Theorem . . . 102
6.9 A bit of history . . . 105
6.10 Loewner's theorem . . . 105

7 Matrix k-tone functions **110**
7.1 Matrix convex functions 110
7.2 Convexity and the second derivative 112
7.3 Matrix k-tone functions 112
7.4 Basic properties 113

8 Trace functions **115**
8.1 Absolute Value 115

9 Representations **117**
9.1 Minkowski–Carathéodory Theorem 118
9.2 Basis k-tone functions 118

Chapter 1

Introduction

1.1 Foreword

This master's thesis delves into the theory of matrix monotone functions. Matrix monotonicity is generalization of standard monotonicity f real functions: now we are just having functions mapping matrices to matrices. Formally, f is *matrix monotone* if for any two matrices A and B such that

$$A \leq B \tag{1.1}$$

we should also have

$$f(A) \leq f(B). \tag{1.2}$$

This kind of function might be more properly called *matrix increasing* but we will mostly stick to the monotonicity for couple of reasons:

- For some reason, that is what people have been doing in the field.
- It doesn't make much difference whether we talk about increasing or decreasing functions, so we might just ignore the latter but try to symmetrize our thinking by choice of words.
- Somehow I can't satisfactorily fill the following table:

monotonic	monotonicity
increasing	?

How very inconvenient.

Of course, it's not really obvious how one should make any sense of these "definitions". One quickly realizes that there two things to understand.

- How should matrices be ordered?
- How should functions act on matrices?

Both of these questions can be (of course) answered in many ways, but for both of them, there is very natural answer. In both cases we can get something more general: instead of comparing matrices we can compare linear maps, and we can apply function to linear mapping.

Just to give a short glimpse of how these things might be defined, we should first fix our ground field (for matrices): let's say it's $\mathbb{R}$, at least for now.

For matrix ordering we should first understand which matrices are *positive*, which here, a bit confusingly maybe, means "at least zero". We can't have everything. For instance it's not very hard to see that it is not possible to give notion of positivity on space of all real $n \times n$ matrices which respects similarity. If we give our space inner product, and restrict to a nice subspace of linear maps, called Hermitian maps, we can have a notion of positivity which respects unitary similarity.

Matrix functions, i.e. "how to apply function to matrix" is bit simpler to explain. Instead of doing something arbitrary the idea is to take real function (a function $f : \mathbb{R} \to \mathbb{R}$, say) and intepret it as function $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, *matrix function.* Polynomials extend rather naturally, and similarly analytic functions, or at least entire. Now, a perverse definition for matrix function for continuous functions would be some kind of a limit when function is uniformly approximated by polynomials (using Weierstrass approximation theorem). This works for Hermitian matrices, but one can do better: apply the function to the eigenvalues of the mapping to get another linear map.

As it turns out, much of the study of matrix monotone and convex functions is all about understanding these definitions of positive maps and matrix functions.

Lastly, one might wonder why should one be interested in the whole business of matrix monotone functions? It's all about point of view. Let's consider a very simple inequality:

For any real numbers $0 < x \leq y$ we have

$$y^{-1} \leq x^{-1}.$$

Of course, this is quite close to the axioms of the real numbers, but there's a rather fruitful interpretation. The function $(x \mapsto \frac{1}{x})$ is decreasing.

Now there's this matrix version of the previous inequality:

For any two matrices $0 < A \leq B$ we have

$$B^{-1} \leq A^{-1}.$$

This is already not trivial, and with previous interpretation in mind, could this be interpreted as the functions $(x \mapsto \frac{1}{x})$ could be *matrix decreasing*? And is this just a special case of something bigger? Yes, and that's exactly what this thesis is about.

1.2 Themes and structure

One of the main themes in this thesis is the notion of *closed salient cone.*

Definition 1.3. Let V be a topological vector space over $\mathbb{R}$ or $\mathbb{C}$ and $C \subset V$. Then C is *closed salient cone* of V if

(i) For every $v \in C$ and $\alpha > 0$ we have $\alpha v \in C$.

(ii) For every $v, w, \in C$ we have $v + w \in C$.

(iii) $0 \in C$.

(iv) If both $v \in C$ and $-v \in C$, then $v = 0$.

(v) C is closed.

Subsets of vector spaces satisfying the first one, two, three and four properties are called *cones*, *convex cones*, *pointed cones* and *salient cones*, respectively. To shorten terminology we call closed salient cones ***proper cones***.

(Pictures of every possibility)

Proper cones are an attempt to generalize the set of non-negative real numbers to a vector space. The first four conditions are pretty much direct translations of the axioms of non-negative numbers to vector spaces. The closedness condition is more of a convenience. In particular, every proper cone gives rise to partial order on the vector space. This makes the vector space (not surprisingly) *ordered topological vector space.*

Definition 1.4. Let V be a topological vector space over $\mathbb{R}$ or $\mathbb{C}$ and $\leq \in V^2$ a relation on V. Then $(V, \leq)$ is *ordered topological vector space* if

(i) $\leq$ is partial order in V, that is

  (a) $v \leq v$ for any $v \in V$.

  (b) If v, w, u such that $v \leq w$ and $w \leq u$, then also $v \leq u$.

  (c) If for some $v, w \in V$ we have both $v \leq w$ and $w \leq u$, then $v = w$.

(ii) If $v, w \in V$ are such that $v \leq w$, then also $v + u \leq w + u$ for any $u \in V$.

(iii) If $v, w \in V$ are such that $v \leq w$, then for any $\alpha \in [0, \infty)$ also $\alpha v \leq \alpha w$.

(iv) The set $\{v \in V | 0 \leq v\}$ is closed.

Of course, one should always hope the order $\leq$ to be total, but this is usually wishful thinking, at least if one wants $\leq$ to be canonical in some way.

Proposition 1.5. *For any proper cone C of topological vector space V the relation $\leq_C$ defined by*

$$v \leq_C w \Leftrightarrow w - v \in C$$

makes $(V, \leq_C)$ ordered topological vector space.

Proof. Easy to check. □

Conversely, for every ordered topological vector space $(V, \leq)$ the set $\{v \in V | 0 \leq v\}$ is a proper cone (which induces $\leq$). Elements of the set $\{v \in V | 0 \leq v\}$ are called the *positive elements* of $(V, \leq)$, or simply *positive*. This terminology is rather unfortunate: if $V = \mathbb{R}$, then the positive elements in $(\mathbb{R}, \leq)$ coincide with the non-negative reals (and not the positive ones). The problem is that with non-total orders the term non-negative should mean something totally different. While potentially confusing, we stick with term positive because of the following reasons:

1. It's short.

The next four chapters are devoted to the four main proper cones of this thesis.

- **Positive maps** (i.e. positive semidefinite maps). These are the main objects of interest and they are used to define the matrix monotone functions, as illustrated in the introduction. As a comprehensive account of these creatures is far beyond the scope of this thesis, only certain aspects, most relevant to the main topic, are discussed. Should the reader lack familiarity with positive maps, the author farmly recommends TODO as a welcome to the topic.

  Nevertheless, chapter should be self-contained given the background topics (in the next section).

- **k-tone functions**. k-tone functions are essentially functions with non-negative k'th derivative. Many of the regularity phenomena discussed in this thesis can be understood through the properties of these functions.

- **Pick functions**. Pick functions are analytic functions on upper half-plane with non-negative imaginary part. It turns out that these functions are intimately linked to matrix monotone functions.

- **Matrix monotone functions**. These are the main objects of the study and all the theory build so far is finally connected.

To be entirely honest, only the first of the four proper cones is really a proper cone (three others fail the fourth condition), but the shortcomings of the others are so minor and canonical that we still call them proper.

The last chapter discusses TODO.

1.3 Plan of attack

This master's thesis is a comprehensive review of the rich theory of matrix monotone functions.

Master's thesis is to be structured roughly as follows.

1. Introduction
    - Introduction to the problem, motivation
    - Brief definition of the matrix monotonicity and convexity
    - Past and present (Is this the right place)
        - Loewner's original work, Loewner-Heinz -inequality
        - Students: Dobsch' and Krauss'
        - Subsequent simplifications and further results: Bendat-Sherman, Wigner-Neumann, Koranyi, etc.
        - Donoghue's work
        - Later proofs: Krein-Milman, general spectral theorem, interpolation spaces, short proofs etc.
        - Development of the convex case
        - Recent simplifications, integral representations
        - Operator inequalities
        - Multivariate case, other variants
        - Further open problems?
    - Scope of the thesis
2. Positive matrices
    - Motivation via restriction, basics
    - Spectral theorem

- Congruence
- Characterizations
- Applications
- Spectrum

3. Divided differences

   - Definition (what kind of?)
   - Mean value theorem
   - Smoothness
   - k-tone functions on $\mathbb{R}$
   - Cauchy's integral formula
   - Regularizations

4. Matrix functions

   - Several definitions: spectral and cauchy
   - Smoothness of matrix functions

5. Pick functions

   - Basic definitions and properties
   - Pick matrices/ determinants
   - Compactness
   - Pick-Nevanlinna interpolation theorem
   - Pick-Nevanlinna representations theorem

6. Monotonic and convex matrix functions

   - Basics
     - Basic definitions and properties (cone structure, pointwise limits, compositions etc.)
     - Classes P_n, K_n and their properties
     - $-1/x$
     - One directions of Loewner's theorem
     - Examples and non-examples

- Pick matrices/determinants vs matrix monotone and convex functions
  - Proofs for (sufficiently) smooth functions
- Smoothness properties
  - Ideas, simple cases
  - General case by induction and regularizations
- Global characterizations
  - Putting everything together: we get original characterization of Loewner and determinant characterization

7. Local characterizations

- Dobsch (Hankel) matrix: basic properties, easy direction (original and new proof)
- Integral representations
  - Introducing the general weight functions for monotonicity and convexity (and beyond?)
  - Non-negativity of the weights
  - Proof of integral representations
- Proof of local characterizations

8. Structure of the classes P_n and K_n, interpolating properties (?)

- Strict inclusions, strict smoothness conditions
- Strictly increasing functions
- Extreme values
- Interpolating properties

9. Loewner's theorem

- Preliminary discussion, relation to operator monotone functions
- Loewner's original proof
- Pick-Nevanlinna proof
- Bendat-Sherman proof
- Krein-Milman proof
- Koranyi proof

- Discussion of the proofs
- Convex case

10. Alternative characterizations (?)

    - Some discussion, maybe proofs

11. Bounded variations (?)

    - Dobsch' definition, basic properties
    - Decomposition, Dobsch' theorems

1.4 How to rewrite this thesis

1. Positive maps: lose all the fat.
2. Divided differences: concentrate on important things, namely relationship between smoothness and k-tone functions.
3. Keep it relatively short, as it is (?)
4. Pick functions: is this the place for these. Start with Schwarz lemma as an rigidity example. Then express Schwarz lemma with contour integrals: generalize, proof by tricks. Notion of Pick points, and finally Pick-Nevanlinna interpolation theorem, some form of it.

1.5 Some random ideas

1. TODO: fix Boor in the references
2. It's easy to see that [Something]. Actually, it's so so easy that we have no excuse for not doing it.
3. When is matrix of the form $f(a_i + a_j)$ positive: f is completely monotone (?).
4. Polynomial regression...
5. TODO: Maximum of two matrices (at least as big), $(a + b)/2 + abs(a - b)/2$
6. If $\langle Ax, y\rangle = 0$ implies $\langle x, Ay\rangle = 0$, then A is constant times hermitian.

7. Angularity preserving functions
8. If subspace of linear maps are diagonalizable with real eigenvalues, is there a inner product such that subspace consists of only Hermitian maps
9. One should be alarmed should one see a positive cone.
10. Make DAG (hopefully) of logical structure of the thesis, colour-coded (with respect to the topic, maybe). Theorem numbers, maybe named theorems with names. To the introduction.
11. Cut the bullshit

1.6 Main TODO -list

1.6.1 Missing proofs

1. Eigenvalues of AB when A and B are positive (?)
2. Symmetric product fail (?)
3. Hindmarsh theorem
4. non-smooth Dobsch char.
5. classes are different
6. matrix k-tone (please expand)

1.6.2 Sections to write

1. Integral representations
2. All “Notes and references” sections
3. More on convex and k-tone matrix functions

1.6.3 Figures to make

1. Proof of spectral theorem
2. Compression
3. Pictures of Peano kernels
4. k-tone functions: visual definition
5. k-tone functions are smooth
6. Pick functions and measures
7. Mean value theorem
8. Pick extension lemma
9. Eigenvalue inequalities: projections and compressions
10. Concrete Pick function extension procedure
11. Change of eigenvalues of $A + tB$
12. Disc lemma
13. Cones in the introduction

Chapter 2

Positive maps

2.1 Motivation

2.1.1 The right definition

Definition 2.1. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and $A \in \mathcal{L}(V)$. We say that A is *positive map*, or simply *positive*, and write $A \geq 0$, if for any $v \in V$ we have

$$\langle Av, v \rangle \geq 0.$$

Why is this the right definition for positivity? Do we really need an inner product to define positivity?

While these are both excellent questions (and one should definitely think about them), there is no way to satisfactorily answer them in the scope of this thesis. Instead, I just try to explain why the definition is pretty damn good.

But first things first: we should check that this notion gives us a proper cone.

Theorem 2.2. *The set*

$$\{A \in \mathcal{L}(V) |\ A \textit{ is positive}\}$$

is a proper cone (of $\mathcal{L}(V)$).

Proof. We verify the five conditions.

(i) Take any $A \geq 0$. Then for every $\alpha > 0$ and $v \in V$ we have $\langle \alpha Av, v \rangle = \alpha \langle Av, v \rangle \geq 0$.

(ii) Take any two $A, B \geq 0$. Then for any $v \in V$ we have $\langle \alpha(A + B)v, v \rangle = \langle Av, v \rangle + \langle Bv, v \rangle \geq 0$.

(iii) For any $v \subset V$ we have $\langle 0v, v\rangle = 0 \geq 0$.

(iv) If $A, -A \geq 0$, then $\langle Av, v\rangle = 0$ for every $v \in V$. We should show that $A = 0$. We'll get back to this in a minute.

(v) We should check that if $(A_i)_{i=1}^{\infty}$ are positive and $\lim_{i\to\infty} A_i = A$, also A is positive. But as the inner product is continuous, for every $v \subset V$ we have $\langle (\lim_{i\to\infty} A_i)v, v\rangle = \lim_{i\to\infty}\langle A_i v, v\rangle \geq 0$.

□

We denote the cone of positive maps by $\mathcal{H}_+(V)$.

Before completing the proof, one should stop and appreciate how elegent and general the construction really is. Indeed, it carries directly to a much more general setting:

Theorem 2.3. *Let V be a topological vector space (over $\mathbb{R}$ or $\mathbb{C}$) and C^* a subset of its continuous dual. Assume that*

$$\{v \in V | w^*(v) = 0 \textit{ for every } w^* \in C^*\} = \{0\}.$$

Then

$$\{v \in V | w^*(v) \geq 0 \textit{ for every } w^* \in C^*\}$$

is a proper convex cone of V.

In our case the subset of the linear functional are the mappings of the form $A \mapsto \langle Av, v\rangle$: they are called *quadratic functionals*. For fixed $A \in \mathcal{L}(V)$ the map $v \mapsto \langle Av, v\rangle$ is the *quadratic form* of A. Now we are left to check the following.

Lemma 2.4 (Injectivity of compression)**.** *If $A \in \mathcal{L}(V)$ and $\langle Av, v\rangle = 0$ for any $v \subset V$, then $A = 0$.*

Proof. The idea is that we can recover the inner product from norm. Indeed, if $v, w \in V$, then $\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\Re(\langle v, w\rangle)$, so knowing the norm, we at least know the real part of the inner product. Doing the same trick with $\|v + iw\|^2$ we can figure out the imaginary part.

How does this help us? By a similar argument $\langle A(v+w), v+w\rangle = \langle Av, v\rangle + \langle Aw, w\rangle + \langle Av, w\rangle + \langle Aw, v\rangle$, so given that the quadratic form is always zero, we have $\langle Av, w\rangle + \langle Aw, v\rangle = 0$ for any $v, w \in V$. Expanding $\langle A(v + iw), v + iw\rangle$ we see that $-i\langle Av, w\rangle + i\langle Av, w\rangle = 0$, which together with the previous observation implies that $\langle Av, w\rangle = 0$ for any $v, w \in V$. Now setting $w = Av$ this implies that $\|Av\|^2 = 0$ for every $v \in V$ so $A = 0$. □

As one would hope, map $v \to \alpha v$, i.e. αI is positive, if and only if $\alpha \geq 0$. In particular in one-dimensional spaces the notion works as expected. Fortunately there are other examples, also. Indeed, any orthogonal projection is positive.

Proposition 2.5. *If $A \in \mathcal{L}(V)$ is a orthogonal projection, then $A \geq 0$.*

Proof. As any orthogonal projetion is sum of one-dimensional orthogonal projections, we can assume that the A is one-dimensional in the first place. It follows that $A = \langle \cdot, v \rangle v / \|v\|^2$ for some $v \in V \setminus \{0\}$. Now for every $w \in V$ we have

$$\langle Aw, w \rangle = \langle \langle w, v \rangle v, w \rangle / \|v, v\|^2 = |\langle w, v \rangle|^2 / \|v\|^2 \geq 0,$$

so A is positive. □

We denote the one-dimensional orthogonal projection to the span of $v \in V \setminus \{0\}$, i.e. the map $\langle \cdot, v \rangle v / \|v\|^2$ by P_v.

Taking positive linear combinations of orhogonal projections leads to large number of examples of positive maps.

2.1.2 Real maps and adjoint

Dual cone thinking lets us also lift other important notions.

Definition 2.6. We say that a map $A \in \mathcal{L}(V)$ is *real*, if

$$\langle Av, v \rangle \in \mathbb{R}$$

for any $v \in V$.

Definition 2.7. We say that a map $A \in \mathcal{L}(V)$ is *imaginary*, if

$$\langle Av, v \rangle \in i\mathbb{R}$$

for any $v \in V$.

The previous two families of maps are usually called Hermitian and Skew-Hermitian and as with positive maps, many of their properties are lifted form usual complex numbers. reals maps will have a special role in our discussion. They form a vector space over $\mathbb{R}$, which is denoted by $\mathcal{H}(V)$. Of course, every imaginary map is just i times real map, and we won't preserve any special notation for such maps.

Interestingly enough, we can also lift the concept of complex conjugate.

Theorem 2.8. *For any $A \in \mathcal{L}(V)$ there exists unique map $A^* \in \mathcal{L}(V)$, called the adjoint of A, for which for any $v \in V$ we have*

$$\langle A^* v, v\rangle = \overline{\langle Av, v\rangle}$$

Proof. The uniqueness of adjoint is immediate from the injectivity of compression. The map $(\cdot)^* : \mathcal{L}(V) \to \mathcal{L}(V)$ should evidently be conjugate linear, so for existence it suffices to find adjoint for suitable basis elements of $\mathcal{L}(V)$: the maps of the form $A = (x \mapsto \langle x, v\rangle w)$ for $v, w \in V$ will do.

Th quadratic form for such map is given by

$$\langle Ax, x\rangle = \langle x, v\rangle\langle w, x\rangle.$$

But if we define $A^* = (x \mapsto \langle x, w\rangle v)$, we definitely have

$$\langle A^* x, x\rangle = \langle x, w\rangle\langle v, x\rangle = \overline{\langle w, x\rangle\langle x, w\rangle} = \overline{\langle Av, v\rangle}.$$

$\square$

In more common terms: a adjoint of linear map $A \in \mathcal{L}(V)$ is the unique map A^* such that

$$\langle Av, w\rangle = \langle v, A^* w\rangle$$

for any $v, w \in V$.

As real maps are their own adjoints, they are often called appropriately *self-adjoint*.

The previous observation makes many of the basic properties of adjoint, which we collect below, evident.

Theorem 2.9. *For any linear maps A and B, with appropriate domains and codomains, and $\lambda \in \mathbb{C}$ we have*

i) *Matrix of A^* with respect to any orthonormal basis is conjugate transpose of matrix of A, i.e. $A^*_{i,j} = \overline{A_{j,i}}$.*

ii) $(A^*)^* = A$

iii) $(A + B)^* = A^* + B^*$

iv) $(\lambda I)^* = \overline{\lambda} I$

v) $(AB)^* = B^* A^*$.

2.1.3 More convincing

Positive maps have many other desirable properties. First of all, eigenvalues of a positive map are non-negative. This fact is a corollary of a more general property.

Definition 2.10. Let $W \subset V$ be a subspace and $A \in \mathcal{L}(V)$. Then the *compression* of A to W, denoted by A_W is the linear map

$$P_W \circ A \circ J_W : W \to W$$

where J_W is the inclusion from W to V and P_W is an orthogonal projection to W.

Lemma 2.11. *Let $W \subset V$ and $A \geq 0$. Then also $A_W \geq 0$. In particular all the eigenvalues of A are non-negative.*

Proof. Note that quadratic form give essentially the one-dimensional compressions. Indeed, if $W = (v)$, then

$$A_W x = \frac{\langle Ax, v\rangle}{\langle v, v\rangle} v = \frac{\langle Av, v\rangle}{\langle v, v\rangle} x$$

for any $x \in (v)$. This means that a map is positive, if and only if its compressions to one-dimensional subspaces are.

Now the trick is that nested compressions work nicely: if $W' \subset W \subset V$ and $A \in \mathcal{L}(V)$, then $(A_W)_{W'} = A_{W'}$. Consequently, if every one-dimensional compression A is positive, same is true for all its compressions.

Now compressing to eigenspace we see that if A is positive, all it's eigenvalues are non-negative. □

In addition, (categorical) sum of two positive map is positive.

Lemma 2.12. *Let $A_1 \in \mathcal{L}(V_1)$ and $A_2 \in \mathcal{L}(V_2)$. Then $A_1 \oplus A_2 \in \mathcal{H}_+(V_1 \oplus V_2)$, if and only if $A_1 \in \mathcal{H}_+(V_1)$ and $A_2 \in \mathcal{H}_+(V_2)$.*

Proof. Recall that one defines $\langle (v_1, v_2), (w_1, w_2)\rangle_{V_1 \oplus V_2} = \langle v_1, w_1\rangle_{V_1} + \langle w_2, w_2\rangle_{V_2}$. Now clearly

$$\langle (A_1 \oplus A_2)(v_1, v_2), (v_1, v_2)\rangle_{V_1 \oplus V_2} = \langle A_1 v_1, v_1\rangle_{V_1} + \langle A_2 v_2, v_2\rangle_{V_2} \geq 0$$

for every $(v_1, v_2) \in V_1 \oplus V_2$, if and only if both $\langle A_1 v_1, v_1\rangle_{V_1} \geq 0$ for every $v_1 \in V_1$ and $\langle A_2 v_2, v_2\rangle_{V_2} \geq 0$ for every $v_2 \in V_2$. □

2.2 Spectral theorem

The most important result in the theory of positive and real maps is the Spectral theorem.

Theorem 2.13. *Let $n = \dim(V)$. Then $A \in \mathcal{L}(V)$ is real if and only there exists real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ and for pairwise orthogonal vectors $v_1, v_2, \ldots, v_n \in V$ such that*

$$A = \sum_{i=1}^{n} \lambda_i P_{v_i}. \tag{2.14}$$

Proof. We first prove the theorem for the positive maps.

We already proved one direction: every map of the previous form is positive.

The other direction is tricky. The idea is to somehow find the vectors v_i. The problem is that such representation is by no means unique. If A is any projection for instance, we could let v_i's by any orthonormal basis of the corresponding subspace and λ_i's all equal to one. There's no vector one has to choose.

But we can think in reverse: there could be many vectors we cannot choose, depending on the map A. If A is any non-identity projection to subspace W, say, we can only choose v_i's in W itself. Indeed, if $x \in W^{\perp}$, we have $Ax = 0$, and hence $\langle Ax, x \rangle = 0$. By comparing the quadratic form it follows $\langle P_{v_i} x, x \rangle = |\langle v_i, x \rangle|^2$ for any $1 \leq i \leq m$. But this means that $v_i \perp W^{\perp}$ and hence $v_i \in W$.

More generally, if it so happens that for some $v \in V$ we have $\langle Av, v \rangle = 0$, we must have $v_i \perp v$ for any $1 \leq i \leq m$. But this means that were there such representation, we should have the following.

Lemma 2.15. *If $A \in \mathcal{H}_+(V)$ and $\langle Av, v \rangle = 0$ for some $v \in V$, then $Av = 0$ and $Aw \perp v$ for any $w \in v$.*

Before proving the Lemma, we complete the proof given the Lemma.

Proof is by induction on n, the dimension of the space. If $n = 0$, the claim is evident. For induction step assume first that there exists $v \in V$ such that $\langle Av, v \rangle = 0$. Then by the Lemma for any $w \in v^{\perp}$ we have $Aw \in v^{\perp} =: W$. But this means that $A = J_W \circ A_W \circ P_W = A$. Now A_W is also positive, and $\dim(W) < n$. By induction assumption we have numbers λ_i and vectors $v_i \in V$ for the map A_W, but such representation for A_W immediately gives representation for A also.

We just have to get rid of the extra assumption on the existence of such v. But for this, note that if $\lambda = \inf_{|v|=1} \langle Av, v \rangle$, consider $B = A - \lambda I$. Now $\inf_{|v|=1} \langle Bv, v \rangle = 0$, and B is hence positive. Also, by compactness, the infimum is attained at some point v, so B satisfies our assumptions. Now cook up a representation for B and add orthonormal basis of V with λ_i's equal to λ: this is required representation for A.

It remains to prove the general case of real map. But there's a rather simple trick: for every real map A the map $A + \|A\|I$ is positive. Indeed, by the Cauchy-Schwarz -inequality one has

$$|\langle Av, v\rangle| \geq -\|Av\|\|v\| \geq -\|A\|\|v\|^2.$$

Now if we manage to the representation for $A + \|A\|I$, we can certainly cook it for A simply subtracting $\|A\|$ from the λ_i's. $\square$

Proof of lemma 2.15. Take any $w \in V$. Now by assumption for any $c \in \mathbb{C}$ we have

$$\langle A(cv + w), cv + w\rangle = |c|^2\langle Av, v\rangle + c\langle Av, w\rangle + \overline{c}\langle Aw, v\rangle + \langle Aw, w\rangle \geq 0$$

But this easily implies that $\langle Av, w\rangle = 0 = \langle Aw, v\rangle$ for any $w \in V$. The first equality implies that $Av = 0$ and the second that $Aw \perp v$ for any $w \in V$. $\square$

Again, as to why such result should be true is a story for another time.

In the representation 2.14 the numbers λ_i are evidently the eigenvalues of A and vectors v_i the corresponding eigenvectors; this is why we call it the *Spectral representation*. As remarked, the representation is of course not unique, but there is a way to make the Spectral representation unique, however. For this we have to change v_i to corresponding eigenspaces.

Theorem 2.16 (Spectral theorem)**.** *Let $A \in \mathcal{H}(V)$. Then there exists unique non-negative integer m, distinct real numbers $\lambda_1, \lambda_2, \ldots, \lambda_m$ and non-trivial orthogonal subspaces of V, $E_{\lambda_1}, E_{\lambda_2}, \ldots E_{\lambda_m}$ with $E_{\lambda_1} + E_{\lambda_2} + \ldots + E_{\lambda_m} = V$, such that*

$$A = \sum_{i=1}^{m} \lambda_i P_{E_{\lambda_i}}.$$

Moreover, this representation is unique.

Proof. Existence of such representation immediately follows from the previous form of Spectral theorem. For uniqueness, note that λ_i's are necessarily the eigenvalues of A and E_{λ_i}'s the corresponding eigenspaces. $\square$

Although the latter version is definitely of theoretical importance, we will mostly stick the former as it only contains one-dimensional projections.

Spectral representation makes many of the properties of real maps obvious. For instance any power of real map is real: indeed, if $A = \sum_{1 \leq i \leq n} \lambda_i P_{v_i}$, then

$$A^2 = \left(\sum_{i=1} \lambda_i P_{v_i}\right)\left(\sum_{j=1} \lambda_j P_{v_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j P_{v_i} P_{v_j} = \sum_{i=1}^{n} \lambda_i^2 P_{v_i},$$

since $P_v P_w = 0$ for $v \perp w$. By induction one can extend the previous for higher powers. In other words: eigenspaces are preserved under compositional powers, and eigenvalues are ones to get powered up. From the original definition this is not all that clear. One could even extend to polynomials. If $p(x) = c_n x^n + c_{n-1}x^{n-1} + \dots c_1 x + c_0$, with $c_i \in \mathbb{R}$, we should write

$$p(A) = c_n A^n + c_{n-1}A^{n-1} + \dots c_1 A + c_0 = \sum_{1\le i\le n} p(\lambda_i) P_{v_i}. \tag{2.17}$$

This implies that if p is the characteristic polynomial of A, then $p(A) = 0$: the special case of Cayley Hamilton theorem. Moreover, the minimal polynomial of A is the polynomial with the eigenvalues of A as single roots.

But even better, if p is polynomial with all except one, say λ_i, of the eigenvalues of A as roots, then $p(A) = p(\lambda_i)P_{E_{\lambda_i}}$. In particular, the projections to eigenspaces of A are actually polynomials of A.

Also, given $A \in \mathcal{H}(V)$, we may write any $x \in V$ in the form $v = \sum_{1\le i\le n} x_i v_i$, where $(v_i)_{i=1}^n$ is a eigenbasis for A and $x_i = \langle x, v_i\rangle$. Now $Ax = \sum_{1\le i\le n} \lambda_i x_i v_i$, so for instance

- $Q_A(x) = \langle Ax, x\rangle = \sum_{1\le i\le n} \lambda_i x_i^2$. Thus Q_A is just a positive linear combination of eigenvalues, and $R(A,\cdot)$ convex combination.
- $\|Ax\|^2 = \langle Ax, Ax\rangle = \sum_{1\le i\le n} \lambda_i^2 x_i^2 \le (\max_{1\le i\le n} \lambda_i^2) \sum_{1\le i\le n} x_i^2 = (\max_{1\le i\le n} \lambda_i^2)\|x\|^2$. It follows that $\|A\| = \max_{1\le i\le n} |\lambda_i|$.

Similarly, if $A \ge 0$, A has a unique positive square root, which we denote by $A^{\frac{1}{2}}$: map such that $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$. Given the spectral reprsentation $A = \sum_{1\le i\le n} \lambda_i P_{v_i}$, we can simply set $A^{\frac{1}{2}} = \sum_{1\le i\le n} \lambda_i^{\frac{1}{2}} P_{v_i}$. As for the uniqueness, note that if B is a positive square root for A and $B = \sum_{1\le i\le n} \lambda_i' P_{v_i'}$, then $B^2 = \sum_{1\le i\le n} \lambda_i'^2 P_{v_i'}$. It follows that eigenvalues of B are simply square roots of eigenvalues of A and the corresponding eigenspaces are equal. Of course, the whole uniqueness argument floats more naturally with unique spectral representation.

Finally, one should note that the lemma 2.15 enjoys following natural generalization.

Proposition 2.18. *If $A \ge 0$ and $A_W = 0$ for some subspace $W \subset V$ then we may decompose $A = A_{W^\perp} \oplus 0_W$.*

Proof. We prove the statement by induction on the dimension of W. Lemma 2.15 took care of the case $\dim(W) = 1$. When $\dim(W) > 1$ we may decompose $W = W' \oplus W''$ where $\dim(W') = 1$. Now $A_{W'} = (A_W)_{W'} = 0_{W'} = 0$, so we may decompose $A = A_{W'^\perp} \oplus 0_{W'}$. But $(A_{W'^\perp})_{W''} = A_{W''} = (A_W)_{W''} = 0_{W''} = 0$, so by the induction hypothesis $A_{W'^\perp} = (A_{W'^\perp})_{W''^\perp} \oplus 0_{W''} = A_{W^\perp} \oplus 0_{W''}$. Consequently $A = A_{W'^\perp} \oplus 0_{W'} = A_{W^\perp} \oplus 0_{W''} \oplus 0_{W'} = A_{W^\perp} \oplus 0_W$, as desired. □

It's easy to see that this property actually characterizes the set of positive and negative maps: we may find kernel of a positive or negative map by finding where the map compresses to zero.

The previous proposition has an useful corollary.

Corollary 2.19. *If $W, W' \subset V$, then $(P_W)_{W'} = 0$ if and only if $W \perp W'$.*

Proof. Assume first that $(P_W)_{W'} = 0$. Then by the lemma 2.18 we have $W = \mathrm{im}(P_W) \subset W'^{\perp}$. The other direction is evident. $\square$

2.2.1 Commuting real maps

Warning! Composition of positive maps need not be positive!

If $A, B \in \mathcal{H}_+(V)$, then, as we noticed, $(AB)^* = B^*A^* = BA$, so for AB to be even real, A and B would at least need to commute. Natural question follows: when do two positive maps commute? Since $(c_1 I + A)$ and $(c_2 I + B)$ commute if and only if A and B do, this is same as asking when do two real maps commute.

It turns out that real maps commute only if they "trivially" commute, in the following sense. If there exists vectors $v_1, v_2, \ldots, v_n$ and numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $\lambda'_1, \lambda'_2, \ldots, \lambda'_n$ such that

$$A = \sum_{1 \leq i \leq n} \lambda_i P_{v_i} \quad \text{and} \quad B = \sum_{1 \leq i \leq n} \lambda'_i P_{v_i},$$

then A and B are said to be *simultaneously diagonalizable*. Simultaneosly diagonalizable maps trivially commute, and it turns out that if two real maps commute, they are indeed simultenously diagonalizable.

To prove this statement, we start with a lemma, simplest non-trivial case of the statement.

Lemma 2.20. *Let $W_1, W_2 \subset V$ be two subspaces. Then P_{W_1} and P_{W_2} commute if and only if there exists orthogonal subspaces U_1, U_2 and U_0 such that*

$$W_1 = U_1 + U_0 \quad \textit{and} \quad W_2 = U_2 + U_0.$$

We then have $P_{W_1} = P_{U_1} + P_{U_0}$ and $P_{W_2} = P_{U_2} + P_{U_0}$, and $U_0 = W_1 \cap W_2$.

Proof. Write $U_0 := W_1 \cap W_2$ and $W_i = U_0 + U_i$ for some $U_i \perp U_0$ for $i \in \{1, 2\}$. Now $P_{W_i} = P_{U_i} + P_{U_0}$ for $i \in \{1, 2\}$ so it suffices to check that $U_1 \perp U_2$. Equivalently, it suffices to prove that if $W_1 \cap W_2 = \{0\}$, and P_{W_1} and P_{W_2} commute, then $W_1 \perp W_2$ or equivalently $P_{W_1} P_{W_2} = 0 = P_{W_2} P_{W_1}$. But for any $v \in V$ we have $W_1 \ni P_{W_1} P_{W_2} v = P_{W_2} P_{W_1} v \in W_2$, so indeed $P_{W_1} P_{W_2} = 0 = P_{W_2} P_{W_1}$. $\square$

Definition 2.21. We say that two $W_1, W_2 \subset V$ subspaces commute if the respective projections commute.

Theorem 2.22. *Let $\mathcal{A} = (A_j)_{j \in J}$ by an arbitrary family of commuting real maps. Then there exists non-trivial orthogonal subspaces of V, $E_1, E_2, \ldots E_m$ with $E_1 + E_2 + \ldots + E_m = V$ and numbers $\lambda_{i,j}$ for $j \in J$ and $1 \leq i \leq n$ such that*

$$A_j = \sum_{1 \leq i \leq m} \lambda_{i,j} P_{E_i}$$

for any $j \in J$.

Proof. The main idea is the following: like in the spectral theorem, we would like to somehow find the subspaces $E_1, E_2, \ldots E_m$. Also, at least for finite families, we could probably use induction, so we should get far just by proving the theorem for a family of only two maps. For two projections we have already proved the statement as lemma 2.20.

Now here's the trick: if two maps commute, so do all their polynomials. Hence if we have two commuting A and B, we know that all the respective eigenspaces commute. Now if we could prove the statement at least for finite families of projections, we could conclude the case of two general maps. Indeed we could write any eigenprojection of A or B as a linear combination of sum finite family of orthogonal (orthogonal) projections, but those projections would then also span A and B.

More generally, if we could prove the statement for arbitrary families of projections, the same argument would yield it for any family of more general linear maps, so we can safely assume that all the maps A_j are projections.

Let's first deal with the finite case by induction. As mentioned, we already dealt with the case $|J| = 2$, but we can draw better conclusions. If we have two commuting projections P_{W_1} and P_{W_2} in $\mathcal{A}$. Now by the lemma we may write $P_{W_1} = P_{U_1} + P_{U_0}$ and $P_{W_2} = P_{U_2} + P_{U_0}$. The nice things is that any map in $\mathcal{A}$ also commutes with $P_{W_1} + P_{W_2} = P_{U_1} + P_{U_2} + 2P_{U_0}$, so also with it's eigenprojections, P_{U_0} and $P_{U_1 + U_2}$. It follows that any map in $\mathcal{A}$ commutes with U_0, U_1 and U_2.

We have split the subspaces W_1 and W_2 in pieces, and we could actually forget W_1 and W_2 altogether and replace them by U_0, U_1 and U_2: note that all the same assumption hold for this new family, and U_0, U_1 and U_2 span W_1 and W_2.

Problem here is of course: it's not clear that the new family, say $\mathcal{A}'$ is any simpler than $\mathcal{A}$! It could well have more elements than $\mathcal{A}$ so we can't just do straightforward induction. What could happen also is that some of the subspaces U_0, U_1, U_2 coincide with the subspaces already present in the family, so the size of the family doesn't increase, and it could even decrease. This will indeed happen. One way to see this is to look at the sum of dimensions of all the projections of the family: if we change the family this sum cannot increase. Moreover, if we pick two subspaces W_1 and W_2 which are not orthogonal, this sum will decrease!

The conclusion is: pick pairs projections with non-orthogonal subspaces and do the replacing procedure as explained before; this process will eventually stop since the sum of dimesions can't drop below zero. But the only reason this process could stop is that all subspaces are pairwise orthogonal in which case we are done. The proof of finite case is complete.

There are many ways to bootstrap the previous argument for arbitrary families. For any finite subfamily we can form the set of generating projections. If add one more map, the set projections get refined: some of the subspaces get split to pieces. Now sizes of all these generating families are bounded by n so we may pick one with most number of elements. Now if A is any projection in $\mathcal{A}$, by maximality, adding it to the family does not refine the generating set. But this means that the generating set generates any element of $\mathcal{A}$ and we are done.

We also see that there exists unique minimal family of generating projections TODO.

Alternative approach to the theorem could be to look at the commutative $\mathbb{R}$-algebra of real maps generated by $\mathcal{A}$: generating projections will be in some sense minimal projections in this algebra. □

The previous theorem sends a very important message.

Philosophy 2.23. Commutativity kills the exciting phenomena.

One would naturally hope that product of positive maps is still positive, but as soon as we try to make such restriction, everything degenerates to $\mathbb{R}^m$, or to diagonal maps. Dealing with diagonal maps is then again just dealing with many real numbers at the same time: of course this makes sense and all, but doesn't lead to very interesting concept.

Conversely, if one wants exciting things to happen, one should make things very non-commutative.

As another corollary of theorem 2.22 we have

Corollary 2.24. *If $A, B \geq 0$ and A and B commute, then $AB \geq 0$.*

Also in the general case we can say something positive:

Proposition 2.25. *If $A, B \geq 0$, then AB is diagonalizable and has non-negative eigenvalues. Conversely, if C is diagonalizable and has non-negative eigenvalues, then it's of the form AB for some positive A and B.*

Proof. TODO (Is this true? Probably not) □

TODO: independence of random variables.

2.2.2 Symmetric product

As normal product doesn't quite work with positivity, next attempt might be symmetrized product

$$S(A, B) = AB + BA,$$

(or maybe with normalizing constant $\frac{1}{2}$ in the front), instead of the usual one. It turns out that even this doesn't fix positivity.

For one dimesional projections things go as badly as they possibly can.

Proposition 2.26. *If $v, w \in V \setminus \{0\}$, then*

$$P_v P_w + P_w P_v \geq 0,$$

if and only if v and w are parallel or orthogonal, i.e. if and only if P_v and P_w commute.

Proof. Since everything is happening in a (at most) two dimensional subspace of V, we may assume that V is two dimensional in the first place. Note that

$$P_v P_w + P_w P_v = (P_v + P_w)^2 - P_v^2 - P_w^2 = (P_v + P_w)^2 - P_v - P_w = A^2 - A = A(A - I),$$

where $A := P_v + P_w$. This is positive, if and only if the eigenvalues of A are outside the interval $(0, 1)$. But since $\mathrm{tr}(A) = 2$ and $A \geq 0$, the only way this can happen is that either A has double eigenvalues 1 or A has eigenvalues 0 and 2. To conclude the claim itself, we are left to do two reality checks:

Lemma 2.27. *If $A = P_v + P_w = I$, then v and w are orthogonal.*

Proof. Note that $I_{(v)} = A_{(v)} = (P_v)_{(v)} + (P_w)_{(v)} = I_{(v)} + (P_w)_{(v)}$, so $(P_w)_{(v)} = 0$. By lemma 2.18 we have $(w) \perp (v)$. □

Lemma 2.28. *If $A = P_v + P_w = 2P_u$ for some $u \in V$, then v, w and u are all parallel.*

Proof. Since $0 = A_{(u)^\perp} = (P_v + P_w)_{(u)^\perp} = (P_v)_{(u)^\perp} + (P_w)_{(u)^\perp} \geq 0$, we have $(P_v)_{(u)^\perp} = (P_w)_{(u)^\perp}$. Now by the lemma 2.18 we have $(v), (w) \subset (u)$: hence the claim. □

□

Moreover, even if one adds positive buffer, things won't work in general.

Proposition 2.29. *Let $\alpha, \beta, \gamma \in \mathbb{C}$ and $n \geq 2$. Then the expression $\alpha A^2 + \beta AB + \overline{\beta} BA + \gamma B^2$ is positive for any $A, B \geq 0$ if and only if $\alpha, \gamma \in [0, \infty)$ and $|\beta|^2 \leq \alpha\gamma$.*

Proof. By easy scaling arguments consideration we may reduce the considerations to the case $\alpha = \gamma = 1$. If $|\beta| \leq 1$, we may write

$$A^2 + \beta AB + \overline{\beta} BA + B^2 = (A + \beta B)^*(A + \beta B) + (1 - |\beta|^2)B^2 \geq 0.$$

If $|\beta| > 1$, we need to find $A, B \geq 0$ such that $A^2 + \beta AB + \overline{\beta} BA + B^2 \ngeq 0$.

TODO (is this true? probably) □

So in some sense, by taking non-commutative products, we really lose most of the structure.

2.3 Congruence

2.3.1 $*$-conjugation

There is one very important way to produce positive maps from others, called congruence. Given any two positive maps A and B, their composition need not be positive, but the map BAB is. First of all, it is real as $(BAB)^* = B^*A^*B^* = BAB$. Also $Q_{BAB}(v) = \langle BABv, v\rangle = \langle A(Bv), (Bv)\rangle \geq 0$ for any $v \in V$. We didn't really need the assumption on the positivity of B, but realness was not that important either. Namely for arbitrary linear B we could consider the product B^*AB instead: this is positive whenever A is. If $C = B^*AB$ for some $B \in \mathcal{L}(V)$, we say that C is $*$-conjugate of A.

Definition 2.30. Let $A, B \in \mathcal{H}$. We say that B is $*$-*conjugate* of A if for some $C \in \mathcal{L}(V)$ we have $B = C^*AC$.

We also see that $Q_{B^*AB} = Q_A \circ B$: conjugation is a change of basis in the quadratic form. This is the main motivation for the definition of the $*$-conjugation. We have already seen that the quadratic form of a map is a good way to characterize many of its good properties, so to some extent to understand maps, we just to need to understand structure of their quadratic forms. By change of basis of the quadratic form we have a good control of what happens. We might however lose some information: if $B = 0$, for instance, the quadratic form after $*$-conjugation by B doesn't tell much about A. But if B is invertible, or equivalently if C and B are $*$-conjugates of each other, we shouldn't lose any information.

Definition 2.31. Let $A, B \in \mathcal{H}$. We say that A and B are *congruent* if they are $*$-conjugates of each other.

It is easily verified that congruence is a equivalence relation.

The construction of $*$-conjugation makes also sense for general linear map A, i.e. we could just as well $*$-conjugate non-positive, or even non-real maps. The result then need not be positive or real, and in general, $*$-conjugation loses its usefulness.

The previous construction can be also performed between two spaces V and W: given any map $B \in \mathcal{L}(V, W)$ and $A \in \mathcal{H}_+(W)/\mathcal{H}(W)/\mathcal{L}(W)$, we note that $B^*AB \in \mathcal{H}_+(V)/\mathcal{H}(V)/\mathcal{L}(V)$. For real maps we can say a lot more: while congruence doesn't in general preserve eigenvalues, it preserves their signs.

Theorem 2.32 (Sylvester's Law of Inertia)**.** *$A, B \in \mathcal{H}(V)$ are congruent, if and only if A and B have equally many positive, negative and zero eigenvalues, counted with multiplicity.*

Proof. Let's start with the "if" part. Let's denote the eigenvalues of A and B by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and $\lambda_1' \leq \lambda_2' \leq \ldots \leq \lambda_n'$, respectively, and the corresponding eigenvectors with $v_1, v_2, \ldots, v_n$ and $v_1', v_2', \ldots, v_n'$. By assumption λ_i and λ_i' have the same sign (or are both zero) for any $1 \leq i \leq n$, so we may find non-zero real numbers $t_1, t_2, \ldots, t_n$ such that $\lambda_i = \lambda_i' t_i^2$. Now consider a linear map C with $Cv_i = t_i v_i'$. C is clearly a surjection and hence a bijection. Also if $v = \sum_{i=1}^n x_i v_i$ $(Q_B \circ C)(v) = Q_B(\sum_{i=1}^n x_i t_i v_i') = \sum_{i=1}^n |x_i|^2 t_i^2 \lambda_i' = \sum_{i=1}^n |x_i|^2 \lambda_i = Q_A(v)$ so $Q_{C^*BC} = Q_B \circ C = Q_A$. It follows that $C^*BC = A$ and hence A and B are congruent.

The "only if" - part is a bit trickier. The idea is to find a good description for the number of positive non-negative eigenvalues. We noticed before that we can write quadratic forms in the form $Q_A(v) = \sum_{i=1}^n \lambda_i |x_i|^2$ if $v = \sum_{i=1}^n x_i v_i$, and v_i are the eigenvectos of A with $\lambda_i' s$ as the corresponding eigenvectors. In particular if say first k eigenvalues are negative, Q_A will be negative on $\text{span}\{v_i | 1 \leq i \leq k\}$, a k-dimensional subspace, minus zero. Similarly, now $n-k$ of the eigenvalues are non-negative, so the quadratic form is non-negative on a subspace of dimension of at least $n-k$. But the dimensions can't be any bigger: if Q_A were for instance negative on some $k+1$ dimesional subspace, this subspace would necessarily intersect a subspace where Q_A is non-negative, which is non-sense.

Congruence preserves the previous notion: if Q_B is negative on a subspace of dimension k, so is $Q_B \circ C$ for any invertible C; namely in the inverse image. Same reasoning holds for the the subspace on which Q_B is non-negative, so again, $Q_B \circ C$ has to have similar structure. We are done. $\square$

In the proof we used the following useful linear algebra fact.

Lemma 2.33. *Let V be n-dimensional and $W_1, W_2 \subset V$ subspaces such that* $\dim(W_1) + \dim(W_2) > n$. *Then $W_1 \cap W_2 \neq \{0\}$.*

Proof. We find non-trivial element $v \in W_1 \cap W_2$. Take bases for W_1 and W_2, say $(e_i)_{i=1}^{n_1}$ and $(f_i)_{i=1}^{n_2}$ with $n_1 + n_2 > n$. Since $(e_i)_{i=1}^{n_1} \cup (f_i)_{i=1}^{n_2}$ can't be linearly independent, as that would mean $\dim(V) \geq \dim(W_1) + \dim(W_2) > n$, we can find non-trivial pair of sequence $(a_i)_{i=1}^{n_1}$'s and $(b_i)_{i=1}^{n_2}$ such that $\sum_{i=1}^{n_1} a_i e_i + \sum_{i=1}^{n_2} b_i f_i = 0$. But $W_1 \ni \sum_{i=1}^{n_1} a_i e_i = v = -\sum_{i=1}^{n_2} b_i f_i \in W_2$, and since sequences are non-trivial, v is non-trivial element in the intersection. $\square$

If n_0, n_- and n_+ denote the number of zero, negative and positive eigenvalues of A, *inertia* of A is the triplet $\{n_0, n_-, n_+\} := \{n_0(A), n_-(A), n_+(A)\}$. The previous theorem can be hence restated, that inertia is invariant under congruence.

The proof also gives a useful characterization for the number of non-negative eigenvalues.

Corollary 2.34. *If $A \in \mathcal{H}(V)$, number of non-negative eigenvalues of A equals largest non-negative integer k such that for some subspace $W \subset V$ of dimension k the quadratic form Q_A is non-negative on W, or equivalently, $A_W \geq 0$.*

Sylvester's Law of inertia gives another proof of the fact that strictly positive maps are exactly the maps congruent to the identity, and positive maps are the maps congruent to some projection. More precisely, the positive maps are partitioned to $n+1$ congruence classes depending on their rank, k:th congruence class containing the projections to k-dimensional subspaces. 0:th class contains only the zero map, the only rank 0 positive map, and the n:th class is the class of strictly positive maps.

If one $*$-conjugates with non-invertible, the inertia may change, but in quite obvious way only: some eigenvalues may move to 0. In particular, we have the following even a bit more general version of the law.

Theorem 2.35 (General Sylvester's Law of Inertia)**.** *For $A, B \in \mathcal{N}(V)$ and A is $*$-conjugate of B, if and only if $n_\pm(A) \leq n_\pm(B)$.*

Proof. The proof is essentially the same. $\square$

This extension draws a picture about the relation of previously mentioned congcruence classes. We can move to the congruence classes of lower indeces by $*$-conjugation, but cannot move up the ladder: the complexity of quadratic forms cannot increase. One could also think that $*$-congruence for linear maps corresponds to multiplication by non-negative real for real numbers.

2.3.2 Block decomposition

Congruence is a convenient to tool to investigate positivity. The idea is that with conguence we can perform sort of a Gaussian elimination. If $n = 2$ for instance, we can write any real map in the matrix form

$$M = \begin{bmatrix} a & b \\ \overline{b} & c \end{bmatrix}$$

for some $a, c \in \mathbb{R}$ and $b \in \mathbb{C}$. Now if $a \neq 0$, we could eliminate with

$$D = \begin{bmatrix} 1 & -\frac{b}{a} \\ 0 & 1 \end{bmatrix}$$

to get

$$MD = \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix} \begin{bmatrix} 1 & -\frac{b}{a} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & 0 \\ \bar{b} & \frac{ac-|b|^2}{a}. \end{bmatrix}$$

The resulting map of course need not be real, but if we also eliminate from the other side by D^*, we get

$$D^*MD = \begin{bmatrix} a & 0 \\ 0 & \frac{ac-|b|^2}{a}. \end{bmatrix} =: M'$$

Now D is evidently invertible, it's determinant being 1, so M and M' are congruent. Sylvester's law of inertia tell's us hence that that if $a > 0$ and $\det(M) \geq 0$, then $M \geq 0$.

We can generalize this thinking. For general n if we have decomposition $V = W_1 \oplus W_2$, then we can decompose any mapping M as

$$M = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix},$$

where A, B and C are the *blocks* of M given by $A = P_{W_1} \circ M \circ J_{W_1} = M_{W_1}$, $B = P_{W_1} \circ M \circ J_{W_2}$ and $C = P_{W_2} \circ M \circ J_{W_2} = M_{W_2}$. Now we can generalize the previous elimination: if A happens to be invertible and we let

$$D = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}$$

then

$$D^* = \begin{bmatrix} I & 0 \\ -B^*A^{-1} & I \end{bmatrix}$$

and

$$D^*MD = \begin{bmatrix} I & 0 \\ -B^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix}. \tag{2.36}$$

The map $(C - B^*A^{-1}B) : W_2 \to W_2$ is called the *Schur complement* of block A of M, or maybe one should say Schur complement of M with respect to W_1. We denote the Schur complement by M/A.

Now again if A is invertible, $M \geq 0$ if and only if $A > 0$ and $M/A \geq 0$.

This observations leads to convenient characterization for strictly positivity, called Sylvester's criterion. If W_2 is 1-dimensional, M/A is just a real number and M is stricly positive if and only if $A > 0$ and this real number is positive. On the other hand, by computing determinants we see that

$$\det(M) = \det\left(\begin{bmatrix} A & 0 \\ 0 & M/A \end{bmatrix}\right) = \det(A)\det(M/A),$$

as $\det(D) = 1$. Hence M is positive if and only if $\det(M)$ is positive and $A > 0$. Applying the same idea inductively we arrive at

Theorem 2.37 (Sylvester's criterion)**.** *$A \in \mathcal{H}(V)$ is stricly positive if and only for some (and then for any) sequence of subspaces $W_1 \subset W_2 \subset \ldots \subset W_{n-1} \subset W_n = V$ with $\dim(W_m) = m$ we have $\det(A_{W_m}) > 0$ for any $1 \leq m \leq n$.*

TODO: Explain what happend with non-strict case.

One can solve M from 2.36 to arrive at so-called *LDL-decomposition* of M:

$$M = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}. \tag{2.38}$$

LDL-decomposition leads to many interesting identities. First of all, (given that A is invertible), M is invertible if and only if $C - B^*A^{-1}B$ is and its inverse is given by

$$\begin{aligned} M^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (C - B^*A^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -B^*A^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} + A^{-1}B(C - B^*A^{-1}B)^{-1}B^*A^{-1} & -A^{-1}B(C - B^*A^{-1}B)^{-1} \\ -(C - B^*A^{-1}B)^{-1}B^*A^{-1} & (C - B^*A^{-1}B)^{-1} \end{bmatrix}. \end{aligned}$$

If one take Schur complement with respect to C instead one arrives at

$$M^{-1} = \begin{bmatrix} (A - BC^{-1}B^*)^{-1} & (A - BC^{-1}B^*)^{-1}BC^{-1} \\ -C^{-1}B^*(A - BC^{-1}B^*)^{-1} & C^{-1} + C^{-1}B^*(A - BC^{-1}B^*)^{-1}BC^{-1} \end{bmatrix},$$

so by comparing blocks we see that for instance

$$A^{-1} + A^{-1}B(C - B^*A^{-1}B)^{-1}B^*A^{-1} = (A - BC^{-1}B^*)^{-1}, \tag{2.39}$$

Woodbury matrix identity. Why might such identity be useful? The idea is that if $\dim(W_2) \ll \dim(W_1)$, the identity is way to connect inverse of $A - BC^{-1}B^*$, low rank update of A, and A. If $\dim(W_2) = 1$ for instance, by setting $C = -1$ for some $c > 0$ and $B = v$ for some $v \in V$ we get

$$A^{-1} - \frac{A^{-1}vv^*A^{-1}}{1 + \langle A^{-1}v, v\rangle} = (A + vv^*)^{-1} :$$

inverse of rank 1 update can be easily calculated if one knows the inverse of the original map.

In a similar vein one obtains formulas for determinants. Starting with $\det(M) = \det(A)\det(C - B^*A^{-1}B)$, if we happen to know determinant of a map and need determinant of a compression, it is sufficient to find it for a schur complement. This is particularly useful when W_2 is low dimensional. If $\dim(W_2) = 1$ and $W_2 = \operatorname{span}(v)$, then

$$\begin{aligned}\det(M) &= \det(A)\,(C - B^*AB)\\ &= \frac{\det(A)|v|^2}{\langle M^{-1}v, v\rangle}:\end{aligned}$$

Schur complement is inverse of compression M to W_2. It follows that if A is invertible, we have

$$\det(A_W) = \det(A)\langle A^{-1}v, v\rangle. \tag{2.40}$$

By comparing determinants from two LDL-decompositions we arrive at

$$\det(A)\det(C - B^*A^{-1}B) = \det(C)\det(A - BC^{-1}B^*), \tag{2.41}$$

matrix determinant lemma. Again, by the choices for $B = v$ and $C = -1$ we arrive at

$$\det(A)\left(1 + \langle A^{-1}v, v\rangle\right) = \det(A + vv^*):$$

determinant of rank 1 update can be also easily calculated.

Of course, once one knows the statements, such identities could also be easily verified by multiplying everything out, for instance, but this is how one might stumble upon them.

2.4 Loewner order

Definition 2.42. If $A, B \in \mathcal{H}(V)$, we write that $A \leq B$ (A is smaller than B) if $B - A \geq 0$, $B - A$ is positive. If $B - A$ is strictly positive, we write $A < B$.

We could of course have made such definition immediately after defining positive maps, but now we have proper tools to investigate such order. Proposition 2.2 tells us that such order is indeed partial order on the $\mathbb{R}$-vector space of real maps. More explicitly, we have the following properties:

Proposition 2.43. *(i) If $A \leq B$ then $\alpha A \leq \alpha B$ for any $\alpha \geq 0$.*

(ii) If $A \leq B$ and $B \leq C$ then $A \leq C$.

(iii) If $A \leq B$ and $B \leq A$ then $A = B$.

(iv) If $\lambda I \leq A$, then all the eigenvalues of A are at least λ. Similarly if $A \leq \lambda I$, all the eigenvalues of A are at most λ.

Example 2.44. If $W_1, W_2 \subset V$ are two subspaces of V we have $P_{W_1} \leq P_{W_2}$ if and only if $W_1 \subset W_2$. Indeed if $W_1 \subset W_2$ then $W_2 = W_1 + W_3$ for some $W_3 \perp W_1$ and hence $P_{W_2} = P_{W_1} + P_{W_3} \geq P_{W_1}$. Conversely if $P_{W_1} \leq P_{W_2}$, then $0 \leq (P_{W_1})_{W_2^\perp} \leq (P_{W_2})_{W_2^\perp} = 0$, so $(P_{W_1})_{W_2^\perp} = 0$. But now lemma 2.19 implies that $W_1 \subset W_2$.

Key thing here is to note what is missing from the standard real ordering: multiplication by positive map doesn't preserve usual ordering. This is the reason many standard arguments don't work for general real maps.

For example if $0 < a \leq b$, with real numbers one could multiply the inequalities by the positive number $(ab)^{-1}$ to get $0 < b^{-1} \leq a^{-1}$. This doesn't quite work with linear maps anymore.

Congruence is way to at least partially fix this deficit: it's almost like multiplying by positive number. We have

Proposition 2.45. *If $A \leq B$, then for any C we have $C^*AC \leq C^*BC$.*

Using the previous we can mimic the previous proof to make it work.

Theorem 2.46. *If $0 < A \leq B$, then $B^{-1} \leq A^{-1}$.*

Proof. As mentioned, we can't really multiply by $(AB)^{-1}$, as it does not preserve the order and doesn't even need to be positive. If A and B commute, this would work though. We can almost multiply by A^{-1}: $*$-conjugate by $A^{-\frac{1}{2}}$. This preserves the order, and we get

$$I \leq A^{-\frac{1}{2}} B A^{-\frac{1}{2}}.$$

Now one would sort of want to multiply B^{-1}; so $*$-conjugate by $B^{-\frac{1}{2}}$, but B is in the middle, so this doesn't quite work. But now we can follow the original strategy: since $I \leq X := A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ we have $X^{-1} \leq I$, that is

$$A^{\frac{1}{2}} B^{-1} A^{\frac{1}{2}} \leq I.$$

This is already almost what we wanted: simply $*$-conjugate by $A^{-\frac{1}{2}}$. $\square$

There's one wee bit non-trivial part in the proof: if $I \leq X$ then $X^{-1} \leq I$. But if $I \leq X$, all the eigenvalues of X are at least 1, so all the eigenvalues of its inverse are at most 1, so $X \leq I$.

Remark 2.47. Alternatively, we could conjugate both sides by $X^{-\frac{1}{2}}$ to arrive at the conclusion. Note that by doing this we have only used $*$-conjugation in the proof: actually we have $*$-conjugated altogether with

$$A^{-\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^{-\frac{1}{2}}A^{-\frac{1}{2}} = (A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^{\frac{1}{2}}A^{\frac{1}{2}})^{-1}.$$

The map $A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^{\frac{1}{2}}A^{\frac{1}{2}}$, which is real, is usually called the geometric mean of A and B. It turns out that this mean, denoted by $G(A, B)$ satifies

$$G(A, B) = G(B, A) \quad \text{and} \quad G(A, B)^{-1} = G(A^{-1}, B^{-1}),$$

and if A and B commute we have $G(A, B) = (AB)^{\frac{1}{2}}$. The defining property of it we used it was that $G(A, B)$ is unique real map with

$$B = G(A, B)A^{-1}G(A, B).$$

The point is: somewhat curiously we can almost do the original proof: just replace multiplication by congruence by square root, and replace square root of product by geometric mean.

To further highlight the importance of congruence, we can use it to change map inequalities to usual real inequalities. For instance, one can generalize so called (two variable) arithmetic-harmonic mean inequality, which states that for any two positive real numbers a and b we have

$$\frac{a+b}{2} \geq \frac{2}{\frac{1}{a} + \frac{1}{b}}.$$

This classic inequality, which can be seen as a restatement of the convexity of the map $x \mapsto \frac{1}{x}$, can be verified for instance by multiplying out the denominator and rewriting it as $\frac{(a-b)^2}{ab} \geq 0$.

To prove the matrix version, namely

$$\frac{A+B}{2} \geq 2(A^{-1} + B^{-1})^{-1}$$

for any $A, B > 0$, we can $*$-conjugate both sides by $A^{-\frac{1}{2}}$ to arrive at

$$\frac{I + A^{-\frac{1}{2}}BA^{-\frac{1}{2}}}{2} \geq 2(I + A^{\frac{1}{2}}B^{-1}A^{\frac{1}{2}})^{-1}.$$

If one writes $X = A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$, this rewrites to

$$\frac{I + X}{2} \geq 2(I + X^{-1})^{-1}.$$

But now since I and X commute, the claim is evident form the scalar inequality. In a similar manner one could also prove that the geometric mean lies between arithmetic and harmonic.

2.5 Eigenvalue inequalities

There's great deal of things to be said about relationship between eigenvalues and Loewner order. Let's denote the eigenvalues of real map A by $\lambda_1(A) \geq \lambda_2 \geq \ldots \geq \lambda_n(A)$. One of the most basic result is the following.

Proposition 2.48. *Assume that $A \leq B$. Then for any $1 \leq k \leq n$ we have $\lambda_k(A) \leq \lambda_k(B)$.*

Proof. We first claim that A has at most as many non-negative eigenvalues as B: if we manage to do this, we can apply the observation for the maps $A - \lambda I$ and $B - \lambda I$ and conclude that B has at least k eigenvalues in $[\lambda_k(A), \infty)$, which implies that $\lambda_k(A) \leq \lambda_k(B)$.

To prove the claim note that if A has k non-negative eigenvalues, by lemma 2.34 it's restriction to some k-dimensional subspace is positive. But then also the compression of B to this subspace is positive, so also B has at least k non-negative eigenvalues. $\square$

In general that's all one can say: if numbers $a_1 \geq a_2 \geq \ldots a_n$ and $b_1 \geq b_2 \geq \ldots \geq b_n$ satisfy $a_k \leq b_k$, then we can definitely find A and B with $A \leq B$ and a_i's and b_i's as eigenvalues: simply take $A = \sum_{i=1}^n a_i P_{e_i}$ and $B = \sum_{i=1}^n b_i P_{e_i}$ where $(e_i)_{i=1}^n$ is an orthonormal basis.

Eigenvalues work also desirably with compression.

Proposition 2.49 (Cauchy interlacing theorem)**.** *If $A \in \mathcal{H}^n(V)$ and $W \subset V$ is of dimension $n-1$, then we have*

$$\lambda_1(A) \geq \lambda_1(A_W) \geq \lambda_2(A) \geq \lambda_2(A_W) \geq \ldots \geq \lambda_{n-1}(A) \geq \lambda_{n-1}(A_W) \geq \lambda_n(A).$$

Proof. We use the same appoach and first prove that A has at least as many non-negative eigenvalues as A_W: again if we know this, we get inequalities of the form $\lambda_k(A) \geq \lambda_k(A_W)$. Then applying the idea for the $-A$, we get the reverse inequalities, and finally the complete chain.

To prove the claim, note again that if A_W has k non-negative eigenvalues, by lemma 2.34 it's compression to some k-dimensional subspace is positive. But then also compression of A to this same subspace is positive and hence it has k non-negative eigenvalues. $\square$

TODO picture of eigenvalues changing when compressed

Again one can prove that this result is strongest possible.

Proposition 2.50. *For any $a_1 \geq b_1 \geq a_2 \geq \ldots \geq b_{n-1} \geq a_n$ we may find $A \in \mathcal{H}^n(V)$ with a_i's as spectra and $(n-1)$-dimensional subspace W of V such that eigenvalues of A_W are the b_i's.*

Before approaching the proof we note an interesting corollary.

Let us call pair $(A, B) \in \mathcal{H}(V)^2$ a *projection pair* if $B - A = vv^*$ for some $v \in V$. Note that such v is always unique up to phase. Let us say that a projection pair (A, B) is strict, if whenever $B - A = vv^*$ then v is not orthogonal to any eigenvector of A.

Corollary 2.51. *Let (A, B) be a projection pair. Then*

$$\lambda_1(B) \geq \lambda_1(A) \geq \lambda_2(B) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(B) \geq \lambda_n(A).$$

(A, B) is strict if and only if all the inequalities are strict.

Proof. By proposition 2.48 $\lambda_k(A) \leq \lambda_k(B)$, so we just need to prove that $\lambda_{k+1}(B) \leq \lambda_k(A)$. Let W be orthocomplement of $\operatorname{span}\{v\}$. Then $A_W = B_W$ and W is $(n-1)$-dimensional. Hence by lemma 2.49 we have $\lambda_{k+1}(B) \leq \lambda_k(B_W) = \lambda_k(A_W) \leq \lambda_k(A)$, which is what we wanted. TODO □

One could now use induction to make similar but more complicated statements about inequalities when compression is to subspace of bigger codimension or when $B - A$ is or larger rank. One could also ask what happens $B - A$ multiple of projection to k-dimensional subspace (TODO: what happens?).

One also has a similar converse as in the compression case.

Proposition 2.52. *For any $b_1 \geq a_1 \geq b_2 \geq a_2 \ldots \geq b_n \geq a_n$ we may find projection pair $A, B \in \mathcal{H}^n(V)$ with a_i's and b_i's as spectra.*

We will first prove this converse. The idea is the following: the eigenvalues of roots of the characteristic polynomial, hence to control eigenvalues, we should control characteristic polyomials. It turns out that if two maps differ by map rank 1, their characteristic polynomials are intimately related.

Lemma 2.53. *Let $A, B \in \mathcal{H}$ be a projection pair. Then*

$$\det(B - zI) = \det(A - zI)\left(1 + \langle (A - zI)^{-1} v, v\rangle\right).$$

Proof. This is just direct application of rank 1 version of matrix determinant lemma 2.41. □

Proof of propostion 2.52. If $a_i = b_j$ for some $1 \leq i, j \leq n$ we can forget a_i and b_j, solve the remaining problem on smaller space to get A' and v' and take $A : V' \oplus \mathbb{C} \to V' \oplus \mathbb{C}$ to be $A' \oplus a_i$ and $v = v' \oplus 0$. We may hence assume that the numbers are distinct.

First take A with the given eigenvalues. By the previous lemma we just want to choose v in such a way that

$$\frac{p_B(z)}{p_A(z)} = 1 + \langle (A - zI)^{-1} v, v\rangle = 1 + \sum_{i=1}^{n} \frac{|\langle v, e_i\rangle|^2}{a_i - z},$$

where e_i's are the eigenvectors of A and p_A and p_B are polynomials with a_i's and b_i's as roots. But this is easily achieveable if can show that the residues of $p_B(z)/p_A(z)$ are negative, which follows easily from the interlacing property.

From the identity we can also easily deduce the other direction. If $\langle v, e_i\rangle \neq 0$ for any $1 \leq i \leq n$ the function

$$z \mapsto 1 + \sum_{i=1}^{n} \frac{|\langle v, e_i\rangle|^2}{a_i - z}$$

has n poles of negative residue so it has a root between any two poles. Also it tends to 1 at infinity so it has also root on (a_1, ∞). □

The proof of 2.50 is similar: the aim to first connect the characteristic polynomials of A and its compression and then do similar observations.

Lemma 2.54. *Let $A \in \mathcal{H}(V)$ and $W \subset V$ a subspace of codimension 1, orthocomplement of subspace spanned by unit vector v. Then*

$$\det(A_W - zI) = \det(A - zI)\langle (A - zI)^{-1} v, v\rangle$$

Proof. This is direct application of 2.40. □

Proof of proposition 2.50. Proof is just an easier version of the proof of 2.52 □

These eigenvalue inequalities have interesting corollaries.

Corollary 2.55. *If $A, B \in \mathcal{H}^n(V)$, then $|\lambda_i(A) - \lambda_i(B)| \leq \sum_{i=1}^n |\lambda_i(A - B)| \leq n\|A - B\|$ for any $1 \leq i \leq n$.*

Proof. If $B - A = \sum_{i=1}^n \lambda_i(B - A)P_{v_i}$, write $A_j = A + \sum_{i=1}^j \lambda_i(B - A)P_{v_i}$. By using lemma 2.52 we may trace how the eigenvalues of A_j change when j increases. We conclude the given bound ... almost: this implies bound for terms $|\lambda_i(A) - \lambda_{\sigma(i)}(B)|$ for some permutation σ of $\{1, 2, \ldots, n\}$. But $\max_{1 \leq i \leq n} |\lambda_i(A) - \lambda_i(B)| \leq \max_{1 \leq i \leq n} |\lambda_i(A) - \lambda_{\sigma(i)}(B)|$ for any permutation σ (as can be seen by simple exchange argument, for instance). The last inequality is trivial, so we are done. □

TODO: change order of compression and projection eigenvalues converses.

2.6 Notes and references

2.7 Ideas

- Normal maps
- Square root of a matrix
- Ellipses map to ellipses
- adjoints of vectors
- Moore-Penrose pseudoinverse
- (canonical, löwdin) orthogonalization, polar decomposition and orthogonal Procrustes problem
- projection matrices
- Hilbert-Schmidt norm ($\to$ matrix functions?) and inner product
- Hilbert spaces
- Real vs. complex
- Positive definite kernels
- Weakly positive matrices
- Hlawka inequality for determinant
- Trace-characterization of positive maps.
- Splitting positive maps to pseudo square roots
- Product of maps
- Exponential formula for geometric mean?
- Maximum of matrices with powerlimit
- If A, B are Hermitian, what eigenvalues AB can have? What if the eigenvalues are known? What about $AB+BA$. What eigenvalues A can have if eigenvalues of $\Re(A)$ are known.

- It seems to be the case that if $n = 2$, and A is Hermitian with $\mathrm{spec}(A) = \{\lambda_1, \lambda_2\}$ ($\lambda_1 \leq \lambda_2$), then there exists linear B such that $\Re(B) = A$, and $\mathrm{spec}(B) = \{\mu_1, \mu_2\}$ if and only if $\Re(\mu_1 + \mu_2) = \lambda_1 + \lambda_2$ and $\lambda_1 \leq \Re(\mu_i) \leq \lambda_2$. In general this is known as Ky-Fan theorem, according to [3].

- Let's define $A \leq_2 B$ if $\mathrm{tr}(A) = \mathrm{tr}(B)$ and for any $t \in \mathbb{R}$ we have $\mathrm{tr}(|A - tI|) \leq \mathrm{tr}(|B - tI|)$. Similarly we can define $A \leq_k B$. This easily (?) defines a partial order on matrices. But know we lose all the data about the eigenvectors? Is there a way to bring it back? Is there some nice interpretation.

- One would like to get such order with restrictions. Maybe this is related to sectional curvature.

- What happens if $n = 2$, $A, B \in \mathcal{H}$, and $\mathrm{tr}(B) = 0$ and $\mathrm{tr}(AB) \geq 0$. What can be said about the relation between A and $A + B$.

- We have up to first order that if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of A, with respective eigenvectors v_i, then we should have

$$\sum_{i=1}^{k} \langle \dot{A} v_i, v_i \rangle \geq 0,$$

  for any $1 \leq k \leq n$ with equality for $k = n$.

- Is the right condition something like: for any $t \in \mathbb{R}$ we should have $A \cdot \chi_{(t,\infty)} \leq B \cdot \chi_{(t,\infty)}$ or something like that.

- Does the following work? We say that $A \leq_2 B$ if for any orthonormal basis $(e_i)_{i=1}^n$ we have

$$(\langle Ae_i, e_i \rangle)_{i=1}^n \prec_2 (\langle Be_i, e_i \rangle)_{i=1}^n.$$

  Does this correspond to the case $n = 1$? This probably doesn't work: if $n = 2$, $\mathrm{tr}(A) = \mathrm{tr}(B) = 0$ and e_1 is in Kernel of B, then the right-hand sequence is zero sequence.

- Lorenz order?

- BMV-conjecture (theorem)

- Proof difficulties

- Proof "sketch" (as in joke)

- Positive linear functions $\mathcal{H} \to \mathbb{R}$.
- What about positive linear functionals form $\mathcal{H}^n \to \mathcal{H}^m$?
- Power series for positivity of inverse function.
- Two notions of positivity: spectral and quadratic form. First works well with functional calculus and second with linear phenomena, but one shouldn't mix these two things.

Chapter 3

k-tone functions

3.1 Motivation

As mentioned in the introduction, k-tone functions should correspond to the functions with non-negative k'th derivative. What should this mean?

We already know the perfect answer for the case $k = 1$: 1-tone functions should be the increasing functions.

Theorem 3.1. *Let $f : (a, b) \to \mathbb{R}$ be differentiable. Then f is increasing, if and only if $f'(x) \geq 0$ for every $x \in (a, b)$.*

Proof. If f is increasing, then all its divided differences, i.e. the quotients of the form

$$\frac{f(x) - f(y)}{x - y}$$

for $x \neq y$ are non-negative. As derivatives are limits of such quotients, also they are non-negative at any point. Conversely, by the mean value theorem for every $x \neq y$ we may find ξ such that

$$\frac{f(x) - f(y)}{x - y} = f'(\xi).$$

Now if the derivatives are non-negative, so are the divided differences, so the function is increasing. □

While this proof by the mean value theorem works in more general setting, if $f \in C^1$, one has more instructive proof.[1]

[1] Of course, the following argument would also work with slightly weaker assumptions, but that's not important to us.

Alternate proof for the theorem 3.1 (in the case $f \in C^1(a, b)$). Note that if $f \in C^1(a, b)$, we may write

$$\frac{f(y) - f(x)}{y - x} = \frac{1}{y - x} \int_x^y f'(t)dt = \int_0^1 f'(tx + (1 - t)y)dt.$$

Note that on the right-hand side one we have average of the derivative over the interval. This means that the claim can be translated to: continuous function is non-negative, if and only if its averages over all intervals are non-negative. But this is clear. □

This is really powerful point of view. While one would like to say the increasing functions are the functions with non-negative derivative, that's a bit of a lie. Instead, one can say that they are the functions whose derivative is non-negative average, and all the problems are gone. This should rougly mean that the derivative defines a positive distribution and it is hence a measure. Thus all increasing functions should be integrals of a positive measure (at least almost everywhere). Although this kind of thinking could be carried out, the details aren't important for us. The main point is that one should that think increasing functions, i.e. the 1-tone functions are functions whose first derivative is a (positive) measure. The divided differences are an averaged (i.e. weak) way of talking about the positivity of the derivative (measure).

This is essentially distributional way of thinking, and we could keep going and end up with the whole business of weak derivatives and stuff. But we don't have to: the plain averages suffice. We write

$$[x, y]_f := \frac{f(x) - f(y)}{x - y},$$

and say that $[\cdot, \cdot]_f$ is the (first) divided difference of f. The domain of $[\cdot, \cdot]_f$ should naturally be $(a, b)^2$ minus the diagonal. And of course, if $f \in C^1$, we should extend $[\cdot, \cdot]_f$ to the diagonal, as the derivative. Divided differences then becomes a continuous function on the whole set $(a, b)^2$. Aside from capturing the first derivative, divided difference has two rather convenient properties.

- For given x and y, $f \mapsto [x, y]_f$ defines a linear map, which is continuous if the domain ($\mathbb{R}^{(a,b)}$) has any reasonable topology (any topology finer than the topology of pointwise convergence, i.e. the product topology will do).

- Divided differences are local in the sense that if f and g agree on $\{x, y\}$, divided differences agree; this observation readily implies the previous continuity claim.

These are the ways the divided differences are compromise between real derivative and the full distributional view. The first point says that one doesn't have worry too much, only about pointwise convergence, while the second says that things are still rather concrete (and it makes the life whole lotta easier).

Now, the real power of this approach comes with larger k. What about the case $k = 2$? Again, we already know the perfect answer: 2-tone functions should be the convex functions.

Theorem 3.2. *Let $f : (a, b) \to \mathbb{R}$ be twice differentiable. Then f is convex, if and only if $f^{(2)}(x) \geq 0$ for every $x \in (a, b)$.*

Proof. While the result is true as stated, let us only proof the case $f \in C^2(a, b)$ (we'll come back to the more general case). Recall that f is convex, if and only if for any $x, y \in (a, b)$ and $t \in [0, 1]$ we have

$$tf(x) + (1 - t)f(y) \geq f(tx + (1 - t)y).$$

This suggest that we may write

$$tf(x) + (1 - t)f(y) - f(tx + (1 - t)y) = \int_x^y w(t)f^{(2)}(t)dt$$

for some weight w. Note that if we manage to find such weight, which is non-negative (and positive enough), we would be done.

How to find the weight w? The idea is rather simple: we want to choose f such that $f^{(2)} = \delta_a$ for $a \in \mathbb{R}$ (in some sense). Now, this should mean that $f(t) = (t - a)_+ + ct + d$ for some $c, d \in \mathbb{R}$, where we write $t_+ = \max(t, 0)$. Plugging this is on the left hand side we get

$$t(x - a)_+ + (1 - t)(y - a)_+ - (tx + (1 - t)y - a)_+ = w(a).$$

TODO: picture

Now, while the steps taken might have contained some leaps of faith, it can be easily verified with partial integration that the given w really works. □

The giveaway is that while the divided differences are a convenient averaged way to talk about first derivative, the quantity $tf(x)+(1-t)f(y)-f(tx+(1-t)y)$ is a convenient averaged way to talk about the second derivative. It captures the fact that the second derivative should be a positive measure – without talking about derivatives. We won't call the quantity the second divided difference, however, as, as it turns out, we can rewrite it in much more convenient form.

If we denote $z = tx + (1-t)y$, we can solve that $t = \frac{z-y}{x-y}$ and express

$$\begin{aligned}
& tf(x) + (1-t)f(y) - f(tx + (1-t)y) \\
= \quad & \frac{z-y}{x-y}f(x) + \frac{x-z}{x-y}f(y) - f(z) \\
= \quad & -(z-y)(z-x)\left(\frac{f(x)}{(x-y)(x-z)} + \frac{f(y)}{(y-z)(y-x)} + \frac{f(z)}{(z-x)(z-y)}\right)
\end{aligned}$$

If $t \notin \{0, 1\}$, $-(z-y)(z-x)$ is positive, so if f is convex,

$$\frac{f(x)}{(x-y)(x-z)} + \frac{f(y)}{(y-z)(y-x)} + \frac{f(z)}{(z-x)(z-y)} \geq 0$$

for any x, y and z such that z is between x and y. This new expression is symmetric in its variables, so actually there's no need to assume anything on the fo x, y and z, just that they're distinct. We can also easily carry this argument to the other direction. If this expression is non-negative any distinct x, y and z, f is convex. This motivates us to define

$$[x, y, z]_f := \frac{f(x)}{(x-y)(x-z)} + \frac{f(y)}{(y-z)(y-x)} + \frac{f(z)}{(z-x)(z-y)},$$

the second divided difference of f.

One would naturally expect that by setting

$$[x_0, x_1, \ldots, x_n]_f := \sum_{i=0}^{n} \frac{f(x_i)}{\prod_{j \neq i}(x_i - x_j)},$$

one obtains something that naturally generalizes divided differences for higher orders. This is indeed the case.

3.2 Divided differences

For $n \geq 1$ define $D_n = \{x \in \mathbb{R}^n | x_i = x_j \text{ for some } 1 \leq i < j \leq n\}$.

Definition 3.3. Let $n \geq 0$. For any real function $f : (a, b) \to \mathbb{R}$ we define the corresponding n'th divided difference $[\cdots]_f : (a, b)^{n+1} \setminus D_{n+1}$ by setting

$$[x_0, x_1, \ldots, x_n]_f = \sum_{i=0}^{n} \frac{f(x_i)}{\prod_{j \neq i}(x_i - x_j)}.$$

We have the following important properties.

Proposition 3.4. *Divided differences are symmetric in the variable, i.e. for any* $f : (a,b) \to \mathbb{R}$ *and pairwise distinct* $a < x_0, x_1, \ldots, x_n < b$ *permutation* $\sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ *we have*

$$[x_1, x_2, \ldots, x_n]_f = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}]_f.$$

Also, if f *is continuous, so is the divided difference. Finally, for fixed (pairwise distinct)* $a < x_0, x_1, \ldots, x_n < b$ *the map* $[x_0, x_1, \ldots, x_n]_\cdot : \mathbb{R}^{(a,b)} \to \mathbb{R}$ *is linear and continuous (when the product is equipped with the product topology).*

Proof. Easy to check. □

The name "divided differences" stems from the fact that the higher order divided differences are itself divided differences of lower order ones.

Proposition 3.5. *For any* $f : (a,b) \to \mathbb{R}$ *and pairwise distinct* $x_0, x_1, \ldots, x_n \in (a,b)$ *we have*

$$[x_0, x_1, \ldots, x_n]_f = \frac{[x_0, x_1, \ldots, x_{n-1}]_f - [x_1, x_2, \ldots, x_n]_f}{x_0 - x_n} = [x_0, x_n]_{[\cdot, x_1, \ldots, x_{n-1}]_f} \tag{3.6}$$

More generally, for any pairwise distinct $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in (a,b)$ *we have*

$$[y_1, y_2, \ldots, y_m]_{[\cdot, x_1, x_2, \ldots, x_n]_f} = [y_1, y_2, \ldots, y_m, x_1, x_2, \ldots, x_n]_f. \tag{3.7}$$

Proof. The simpler case is easy to check directly. For more general case note that both

$$[y_1, y_2, \ldots, y_m]_{[\cdot, x_1, x_2, \ldots, x_n]_f} \text{ and } [y_1, y_2, \ldots, y_m, x_1, x_2, \ldots, x_n]_f$$

satisfy the simpler case (as a function of the y's) and they agree when $m = 1$. □

We call 3.7 the *nesting property* of divided differences. Although the analogy isn't perfect, one could think that this identity says that m'th derivative of the n'th derivative is the $(n+m)$'th derivative.

The following observation tells us that the divided differences work as n'th derivative insomuch that it kills polynomials of degree less than n and works with degree n as expected.

Proposition 3.8. *We have* $[x_0, x_1, \ldots, x_n]_{(x \mapsto x^n)} = 1$ *and* $[x_0, x_1, \ldots, x_n]_p = 0$ *for any polynomial of degree at most* $n-1$*. In other words,* $[x_0, x_1, \ldots, x_n]_f$ *is the leading coefficient of the Lagrange interpolation polynomial on pairs* $(x_0, f(x_0)), (x_1, f(x_1), \ldots, (x_n, f(x_n))$.

Proof. As the Lagrange interpolation polynomial of a polynomial of degree at most n on a dataset of $(n+1)$ pairs is the polynomial itself, the second claim readily implies the first. Recall that the Lagrange interpolation polynomial of a dataset $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ is given by

$$\sum_{i=0}^{n} y_i \frac{\prod_{j\neq i}(x - x_j)}{\prod_{j\neq i}(x_i - x_j)},$$

and the leading coefficient of this polynomial is exactly the divided difference. □

Coming back to the original motivation, divided difference enjoys a integral representation for larger n, albeit somewhat more complicated.

Theorem 3.9. *If $f \in C^n(a, b)$, then for any pairwise distinct $a < x_0, x_1, x_2, \dots, x_n < b$ we have*

$$[x_0, x_1, \dots, x_n]_f = \int_{\mathbb{R}} f^{(n)}(t) w(t) dt, \tag{3.10}$$

where

$$w(t) := w_{x_0,x_1,\dots,x_n}(t) = \frac{1}{(n-1)!} \sum_{i=0}^{n} \frac{((x_i - t)_+)^{n-1}}{\prod_{j\neq i}(x_i - x_j)}.$$

In addition, w is non-negative, supported on $[\min(x_i), \max(x_i)]$ and integrates to $(n!)^{-1}$.

Proof. Note that the weight is simply the n'th divided difference of the map $g_{t,n} : x \mapsto \frac{1}{(n-1)!}((x - t)_+)^{n-1}$. This is not very surprising: one should think that w is the function whose n'th derivative is δ_t. Now if $f = \sum_{i=1}^{m} c_i g_{t_i,n}$, by linearity we have

$$\begin{aligned} [x_0, x_1, \dots, x_n]_f &= \sum_{i=1}^{m} c_i [x_0, x_1, \dots, x_n]_{g_{t_i,n}} = \sum_{i=1}^{m} c_i w(t_i) \\ &= \int_{\mathbb{R}} c_i \delta_{t_i} w(t) dt = \int_{\mathbb{R}} f^{(n)}(t) w(t) dt. \end{aligned}$$

While the previous argument could be pushed through, we take safer route. To prove that the formula even makes sense, we should prove the claim on the support. It is clear that w is zero whenever $t \geq \max(x_i)$. If on the other hand $t \leq \min(x_i)$, $w(t)$ is simply a n'th divided difference of the map $x \mapsto \frac{1}{(n-1)!}(x - t)^{n-1}$, which is zero by the proposition 3.8.

We may hence repeatedly partially integrate the right-hand side:

$$\begin{aligned}\int_{\mathbb{R}} f^{(n)}(t)w(t)dt &= \int_{\mathbb{R}} f^{(n-1)}(t)(-1)w'(t)dt \\ &= \int_{\mathbb{R}} f^{(n-2)}(t)w^{(2)}(t)dt \\ &= \dots \\ &= \int_{\mathbb{R}} f^{(1)}(t)(-1)^n w^{(n-1)}(t)dt,\end{aligned}$$

where

$$(-1)^n w^{(n-1)}(t) = \sum_{i=0}^{n} \frac{\chi_{(t,\infty)}(x_i)}{\prod_{j\neq i}(x_i - x_j)}.$$

Note that $w^{(j)}$ is continuous, piecewise C^1, and compactly supported for every $0 \leq j < n-1$, so the partial integration is legitimate. The final step is a easy calculation.

Applying the identity to $x \mapsto x^n$ shows the claim on the integral of w, so it remains to be shown that w is non-negative.

This is very important property. It allows us to conclude that the n'th divided differences are really weighted averages the n'th derivatives. The property is also by no means trivial with large n.

We prove the non-negativity by induction on n. The case $n = 1$ is clear. The idea is rather simple: we should prove that the functions $g_{t,n}$ has non-negative divided differences, which should rougly mean it has non-negative n'th derivative (being δ_t). By the nesting property we have

$$[x_0, x_1, \dots, x_n]_{g_{t,n}} = [x_0, x_1, \dots, x_{n-1}]_{[\cdot, x_n]_{g_{t,n}}}.$$

Now if we could replace $[\cdot, x_n]_{g_{t,n}}$ with the derivative of $g_{t,n}$, which is conveniently $g_{t,n-1}$, we would be done by the induction hypothesis. Note that while these functions aren't the same in general, they agree (up to constant) if $x_n = t$. But if $x_n \neq t$, we can play the same game as before: $[\cdot, x_n]_{g_{t,n}}$ is weighted average of the derivative $g'_{t,n} = g_{t,n-1}$. Indeed, as

$$[\cdot, x_n]_{g_{t,n}} = \int_0^1 g_{t,n-1}(s \cdot + (1-s)x_n)ds,$$

we have

$$[x_0, x_1, \dots, x_n]_{[\cdot, x_n]_{g_{t,n}}} = \int_0^1 [x_0, x_1, \dots, x_{n-1}]_{g_{t,n-1}(s\cdot + (1-s)x_n)}ds,$$

Now since all the divided differences of $g_{t,n-1}$ are non-negative, the same is clearly true for $g_{t,n-1}(s \cdot +(1-s)x_n)$, so we are done. □

TODO: pictures of Peano Kernels

As an very important corollary we get the following.

Theorem 3.11 (Mean value theorem for divided differences)**.** *Let $n \geq 1$ and $f \in C^n(a,b)$. Then for any pairwise distinct $x_0, x_1, \ldots, x_n \in (a,b)$ we have $\min_{0 \leq i \leq n}(x_i) < \xi < \max_{0 \leq i \leq n}(x_i)$ such that*

$$[x_0, x_1, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}. \tag{3.12}$$

Proof. This follows immediately from 3.9.

One can also give a proof using mean value theorem and with slightly weaker assumptions: it suffices to assume that f is n times differentiable.

By linearity and proposition 3.8 it suffices to verify the claim for the case $f(x_i) = 0$ for $0 \leq i \leq n$, or to verify the following claim.

Lemma 3.13. *If f is n times differentiable, and has $n+1$ roots, then $f^{(n)}$ has a root (in the interior of the convex hull of the roots).*

Proof. If f has $n+1$ roots, by the mean value theorem its derivative has n roots (in the interior of the convex hull of the roots of f) and is $(n-1)$ times differentiable. Since the derivative satisfies the same assumptions for $n-1$, the claim follows by induction. □

□

The mean value theorem could be also used to prove the non-negativity of the weight w: if w were somewhere negative, one could construct function with non-negative derivative and negative divided difference, which would contradict 3.12.

As in the case $n = 1$, if for $n > 1$ we can continuously extend divided differences to the set D_{n+1}, we should do that, and we identify the resulting function with the original one. We will later proof that, as expected, this can be done, if and only $f \in C^n(a,b)$. In this case by 3.12 the extesion satisfies

$$[x_0, x_0, \ldots, x_0]_f = \frac{f^{(n)}(x_0)}{n!},$$

which together with 3.6 is enough to describe the divided differences with values of the function and its derivative.

Many of the familiar identities for the derivatives have analogs with divided differences. We won't really need these formulas, but it's nevertheless nice to know that there are such. Also, they are not really more complicated than the derivative counterparts, on the contrary; the author honestly thinks that they are in fact easier to remember. One of the downsides of the divided differences identities is however that they are usually not symmetric with respect to the sequence $x_0, x_1, \ldots, x_n$ anymore. That's life.

Proposition 3.14. *Let n, k, f, g, $f_1, f_2, \ldots, f_k$ and $x_0, x_1, \ldots, x_n$ be such that the following identities make sense.*

(i) (Newton expansion)

$$\begin{aligned} f(x) = [x_0]_f + [x_0, x_1]_f(x - x_0) + [x_0, x_1, x_2]_f(x - x_0)(x - x_1) + \ldots \\ + [x_0, x_1, \ldots, x_n]_f(x - x_0)(x - x_1) \cdots (x - x_{n-1}) \\ + [x, x_0, x_1, \ldots, x_n]_f(x - x_0)(x - x_1) \cdots (x - x_n), \end{aligned} \tag{3.15}$$

in particular, if the points coincide we get the familiar Taylor expansion

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} + [x, x_0, x_0, \ldots, x_0]_f(x - x_0)^n, \tag{3.16}$$

(ii) (Product rule)

$$[x_0, x_1]_{fg} = [x_0]_f[x_0, x_1]_g + [x_0, x_1]_f[x_1]_g.$$

(iii) (Leibniz rule)

$$\begin{aligned} [x_0, x_1, \ldots, x_n]_{fg} = [x_0]_f[x_0, \ldots, x_n]_g + [x_0, x_1]_f[x_1, \ldots, x_n]_g + \ldots \\ + [x_0, x_1, \ldots, x_{n-1}]_f[x_{n-1}, x_n]_g + [x_0, x_1, \ldots, x_n]_f[x_n]_g. \end{aligned} \tag{3.17}$$

More generally

$$[x_0, x_1, \ldots, x_n]_{f_1 f_2 \cdots f_k} = \sum_{0=i_0<i_1<i_2<\ldots<i_{k-1}<i_k=n} \prod_{j=1}^{k} [x_{i_{j-1},\ldots,x_{i_j}}]_{f_j}$$

(iv) (Chain rule)

$$[x_0, x_1]_{f \circ g} = [g(x_0), g(x_1)]_f[x_0, x_1]_g$$

(v) (Faà di Bruno formula)

$$\begin{aligned} &[x_0, x_1, \ldots, x_n]_{f \circ g} \\ &= \sum_{k=1}^{n} \sum_{0=i_0<i_1<i_2<\ldots<i_{k-1}<i_k=n} [g(x_{i_0}), g(x_{i_1}) \ldots, g(x_{i_k})]_f \prod_{j=1}^{k} [x_{i_{j-1},\ldots,x_{i_j}}]_g \end{aligned}$$

Proof sketches. (i) Easy induction using 3.6. Notice that this formula makes it also clear that the divided difference agrees with the degree n coefficient of the interpolating polynomial.

(ii) Easy to check.

(iii) Induction using the product rule (i.e. the case $n = 1$) and the nesting rule 3.7. Alternatively one could write Newton expansions of both f and g with sequences $(x_0, x_1, \ldots, x_n)$ and $(x_n, x_{n-1}, \ldots, x_0)$ and notice that the given sum gives exactly the leading term of the interpolating polynomial of fg. The more general case follows from the case of two functions by induction.

(iv) Easy to check.

(v) A bit tedious induction using the Leibniz rule and 3.6.

□

Remark 3.18. As Taylor expansion lead to Taylor series, one might wonder under which conditions do Newton expansions lead to Newton series. That is, under which conditions for analytic $f : U \to \mathbb{C}$ and a sequence $z_0, z_1, \ldots, z_n, \ldots \in \mathbb{C}$ and $z \in \mathbb{C}$ the following converges

$$\begin{aligned} f(z) = {} & [z_0]_f \\ & + [z_0, z_1]_f (z - z_0) \\ & + [z_0, z_1, z_2]_f (z - z_0)(z - z_1) \\ & + \ldots \\ & + [z_0, z_1, z_2, \ldots, z_n]_f (z - z_0)(z - z_1) \cdots (z - z_{n-1}) \\ & + \ldots \end{aligned}$$

If the points coincide we recover the usual Taylor expansion and from the 3.16 we see that the Taylor series converges on disc $\mathbb{D}(z_0, r)$ if $\left|\frac{f^{(n)}}{n!}\right| \leq C/r^n$ for some $C > 0$. If f is entire, we have such bound for every r and the series converges everywhere. In a similar vein, if f is entire and $Z = (z_i)_{i \geq 0}$ is bounded, also Newton series converges for every $z \in \mathbb{C}$, but if Z is not bounded, series need not converge for any z outside Z.

For other domains the question is more subtle, and it's closely related to logarithmic potentials and subharmonic functions. (?)

3.3 Cauchy's integral formula

Complex analysis offers a nice view on divided differences: if f is analytic, we may interpret divided differences contour integrals.

Lemma 3.19 (Cauchy's integral formula for divided differences)**.** *If γ is a closed counter-clockwise curve enclosing the numbers $x_0, x_1, \ldots, x_n$, we have*

$$[x_0, x_1, \ldots, x_n]_f = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{(z - x_0)(z - x_1) \cdots (z - x_n)} dz.$$

Proof. Easy induction, verifying 3.6, by taking Cauchy's integral formula as a base case. Alternatively, the claim is a direct consequence of the Residue theorem.

There's another rather instructive proof for the statement. Write Newton expansion for f and integrate both sides along γ. We get

$$\begin{aligned}
\frac{1}{2\pi i} \int_\gamma \frac{f(z)}{(z - x_0)(z - x_1) \cdots (z - x_n)} dz = & \frac{1}{2\pi i} \int_\gamma \frac{[x_0]_f}{(z - x_0)(z - x_1) \cdots (z - x_n)} dz \\
& + \frac{1}{2\pi i} \int_\gamma \frac{[x_0, x_1]_f}{(z - x_1) \cdots (z - x_n)} dz \\
& + \frac{1}{2\pi i} \int_\gamma \frac{[x_0, x_1, x_2]_f}{(z - x_2) \cdots (z - x_n)} dz \\
& + \ldots \\
& + \frac{1}{2\pi i} \int_\gamma \frac{[x_0, x_1, x_2, \ldots, x_{n-1}]_f}{(z - x_{n-1})(z - x_n)} dz \\
& + \frac{1}{2\pi i} \int_\gamma \frac{[x_0, x_1, x_2, \ldots, x_n]_f}{(z - x_n)} dz \\
& + \frac{1}{2\pi i} \int_\gamma [z, x_0, x_1, x_2, \ldots, x_n]_f dz
\end{aligned}$$

As $z \mapsto [z, x_0, x_1, x_2, \ldots, x_n]_f$ is analytic, the last integral vanishes. First n integrals vanish also, since the integrands decay at least as $|z|^{-2}$. Finally, the $(n+1)$:th term gives exactly what we wanted. $\square$

If all the points coincide, we get the familiar formula for the n'th derivative. Also, if f is polynomial of degree at most $n - 1$, the integrand decays as $|z|^{-2}$ and hence the divided differences vanish. Also, for $z \mapsto z^n$ one can use the formula to calculate the n'th divided difference with a residue at infinity. Formula is slightly more concisely expressed by writing for a sequence $X = (x_i)_{i=0}^n$ $p_X(x) = \prod_{i=0}^n (x - x_i)$. Also if one shortens $[X]_f = [x_0, x_1, \ldots, x_n]_f$, we have

$$[X]_f = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{p_X(z)} dz.$$

Cauchy's integral formula is a convenient way to think about severel identities.

Example 3.20. We may express the Lagrange interpolation polynomial of a analytic function f and sequence $X = (x_i)_{i=0}^n$ by

$$P_X(x) = \frac{1}{2\pi i}\int_\gamma \frac{p_X(x) - p_X(z)}{x - z}\frac{f(z)}{p_X(z)}dz = [X]_{f[x,\cdot]p_X}.$$

More generally, if some of the points coincide, we get the Hermite interpolation polynomial. In general, if one wants to find a polynomial which vanishes at points TODO

As an another example one can give variant of the proof the Leibniz rule using the ideas from complex analysis.

Alternate proof for the Leibniz rule for the divided differences. Write Newton expansion for f and g with points reversed for g. The rest follows as in the final proof of Theorem (3.19). $\square$

Actually, we are not quite done yet. Cauchy's integral formula only works for analytic functions. We can however extend the prove with the following useful observation.

Lemma 3.21. *Let $n, m \geq 0$. Assume that for some constants $c_{i,j}$ and $a_{i,j} \in (a, b)$ we have*

$$\sum_{\substack{0 \leq i \leq n \\ 1 \leq j \leq m}} c_{i,j} f^{(i)}(a_{i,j}) = 0$$

for every polynomial. Then the numbers $c_{i,j}$ are all zeroes.

Proof. By Hermite interpolation TODO we can find for any pair (i, j) polynomial with $f^{(i)}(a_{i,j}) = 1$ and $f^{(j')}(a_{i',j'})$ for every other pair (i', j'). Consequently $c_{i,j} = 0$ and we have the claim. $\square$

Of course there's nothing really special about the functional being linear, but the point is: if the $F : C^n(a, b) \to \mathbb{R}$ depends only f and it's derivatives up to some fixed order at some finite set of fixed points, then we know F just by knowing the values at polynomials.

Rest of the alternate proof. Note that if we expand the divided differences, we are almost in the situation of the lemma 3.21; now we just have product of two functions instead. Story is the same. $\square$

3.4 k-tone functions

Definition 3.22. $f : (a, b) \to \mathbb{R}$ is called k-tone if for any $X = (x_i)_{i=0}^n$ of distinct points we have

$$[X]_f \geq 0,$$

i.e. the n'th divided difference is non-negative.

As we noticed, 1-tone and 2-tone functions are exactly the monotone increasing and convex functions. The terminology is not very established, and such functions are also occasionally called k-monotone or k-convex.

Mean value theorem for divided differences tells us that C^k k-tone functions are exactly the functions with non-negative k'th derivative. It turns out that this almost true in general case, namely we have the following result.

Theorem 3.23. *Let $f : (a, b) \to \mathbb{R}$ and $k \geq 2$. Then f is k-tone, if and only if $f \in C^{k-2}(a, b)$, $f^{(k-2)}(x)$ is convex.*

We will postpone the proof.

In some sense the further smoothness assumption is not that much of a game changer. It turns out k-tone functions are always k times differentiable in a weak sense (?), and the weak derivative is non-negative.

One can also usually use regularization techniques discussed in ? to reduce a problem about general k-tone functions to smooth k-tone functions. In general:

Philosophy 3.24. One should not worry about smoothness issues.

We will not resort to such sorcery, however, but try to understand the true reasons behind the smoothness.

We denote the space of k-tone functions by on interval (a, b) by $P^{(k)}(a, b)$. k-tone functions the following enjoy the following useful properties.

Proposition 3.25. *For any positive integer k and open interval (a, b) $P^{(k)}(a, b)$ is a closed (under pointwise convergence) convex cone.*

Proof. Convex cone property is immediate form the linearity of divided differences. Also, if $f_i \to f$ pointwise, the respective divided differences converge, so also the closedness is clear. $\square$

Proposition 3.26. *$P^{(k)}$ is a local property i.e. $P^{(k)}(a, b) \cap P^{(k)}(c, d) \subset P^{(k)}(a, d)$ for any $-\infty \leq a \leq c < b \leq d \leq \infty$. To be more precise, if $f : (a, d) \to \mathbb{R}$ such that $f\big|_{(a,b)} \in P^{(k)}(a, b)$ and $f\big|_{(c,d)} \in P^{(k)}(c, d)$, then $f \in P^{(k)}(a, d)$.*

Proof. For $f \in C^k$ the statement is immediate form the mean value theorem. In general we can argue bit similarly as in the case $k = 1$. For $k = 1$ note that if $a < x < c < b < z < d$ we may take $c < y < b$. Now

$$f(z) - f(x) = (f(z) - f(y)) + (f(y) - f(x)) \geq 0.$$

In terms of divided differences

$$[x, z]_f = [z, y]_f \frac{z - y}{z - x} + [x, y]_f \frac{y - x}{z - x}.$$

The point is that we can express divided differences as weighted sums of divided differences of tuples with smaller supports. More generally, if $a < x_0 < \ldots < x_k < d$ with $x_0 < c$ and $d < x_k$, take any $y \in (c, b)$ distinct from x_i's and we have

$$[x_0, \ldots, x_k]_f = [x_1, \ldots, x_k, y]_f \frac{x_k - y}{x_k - x_0} + [x_0, \ldots, x_{k-1}, y]_f \frac{y - x_0}{x_k - x_0}.$$

This identity is easily verified by applying the previous version for the function $x \mapsto [\cdot, x_1, x_2, \ldots, x_{n-1}]_f$. By repeating this process, we will end up with divided differences of tuples completely lying on (a, b) or (c, d). Formally, we could induct on l number of x_i's outside (c, b) and note that if tuple isn't good already, the two new tuples have lower number l. □

3.5 Smoothness

TODO: prove the following with the locally uniform continuity trick.

Theorem 3.27. *Let $f : (a, b) \to \mathbb{R}$ and $n \geq 1$. Then $f \in C^n(a, b)$, if and only if n:th divided difference of f extends to continuous function on $(a, b)^{n+1}$.*

Actually we can prove a slightly better statement. Let $n, m \geq 1$ and and denote

$$D_{n,m} = \{x \in \mathbb{R}^n | x_{i_1} = x_{i_2} = \ldots = x_{i_k} \text{ for some } 1 \leq i_1 < i_2 < \ldots < i_m \leq n\}.$$

Theorem 3.28. *Let $f : (a, b) \to \mathbb{R}$, $0 \leq m \leq n$. Then $f \in C^m(a, b)$, if and only if n:th divided difference of f extends to continuous function to $(a, b)^{n+1} \setminus D_{n+1,m+2}$. Moreover, this extension is unique and satisfies 3.6 and 3.12.*

Proof. We first prove the "only if"-direction by induction on m.

The case $m = 0$ is clear. Now fix $m > 0$. Let us prove the statement for this m by induction on n.

Consider the base case case $n = m$. Take any sequence $a < x_0 \leq x_1 \leq x_2 \ldots \leq x_n < b$. By induction hypothesis for the pair $(n, m-1)$ we can extend the divided difference to this point if $x_0 < x_n$. If $x_0 = x_1 = \ldots = x_n$, we will extend the divided difference as

$$[x_0, x_1, \ldots, x_n] = \frac{f^{(n)}(x_0)}{n!}.$$

But the mean value theorem for the divided difference immediately implies this extension is continuous at $(x_0, x_1, \ldots, x_n)$.

If $n > m$, and $x \in \mathbb{R}^{n+1} \setminus D_{n+1,m+2}$, we necessarily have $x_0 < x_n$ and hence we can extend the divided differences using 3.6. By construction, our extension satisfies 3.6 and 3.12 and since $\mathbb{R}^{n+1} \setminus D_{n+1}$ is dense in $\mathbb{R}^{n+1} \setminus D_{n+1,m+2}$, the extension is necessarily unique.

Let us then prove the "if"-direction. We start with a lemma.

Lemma 3.29. *Let $n \geq 1$ and $m \geq 0$. If the n:th divided difference of f has continuous extension to the set $\mathbb{R}^{n+1} \setminus D_{n+1,m+2}$, then there also is a continuous extension for the $(n-1)$:th divided difference of f to the set $\mathbb{R}^n \setminus D_{n,m+2}$.*

Proof. Take any $x \in \mathbb{R}^n \setminus D_{n,m+2}$. We induct downward on l, the number of distinct components of x. If $l = n$, the statement is clear. Assume then that $l < n$. Now there exist $x' \in \mathbb{R}^n \setminus D_{n,m+2}$ with $l+1$ distinct components, which differs from x by exactly one component, say i'th one. We then extend

$$[x_1, x_2, \ldots, x_n]_f := [x_1, x_2, \ldots, x'_i, \ldots, x_n]_f + (x'_i - x_i)[x_1, x_2, \ldots, x_i, x'_i, \ldots, x_n]_f.$$

Now if $x^{(n)} \to x$, then by comparing both sides also $[x_1, x_2, \ldots, x_n]_f$'s converge, and we have the continuous extension. □

Now let us continue with the proof. We induct on m. The case $m = 0$ is clear. The case $n = m = 1$ is also rather clear, the diagonal $[x, x]_f$ given the derivative of f and continuity of the derivative is implied by the continuity of $[\cdot, \cdot]_f$ along the diagonal.

Lemma 3.30. *If $f \in C^1(a, b)$ and $a < x_1, x_2, \ldots, x_n < b$ are distinct, then*

$$[x_1, x_2, \ldots, x_n]_{f'} = \sum_{1 \leq i \leq n} [x_1, x_2, \ldots, x_{i-1}, x_i, x_i, x_{i+1}, \ldots, x_n]_f$$

Proof. We induct on n: the case $n = 1$ is clear. When $n > 1$, by 3.6 we have

$$\begin{aligned}
[x_1, x_2, \ldots, x_n]_{f'} &= \frac{[x_1, \ldots, x_{n-1}]_{f'} - [x_2, \ldots, x_n]_{f'}}{x_1 - x_n} \\
&= \frac{\left(\sum_{1 \le i \le n-1} [x_1, \ldots, x_i, x_i, \ldots, x_{n-1}]_f\right) - \left(\sum_{2 \le i \le n} [x_2, \ldots, x_i, x_i, \ldots, x_n]_f\right)}{x_1 - x_n} \\
&= \sum_{2 \le i \le n-1} \frac{[x_1, \ldots, x_i, x_i, \ldots, x_{n-1}]_f - [x_2, \ldots, x_i, x_i, \ldots, x_n]_f}{x_1 - x_n} \\
&+ \frac{[x_1, x_1, x_2, \ldots, x_{n-1}]_f - [x_2, \ldots, x_n, x_n]_f}{x_1 - x_n} \\
&= \sum_{2 \le i \le n-1} [x_1, \ldots, x_i, x_i, \ldots, x_n]_f \\
&+ \frac{[x_1, x_1, \ldots, x_{n-1}]_f - [x_1, \ldots, x_n]_f}{x_1 - x_n} + \frac{[x_1, \ldots, x_n]_f - [x_2, \ldots, x_n, x_n]_f}{x_1 - x_n} \\
&= \sum_{1 \le i \le n} [x_1, \ldots, x_i, x_i, \ldots, x_n]_f
\end{aligned}$$

□

Let us then take $2 \le m = n$. By the lemma 3.29 and the case $m = 1$ we see that $f \in C^1(a, b)$. But by the lemma 3.30 $(n - 1)$:th divided differences of f' extend to continuous function to $\mathbb{R}^n$. Hence by the induction hypothesis $f' \in C^{(n-1)}(a, b)$ and hence $f \in C^n(a, b)$.

Finally the lemma 3.29 immediately implies the remaining cases $\le m < n$, by induction on n. □

There is however more interesting equivalence to be made.

Theorem 3.31. *Let $f : (a, b) \to \mathbb{R}$ and $n \ge 1$. Then $f \in C^{n-1}(a, b)$ and $f^{(n-1)}$ is Lipschitz, if and only if n:th divided difference of f is bounded. Moreover,*

$$\sup_{a < x_0 < x_1 < \ldots < x_n < b} |[x_0, x_1, \ldots, x_n]_f| = \frac{\mathrm{Lip}(f^{(n-1)})}{n!}$$

Proof. We first prove the "if"-direction.

The case $n = 1$ is clear. In the case $n = 2$ note that

$$[x, y, z] = \frac{[y, x]_f - [z, x]_f}{y - z}$$

Since this quantity is bounded, it follows that $[\cdot, x]_f$ has a limit at x, which means exactly that f is differentiable at x. Note that in addition for any x, x', y, y' we have

$$[x', x, y]_f + [y', y, x] = \frac{([x', x]_f - [x, y]_f) - ([y', y]_f - [x, y]_f)}{x - y} = \frac{[x', x] - [y, y']}{x - y}.$$

Letting $x' \to x$ and $y' \to y$ we see that f' is Lipschitz and hence also $f \in C^1(a, b)$.

Now for general n we argue by induction on n. Let $n > 2$. Note that since

$$[x_0, x_1, x_2, \ldots, x_n]_f = \frac{[x_0, x_1, \ldots, x_{n-1}]_f - [x_1, \ldots, x_n]_f}{x_0 - x_n}.$$

The map $x \mapsto [x, x_1, x_2, \ldots, x_{n-1}]$ is 1-Lipschitz for any $x_1, x_2, \ldots, x_{n-1}$ and hence $(n-1)$:th divided difference is Lipschitz and consequently bounded. By induction hypothesis f is $C^{n-2}(a, b)$ and hence at least $C^1(a, b)$. But now since

$$[x_1, x_2, \ldots, x_n]_{f'} = \sum_{1 \leq i \leq n} [x_1, x_2, \ldots, x_{i-1}, x_i, x_i, x_{i+1}, \ldots, x_n]_f$$

f' has bounded $(n-1)$:th divided differences, and by the induction hypothesis, $f \in C^{(n-1)}(a, b)$ and $f^{(n-1)}$ is Lipschitz. Induction also immediately gives

$$\frac{\mathrm{Lip}(f^{(n-1)})}{n!} \leq \sup_{a < x_0 < x_1 < \ldots < x_n < b} |[x_0, x_1, \ldots, x_n]_f|.$$

Let us then prove the "only if"-direction. Take any $a < x_0 < x_1 < \ldots < x_n < b$. By the mean-value theorem for divided differences we have

$$\begin{aligned}
|[x_0, x_1, \ldots, x_n]_f| &= \left| \frac{[x_0, x_1, \ldots, x_{n-1}]_f - [x_1, x_2, \ldots, x_n]_f}{x_0 - x_n} \right| \\
&= \frac{1}{(n-1)!} \left| \frac{f^{(n-1)}(\xi_1) - f^{(n-1)}(\xi_2)}{x_0 - x_n} \right| \\
&\leq \frac{\mathrm{Lip}(f^{(n-1)})}{(n-1)!} \left| \frac{\xi_1 - \xi_2}{x_0 - x_n} \right|
\end{aligned}$$

for some $x_0 \leq \xi_1 \leq \xi_2 \leq x_n$. Hence n:th divided difference is bounded, but the inequality is not quite sharp enough.

We can make it sharp with some tricks. Firstly, when considering the supremum, we only need to consider tuples where all but one of entries are equal. Indeed pick any $a < x_0 < x_1 < x_2 < \ldots < x_n$. Now consider the map $g(x) = [x, x_0]_f$. This is $C^{n-1}(x_0, b)$ so by the mean-value theorem we have

$$[x_0, x_1, \ldots, x_n]_f = [x_1, x_2, \ldots, x_n]_g = \frac{g^{(n-1)}(\xi)}{(n-1)!} = [\xi, \xi, \ldots, \xi]_g = [x_0, \xi, \xi, \ldots]_f.$$

Hence it suffices to consider only tuples $(x, x, x, \ldots, y)$ where x appears n times. Now by the Taylor's theorem

$$
\begin{aligned}
|[x, x, \ldots, y]_f| &= \left| \frac{[x, x, \ldots, x]_f - [x, x, \ldots, y]}{x - y} \right| \\
&= \left| \frac{\frac{f^{(n-1)}(x)}{(n-1)!} - \int_x^y \frac{f^{(n-1)}(t)(y-t)^{n-2}}{(n-2)!} dt}{x - y} \right| = \\
&= \frac{1}{(n-2)!|y-x|^{n-1}} \left| \int_x^y (f^{(n-1)}(x) - f^{(n-1)}(t))(y-t)^{n-2} dt \right| \\
&\leq \frac{\mathrm{Lip}(f^{(n-1)})}{(n-2)!} \int_x^y |x-t||y-t|^{n-2} dt \\
&= \frac{\mathrm{Lip}(f^{(n-1)})}{n!},
\end{aligned}
$$

and we are done. □

With such tools we are ready to tackle the regularity of k-tone functions

Proof of the theorem 3.23. We start with a lemma.

Lemma 3.32. *If $k \geq 1$ and $f : (a, b) \to \mathbb{R}$ is k-tone, then the $(k-1)$:th divided differences of f are locally bounded, i.e. bounded on every closed subinterval of (a, b).*

Proof. We induct on k. The case $k = 1$ is rather clear: for any $a < c < x < d < b$ we have $f(x) \in [f(c), f(d)]$. Take then $k > 1$ and any closed interval $[c, d] \subset (a, b)$. Take $a < x_0 < c$. The map $g = [\cdot, x_0]_f$ is $(k-1)$-tone, so by induction hypothesis the $(k-2)$:th divided differences of g are bounded on $[c, d]$. Now for any $c \leq x_1 < x_2 < \ldots < x_k \leq d$ we have

$$[x_1, x_2, \ldots, x_k]_f = (x_k - x_0)[x_0, x_1, \ldots, x_k]_f + [x_0, x_1, \ldots, x_{k-1}]_f \geq [x_0, x_1, \ldots, x_{k-1}]_f$$

But $[x_0, x_1, \ldots, x_{k-1}]_f = [x_1, \ldots, x_{k-1}]_g$ is bounded, so we have lower bound for $(k-1)$:th divided differences of f. Similarly, by taking $d < x_0 < b$ we get upper bound, and we are done. □

Combining the lemma with theorem 3.31 gives the right smoothness. Convexity condition is implied by the lemma 3.30 combined with induction on k. □

3.6 Analyticity and Bernstein's theorems

By requiring (some kind of) regularity for the divided differences of all orders, occasionally we get more than smoothness, namely analyticity. Most basic result of this kind is the following.

Theorem 3.33. *Let $f : (a, b) \to \mathbb{R}$. Then f is real analytic, if and only if for every closed subinteval $[c, d]$ of (a, b) there exists constant C such that for any $n \geq 1$*

$$\sup_{a<x_0<x_1<\ldots<x_n<b} |[x_0, x_1, \ldots, x_n]_f| \leq C^{n+1}.$$

Proof. Let's first prove that "if"-direction. We need to prove that the for any $x_0 \in (a, b)$ Taylor series at x_0 converges in some neighbourhood of x_0. As observed before, the n:th error term in Taylor series is given by

$$[x, x_0, x_0, \ldots, x_0]_f (x - x_0)^n$$

with n x_0's. Now choose $a < c < x_0 < d < b$ and take any x with $x \in [c, d]$ and $|x - x_0|C < 1$, where C is given by the assumption for interval c, d. But then the error term tends to zero and we are done.

For the other direction note that if $x_0 \in (a, b)$ and f extends to analytic function on $\mathbb{D}(x_0, r)$, we definititely have $\left|\frac{f^{(n)}(x_0)}{n!}\right| \leq C^{n+1}$ for some C. If $|x - x_0| < r$ we have

$$\frac{f^{(k)}(x)}{k!} = \sum_{n=k}^{\infty} \binom{n}{k} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^{n-k},$$

which may be estimated by

$$\left|\frac{f^{(k)}(x)}{k!}\right| \leq C^{k+1} \sum_{n=k}^{\infty} \binom{n}{k} C^{n-k} (x - x_0)^{n-k} = \frac{C^{k+1}}{(1 - |x - x_0|C)^k},$$

whenever $|x - x_0|C < 1$. By the mean value theorem for divided differences it follows that we get required bound for some neighbourhood of x_0 and consequently, by compactness for any closed subinteval of (a, b). □

Of course, we could just as well replace the closed inteval by any compact subset of (a, b). The previous result is some kind of relative of 3.31. Also theorem 3.23 has rather interesting relative.

Theorem 3.34 (Bernstein's little theorem)**.** *If $f : (a, b) \to \mathbb{R}$ is k-tone for every $k \geq 0$, then f is real-analytic on (a, b).*

Proof. We prove that the conditions of the theorem 3.33 are satisfied. Pick any $a < x_0 < x < b$. Now for any $n \geq 0$ we have

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + [x, x_0, x_0, \ldots, x_0]_f (x - x_0)^{n+1}.$$

Note that all the terms on the right-hand side are non-negative, and hence

$$0 \leq \frac{f^{(n)}(x_0)}{n!} \leq f(x)(x - x_0)^{-n}.$$

Now given any interval $[c, d] \subset (a, b)$ we can make such estimate uniform over $x_0 \in [c, d]$ simply by picking $x \in (d, b)$, and we are done. □

We could further conclude that f in the previous theorem extends to a complex analytic function to $\mathbb{D}(a, |b - a|)$.

Usually by Bernstein's little theorem one means slightly weaker statement: if $f : (0, \infty) \to \mathbb{R}$ is smooth such that $(-1)^n f^{(n)}(t) \geq 0$ for any $t > 0$, then f extends to analytic function to right half-plane. This is readily implied by the previous observation. Latter version has however a considerable strenghtening.

Theorem 3.35 (Bernstein's big theorem)**.** *If $f : (0, \infty) \to \mathbb{R}$ is smooth such that $(-1)^n f^{(n)}(t) \geq 0$ for any $t > 0$, then f is Laplace transform of a radon measure μ on $[0, \infty)$, that is we have*

$$f(x) = \int_0^\infty e^{-xt} d\mu(t)$$

for every $x > 0$.

We will postpone the proof.
TODO:

- Mean value theorem, coefficient of the interpolating polynomial
- Basic properties, product rule.
- k-tone functions, smoothness, and representation
- Majorization, Jensen and Karamata inequalities, generalizations, and corollaries concerning spectrum and trace functions. Schur-Horn conjectures and Honey-Combs
- Tohoku contains nice proof of Lidskii inequality

- How to understand the inequalities arising from k-tone functions: is there nice way to parametrize the tuples coming from the k-majorization.
- For $k = 3$ and 3 numbers, it's all about the biggers number: one with the largest largest number dominates.
- The previous probably generalizes: for k-tone functions and k numbers on both sides, with all polynomials of degree less than k vanishing on both tuples, one with largest largest value dominates, or equivalently, it's all about the constant term. This is clearly necessary, by is it also sufficient? Should be: express the whole thing as an integral, differentiate with respect to the constant term, and finally interpret as a divided difference.
- What if we add more terms: is there simple characterization? Why have similar integral representation, and can probably differentiate: Maybe not, or one has to be really careful. Is there characterization with linear inequalities (in addition to the equalities)?
- Peano Kernels: Smoothness properties, Bernstein (?) polynomials as examples.
- Opitz formula
- Regularization techniques
- Notion of midpoint-convexity should generalize by regularization techniques.
- Should Legendre transform generalize to higher orders? For smooth enough functions probably with derivatives being inverses of each other, but what is the correct definition? And is it of any use? Maybe differentiating $k-2$ times and then having similar characterization. Is there higher order duality?
- Is there elementary transformations for k-tone Karamata?
- Divided-difference series for entire functions (Newton expansion)? For analytic function? When does it converge? When does it converge to the right function?
- Given domain $U \subset \mathbb{C}$ and analytic function $f : U \to \mathbb{C}$, determine all subsets $V \subset U$ such that there exists Newton series with some sequence $x_1, x_2, \ldots$ converging in V. This is very much related to logarithmic potentials and subharmonic functions: sequence, if say bounded for starters, corresponds to a radon measure. Indeed, take weak limit of radon measures averaged exprimental measures of first elements in the sequence, if the limit exists (if not...). Now if $f = \frac{1}{z}$ for starters, we have the logarithmic potential $U(z)$ and the Newton series converges whenever $U(z) < U(0)$.

- Harnack-type inequalities for derivatives of Pick functions?
- Smooth function is in $P(0, \infty)$ if it's negative of Laplace transform of Laplace transform of a measure on $[0, \infty)$?
- Are there better bounds for theorem 3.31?

Chapter 4

Matrix functions

4.1 Functional calculus

Definition 4.1. For any $-\infty \leq a < b \leq \infty$ $f : (a, b) \to \mathbb{R}$ the associated matrix function on V is the map $f_V : \mathcal{H}_{(a,b)}(V) \to \mathcal{H}(V)$ given by

$$f_V(A) = \sum_{\lambda \in \mathrm{spec}(A)} f(\lambda) P_{E_\lambda}$$

if $A = \sum_{\lambda \in \mathrm{spec}(A)} \lambda P_{E_\lambda}$.

Hence to calculate the matrix function we just apply the function to the eigenvalues of the map and leave the eigenspaces as they are. Note as the spectral representation is unique this definition makes sense.

We have already discussed four types of matrix functions: inverse, polynomials, square root and absolute value. All these notion coincide with the general notion of matrix function for real maps, as notion in (2.17) and TODO.

Matrix functions enjoy many natural and useful properties.

Proposition 4.2. *Let $f : (a, b) \to \mathbb{R}$ and $A \in \mathcal{H}_{(a,b)}$*

1. *If $f[(a, b)] \subset (c, d)$ then $f_V(A) \in \mathcal{H}_{(c,d)}$.*
2. *If also $g : (a, b) \to \mathbb{R}$ then $(f + g)_V = f_V + g_V$ and $(fg)_V = f_V g_V$.*
3. *$f_{V_1 \oplus V_2} = f_{V_1} \oplus f_{V_2}$.*
4. *If $g : (a, b) \to \mathbb{R}$ and f and g agree on spectrum of A, then $f(A) = g(A)$.*
5. *If $f[(a, b)] \subset (c, d)$ and $g : (c, d) \to \mathbb{R}$ then $(g \circ f)_V = g_V \circ f_V$.*

6. *If $f_n : (a,b) \to \mathbb{R}$ converge pointwise to f, then the same holds true for $(f_n)_V$'s.*

These properties make it clear that such definition is natural. We will drop the subscript V and identify f with its matrix function f_V if V is clear from context.

4.2 Holomorphic functional calculus

If f is entire, there's another way to appoach matrix functions. As f can be written as

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

power series convergent whole any $z \in \mathbb{C}$, we should have

$$f_V(A) = \sum_{n=0}^{\infty} a_n A^n.$$

This matrix power series indeed converges as $\|A^n\| \leq \|A\|^n$. Also, this definition coincides with the spectral one. Indeed, if one writes $f_n : z \mapsto \sum_{k=0}^{n} a_n z^n$, then we have, by definition,

$$\sum_{n=0}^{\infty} a_n A^n = \lim_{n\to\infty} [(f_n)_V(A)] = f_V(A),$$

by point (6) of proposition (4.2).

Note however that the power series definition makes perfect sense even if $a_n \notin \mathbb{R}$ and even better, A need not be real.

If f is not entire, the power series might not converge every $A \in \mathcal{H}_{(a,b)}(V)$. Instead, we can more generally use Cauchy's integral formula for matrix functions.

$$f_V(A) = \int_\gamma (zI - A)^{-1} f(z) dz,$$

where γ is simple closed curve enclosing the spectrum of A. This formula is immediate when viewed in a eigenbasis of A. Again, this formula makes perfect sense even for non-real A, given that spectrum of A lies in the domain of f.

4.3 Derivative of a matrix function

If f is analytic, for suitable γ we have

$$\begin{aligned} f(B) - f(A) &= \frac{1}{2\pi i}\int_\gamma (zI - B)^{-1} f(z)dz - \int_\gamma (zI - A)^{-1} f(z)dz \\ &= \frac{1}{2\pi i}\int_\gamma (zI - B)^{-1}(B - A)(zI - A)^{-1} f(z)dz. \end{aligned}$$

Writing $B = A + tH$, and letting $t \to 0$ we get

$$\begin{aligned} \lim_{t\to 0} \frac{f(A + tH) - f(A)}{t} &= \lim_{t\to 0}\int_\gamma (zI - A - tH)^{-1} H(zI - A)^{-1} f(z)dz \\ &= \int_\gamma (zI - A)^{-1} H(zI - A)^{-1} f(z)dz. \end{aligned}$$

Derivative of f at A is hence the linear map

$$H \mapsto \int_\gamma (zI - A)^{-1} H(zI - A)^{-1} f(z)dz.$$

If we write everything in the eigenbasis of A, $A = (\lambda_i \delta_{i,j})_{1\le i,j\le n}$ and $H = (H_{i,j})_{1\le i,j\le n}$, we have

$$\begin{aligned} \int_\gamma (zI - A)^{-1} H(zI - A)^{-1} f(z)dz &= \left(H_{i,j}\int_\gamma (z - \lambda_i)^{-1}(z - \lambda_j)^{-1} f(z)dz\right)_{1\le i,j\le n} \\ &= (H_{i,j}[\lambda_i, \lambda_j]_f)_{1\le i,j\le n} \\ &= H \circ ([\lambda_i, \lambda_j]_f)_{1\le i,j\le n}. \end{aligned}$$

Here $\circ$ denotes the Hadamard product of matrices, given by $(A \circ B)_{i,j} = A_{i,j} \circ B_{i,j}$.

This formula holds even if f is not analytic, namely as long as $f \in \mathbb{C}^1(a, b)$. Indeed, by polynomial interpolation it is sufficient to prove the following lemma.

Lemma 4.3. *If $f \in C^1(a, b)$, $A \in \mathcal{H}_{(a,b)}$ such that $f(\lambda_i) = 0 = f'(\lambda_i)$ for $1 \le i \le n$, then*

$$\|f(A + H)\| = o(\|H\|).$$

Proof. By lemma 2.55 we see that $|\lambda_i(A+H)-\lambda_i(A)| \le n\|H\|$. Now by Taylor expansion $|f(\lambda_i(A+H))| = n\|H\|f'(\xi_i)$ for some $\xi_i = \xi_{i,A,H}$ on $(\lambda_i(A)-n\|H\|, \lambda_i(A)+n\|H\|)$. Now

$$\|f(A + H)\| \le \sum_{i=1}^{n} |\lambda_i(A + H)|\|P_{v_i}\| \le n\|H\| \sum_{i=1}^{n} |f'(\xi_{i,A,H})|.$$

But by continuity of f' the sum tends to zero as $\|H\| \to 0$. $\square$

Derivative of matrix function can be also approached via power series. Since

$$(A+tH)^k = \sum_{j=0}^{k} t^j \sum_{\substack{i_0,i_1,\ldots,i_j\geq 0\\ i_0+i_1+\ldots+i_j=k-j}}^{k-1} A^{i_0}HA^{i_1}H\cdots HA^{i_j},$$

we have

$$\lim_{t\to 0}\frac{(A+tH)^k - A^k}{t} = \sum_{j=0}^{k-1} A^j H A^{k-1-j}.$$

With the same notation as before, we have for eigenbasis of A

$$\sum_{j=0}^{k-1} A^j H A^{k-1-j} = \sum_{j=0}^{k-1}\left(\lambda_k^j H_{k,l}\lambda_l^{k-1-j}\right)_{1\leq k,l\leq n} = \left(H_{k,l}[\lambda_k,\lambda_l]_{(x\mapsto x^k)}\right)_{1\leq k,l\leq n}.$$

Summing this identity over the Taylor terms yields the derivative formula for entire functions.

4.4 Higher derivatives

TODO

TODO:

- Basic definition
- Equivalent definitions
- Continuity properties
- Examples
- Calculating with matrix functions
- Smoothness properties, derivative formulas, Hadamard product
- Cauchy's integral formula
- Jordan block formula
- How to extend functions $f : (a,b)^2 \to \mathbb{R}^2$ to a matrix function taking two entries? What is $f(A,B)$? If A and B commute, there exists $h_A, h_B : \mathbb{R} \to (a,b)$, $C \in \mathcal{H}$ such that $h_A(C) = A$ and $h_B(C) = B$ and we should hence define $f(A,B) = f(h_A(C), h_B(C))$. What about the general case?

Chapter 5

Pick-Nevanlinna functions

Pick-Nevanlinna function is an analytic function defined in upper half-plane with a non-negative real part. Such functions are sometimes also called Herglotz or $\mathbb{R}$ functions but we will call them just *Pick functions*. The class of Pick functions is denoted by $\mathcal{P}$.

Pick functions have many interesting properties related to positive matrices and that is why they are central objects to the theory of matrix monotone functions.

5.1 Basic properties and examples

Most obvious examples of Pick functions might be functions of the form $\alpha z + \beta$ where $\alpha, \beta \in \mathbb{R}$ and $\alpha \geq 0$. Of course one could also take $\beta \in \overline{\mathbb{H}}_+$. Actually real constants are the only Pick functions failing to map $\mathbb{H}_+ \to \mathbb{H}_+$: non-constant analytic functions are open mappings.

Sum of two Pick functions is a Pick function and one can multiply a Pick function by non-negative constant to get a new Pick function. Same is true for composition.

The map $z \mapsto -\frac{1}{z}$ is evidently a Pick function. Hence are also all functions of the form

$$\alpha z + \beta + \sum_{i=1}^{N} \frac{m_i}{\lambda_i - z},$$

where N is non-negative integer, $\alpha, m_1, m_2, \ldots, m_N \geq 0$, $\beta \in \mathbb{H}_+$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{H}_+$. So far we have constructed our function by adding simple poles to the closure of lower half-plane. We could further add poles of higher order at lower half plane, and change residues and so on, but then we have to be a bit more careful.

There are (luckily) more interesting examples: all the functions of the form x^p where $0 < p < 1$ are Pick functions. To be precise, one should choose branch for the previous so that they are real at positive real axis. Also log yields Pick function when branch is chosen properly i.e. naturally again. Another classic example is tan. Indeed, by the addition formula

$$\begin{aligned}\tan(x+iy) &= \frac{\tan(x)+\tan(iy)}{1-\tan(x)\tan(iy)} = \frac{\tan(x)+i\tanh(y)}{1-i\tan(x)\tanh(y)} \\ &= \frac{\tan(x)(1+\tanh^2(y))}{1+\tan^2(x)\tanh^2(y)} + i\frac{(1+\tan^2(x))\tanh(y)}{1+\tan^2(x)\tanh^2(y)},\end{aligned}$$

and y and $\tanh(y)$ have the same sign.

We observe the following useful fact.

Proposition 5.1. *If $(\varphi_i)_{i=1}^{\infty}$ is a sequence of Pick functions converging locally uniformly, the limit function is also a Pick function.*

Proof. Locally uniform limits of analytic functions are analytic. Also the limit function has evidently non-negative imaginary part. □

This is one of the main reasons we include real constants to Pick functions, although they are exceptional in many ways. Note that for any $z \in \mathbb{H}_+$ we have $\log(z) = \lim_{p\to 0^+}(z^p - 1)/p$: log can be understood as a limit of Pick functions.

There's a considerable strengthening of the previous result.

Proposition 5.2. *If $(\varphi_i)_{i=1}^{\infty}$ is a sequence of Pick functions converging pointwise, the limit function is also a Pick function.*

Note that this result far from clear: pointwise limits of analytic functions need not in general be analytic. We will not prove this result yet, but it strongly suggests that the class of Pick functions is very rigid in some sense.

5.2 Rigidity

5.2.1 Schur functions

To understand the rigidity phenomena we take look at the close relative to Pick functions, *Schur functions.* Schur funtions are analytic maps from open unit disc to closed unit disc. These functions functions include for instance power functions $z \mapsto z^n$ and more generally, as one may check, any Blaschke products, that is products of the terms of the form

$$\rho_{a,\omega}(z) = \omega\frac{a-z}{1-\overline{a}z},$$

Blaschke factors. Classic fact about these functions is the Schwarz lemma.

Theorem 5.3 (Schwarz lemma)**.** *Let $\psi : \mathbb{D} \to \mathbb{D}$ be analytic such that $\psi(0) = 0$. Then $|\psi(z)| \leq |z|$ for any $z \in \mathbb{D}$ and hence also $|\psi'(0)| \leq 1$. If $|\psi(z)| = |z|$ for some $z \in \mathbb{D} \setminus \{0\}$ or $|\psi'(0)| = 1$, $\psi(z) = \omega z$ for some $\omega \in \mathbb{S}$.*

Fixing value of Schur function at 0 restricts function a whole lot.

The usual proof is by cleverly using maximum modulus principle for $\psi(z)/z$. Maximum modulus principle itself is consequence of the Cauchy's integral formula. There's also more symmetric form for Schwarz lemma, called Schwarz-Pick theorem.

Theorem 5.4 (Schwarz-Pick theorem)**.** *Let $\psi : \mathbb{D} \to \mathbb{D}$ be analytic. Then for any $z_1, z_2 \in \mathbb{D}$ we have*

$$\left| \frac{\psi(z_1) - \psi(z_2)}{1 - \overline{\psi(z_1)}\psi(z_2)} \right| \leq \left| \frac{z_1 - z_2}{1 - \overline{z_1} z_2} \right|,$$

and

$$\frac{|\psi'(z_1)|}{1 - |\psi(z_1)|^2} \leq \frac{1}{1 - |z_1|^2}.$$

If the equality holds in one of the inequalities, ψ is an Blaschke factor.

Note that one obtains the usual Schwarz lemma if $z_1 = 0 = \psi(z_1)$. One may check that if ψ is Blaschke factor, the inequalities hold as equalities.

Proof. Consider the map $\psi_1 = \rho_{\psi(z_1)} \circ \psi \circ \rho_{z_1}$. The claim follows by using the previous form of the Schwarz lemma for the ψ_1 and point z_2. □

The previous result shows that Blaschke factors are exactly the analytic bijections $\mathbb{D} \to \mathbb{D}$.

One can also make weaker estimates straight from Cauchy's integral formula. We can for instance write

$$\psi(a) = \frac{1}{2\pi i} \int_\gamma \frac{\psi(z)}{z - a} dz$$

for suitable γ. Now if ψ extends over unit circle, letting γ trace unit circle we have

$$|\psi(a)| \leq \left(\max_{|z|=1} |\psi(z)| \right) \frac{1}{2\pi} \int_0^{2\pi} \frac{dt}{|e^{it} - a|}.$$

This is not very strong estimate; the averaged integral tends to infinity as $|a| \to 1$, but the point is that by a simple argument we have can some bound on analytic function on a disc given only bound on its boundary values.

The point in looking at Schur functions is that we can directly bring the claims for Schur functions to Pick functions with maps

$$\xi : \mathbb{D} \to \mathbb{H}_+ \qquad \xi(z) = i\frac{1-z}{1+z}$$
$$\eta : \mathbb{H}_+ \to \mathbb{D} \qquad \eta(z) = \frac{i-z}{i+z}.$$

If ψ is Schur function then $\xi \circ \psi \circ \eta$ is a Pick function, and conversely every Pick function φ gives rise to Schur function $\eta \circ \varphi \circ \xi$. We can directly translate Schwarz-Pick theorem to Pick functions.

Theorem 5.5 (Schwarz-Pick theorem for the upper half-plane)**.** *Let $\varphi : \mathbb{H}_+ \to \mathbb{H}_+$ be analytic. Then for any $z_1, z_2 \in \mathbb{H}_+$ we have*

$$\left|\frac{\varphi(z_1) - \varphi(z_2)}{\varphi(z_1) - \overline{\varphi(z_2)}}\right| \leq \left|\frac{z_1 - z_2}{z_1 - \overline{z_2}}\right|.$$

Correspondingly analytic bijections in $\mathcal{P}$ are exactly function of the form

$$z \mapsto \frac{az+b}{cz+d},$$

where $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. Among these mapping the map

$$f : z \mapsto \frac{z\Re(w_0) - (|w_0|^2 + \Im(w_0))}{(1 + \Im(w_0))z - \Re(w_0))}$$

satisfies $f(i) = w_0$.

We denote

$$\tilde{d}(z_1, z_2) = \left|\frac{z_1 - z_2}{z_1 - \overline{z_2}}\right|.$$

Schwarz-Pick theorem tells us that $\tilde{d} \circ \varphi \leq \tilde{d}$ for any Pick functions φ. Consequently $\tilde{d} \circ \varphi = \tilde{d}$ if φ is an bijection.

5.2.2 Pick matrices

We can also arrive at the previous estimate directly using Cauchy's integral theorem, a bit similarly as with unit disc. Assume first that φ is a bounded Pick-function extending over real line. We can start by proving that if φ has positive imaginary part on real line, then it has positive real part on whole upper half-plane The idea to consider

$$\varphi(a) = \frac{1}{2\pi i}\int_\gamma \frac{\varphi(z)}{z-a}dz,$$

again for suitable closed curve γ. If φ decays fast enough at infinitity, we can deform the contour γ to coincide with real axis. The unfortunate thing is that the we can't say much about the real or imaginary part of $\frac{\varphi(z)}{z-a}$. But there's a trick: we can fix our problem by considering

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\varphi(z)}{(z-a)(z-\overline{a})}dz.$$

This expression has positive real part, and by using the residue theorem, the real part equals

$$\frac{\varphi(a) - \overline{\varphi(a)}}{a - \overline{a}},$$

hence the claim. Also, boundedness is enough for this computation.

To arrive at Schwarz-Pick theorem consider integral

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} \varphi(z)\left(\frac{c_1}{z-z_1} + \frac{c_2}{z-z_2}\right)\left(\frac{\overline{c_1}}{z-\overline{z_1}} + \frac{\overline{c_2}}{z-\overline{z_2}}\right)dz.$$

Again, the point is that writing $h(z) = \frac{c_1}{z-z_1} + \frac{c_2}{z-z_2}$, h is meromorphic in upper half-plane expect for the simple poles at z_1 and z_2, thus giving information about $\varphi(z_1)$ and $\varphi(z_2)$, $\overline{h(\overline{z})}$ is analytic in the upper halfplane and $h(z)\overline{h(\overline{z})}$ is real on the real axis.

Now this expression should have positive real part for any $c_1, c_2 \in \mathbb{C}$, and computing the real part using the residue theorem, we arrive at

$$\begin{gathered}[z_1, \overline{z_1}]_\varphi |c_1|^2 + [z_1, \overline{z_2}]_\varphi c_1\overline{c_2} \\ [z_2, \overline{z_1}]_\varphi c_2\overline{c_1} + [z_2, \overline{z_2}]_\varphi |c_2|^2,\end{gathered}$$

where we abuse the notation a little by writing $\varphi(\overline{z}) = \overline{\varphi(z)}$.

But this is to say that the matrix

$$([z_i, \overline{z_j}]_\varphi)_{1\le i,j\le 2} = \begin{bmatrix}[z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi\end{bmatrix}$$

is positive. Now, it just so turns out that “determinant of the matrix is non-negative” is equivalent to Schwarz-Pick theorem.

Indeed, straightforward although laborious computation shows that

$$\det\left(\begin{bmatrix}[z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi\\ [z_2,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi\end{bmatrix}\right)=\frac{|\varphi(z_1)-\varphi(z_2)|^2}{4\Im(z_1)\Im(z_2)}\left(d(z_1,z_2)^2-d(\varphi(z_1),\varphi(z_2)^2)\right). \tag{5.6}$$

It is not very hard to generalize the previous argument larger matrices, and we have arrived to

Theorem 5.7. *If φ is Pick function and $z_1, z_2, \ldots, z_n$ are any points in the upper half-plane, then the matrix*

$$\begin{bmatrix}[z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi & \cdots & [z_1,\overline{z_n}]_\varphi\\ [z_2,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi & \cdots & [z_2,\overline{z_n}]_\varphi\\ \vdots & \vdots & \ddots & \vdots\\ [z_n,\overline{z_1}]_\varphi & [z_n,\overline{z_2}]_\varphi & \cdots & [z_n,\overline{z_n}]_\varphi\end{bmatrix} \tag{5.8}$$

is positive.

The matrix 5.8 is called *Pick matrix.*

Proof. We have already proved the theorem in the case that φ is bounded and extends analytically over the real line. For general case consider the sequence g_n of Pick functions given by

$$g_n(z)=\frac{(1-\frac{1}{n})x+\frac{i}{n}}{(1-\frac{1}{n})-i\frac{x}{n}}=\xi\circ\left(z\mapsto\left(1-\frac{2}{n}\right)z\right)\circ\eta.$$

Now

1. $g_n(z)\to z$ pointwise.
2. g_n's extend analytically over real line and $g_n(\overline{\mathcal{H}_+})$ is compact subset of $\mathcal{H}_+$ for every $n\geq 1$.

It follows $\varphi\circ g_n\to\varphi$ pointwise and $\varphi\circ g_n$'s satisfy the already proven case. Finally, also the corresponding Pick-matrices of $\varphi\circ g_n$'s converge to Pick matrix of φ, hence the general case. □

5.3 Weak characterization

Theorem 5.7 has a converse.

Theorem 5.9. *If $\varphi : \mathbb{H}_+ \to \overline{\mathbb{H}}_+$ such that for any $n \geq 1$ and $z_1, z_2, \ldots, z_n \in \mathbb{H}_+$ the respective Pick matrix*

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi & \cdots & [z_1, \overline{z_n}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi & \cdots & [z_2, \overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_n, \overline{z_1}]_\varphi & [z_n, \overline{z_2}]_\varphi & \cdots & [z_n, \overline{z_n}]_\varphi \end{bmatrix}$$

is positive, then φ is a Pick function.

Note that if all the Pick matrices are positive, function clearly has non-negative imaginary part. Thus we "only" need to verify analyticity.

Let's first check continuity. For this we only need positivity on 2×2-matrices.

Theorem 5.10. *Let $A \subset \mathbb{H}_+$ and $\varphi : A \to \overline{\mathbb{H}}_+$ such that for any $z_1, z_2 \in \mathbb{H}_+$ Pick matrix*

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi \end{bmatrix}$$

is positive. Then f is locally Lipschitz, in particular continuous.

Proof. It follows from 5.6 that we have

$$\left| \frac{\varphi(z_1) - \varphi(z_2)}{\varphi(z_1) - \overline{\varphi(z_2)}} \right| \leq \left| \frac{z_1 - z_2}{z_1 - \overline{z_2}} \right|.$$

for any $z_1, z_2 \in A$; we can draw the same conclusion as in the Schwarz lemma. Now it clearly suffices to check that φ is bounded on any compact set. For any compact K there exists ε_K such that the expression

$$\left| \frac{z_1 - z_2}{z_1 - \overline{z_2}} \right| < 1 - \varepsilon_K,$$

whenever $z_1, z_2 \in K$. If we fix $z_2 \in K$, it is easily seen that we obtain uniform bound on $|\varphi(z_1)|$, as

$$\lim_{z \to \infty} \left| \frac{z - w}{z - \overline{w}} \right| = 1.$$

(Here ∞ denotes the complex infinity). □

Curiously enough, φ is analytic as long as all its 3×3 Pick matrices are positive. This result is known as Hindmarsh's theorem.

Theorem 5.11. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$ such that for every $z_1, z_2, z_3 \in \mathbb{H}_+$ Pick matrix*

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi & [z_1, \overline{z_3}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi & [z_2, \overline{z_3}]_\varphi \\ [z_3, \overline{z_1}]_\varphi & [z_3, \overline{z_2}]_\varphi & [z_3, \overline{z_3}]_\varphi \end{bmatrix}$$

is positive. Then φ is analytic.

Proof. We are going to check that the quantity $[z_1, z_2, z_3]_\varphi$ is locally bounded, which implies the claim as in the proof of theorem 3.31. □

As an immediate corollary we get theorem 5.9.

Proof of theorem 5.9. If all Pick matrices are positive, so are all 3×3 Pick matrices. □

We can hence characterize Pick functions purely with Pick matrices, without limits and concerns of regularity, "weakly". As an immediate corollary we get proposition 5.2.

Proof of theorem 5.2. Pointwise limits preserve positivity of the Pick matrices. □

One might still go even further and argue that one does not need Pick matrices larger than 3×3 to talk about Pick functions. They however carry interesting "global" information.

Theorem 5.12. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$ such that for any $n \geq 1$ and $z_1, z_2, \ldots, z_n \in U$ the respective Pick matrix*

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi & \cdots & [z_1, \overline{z_n}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi & \cdots & [z_2, \overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_n, \overline{z_1}]_\varphi & [z_n, \overline{z_2}]_\varphi & \cdots & [z_n, \overline{z_n}]_\varphi \end{bmatrix}$$

is positive. Then φ is a restriction of an unique Pick function.

This tells us that we may recognise Pick functions from local information. To prove the theorem, we first introduce the notion of *Pick point.*

Definition 5.13. Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$. We say that $z \in U$ is Pick point of φ, if φ is analytic at z and the $n \times n$ matrix

$$\begin{bmatrix} [z, \overline{z}]_\varphi & [z, \overline{z}, \overline{z}]_\varphi & \cdots & [z, \overline{z}, \overline{z}, \ldots, \overline{z}]_\varphi \\ [z, z, \overline{z}]_\varphi & [z, z, \overline{z}, \overline{z}]_\varphi & \cdots & [z, z, \overline{z}, \overline{z}, \ldots, \overline{z}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z, z, \ldots, z, \overline{z}]_\varphi & [z, z, \ldots, z, \overline{z}, \overline{z}]_\varphi & \cdots & [z, z, \ldots, z, \overline{z}, \overline{z}, \ldots, \overline{z}]_\varphi \end{bmatrix}$$

is positive for every n.

Where does this definition come from? The idea is to answer the question: what does it mean the Pick matrices to be non-negative at a single point, say z_0? In the definition of Pick matrix we could let all the variables be equal and conclude that the matrix

$$\begin{bmatrix} [z_0,\overline{z_0}]_\varphi & [z_0,\overline{z_0}]_\varphi & \cdots & [z_0,\overline{z_0}]_\varphi \\ [z_0,\overline{z_0}]_\varphi & [z_0,\overline{z_0}]_\varphi & \cdots & [z_0,\overline{z_0}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_0,\overline{z_0}]_\varphi & [z_0,\overline{z_0}]_\varphi & \cdots & [z_0,\overline{z_0}]_\varphi \end{bmatrix}$$

is positive, but this would only tell us that $[z_0,\overline{z_0}]_\varphi$ is non-negative. We need derivatives.

The idea is to $*$-conjugate the Pick matrix first. Say $n=2$. If we subtract first row from the second in the Pick matrix

$$\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi \\ [z_2,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi \end{bmatrix}$$

we get

$$\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi \\ [z_2,\overline{z_1}]_\varphi-[z_1,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi-[z_1,\overline{z_2}]_\varphi \end{bmatrix}=\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi \\ (z_2-z_1)[z_1,z_2,\overline{z_1}]_\varphi & (z_2-z_1)[z_1,z_2,\overline{z_2}]_\varphi \end{bmatrix}.$$

Now subtracting first column from the second results in

$$\begin{aligned}&\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi-[z_1,\overline{z_1}]_\varphi \\ (z_2-z_1)[z_1,z_2,\overline{z_1}]_\varphi & (z_2-z_1)[z_1,z_2,\overline{z_2}]_\varphi-(z_2-z_1)[z_1,z_2,\overline{z_1}]_\varphi \end{bmatrix}\\ =&\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & \overline{(z_2-z_1)}[z_1,\overline{z_1},\overline{z_2}]_\varphi \\ (z_2-z_1)[z_1,z_2,\overline{z_1}]_\varphi & (z_2-z_1)\overline{(z_2-z_1)}[z_1,z_2,\overline{z_1},\overline{z_2}]_\varphi \end{bmatrix}.\end{aligned}$$

In the language of matrices this really says that

$$\begin{aligned}&\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi \\ [z_2,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\\ =&\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & \overline{(z_2-z_1)}[z_1,\overline{z_1},\overline{z_2}]_\varphi \\ (z_2-z_1)[z_1,z_2,\overline{z_1}]_\varphi & (z_2-z_1)\overline{(z_2-z_1)}[z_1,z_2,\overline{z_1},\overline{z_2}]_\varphi \end{bmatrix}\\ =&\begin{bmatrix} 1 & 0 \\ 0 & (z_2-z_1) \end{bmatrix}\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_1},\overline{z_2}]_\varphi \\ [z_1,z_2,\overline{z_1}]_\varphi & [z_1,z_2,\overline{z_1},\overline{z_2}]_\varphi \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \overline{(z_2-z_1)} \end{bmatrix}:\end{aligned}$$

matrices

$$\begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_2}]_\varphi \\ [z_2,\overline{z_1}]_\varphi & [z_2,\overline{z_2}]_\varphi \end{bmatrix} \text{ and } \begin{bmatrix} [z_1,\overline{z_1}]_\varphi & [z_1,\overline{z_1},\overline{z_2}]_\varphi \\ [z_1,z_2,\overline{z_1}]_\varphi & [z_1,z_2,\overline{z_1},\overline{z_2}]_\varphi \end{bmatrix}$$

are congruent.

Generalizing this argument we see that the matrices

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_2}]_\varphi & \cdots & [z_1, \overline{z_n}]_\varphi \\ [z_2, \overline{z_1}]_\varphi & [z_2, \overline{z_2}]_\varphi & \cdots & [z_2, \overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_n, \overline{z_1}]_\varphi & [z_n, \overline{z_2}]_\varphi & \cdots & [z_n, \overline{z_n}]_\varphi \end{bmatrix}$$

and

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \\ [z_1, z_2, \overline{z_1}]_\varphi & [z_1, z_2, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, z_2, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_1, z_2, \dots, z_n, \overline{z_1}]_\varphi & [z_1, z_2, \dots, z_n, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, z_2, \dots, z_n, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \end{bmatrix}$$

are congruent. We hence conclude the following.

Lemma 5.14. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$. Let $z_0 \in U$. Assume that for some $r > 0$, for every $n \geq 1$ and $z_1, z_2, \dots, z_n \in \mathbb{D}(z_0, r) \cap U$ the respective Pick matrix is positive. Then z_0 is Pick point of φ.*

Proof. By theorem 5.11 φ is analytic at z_0. By previous observation all the matrices of the form

$$\begin{bmatrix} [z_1, \overline{z_1}]_\varphi & [z_1, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \\ [z_1, z_2, \overline{z_1}]_\varphi & [z_1, z_2, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, z_2, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_1, z_2, \dots, z_n, \overline{z_1}]_\varphi & [z_1, z_2, \dots, z_n, \overline{z_1}, \overline{z_2}]_\varphi & \cdots & [z_1, z_2, \dots, z_n, \overline{z_1}, \overline{z_2}, \dots, \overline{z_n}]_\varphi \end{bmatrix} \tag{5.15}$$

are positive. By letting $z_1, z_2, \dots, z_n \to z_0$ we get the claim. $\square$

Matrix 5.15 is called *extended Pick matrix.*

Again, one can, and should, interpret the extended Pick matrix as some kind of change of basis. In terms of a Cauchy's integral formula the original Pick matrix says that

$$\Re\left(\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^n \frac{c_j}{z - \lambda_j}\right)\left(\sum_{j=1}^n \frac{\overline{c_j}}{z - \overline{\lambda_j}}\right)\varphi(z)dz\right) \geq 0,$$

for any $c_1, c_2, \dots, c_n \in \mathbb{C}$ and $\lambda_1, \lambda_2, \dots, \lambda_n \in U$ while the extended Pick matrix says that

$$\Re\left(\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^n \frac{c_j}{\prod_{1 \leq j' \leq j}(z - \lambda_{j'})}\right)\left(\sum_{j=1}^n \frac{\overline{c_j}}{\prod_{1 \leq j' \leq j}(z - \overline{\lambda_{j'}})}\right)\varphi(z)dz\right) \geq 0$$

for any $c_1, c_2, \ldots, c_n \in \mathbb{C}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in U$. This interpretation makes it clear that two different matrices are really about

$$\frac{1}{z-\lambda_1}, \frac{1}{z-\lambda_2}, \ldots, \frac{1}{z-\lambda_n}$$

and

$$\frac{1}{z-\lambda_1}, \frac{1}{(z-\lambda_1)(z-\lambda_2)}, \ldots, \frac{1}{(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_n)}$$

spanning the same set of rational functions (whenever λ_i's are pairwise distinct).

The idea of the proof of theorem 5.12 is the following: we try to make every point in $\mathbb{H}_+$ is Pick point of φ. For this we need two lemmas. first one says that if point is a Pick point of φ, then φ can be extended to a reasonable large disc around z_0, which is really to say that the Taylor coefficients of φ at z_0 don't grow too fast. Second one tells us that if we managed to extend function to such disc, then all the points in the disc are Pick points.

Lemma 5.16. *There exists absolute constant c_0 with the following property: Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$. Assume that $z_0 \in U$ is a Pick point of U. Then the Taylor series of φ at z_0 converges in $\mathbb{D}(z_0, c_0\Im(z_0))$. One may take $c_0 = \frac{1}{9}$.*

Proof. By affine transformation we may assume that $z_0 = i$ and $\varphi(i) = i$. We should prove that there exists an absolute constant C_0 such that $\left|\frac{\varphi^{(n)}(i)}{n!}\right| \leq C_0^{n+1}$ for any n. Write $a_k = \varphi^{(k)}(i)/k!$. The claim follows quite easily if we can prove that for some constant C_0'. We have

$$|a_k| \leq C_0' \sum_{i=0}^{k-1} |a_i|$$

for $k \geq 1$.

We are going to express the entries of the Dobsch matrix with respect to the Taylor coefficients.

For convenience, write

$$b_k = a_k(-2i)^{k+1}$$

By staightforvard induction one sees that

$$\begin{aligned}&[i, i, \ldots, i, -i, -i, \ldots, -i]_\varphi \\ &= \frac{1}{(-2i)^{k_1}(2i)^{k_2}} \left(\sum_{j=0}^{n-1} b_j \binom{m+n-2-j}{m-1} + \sum_{j=0}^{m-1} \overline{b_j} \binom{m+n-2-j}{n-1} \right).\end{aligned}$$

where i and $-i$ appear k_1 and k_2 times in the divided difference.

The idea is to look at the $n \times n$ matrix and conclude the inequality for b_{n-1}. It turns out that it is sufficient to look at the submatrix consisting of only first and the last row and column. We may interpret the resulting matrix as sum of n matrices A_j, $0 \leq j \leq n-1$, namely

$$A_j = \begin{cases} \begin{bmatrix} \frac{1}{4}(b_0 + \overline{b_0}) & \frac{1}{(-2i)^n(2i)}(b_0 + \overline{b_0}) \\ \frac{1}{(-2i)(2i)^n}(b_0 + \overline{b_0}) & \frac{\binom{2n-2}{n-1}}{2^{2n}}(b_0 + \overline{b_0}) \end{bmatrix} & \text{if } j = 0 \\ \begin{bmatrix} 0 & \frac{1}{(-2i)^n(2i)} b_j \\ \frac{1}{(-2i)(2i)^n}\overline{b_j} & \frac{\binom{2n-2-j}{n-1}}{2^{2n}}(b_j + \overline{b_j}) \end{bmatrix} & \text{otherwise.} \end{cases}$$

Using the trivial inequality $\binom{2n-2-j}{n-1} \leq 2^{2n-2-j}$ one easily sees that $\|A_j\| \leq 2^{-j}|b_j|$ for any $0 \leq j \leq n-1$. Additionally it is easy to verify that A_{n-1} has negative eigenvalue with absolute value at least $2^{-n-2}|b_{n-1}|$.

It follows that

$$\frac{|b_{n-1}|}{2^{n-1}} \leq 8 \sum_{j=0}^{n-2} \frac{|b_j|}{2^j},$$

which implies that

$$|a_{n-1}| \leq 8 \sum_{j=0}^{n-2} |a_j|.$$

Since $a_0 = 1$, we see by induction that $|a_j| \leq 9^j$, and we may hence take $c_0 = \frac{1}{9}$. □

Lemma 5.17. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}_+}$. Let $z_0 \in U$ and $0 < r \leq \Im(z_0)$. Assume that z_0 is Pick point of φ and φ is analytic in $\mathbb{D}(z_0, r)$. Then all Pick matrices of φ are positive on $\mathbb{D}(z_0, r)$. Consequently, all the points in $\mathbb{D}(z_0, r)$ are Pick points of z_0.*

Proof. Rewriting things with Cauchy's integral formula, our goal is to prove that for any $c_1, c_2, \ldots, c_n \in \mathbb{C}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{D}(z_0, r)$ we have

$$\Re\left(\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^n \frac{c_j}{z - \lambda_j}\right)\left(\sum_{j=1}^n \frac{\overline{c_j}}{z - \overline{\lambda_j}}\right)\varphi(z)dz\right) \geq 0$$

where γ traces a circle centered at z_0, enclosing the λ's. What we know, again, in terms of Cauchy's integral formula, is that for any $c'_1, c'_2, \ldots, c'_n \in \mathbb{C}$ we have

$$\Re\left(\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^n \frac{c'_j}{(z - z_0)^j}\right)\left(\sum_{j=1}^n \frac{\overline{c'_j}}{(z - \overline{z_0})^j}\right)\varphi(z)dz\right) \geq 0.$$

The idea is to simply approximate upper integrand: if we manange to approximate terms of the form $\frac{1}{z-\lambda}$ uniformly on γ by terms of the form $\sum_{j=1}^{n} \frac{c'_j}{(z-z_0)^j}$, then we can certainly make our goal expression as close to a non-negative number as we wish. But this is exactly what Laurent expansions are for.

One should not forget that also the conjugate part should be approximated, but since $|\overline{z} - z_0| > |z - z_0|$ for any $z \in \mathbb{H}_+$, the conjugated approximation is even better. $\square$

Proof of theorem 5.12. Consider all open sets $U \subset V \subset \mathbb{H}_+$ such that φ may be extended to V so that all the points of V are Pick points of φ (or the extension thereof). These sets trivially satisfy conditions of Zorn's lemma (where partial order is given by inclusion) so we may Pick maximal such set, V. We claim that this set is $\mathbb{H}_+$. If not, we may pick a point z_0 in $\mathbb{H}_+ \cap \partial V$. Now pick $z \in V$ such that $|z - z_0| < \Im(z_0)c_0$: we may extend φ now further to $\mathbb{D}(z_0, c_0\Im(z_0))$ by Taylor series and all points in the extension are Pick points, which contradicts the maximality. $\square$

There is fundamental flaw with the second argument: when we extend φ with Taylor series, how do we know that these extensions are consistent? If original set U was, say, disjoint union of two discs, and Taylor series centered at a point of one disc would converge also at (some part of) other disc, how do we know that the Taylor series converges to the predefined values in the one disc? This is not at all clear.

If U is disc or more generally domain, we can use the monodromy theorem to show that this kind of extension is indeed possible. Indeed, upper half-plane is simply connected and by the previous lemmas we can continue φ along any path. This, however, still doesn't fix the problem with two disjoint discs.

We can salvage the proof by improving the second lemma. We have to somehow remember the information of the positivity of all the Pick matrices, otherwise we might run into problem. Let's have notion for this.

Definition 5.18. Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$. We call φ *weakly Pick*, or say it is *weak Pick function* if all its Pick matrices are positive.

Lemma 5.19. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$ weakly Pick. Assume that for some $z_0 \in U$ and $r > 0$ the Taylor series of φ at z_0 converges in $\mathbb{D}(z_0, r)$. Then the values of Taylor series in $\mathbb{D}(z_0, r) \cap U$ coincide with φ and the resulting extension is weakly Pick in $U \cup \mathbb{D}(z_0, r)$.*

Proof. Let us first show that the extension coincides with the original function. Note that this not all that obvious and fails if without the weakly Pick -condition.

To do that, let $\tilde{\varphi} : \mathbb{D}(z_0, r) \to \mathbb{C}$ be the extension and let $z_1 \in \mathbb{D}(z_0, r)$ and $z_2 \in U$ be distinct. We prove that the matrix

$$\begin{bmatrix} \frac{\tilde{\varphi}(z_1)-\overline{\tilde{\varphi}(z_1)}}{z_1-\overline{z_1}} & \frac{\tilde{\varphi}(z_1)-\overline{\varphi(z_2)}}{z_1-\overline{z_2}} \\ \frac{\varphi(z_2)-\overline{\tilde{\varphi}(z_1)}}{z_2-\overline{z_1}} & \frac{\varphi(z_2)-\overline{\varphi(z_2)}}{z_2-\overline{z_2}} \end{bmatrix}$$

is positive. As observed before, this shows that $|\tilde{\varphi}(z_1) - \varphi(z_2)| \leq C_{\varphi,z_2}|z_1 - z_2|$ which implies the claim. We could also consider the matrix where z_1 and z_2 are equal, but this is just for notational convenience.

The idea is to try to do the same thing as in the proof of lemma 5.17. The problem is that we can't a priori express both $\varphi(z_1)$ and $\tilde{\varphi}(z_1)$ with Cauchy's integral formula with the same contour. But we can still follow the same approximation scheme, one just has to be more careful about error terms.

To make the notation slightly simpler

If we already knew that the result was true, the positivity of matrix desired would be same as positivity of the real part of

$$\frac{1}{2\pi i}\int_\gamma \left(\frac{c_1}{z-z_1} + \frac{c_2}{z-z_2}\right)\left(\frac{\overline{c_1}}{z-\overline{z_1}} + \frac{\overline{c_2}}{z-\overline{z_2}}\right)\varphi(z)dz.$$

Plugging the identity

$$\frac{1}{z-z_1} = \sum_{j=1}^{n} \frac{(z_1-z_0)^{j-1}}{(z-z_0)^j} + \frac{(z_1-z_0)^n}{(z-z_1)(z-z_0)^n}$$

in the previous integral we arrive at

$$\begin{aligned}
&\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^{n} \frac{c_1(z_1-z_0)^{j-1}}{(z-z_0)^j} + \frac{c_2}{z-z_2}\right)\left(\sum_{j=1}^{n} \frac{\overline{c_1}(\overline{z_1}-\overline{z_0})^{j-1}}{(z-\overline{z_0})^j} + \frac{\overline{c_2}}{z-\overline{z_2}}\right)\varphi(z)dz \\
+&\frac{1}{2\pi i}\int_\gamma \left(\frac{c_1(z_1-z_0)^n}{(z-z_1)(z-z_0)^n}\right)\left(\sum_{j=1}^{n} \frac{\overline{c_1}(\overline{z_1}-\overline{z_0})^{j-1}}{(z-\overline{z_0})^j}\right)\varphi(z)dz \\
+&\frac{1}{2\pi i}\int_\gamma \left(\frac{c_1(z_1-z_0)^n}{(z-z_1)(z-z_0)^n}\right)\left(\frac{\overline{c_2}}{z-\overline{z_2}}\right)\varphi(z)dz \\
+&\frac{1}{2\pi i}\int_\gamma \left(\frac{c_1(z_1-z_0)^n}{(z-z_1)(z-z_0)^n}\right)\left(\frac{\overline{c_1}(\overline{z_1}-\overline{z_0})^n}{(z-\overline{z_1})(z-\overline{z_0})^n}\right)\varphi(z)dz \\
+&\frac{1}{2\pi i}\int_\gamma \left(\sum_{j=1}^{n} \frac{c_1(z_1-z_0)^{j-1}}{(z-z_0)^j}\right)\left(\frac{\overline{c_1}(\overline{z_1}-\overline{z_0})^n}{(z-\overline{z_1})(z-\overline{z_0})^n}\right)\varphi(z)dz \\
+&\frac{1}{2\pi i}\int_\gamma \left(\frac{c_2}{z-z_2}\right)\left(\frac{\overline{c_1}(\overline{z_1}-\overline{z_0})^n}{(z-\overline{z_1})(z-\overline{z_0})^n}\right)\varphi(z)dz.
\end{aligned}$$

Note all this could be expanded as sum linear combination of $\varphi(z_1), \varphi(z_2)$ and $\varphi^{(j)}(z_0)$ for $0 \leq j \leq n-1$. Should one do that, the resulting identity would hold without any additional (to the lemma statement) smoothness assumptions. The point again is that the first row gives the main term, and we know that it is non-negative. It hence suffices to prove that remaining terms tend to zero as $n \to \infty$.

Terms without z_2 are easy to estimate directly as they make also make honest sense. We can take γ to be circle of radius r' for some $|z_1 - z_0| < r' < r$. Then for instance the first error term can be bounded by

$$\frac{|c_1 c_2||z_1 - z_0|^n (r - r')}{r'^n} \left(\sum_{j=1}^{n} \frac{|z_1 - z_0|^{j-1}}{r'^j} \right) \sup_{z \in \partial \mathbb{D}(z_0, r')} |\varphi(z)|,$$

which clearly tends to 0.

With z_2 terms one has to be bit more careful. First, write

$$\begin{aligned}
&\frac{1}{2\pi i} \int_\gamma \left(\frac{c_1 (z_1 - z_0)^n}{(z - z_1)(z - z_0)^n} \right) \left(\frac{\overline{c_2}}{z - \overline{z_2}} \right) \varphi(z) dz \\
=&\frac{c_1 \overline{c_2}}{z_1 - \overline{z_2}} \frac{1}{2\pi i} \int_\gamma \left(\frac{(z_1 - z_0)^n}{(z - z_1)(z - z_0)^n} \right) \varphi(z) dz \\
-&\frac{c_1 \overline{c_2}}{z_1 - \overline{z_2}} \frac{1}{2\pi i} \int_\gamma \left(\frac{c_1 (z_1 - z_0)^n}{(z - \overline{z_2})(z - z_0)^n} \right) \varphi(z) dz.
\end{aligned}$$

First term is again easily estimated. The second integrand, so far formal, can be written as a sum

$$\begin{aligned}
&\frac{(z_1 - z_0)^n}{(z - \overline{z_2})(z - z_0)^n} \varphi(z) \\
=&\frac{(z_1 - z_0)^n}{(z - \overline{z_2})(z_0 - \overline{z_2})^n} \varphi(z) - \sum_{i=1}^{n} \frac{(z_1 - z_0)^n}{(z - z_0)^j (z_0 - \overline{z_2})^{n+1-j}} \varphi(z).
\end{aligned}$$

First term evaluates to 0 and the other terms are again easily estimated. For the another z_2 only

$$\frac{(z_1 - z_0)^n}{(z_0 - \overline{z_2})^n} \varphi(z_0)$$

remains.

This argument also essentially proves that the extension is weakly Pick. We should prove that for any $n_1, n_2 \geq 0$ and $z_1, \ldots, z_{n_1}, w_1, \ldots, w_{n_2} \in \mathbb{C}$ the matrix

$$\begin{bmatrix} \frac{\tilde{\varphi}(z_1)-\overline{\tilde{\varphi}(z_1)}}{z_1-\overline{z_1}} & \cdots & \frac{\tilde{\varphi}(z_1)-\overline{\tilde{\varphi}(z_{n_1})}}{z_1-\overline{z_{n_1}}} & \frac{\tilde{\varphi}(z_1)-\overline{\varphi(w_1)}}{z_1-\overline{w_1}} & \cdots & \frac{\tilde{\varphi}(z_1)-\overline{\varphi(w_{n_2})}}{z_1-\overline{w_{n_2}}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\tilde{\varphi}(z_{n_1})-\overline{\tilde{\varphi}(z_1)}}{z_1-\overline{z_1}} & \cdots & \frac{\tilde{\varphi}(z_{n_1})-\overline{\tilde{\varphi}(z_{n_1})}}{z_1-\overline{z_{n_1}}} & \frac{\tilde{\varphi}(z_{n_1})-\overline{\varphi(w_1)}}{z_1-\overline{w_1}} & \cdots & \frac{\tilde{\varphi}(z_{n_1})-\overline{\varphi(w_{n_2})}}{z_1-\overline{w_{n_2}}} \\ \frac{\varphi(w_1)-\overline{\tilde{\varphi}(z_1)}}{w_1-\overline{z_1}} & \cdots & \frac{\varphi(w_1)-\overline{\tilde{\varphi}(z_{n_1})}}{w_1-\overline{z_{n_1}}} & \frac{\varphi(w_1)-\overline{\varphi(w_1)}}{w_1-\overline{w_1}} & \cdots & \frac{\varphi(w_1)-\overline{\varphi(w_{n_2})}}{w_1-\overline{w_{n_2}}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\varphi(w_{n_2})-\overline{\tilde{\varphi}(z_1)}}{w_{n_2}-\overline{z_1}} & \cdots & \frac{\varphi(w_{n_2})-\overline{\tilde{\varphi}(z_{n_1})}}{w_{n_2}-\overline{z_{n_1}}} & \frac{\varphi(w_{n_2})-\overline{\varphi(w_1)}}{w_{n_2}-\overline{w_1}} & \cdots & \frac{\varphi(w_{n_2})-\overline{\varphi(w_{n_2})}}{w_{n_2}-\overline{w_{n_2}}} \end{bmatrix}$$

is positive. Formally, this would mean that for any $c_1, \ldots, c_{n_1}, d_1, \ldots, d_{n_2} \in \mathbb{C}$ the real part of the integral

$$\frac{1}{2\pi i} \int_\gamma \left(\sum_{i=1}^{n_1} \frac{c_i}{z - z_i} + \sum_{i=1}^{n_2} \frac{d_i}{z - w_i} \right) \left(\sum_{i=1}^{n_1} \frac{\overline{c_i}}{z - \overline{z_i}} + \sum_{i=1}^{n_2} \frac{\overline{d_i}}{z - \overline{w_i}} \right) \varphi(z) dz$$

is positive. But one can simply do the same Laurent expansion, and error terms are estimated completely analogously. □

Fix of the proof of theorem 5.12. Now the Zorn's lemma works if we change the condition a bit: we look at all the weakly Pick extensions of φ to open subsets of upper half-plane.

Zorn's lemma is not really necessary here: one could write explicit extension scheme (TODO: picture) and the lemmas would guarantee that we can always both extend further and extensions are always consistent and weakly Pick. □

5.4 Pick-Nevanlinna Interpolation theorem

One can still considerably strenghten theorem 5.12: instead of open set, domain of φ could be any set. Then we don't in general have unique extension. This is the content of Pick-Nevanlinna interpolation theorem.

Theorem 5.20 (Pick-Nevanlinna interpolation theorem). *Let $A \subset \mathcal{H}_+$ and $\varphi : A \to \overline{\mathbb{H}}_+$ weakly Pick. Then φ is a restriction of a Pick function to A.*

Note that the notion of weak Pick functions makes perfect sense in this more general setting. In the general setting we can't use Taylor series to extend the function anymore, but it turns out that this is not too big of a problem. The idea is again to extend φ to larger and larger sets. Note that to do this we only need to be able to extend φ to one new point so that all Pick matrices are again positive: we can then use the Zorn's lemma -trick again.

Before the proof we need an interesting fact[1].

Lemma 5.21. *Let $\mathcal{D}$ be a non-empty family of (non-empty) closed discs such that for any $\mathbb{D}_1, \mathbb{D}_2 \in \mathcal{D}$ there exists $\mathcal{D} \ni \mathbb{D}_3 \subset \mathbb{D}_1 \cap \mathbb{D}_2$. Then*

$$\bigcap_{\mathbb{D} \in \mathcal{D}} \mathbb{D}$$

is a (non-empty) closed disc.

Proof. Denote the intersection by K. Since $\mathcal{D}$ is family of compact sets with finite intersection property, K is non-empty compact set. Hence we may take two points in K with largest distance. If the points are equal, the intersection is a single points, and we are done. If not, without loss of generality we may assume that the points are -1 and 1. Our aim is to prove that K equals $\overline{\mathbb{D}}(0,1)$.

Fix $r > 1$. We claim that there exists a disc $\mathbb{D} \in \mathcal{D}$ such that $\mathbb{D}$ is contained in the strip $S_r := \{z \in \mathbb{C} | |\Re(z)| < r\}$. For this note that for any $M > 0$ the family $\{[r - iM, r + iM] \cap \mathbb{D} | \mathbb{D} \in \mathcal{D}\}$ (where $[r - iM, r + iM]$ denotes the closed interval from $r - iM$ to $r + iM$) is family of compact sets with empty intersection, so they have finite subfamily with empty intersection. By the main condition if follows that we find a disc in $\mathcal{D}$ with empty intersection with the interval. Taking M large enough we see that one of the discs has empty intesection with the line $\{z \in \mathbb{C} | \Re(z) = r\}$. Playing the same game with $-r$ and taking the intersection we find the required disc.

Next we claim that for any $r' < 1$ any disc in the family contains $\mathbb{D}(0, r')$. Note that since all the discs in the family have radius at least 1, this is certainly true for any disc with center close enough to 0. If $\mathbb{D}_0 \in S_r$ has center z_0, we have $2r^2 \geq |1 - z_0|^2 + |1 + z_0|^2 = 2 + 2|z_0|^2$. Taking r close to 1 we see that z_0 is close to 0 and hence $\mathbb{D}_0$ contains $\mathbb{D}(0, r')$. Since by the intersection property every disc in $\mathcal{D}$ contains disc in S_r, also the second claim is proven.

Now as K as compact, it contains the closed disc $\overline{\mathbb{D}}(0,1)$. But since K has diameter 2, it can't contain anything else, so $K = \overline{\mathbb{D}}(0,1)$ and we are done.

TODO picture □

Not surprisingly, the result remains true also in higher dimensions, as can be seen for instance by intersecting the closed balls with two dimensional planes and applying the 2d case.

[1] We don't really need, but it is an interesting fact nonetheless.

Lemma 5.22. *Let $A \subset \mathcal{H}_+$ and $\varphi : A \to \overline{\mathbb{H}}_+$ weakly Pick. Then if $z_0 \in \mathbb{H}_+ \setminus A$ there exists $w_0 \in \overline{\mathbb{H}}_+$ such that if we extend φ to z_0 by setting $\varphi(z_0) = w_0$, also the extension is weakly Pick.*

Proof. Let's first consider the case of finite A, say $A = \{z_1, z_2, \ldots, z_n\}$ and denote $w_i = \varphi(z_i)$. We should find w_0 such that the matrix

$$\begin{bmatrix} \frac{w_0 - \overline{w_0}}{z_0 - \overline{z_0}} & \frac{w_0 - \overline{w_1}}{z_0 - \overline{z_1}} & \cdots & \frac{w_0 - \overline{w_n}}{z_0 - \overline{z_n}} \\ \frac{w_1 - \overline{w_0}}{z_1 - \overline{z_0}} & \frac{w_1 - \overline{w_1}}{z_1 - \overline{z_1}} & \cdots & \frac{w_1 - \overline{w_n}}{z_1 - \overline{z_n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n - \overline{w_0}}{z_n - \overline{z_0}} & \frac{w_n - \overline{w_1}}{z_n - \overline{z_1}} & \cdots & \frac{w_n - \overline{w_n}}{z_n - \overline{z_n}} \end{bmatrix}$$

is positive. Denote the matrix consisting of the last n rows and columns of the matrix by n. We first assume that M is invertible. Then the original matrix is positive if and only the the Schur complement with respect to the last n rows is positive, that is

$$\begin{aligned} 0 \leq & \frac{w_0 - \overline{w_0}}{z_0 - \overline{z_0}} - (w_0 v - w)^* M^{-1} (w_0 v - w) \\ = & -\langle M^{-1} v, v \rangle |w_0|^2 + \left(\frac{1}{z_0 - \overline{z_0}} + \langle M^{-1} v, w \rangle \right) w_0 \\ & + \left(-\frac{1}{z_0 - \overline{z_0}} + \langle M^{-1} w, v \rangle \right) \overline{w_0} - \langle M^{-1} w, w \rangle \\ = & -\frac{1}{\langle M^{-1} v, v \rangle} \left| \langle M^{-1} v, v \rangle w_0 + \frac{1}{z_0 - \overline{z_0}} - \langle M^{-1} w, v \rangle \right|^2 \\ & + \frac{\left| \frac{1}{z_0 - \overline{z_0}} - \langle M^{-1} w, v \rangle \right|^2 - \langle M^{-1} v, v \rangle \langle M^{-1} w, w \rangle}{\langle M^{-1} v, v \rangle} \end{aligned}$$

where $v = (\frac{1}{z_0 - \overline{z_i}})_{i=1}^n$ and $w = (\frac{\overline{w_i}}{z_0 - \overline{z_i}})_{i=1}^n$. Now if we manage to prove that

$$\left| \frac{1}{z_0 - \overline{z_0}} - \langle M^{-1} w, v \rangle \right|^2 - \langle M^{-1} v, v \rangle \langle M^{-1} w, w \rangle \geq 0,$$

we can choose suitable w_0 and moreover set of suitable w_0's is a closed disc.

The positivity follows from the following formula:

$$\left| \frac{1}{z_0 - \overline{z_0}} - \langle M^{-1} w, v \rangle \right|^2 - \langle M^{-1} v, v \rangle \langle M^{-1} w, w \rangle = \prod_{i=1}^n \left| \frac{z_i - z_0}{z_i - \overline{z_0}} \right|^2 = \prod_{i=1}^n \tilde{d}(z_1, z_0)^2$$

Where does this come from? The idea is that the Pick matrix is of a really special form and as one easily checks, one has

$$\frac{w_k - \overline{w_l}}{z_k - \overline{z_l}} - \frac{w_k - \overline{z_0}}{z_k - \overline{z_0}}\frac{\overline{w_l} - z_0}{\overline{z_l} - z_0} + \frac{w_k - z_0}{z_k - \overline{z_0}}\frac{\overline{w_l} - \overline{z_0}}{\overline{z_l} - z_0} = \frac{z_k - z_0}{z_k - \overline{z_0}}\frac{\overline{z_l} - \overline{z_0}}{\overline{z_l} - z_0}\frac{w_k - \overline{w_l}}{z_k - \overline{z_l}}.$$

This identity can be interpreted as $(M - v_1 v_1^*) + v_2 v_2^* = D^* M D$ where v_1, v_2 are some vectors and D a diagonal matrix. Calculating the determinant from both sides, by multiplicativity on the right-hand side and using rank 1 update formulas in the bracketed order on the left-hand side, one will end up with the sought formula.

If M is singular, one has to be bit careful. The trick is to shift the tuple a bit.

The matrix $\left(\frac{2i}{z_k - \overline{z_l}}\right)_{1 \leq k,l \leq n}$ is strictly positive for any tuple of distinct numbers $z_1, \ldots, z_n \in \mathbb{H}_+$. Indeed, by the residue theorem we have

$$\sum_{1 \leq k,l \leq n} c_k \overline{c_l} \frac{2i}{z_k - \overline{z_l}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\sum_{k=1}^{n} \frac{c_k}{t - z_k}\right|^2 dt$$

and if the integrand is zero almost everywhere, and hence on at least n points, then c_i's are all zero.

Consequently, consider the extension problem for the values $w_k + \varepsilon i$ for some $\varepsilon > 0$, and let $D_\varepsilon + \varepsilon i$ be the respective set of suitable w_0's for given ε. Since clearly $D_\varepsilon \subset D_{\varepsilon'}$ for any $\varepsilon \leq \varepsilon'$ and D_ε's are closed discs, by lemma 5.21 the intersection

$$\cap_{\varepsilon > 0} D_\varepsilon =: D_0$$

is also a closed disc, and one easily sees that it is exactly the set of suitable w_0's.

Let's now consider case of arbitrary A. For every finite subset $F \subset A$ we know that the set of suitable w_0's, say W_F, is a closed disc. We also know that $W_{F_1 \cup F_2} \subset W_{F_1} \cap W_{F_2}$. But now it follows that the family

$$\{W_F \subset \mathbb{H}_+ | F \text{ is finite (non-empty) subset of } A\}$$

satisfies the conditions of the lemma 5.21 and hence the intersection is a closed disc. This intersection is exactly the set of suitable w_0's.

TODO: express this as a hyperbolic disc with canonical center. □

Proof of theorem 5.20. Let us consider the set of all weakly Pick extensions of φ, ordered by restriction. This family trivially satisfies conditions of the Zorn's lemma, so there's a maximal element, φ_0. But by the previous lemma φ_0 must be defined in the whole upper half-plane, since if not, we could extend it to one more point. Finally, by theorem 5.9 the resulting map is Pick function. □

Again, one doesn't really need the Zorn's lemma.

Alternate proof of theorem 5.20. If $\mathbb{H}_+ \setminus A$ is finite, simply apply lemma 5.22 repeatedly. If not, we may pick a countable dense set in $C \subset \mathbb{H}_+ \setminus A$ and extend φ there. As a result we get a map in a dense subset of upper half-plane, which is continuous by lemma 5.10. This means that we have continuous extension to whole upper half-plane, and by continuity also all the Pick matrices of the extension are positive. Finally, by theorem 5.9 the final extension is a Pick function. □

5.5 Schur transform

Properties of Pick functions translate nicely to those of positive maps. Most obvious of these properties is the cone structure: Pick matrix is linear in the function. The automorphims of the upper half-plane correspond to one dimensional projections. Indeed, Pick matrix of map of the form $z \mapsto \frac{az+b}{cz+d}$ is given by

$$\left[\frac{\frac{az_i+b}{cz_i+d} - \frac{a\overline{z_j}+b}{c\overline{z_j}+d}}{z_i - \overline{z_j}}\right]_{i,j=1}^n = \left[\frac{ad-bc}{(cz_i+d)(c\overline{z_i}+d)}\right]_{i,j=1}^n,$$

which is of rank 1. Composition corresponds to Hadamard product: if $\varphi = \varphi_2 \circ \varphi_1$, we have

$$\left[\frac{\varphi(z_i) - \overline{\varphi(z_j)}}{z_i - \overline{z_j}}\right]_{i,j=1}^n = \left[\frac{\varphi_2(\varphi_1(z_i)) - \overline{\varphi_2(\varphi_1(z_j))}}{\varphi_1(z_i) - \overline{\varphi_1(z_j)}}\right]_{i,j=1}^n \circ \left[\frac{\varphi_1(z_i) - \overline{\varphi_1(z_j)}}{z_i - \overline{z_j}}\right]_{i,j=1}^n.$$

There's however more subtle connection, one between Schur transform and, rather appropriately, Schur complement.

If ψ is Schur function such that $\psi(0) = 0$, then, by the Schwarz lemma, also $\psi(z)/z$ is a Schur function. One may again translate this to Pick functions, and get an interesting corollary: if φ is a Pick function such that $\varphi(i) = i$, then also

$$\frac{\varphi(z) - z}{1 + z\varphi(z)}$$

is a Pick function. Actually, this gives a bijection between Pick functions and Pick functions with $\varphi(i) = i \ldots$ almost: $z \mapsto -\frac{1}{z}$ would like to map to constant infinity.

We could form similar bijection for any pair $(z, w) \in \mathcal{H}_+^2$: Pick functions for which $\varphi(z) = w$.

The previous bijection translates nicely to Pick matrices. Take sequence of points $z_1, z_2, \dots, z_n \in \mathcal{H}_+ \setminus \{i\}$. Theorem 5.7 for φ and points $i, z_1, \dots, z_n$ implies that the matrix

$$\begin{bmatrix} [i,-i]_\varphi & [i,\overline{z_1}]_\varphi & \cdots & [i,\overline{z_n}]_\varphi \\ [z_1,-i]_\varphi & [z_1,\overline{z_1}]_\varphi & \cdots & [z_1,\overline{z_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [z_n,-i]_\varphi & [z_n,\overline{z_1}]_\varphi & \cdots & [z_n,\overline{z_n}]_\varphi \end{bmatrix}$$

is positive. Since $[i,-i]_\varphi = 1$ is positive, taking Schur-complement with respect to upper-left corner this is equivalent to the matrix

$$\begin{aligned} &([z_i,\overline{z_j}]_\varphi - [z_i,-i]_\varphi[i,\overline{z_j}]_\varphi)_{1\le i,j\le n} \\ &= \left(\left(\frac{\varphi(z_i)-\overline{\varphi(z_j)}}{z_i-\overline{z_j}}\right) - \left(\frac{\varphi(z_i)+i}{z_i+i}\right)\left(\frac{i-\overline{\varphi(z_j)}}{i-\overline{z_j}}\right)\right)_{1\le i,j\le n} \\ &= \left(\frac{[z_i,\overline{z_j}]_\varphi(z_i\overline{z_j}+1)-1-\varphi(z_i)\overline{\varphi(z_j)}}{(z_i+i)(\overline{z_j}-i)}\right)_{1\le i,j\le n} \end{aligned}$$

being positive. But if one applies the theorem to the function $\frac{\varphi(z)-z}{1+z\varphi(z)}$ and points $z_1, z_2, \dots, z_n$, one arrives at the matrix

$$\left(\frac{[z_i,\overline{z_j}]_\varphi(z_i\overline{z_j}+1)-1-\varphi(z_i)\overline{\varphi(z_j)}}{(1+z_i\varphi(z_i))(1+\overline{z_j}\overline{\varphi(z_j)})}\right)_{1\le i,j\le n}.$$

Two resulting matrices are evidently congruent.

This line of thinking leads to alternate proof for the Pick-Nevanlinna interpolation theorem for finite domain A.

Alternate proof of 5.20 for finite A. We proceed by induction. Without loss of generality we may assume that $z_1 = i = w_1$. We just noted that this matrix being positive is equivalent to smaller matrix, namely

$$\left(\frac{\frac{w_i-z_i}{1+z_iw_i} - \frac{\overline{w_i}-\overline{z_i}}{1+\overline{z_i}\overline{w_i}}}{z_i-\overline{z_j}}\right)_{2\le i,j\le n}$$

being positive. By inductive hypothesis this then means that there exists a Pick function $\tilde{\varphi}$ with $\tilde{\varphi}(z_i) = \frac{w_i-z_i}{1+z_iw_i}$. But then $\varphi(z) = \frac{\tilde{\varphi}(z)+z}{1-z\tilde{\varphi}(z)}$ fits the bill. $\square$

5.6 Compactness

One can lift the previous argument of the Pick-Nevanlinna interpolation theorem to infinite sets by the following compactness result.

Theorem 5.23. *Let $F = \{\varphi_j | j \in J\}$ be a family of Pick functions uniformly bounded in a point. Then F is compact under the topology of pointwise convergence.*

Proof. Note that respective claim holds for Schur functions by Montel's theorem, so hence also holds for Pick functions ... with one small caveat: the Schur function constant -1 corresponds to constant ∞, which isn't proper Pick function. But by the boundedness condition this is not a problem. ☐

There's really nothing special about Pick functions here: in the same way one could prove any family of analytic maps with common (non-whole-of-$\mathbb{C}$) simply connected codomain is compact. Of course we even get locally uniform convergence but that won't be important for us.

One could also give "weaker" proof.

Alternate proof for the theorem 5.23. Fix sequence in F. By lemma 5.10 and the condition values $\sup_{z \in K} |\varphi_j(z)|$ are uniformly bounded for every compact set K. The proof of the lemma also easily implies that φ_j's are uniformly Lipschitz on compact sets. By Arzelà-Ascoli theorem for every compact set K we can find a subsequence convergent in K and exhausting upper half-plane by sequence of nested compact sets and taking the diagonal sequence we get the claim. By proposition 5.2 the limit is also Pick function. ☐

Of course, one could use Arzelà-Ascoli to prove the Montel's theorem in the first place, but the idea is that we can also forget the analyticity and work with Pick matrices on the weak level.

5.7 Pick-Nevanlinna-Herglotz representation theorem

One cannot simply talk about Pick functions without discussing Pick-Nevanlinna-Herglotz representation theorem.[2]

Theorem 5.24. *$\varphi \in \mathcal{P}$, if and only*

$$\varphi(z) = \alpha z + \beta + \int_{-\infty}^{\infty} \left(\frac{1}{\lambda - z} - \frac{\lambda}{\lambda^2 + 1} \right) d\mu(\lambda) \tag{5.25}$$

for some $\alpha \geq 0$ and $\beta \in \mathbb{R}$ and a Borel measure μ with $\int_{-\infty}^{\infty} (\lambda^2 + 1)^{-1} d\mu(\lambda) < \infty$.

[2] TODO: (other names) Schwarz integral formula, Poisson formula

Where does such representation come from? The representation 5.25 tells us that every Pick function is essentially a sum functions of the form $\frac{1}{\lambda - z}$ (and an affine term). If φ is bounded and extends over the real axis, this looks almost like Cauchy's integral formula but the beauty is that μ is real, while with Cauchy's integral formula we can't control the phase. TODO

Theorem 5.26 (Supplement to theorem 5.24)**.** *In the representation 5.25 we have* $\alpha = \lim_{y\to\infty} \frac{\varphi(iy)}{y}$*,* $\beta = \Re(\varphi(i))$ *and we can recover the measure* μ *as*

$$\mu([a,b]) = \frac{1}{\pi} \lim_{x\to 0^+} \lim_{y\to 0^+} \int_{a-x+iy}^{b+x+iy} \Im(\varphi(z)) dz. \tag{5.27}$$

Proof. TODO □

We prove the statement in the chapter TODO. (Intuition on deforming contour, split the function to three pieces)

The integral representation makes many of the properties of the Pick functions obvious. TODO

5.8 Examples of representing measures

5.9 Notes and references

TODO:

- Poincaré metric: discs are discs, Apollonius circle
- Examples of representing measures behind functions and functions behind representing measures
- Spectral commutant lifting theorem
- Use Morera's theorem to prove weak Hindmarsh's theorem

How to rewrite majority of this section:

- Concentrate on Pick-integrals
- Proof open Pick interpolation by extending thinking about Pick-integrals and then showing that the extension is unique, thus showing that everything works.
- Split the Pick measure weakly with $1/(x^2+1)$ (How to do the same thing with k-tone functions?)

Chapter 6

Monotone matrix functions

We already introduced monotone matrix functions in the introduction, but now that we have properly defined and discussed underlying structures we should take a deeper look. As mentioned, monotone matrix functions are sort of generalizations for the standard properties of reals, and this is why we should undestand which of the phenomena for the real functions carry to matrix functions and which do not.

6.1 Basic properties

We first state the definition.

Definition 6.1. Let $(a, b) \subset \mathbb{R}$ be an open, possibly unbounded interval and n positive integer. We say that $f : (a, b) \to \mathbb{R}$ is n-monotone or matrix monotone of order n, if for any $A, B \in \mathcal{H}^n_{(a,b)}$, such that $A \leq B$ we have $f(A) \leq f(B)$.

We will denote the space of n-monotone functions on open interval (a, b) by $P_n(a, b)$. Note that in the notation we don't specify the space V; it doesn't really matter.

Proposition 6.2. *If* $\dim(V) = \dim(V')$, *then f is n-monotone in V if and only if it is n-monotone in V'.*

Proof. The reason is rather clear: inner product spaces of same dimension are isometric. □

One immediately sees that that all the matrix monotone functions are monotone as real functions.

Proposition 6.3. *If $f \in P_n(a, b)$, then f is increasing.*

Proof. Take any $a < x \leq y < b$. Now for $xI, yI \in \mathcal{H}^n_{(a,b)}$ we have $xI \leq yI$ so by definition

$$f(x)I = f(xI) \leq f(yI) = f(y)I,$$

from which it follows that $f(x) \leq f(y)$. This is what we wanted. $\square$

Actually, increasing functions have simple and expected role in n-monotone matrices.

Proposition 6.4. *Let (a, b) be an open interval and $f : (a, b) \to \mathbb{R}$. Then the following are equivalent:*

(i) *f is increasing.*

(ii) *$f \in P_1(a, b)$.*

(iii) *For any positive integer n and commuting $A, B \in \mathcal{H}^n_{(a,b)}$ such that $A \leq B$ we have $f(A) \leq f(B)$.*

Proof. Since 1×1 matrices are for our purposes just reals, $(i) \Leftrightarrow (ii)$ is clear. Also if (iii) holds, since in particular xI and yI commute for every x, y, if $x \leq y$, then $xI \leq yI$, and by assumption hence $f(x)I = f(xI) \leq f(yI) = f(y)I$ so $f(x) \leq f(y)$, which is to say that f is increasing.

Let us then prove that $(i) \Rightarrow (iii)$. If $A \leq B$ and A and B commute, by theorem 2.22 we may write $A = \sum_{i=1}^n a_i P_{v_i}$ and $B = \sum_{i=1}^n b_i P_{v_i}$ for some $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ and $v_1, v_2, \ldots, v_n$, orthonormal basis of V, with $a_i \leq b_i$. But now $f(A) = \sum_{i=1}^n f(a_i) P_{v_i}$ and $\sum_{i=1}^n f(b_i) P_{v_i}$ so

$$f(B) - f(A) = \sum_{i=1}^{n} (f(b_i) - f(a_i)) P_{v_i}$$

is positive, as f is increasing. $\square$

The equivalence of the first two is almost obvious and from this point on we shall identify 1-monotone and increasing functions. But the third point is very important: it is exactly the non-commutative nature which makes the classes of higher order interesting.

Let us then have some examples.

Proposition 6.5. *For any positive integer n, open interval (a, b) and $\alpha, \beta \in \mathbb{R}$ such that $\alpha \geq 0$ we have that $(x \mapsto \alpha x + \beta) \in P_n(a, b)$.*

Proof. Assume that for $A, B \in \mathcal{H}_{(a,b)}$ we have $A \leq B$. Now

$$f(B) - f(A) = (\alpha B + \beta I) - (\alpha A + \beta I) = \alpha(B - A).$$

Since by assumption $B - A \geq$ and $\alpha \geq 0$, also $\alpha(B - A) \geq 0$, so by definition $f(B) \geq f(A)$. This is exactly what we wanted. $\square$

That was easy. It's not very easy to come up with other examples, though.

One of the main properties of the classes of matrix monotone functions has still avoided our discussion, namely the relationship between classes of different orders. We already noticed that matrix monotone functions of all orders all monotonic, or $P_n(a, b) \subset P_1(a, b)$ for any $n \geq 1$. It should not be very surprising that we can make much more precise inclusions.

Proposition 6.6. *For any open interval (a, b) and positive integer n we have $P_{n+1}(a, b) \subset P_n(a, b)$.*

Proof. The idea is that if $\dim(V) \leq \dim(V')$, we can essentially find copy of V in V'. If $A, B \in \mathcal{H}^n(V)$, we can augment A and B to $V' = V \oplus \mathbb{C}$ by setting $A' = A \oplus c$ for any $c \in \mathbb{R}$.

Now if $A \leq B$, by picking any $c \in \mathbb{R}$ we see that $(A \oplus c) \leq (B \oplus c)$. Consequently if $f \in P_{n+1}(a, b)$, we have

$$f(A) \oplus f(c) = f(A \oplus c) \leq f(B \oplus c) = f(B) \oplus f(c),$$

which implies that $f(A) \leq f(B)$. □

One might ask whether these inclusions are strict. It turns out they are, as long as our interval is not the whole $\mathbb{R}$. We will come back to this.

There are also more trivial inclusions: $P_n(a, b) \subset P_n(c, d)$ for any $(a, b) \supset (c, d)$. More interval, more matrices, more restrictions, less functions. To be precise, we only allowed functions with domain (a, b) to the class $P_n(a, b)$, so maybe one should say instead something like: if $(a, b) \supset (c, d)$ and $f \in P_n(a, b)$, then also $f\big|_{(c,d)} \in P_n(c, d)$. We will try not to worry too much about these technicalities.

6.2 Successes

As expected, certain basic properties of monotone functions carry to matrix monotone.

Proposition 6.7. *For any positive integer n and open interval (a, b) the set $P_n(a, b)$ is a convex cone, i.e. it is closed under taking summation and multiplication by non-negative scalars.*

Proof. This is easy: closedness under summation and scalar multiplication with non-negative scalars correspond exaclty to the same property of positive matrices. □

We shall be using especially the previous result a lot.

Similarly we have composition and pointwise limits.

Proposition 6.8. *If $f : (a,b) \to (c,d)$ and $g : (c,d) \to \mathbb{R}$ are n-monotone, so is $g \circ f : (a,b) \to \mathbb{R}$.*

Proof. Fix any $A, B \in \mathcal{H}^n_{(a,b)}$ with $A \leq B$. By assumption $f(A) \leq f(B)$ and $f(A), f(B) \in \mathcal{H}^n_{(c,d)}$ so again by assumption, $g(f(A)) \leq g(f(B))$, our claim. $\square$

Proposition 6.9. *If n-monotone functions $f_i : (a,b) \to \mathbb{R}$ converge pointwise to $f : (a,b) \to \mathbb{R}$ as $i \to \infty$, also f is n-monotone.*

Proof. As always, fix $A, B \in \mathcal{H}^n_{(a,b)}$ with $A \leq B$. Now by assumption

$$f(B) - f(A) = \lim_{i\to\infty} f_i(B) - \lim_{i\to\infty} f_i(A) = \lim_{i\to\infty} (f_i(B) - f_i(A)) \geq 0,$$

so also $f \in P_n(a,b)$. $\square$

6.3 Failures

Most of the common monotone functions fail to be matrix monotone. Let's try some non-examples.

Proposition 6.10. *Function $(x \mapsto x^2)$ is not n-monotone for any $n \geq 2$ and any open interval $(a,b) \subset \mathbb{R}$.*

Proof. Let us first think what goes wrong with the standard proof for the case $n = 1$.

Note that if $A \leq B$,

$$B^2 - A^2 = (B - A)(B + A)$$

is positive as a product of two positive matrices (real numbers).

There are two fatal flaws here when $n > 1$.

- $(B - A)(B + A) = B^2 - A^2 + (BA - AB)$, not $B^2 - A^2$.
- Product of two positive matrices need not be positive.

Note that both of these objections result from the non-commutativity and indeed, both would be fixed should A and B commute.

Let's write $B = A + H$ $(H \geq 0)$. Now we are to investigate

$$(A + H)^2 - A^2 = AH + HA + H^2.$$

Note that $H^2 \geq 0$, but as we have seen in proposition 2.29, $AH + HA$ need not be positive! Also, if H is small enough, H^2 is negligible compared to $AH + HA$. We are ready to formulate our proof strategy: find $A \in \mathcal{H}^n_{(a,b)}$ and $\mathbb{H}^n_+$ such that $AH + HA \not\geq 0$. Then choose parameter $t > 0$ so small that $A + tH \in \mathcal{H}^n(a,b)$ and

$$(A + tH)^2 - A^2 = t(AH + HA + tH^2) \not\geq 0$$

and set the pair $(A, A + tH)$ as the counterexample. TODO (arbitrary intervals) □

As a corollary with get

Corollary 6.11. *The function $\chi_{(0,\infty)}$ is not n-monotone for any $n \geq 2$.*

Proof. If $\chi_{x>0}$ were n-monotone so would be

$$x^2 = \int_0^\infty 2t\chi_{(t,\infty)}(x)dt.$$

□

The function $\chi_{(0,\infty)}$ is in some sense canonical counterexample: every increasing function is more or less positive linear combination of its tranlates, so if monotone functions are not all matrix monotone, the reason is that it is not matrix monotone. For this reason we should really understand why it is not n-monotone for any $n > 1$.

The idea is the following: we are going to construct $A, B \in \mathcal{H}^2$ with the following properties:

1. $A \leq B$
2. A and B have both exactly one positive eigenvalue
3. A and B don't commute

If we can do this, A and B work as counterexamples. Indeed then $\chi_{(0,\infty)}(A) = P_{v_1}$ and $\chi_{(0,\infty)}(B) = P_{w_1}$ where eigenvectors v_1 and w_1 are eigenvectors of A and B corresponding to positive eigenvalues. But $\chi_{(0,\infty)}(A) \not\leq \chi_{(0,\infty)}(B)$ by 2.44.

Constructing such pair is very easy: just take A with eigenvalues -1 and 1 and consider B of the form $A + tH$ for some $H \geq 0$, $t > 0$ and such that A and H do not commute. For small enough H all of the conditions are easily satisfied.

As many properties of real numbers break with real maps, similarly many properties of monotone functions break when $n > 1$. As we saw with the square function example, product of two n-monotone functions need not be n-monotone in general, even if they are both positive functions. Similarly, taking maximums doesn't preserve monotonicity.

Proposition 6.12. *Maximum of two n-monotone functions need not be n-monotone for $n \geq 2$.*

Proof. Again, let's think what goes wrong with the standard proof for $n = 1$.

Fix open interval (a, b), positive integer $n \geq 2$ and two functions $f, g \in P^n(a, b)$. Take any two $A, B \in \mathcal{H}^n_{(a,b)}$ with $A \leq B$. Now $f(A) \leq f(B) \leq \max(f, g)(B)$ and $f(A) \leq f(B) \leq \max(f, g)(B)$. It follows that

$$\max(f, g)(A) = \max(f(A), g(A)) \leq \max(f, g)(B),$$

as we wanted.

Here the flaw is in the expression $\max(f(A), g(A))$: what is maximum of two matrices? This is an interesting question and we will come back to it a bit later, but it turns out that however you try to define it, you can't satisfy the above inequality.

We still need proper counterexamples though. Let's try $f \equiv 0$ and $g = \mathrm{id}$. So far the only n-monotone functions we know are affine functions so that's essentially our only hope for counterexamples.

But now it is rather easy to see that we can take same pair as with $\chi_{(0,\infty)}$ as our counterexample. $\square$

The maximum problem is not too bad; similar statement doesn't hold for k-tone functions in general either and maybe it's more of a pleasent surprise that it holds for usual monotone functions, anyway. But there is very fundamental problem hidden in the square example.

Proposition 6.13. *There exists no $\alpha > 0$, and an open interval $(a, b) \subset \mathbb{R}$ such that $\alpha x + x^2 \in P_n(a, b)$.*

Proof. Adding linear term means just translating domain and codomain, which is not going to help: $x^2 + \alpha x = (x + \frac{\alpha}{2})^2 - \frac{\alpha^2}{4}$. $\square$

Why is this bad? If $f : (a, b) \to \mathbb{R}$ is not too bad (say Lipschitz), for large enough α the function defined by $g(x) = f(x) + \alpha x$ is increasing. But we can't do necessarily do the same thing in the matrix setting even for smooth or analytic functions. Although this might not be such a big surprise or a bad thing in the first place, it is worthwhile to investigate the underlying reason.

Consider the case of entire f and take $A, H \in \mathcal{H}^n_{(a,b)}$ with $H \geq 0$. As observed earlier, we have

$$\begin{aligned}\lim_{t \to 0} \frac{f(A + tH) - f(A)}{t} &= a_1 H \\ &+ a_2(HA + AH) \\ &+ a_3(HA^2 + AHA + A^2H) \\ &+ \ldots .\end{aligned}$$

In the real setting we could just increase a_1 to make the previous expression positive. In the matrix setting there is a problem: note that if H is of rank 1, increasing a_1 means "increasing the right-hand side only to one direction". The point is that if the right-hand side is not positive (map) in the first place, it might be (a priori) non-positive in a big subspace, so rank 1 machinery is not going to save the day.

On the other hand if $n = 2$, for instance, there is not too much room for things to go south. We still, a priori, can't guarantee positivity with a_1, but buffing first few Taylor coefficients starts to somehow affect the whole space.

When n gets larger we have more and more space to worry about, so we should start worrying about more and more Taylor coefficients.

This is all just heuristics, but it leads us to expect two things:

1. Something about things getting worse when n get larger
2. Something about things being not too bad when n is fixed as there is not too much space.

We will later see that both of these phenomena occur.

6.4 Pick functions are matrix monotone

Fortunately the affine functions are not the only matrix monotone functions. We have already discussed a second example.

Proposition 6.14. *We have $(x \mapsto -x^{-1}) \in P_n(a, b)$ for any $(a, b) \not\ni 0$ and $n \geq 1$.*

Proof. The result follows immediately from the theorem 2.46. $\square$

Now also the tranlations $(x \mapsto \frac{1}{\lambda - x}) \in P_n(a, b)$ for any $n \geq 1$ and $(a, b) \not\ni \lambda$ so by the cone property also all the functions of the form

$$x \mapsto \alpha x + \beta + \sum_{i=1}^{m} \frac{t_i}{\lambda_i - x} \tag{6.15}$$

are n-monotone on (a, b) for any $n \geq 1$, $\alpha, t_1, t_2, \ldots, t_m \geq 0$ and $\beta, \lambda_1, \lambda_2, \ldots, \lambda_m$ where $\lambda_1, \lambda_2, \ldots, \lambda_m \notin (a, b)$.

Taking pointwise limits we arrive at

Theorem 6.16. *If $f \in P(a, b)$, then $f \in P_n(a, b)$ for every $n \geq 1$.*

Proof. The reason is of course that every element of $P(a, b)$ is pointwise limit of the functions of the form 6.15. Why is that exactly? One should be a bit careful as things aren't compactly supported. TODO (explain why everything is compact afterall) $\square$

6.5 Derivative and Loewner's characterization

As in the real case, also in the matrix world we may characterize monotonicity with derivatives.

Theorem 6.17. *Let $f \in C^1(a,b)$ and $n \geq 1$. Then the following are equivalent:*

(i) $f \in P_n(a,b)$.

(ii) For any $A \in \mathcal{H}^n_{(a,b)}$ *and* $H \geq 0$ *we have*

$$D^1_n f_A(H) \geq 0.$$

(iii) For any $A \in \mathcal{H}^n_{(a,b)}$ *and P one dimensional (orthogonal) projection we have*

$$D^1_n f_A(P) \geq 0.$$

(iv) For any $A \in \mathcal{H}^n_{(a,b)}$, $H \geq 0$ *and* $v \in V$ *the map*

$$t \mapsto \langle f(A + tH)v, v\rangle$$

is increasing.

(v) For any $A \in \mathcal{H}^n_{(a,b)}$, *$P$ one dimensional (orthogonal) projection and* $v \in V$ *the map*

$$t \mapsto \langle f(A + tP)v, v\rangle$$

is increasing.

Proof. $(i) \Leftrightarrow (ii)$ follows from Spectral theorem linearity of D^1_n, and $(iii) \Leftrightarrow (iv)$ from Spectral theorem. One also easily sees that the derivative of the map

$$t \mapsto \langle f(A + tH)v, v\rangle$$

is the map

$$t \mapsto \langle D^1_n f_A(H)v, v\rangle$$

so also $(i) \Leftrightarrow (iii)$ is clear. □

We already noticed that we can express $D^1_n f_A(H) = ([\lambda_i, \lambda_j]_f)_{1 \leq i,j \leq n} \circ H$, where Hadamard product is taken along eigenbasis of A. We can however make the following simple observation:

Lemma 6.18. *Let $A \in \mathcal{H}$. Then $A \geq 0$, if and only if $A \circ B$ for every $B \geq 0$.*

Proof. If the Hadamard product is along $(e_i)_{1\leq i\leq n}$, we have $A = A \circ \left(\sqrt{n} P_{\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} e_i}\right)$, and hence have the "if". Note that for only if we only need to verify the inequality for A and B both one dimensional projections. But now one easily sees that $A \circ B$ is non-negative multiple of projection. $\square$

We hence have the following characterization.

Theorem 6.19. *$f \in P_n(a, b) \cap C^1(a, b)$, if and only if the matrix*

$$([\lambda_i, \lambda_j]_f)_{i,j} \geq 0 \tag{6.20}$$

for any $\lambda_1, \lambda_2, \ldots, \lambda_n \in (a, b)$.

This is the original characterization by Loewner, and it is pretty much just saying that function is matrix monotone if its (matrix) derivative is positive. The matrix 6.20 is called, appropriately, Loewner matrix (of function f on points $\lambda_1, \lambda_2, \ldots, \lambda_n$). Using the characterization it is in general not very easy to check that the function is n-monotone: we would have to check positivity of the matrix for any tuple on the interval. Also the characterization is not local one: in order to check monotonicity we need to know the behaviour on the whole interval. This is just a reflection of the fact that the space in which we are working on, space of real maps, is itself in a way spread around the interval.

6.6 Local characterization

It nevertheless turns out that n-monotonicity is a local property.

Proposition 6.21. *For any $n \geq 1$, P_n is a local property meaning that whenever $f \in P_n(a, b)$ and $P_n(c, d)$ for some $a < c < b < d$, then also $f \in P_n(a, d)$.*

The reason for this is hidden in the Loewner matrix.

Note that Loewner matrix is essentially something we saw before: it is just a Pick matrix when all the points are on the real line. We observed before that positivity of the Pick matrix was some kind of manifestation of the strength of the Cauchy's integral formula. Namely, if f happens to analytic, in some suitable set, we can write

$$([\lambda_i, \lambda_j]_f)_{i,j} = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{(z - \lambda_i)(z - \lambda_j)} dz.$$

Now the positivity the matrix means that for any $c_1, c_2, \ldots, c_n \in \mathbb{C}$ the quantity

$$\sum_{1\leq i,j\leq n} c_i \overline{c_j} [\lambda_i, \lambda_j]_f = \frac{1}{2\pi i} \int_\gamma f(z) \left(\sum_{1\leq i\leq n} \frac{c_i}{z - \lambda_i}\right) \left(\sum_{1\leq i\leq n} \frac{\overline{c_i}}{z - \lambda_i}\right) dz.$$

But here's the trick: we may write

$$\sum_{1\le i\le n} \frac{c_i}{z-\lambda_i} = \frac{q(z)}{\prod_{1\le i\le n}(z-\lambda_i)} = \frac{q(z)}{p_\Lambda(z)}$$

for some polynomial of degree less than n, and indeed, if the λ's are distinct there's a one-to-one correspondence between polynomials q and the λ's. It follows that we may rewrite the integral as

$$\frac{1}{2\pi i}\int_\gamma f(z)\frac{q(z)\overline{q(\overline{z})}}{p_\Lambda(z)^2}dz.$$

Note that $z \mapsto q(z)\overline{q(\overline{z})})$ is a polynomial of degree at most $(2n-2)$ non-negative on the real line. Easy application of the Fundamental theorem of algebra reveals that all such polynomials are actually of the previous form.

Lemma 6.22. *h is polynomial of degree at most $(2n-2)$ non-negative on $\mathbb{R}$, if and only if it is of the form $p(z)\overline{p(\overline{z})}$ for some complex polynomial of degree of at most $(n-1)$.*

Proof. It is easy to see that all of the polynomials of the specific form fit the bill. Conversely, if h is non-negative on real axis, it's roots all appear in pairs: either with strict complex conjugate pairs, of pairs of double real roots. We may take p to be $\sqrt{a_n}\prod(z-z_i)$ where z_i range over representatives of all the pairs and a_n is the leading coefficient of h. $\square$

Write $h(z) = q(z)\overline{q(\overline{z})})$.

Finally note that resulting expression,

$$\frac{1}{2\pi i}\int_\gamma f(z)\frac{h(z)}{p_\Lambda(z)^2}dz$$

is nothing but the divided difference of the function fh at points $\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n, \lambda_n$. By 3.21 this extends to all $C^1(a,b)$. We would like to conclude that fh is $(2n-1)$-tone, but unfortunately we only know the non-negativity of the divided differences of order $(2n-1)$ special sets of tuples. It however turns out that this is enough, as can be seen by using the same trick as in the proof of theorem 3.31.

Lemma 6.23. *Let k and n be positive integers and $d_1, d_2, \ldots, d_m$ be positive integers with $d_1 + d_2 + \ldots + d_m = k+1$. Assume that $n \ge (\max_{1\le i\le m} d_i) - 1$. Let $f \in C^n(a,b)$. Then the following are equivalent.*

(i) f is k-tone.

(ii) For any $a < x_1 < x_2 < \ldots < x_m < b$ we have

$$[x_1, x_1, \ldots, x_1, x_2, \ldots, x_2, \ldots, x_m, \ldots, x_m]_f \geq 0$$

where x_i appears d_i times.

Proof. $(i) \Rightarrow (ii)$ is clear.

For the other direction let's take any $a < y_1 < y_2 < \ldots < y_n < y_{n+1} < b$. We need to prove that

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f \geq 0.$$

The idea is to bunch the variables together using the mean value theorem.

Let's consider the function $g_1(x) = [x, y_{d_1+1}, y_{d_1+2}, \ldots, y_{k+1}]_f$. In terms of g_1 we need to prove that

$$[y_1, y_2, \ldots, y_{d+1}]_{g_1} \geq 0.$$

Since $f \in C^{d_1-1}(a, b)$, $g_1 \in C^{d_1-1}(a, y_{d_1+1})$ and hence by the mean value theorem we have

$$[y_1, y_2, \ldots, y_{d_1}]_g = [x_1, x_1, \ldots, x_1]_{g_1}$$

for some $a < x_1 < y_{d+1}$. Consequently,

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f = [x_1, x_1, \ldots, x_1, y_{d_1+1}, \ldots, y_{k+1}]_f.$$

Next step is to bunch together the next d_2 terms: consider now the map $g_2(x) = [x_1, x_1, \ldots, x_1, x, y_{d_1+d_2+1,\ldots,y_{k+1}}]_f$ and observe that we are to verify that

$$[y_{d_1+1}, \ldots, y_{d_1+d_2}]_{g_2} \geq 0.$$

Again use mean value theorem to replace $y_{d_1+1}, \ldots, y_{d_1+d_2}$ by x_2's.

One should be bit careful here: the number x_1 certainly depends on all the y's, so once we have fixed it we can't say that

$$[y'_{d_1+1}, y'_{d_1+2}, \ldots, y'_{d_1+d_2}]_{g_2} = [y_1, \ldots, y_{d_1}, y'_{d_1+1}, \ldots, y'_{d_1+d_2}, y_{d_1+d_2+1}, \ldots, y_{k+1}]_f,$$

for instance, anymore. This is of course not a problem.

Making m steps of the previous form we finally find numbers $x_1, x_2, \ldots, x_m$ such that

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f = [x_1, x_1, \ldots, x_1, x_2, \ldots, x_2, \ldots, x_m, \ldots, x_m]_f \geq 0$$

and we are done. $\square$

We have hence the following.

Theorem 6.24. *Let $n \geq 1$ and $f \in C^1(a,b)$. Then $f \in P_n(a,b)$, if and only if fh is $(2n-1)$-tone whenever h is polynomial of degree at most $(2n-2)$, non-negative on real line.*

This result has a curious corollary.

Corollary 6.25. *If $n \geq 1$ and fh is $(2n+1)$:tone for every polynomial h of degree at most $2n$, then fh is $(2n-1)$-tone for every polynomial h of degree at most $(2n-2)$. In particular f is k-tone for every $k = 1, 3, 5, \ldots, 2n-3, 2n-1, 2n+1$.*

Although we strictly speaking only proved this corollary for $f \in C^1$, it holds true without extra assumptions, as can be seen with the following alternate proof.

Proof of corollary 6.25. Take any h, a polynomial of at most $(2n-2)$ non-negative on real axis and points $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n} < b$. We should prove that

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \geq 0.$$

The idea is the following: if f is C^1, we have

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} = [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, t, t]_{fh(\cdot - t)^2} \geq 0,$$

Now, actually $fh(\cdot - t)^2$ is always differentiable at t, so the previous at least should hold without the smoothness TODO, but one can take safer route. For any $a < t, s < b$ we have

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}, t, s]_{fh(\cdot - t)^2} \geq 0.$$

Expanding this Leibniz rule leads to

$$\begin{aligned} 0 \leq\ & [\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \\ & + (s-t)[\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s]_{fh} \end{aligned}$$

But by choosing t and s suitably we can definitely make the second term non-positive, so the first term is non-negative, as we wanted. Indeed choose first arbitrary s and then choose t on (a,s) or on (s,b), depending on the sign of the divided difference. Or if one so prefers, we have

$$\begin{aligned} & 2[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \\ =& [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s+r, s]_{fh(\cdot - s - r)^2} \\ +& [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s-r, s]_{fh(\cdot - s + r)^2} \end{aligned}$$

for small enough r.

Final claim follows by setting $h \equiv 1$. $\square$

6.7 Regularity

Finally, one might ask whether we can get rid of C^1-assumption, and it turns out that we can. This leads to question: why should matrix monotone functions be C^1? Why should they be even continuous?

The continuity question we actually almost answered previously. There are not very many ways increasing function can be discontinuous. For every increasing function we have $\lim_{t\to 0^-} f(t) \leq f(0) \leq \lim_{t\to 0^+} f(t)$, and discontinuity means that at least one of the inequalities is strict. Thus every discontinuous increasing function looks like step function at discontinuity.

Lemma 6.26. *Let $f \in P_n(a, b)$ for some $n > 1$. Then f is continuous.*

Proof. Without loss of generality it suffices to check the continuity at 0. We may decompose $f(t) = c(t) + c_1\chi_{\{x>0\}}(t) + c_2\chi_{\{x=0\}}(t)$, where c is continuous at 0, and $0 \leq c_2 \leq c_1$. We aim to prove that $c_1 = 0$, in which case f is also continuous at 0.

As in the proof of corollary 6.11 we $A \leq B$ (with non-zero eigenvalues) such that $\chi_{x>0}(A) \not\leq \chi_{x>0}(B)$. Now for small enough $t > 0$, we have $tA, tB \in \mathcal{H}_{(a,b)}$ and we have

$$\begin{aligned} f(tB) - f(tA) = c(tB) - c(tA) \\ + c_1\left(\chi_{\{x>0\}}(tB) - \chi_{\{x>0\}}(tA)\right) \\ + c_2\left(\chi_{\{x=0\}}(tB) - \chi_{\{x=0\}}(tA)\right) \end{aligned}$$

The last difference vanishes as tA and tB have non-zero eigenvalues. By continuity of c the first difference tends to zero with decreasing t. The second difference is independent of t, so as $f \in P_n(a, b)$, we conclude that

$$c_1\left(\chi_{\{x>0\}}(B) - \chi_{\{x>0\}}(A)\right) \geq 0,$$

which is only possible if $c_1 = 0$. $\square$

Differentiablity is much more subtle however. Using similar ideas one could prove the matrix monotone function cannot have corners, i.e. cannot look locally like multiple of absolute value + affine function (up to first order), but there are far worse ways in which continuous function can be non-differentiable.

This is the point where divided differences come in. If one interprets matrix monotonicity solely with formulas, function being matrix monotone just means that some special set of linear combinations of the function values are positive. But this is precisely the type of condition we had with k-tone functions, and regularity started appearing. It is hence not such a big surprise that there should be some kind smoothness going on.

6.8 Main Theorem

Theorem 6.27. *Let $n \geq 1$. Then $f \in P_n(a, b)$, if and only if fh is $(2n-1)$-tone whenever h is polynomial of degree at most $(2n-2)$, non-negative on the real line.*

Proof. For the version with extra assumption, the starting point was to take derivative of the matrix function. Although we now cannot do that, we can try to replicate the proof otherwise.

Instead of proving that

$$[\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n, \lambda_n]_{fh} \geq 0$$

for any $a < \lambda_1, \lambda_2, \ldots, \lambda_{n-1}, \lambda_n < b$, we should prove that

$$[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh} \geq 0.$$

λ's should be eigenvalues of some map, but now there are $2n$ of them. Natural guess would be that they are eigenvalues of two maps, A and B.

But now everything starts to make sense: whenever A, B with $A \leq B$ and $w \in V$ the quantity

$$\langle (f(B) - f(A))w, w \rangle$$

is non-negative. On the other hand this can be expanded as some kind of linear combination of values of f at eigenvalues of A and B. Same is true for the divided differences, so there might be a chance to choose A, B and w such that

$$\langle (f(B) - f(A))w, w \rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}.$$

Moreover, should we find some kind of correspondence between triplets (A, B, w) and pairs $((\lambda_i)_{i=1}^{2n}, h)$, we would be done. This is the content of the main lemma.

Lemma 6.28. *If $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ and h is polynomial of degree at most $(2n-2)$ non-negative on the real line, we may find a strict projection pair (A, B) such that*

$$\langle (f(B) - f(A))w, w \rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}$$

for any $f : (a, b) \to \mathbb{R}$.

Conversely, if (A, B) is a strict projection pair and $w \in V$, then there exists $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ and polynomial h of degree at most $(2n-2)$, non-negative on the real line such that for any $f : (a, b) \to \mathbb{R}$ we have

$$\langle (f(B) - f(A))w, w \rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}.$$

Before proving the lemma we show how it implies the theorem.

Assume first that $f \in P_n(a,b)$. We need to prove that fh is $(2n-1)$-tone for any h polynomial of degree at most $(2n-2)$ non-negative on real line. But any divided difference of such fh can be expressed by the main lemma 6.28 as $\langle (f(B)-f(A))w, w\rangle$ for some projection pair (A,B), and the previous is non-negative by the assumption.

Conversely, assume that fh is $(2n-1)$-tone for any suitable h and take any $A \leq B$. Write $B - A = \sum_{i=1}^{n} c_i P_{v_i}$ for some $c_i \geq 0$. To prove that $f(B)-f(A) \geq 0$ we simply need to prove that $f(A + \sum_{i=1}^{k} c_i P_{v_i}) - f(A + \sum_{i=1}^{k-1} c_i P_{v_i}) \geq 0$ for any $1 \leq k \leq n$, as $f(B)-f(A)$ is sum of such terms. We may hence assume that (A,B) projection pair.

We may also assume that (A,B) is strict. Indeed, if this would not be the case, we could decompose $V = \text{span}\{v_1\} \oplus V'$, where v_1 is the eigenvector, and factorize $A = A_{\text{span}\{v_1\}} \oplus A_{V'}$ and $P_w = 0 \oplus (P_w)_{V'}$. But now checking that $f(B)-f(A) \geq 0$ boils down to checking that $f(B_{V'}) - f(A_{V'}) \geq 0$, which would follow if we could prove that $f \in P_{n-1}(a,b)$. But this follows if we add the sentence "We induct on n." as the first sentence of this proof and use lemma 6.25.

Finally in this case, by the lemma 6.28 we may find $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ such that

$$\langle (f(B)-f(A))w, w\rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh} \geq 0$$

and we are finally done.

In the "if"-direction we could alternatively make use of the continuity of f, which is guaranteed by the lemma 3.23

□

Let us then complete proof by proving the lemma 6.28.

Proof of lemma 6.28. The proof is based on lemmas 2.51 and 2.52. To find the connection we first assume f is entire. Then if and (A,B) is a strict projetion pair with $B - A = vv^*$ for some $v \in V$ and $w \in V$ we have

$$\begin{aligned}
&= \langle (f(B)-f(A))w, w\rangle \\
&= \frac{1}{2\pi i}\int_\gamma \langle (zI-B)^{-1}v, w\rangle \langle (zI-A)^{-1}w, v\rangle f(z)dz \\
&= \frac{1}{2\pi i}\int_\gamma \frac{\det(zI-A)\langle (zI-B)^{-1}v, w\rangle \det(zI-B)\langle (zI-A)^{-1}w, v\rangle}{\det(zI-A)\det(zI-B)} f(z)dz.
\end{aligned}$$

The integrand equals

$$\frac{h(z)}{\prod_{i=1}^{n}(z-\lambda_i(A))\prod_{i=1}^{n}(z-\lambda_i(B))} f(z),$$

where $h(z) = \det(zI - B)\langle (zI - B)^{-1}v, w\rangle \det(zI - A)\langle (zI - A)^{-1}w, v\rangle$ and hence

$$\langle (f(B) - f(A))w, w\rangle = [\lambda_1(A), \ldots, \lambda_n(A), \lambda_1(B), \ldots, \lambda_n(B)]_{fh}.$$

Note that this identity evidently holds without any extra smoothness assumptions.

Now when (A, B) ranges over all strict projection pairs, the permutations of tuples

$$(\lambda_1(A), \ldots, \lambda_n(A), \lambda_1(B), \ldots, \lambda_n(B)) \tag{6.29}$$

range over all tuples of distinct numbers on (a, b). Hence to prove the lemma, we should prove that for fixed strict projection pair (A, B), as w ranges over V, h ranges over all polynomials of degree at most $(2n-2)$, non-negative on $\mathbb{R}$. This follows from lemma 6.22 and the following observation.

Lemma 6.30. *If (A, B) is a projection pair with $B - A = vv^*$ then*

$$\det(zI - A)(zI - A)^{-1}v = \det(zI - B)(zI - B)^{-1}v$$

Proof. As $zI - A = zI - B + vv^*$, multiplying both sides from left by $(zI - A)$ leads to the equivalent

$$\det(zI - A)v = \det(zI - B)(1 + \langle (zI - B)^{-1}v, v\rangle)v$$

which follows from 2.53. □

It follows that if $p(z) = \det(zI - B)\langle (zI - B)^{-1}v, w\rangle$, $h(z) = p(z)\overline{p(\overline{z})}$, so to finish the proof, we need only need to observe that when w ranges over V, $\det(zI - B)\langle (zI - B)^{-1}v, w\rangle$'s range over all complex polynomials of degree at most $(n-1)$. But this is clear as components of $\det(zI - A)(zI - A)^{-1}v$ with respect to eigenbasis of A, $(e_i)_{i=1}^n$ are $p_j(z) = \prod_{i\neq j}(z - \lambda_i(B))\langle v, e_i\rangle$, which are clearly linearly independent polynomials over $\mathbb{C}$.

To recap, the map

$$\begin{aligned} V &\to P_{n-1}(\mathbb{C}) = \{\text{Complex polynomials of degree at most } (n-1)\} \\ w &\mapsto \det(zI - A)\langle (zI - A)^{-1}v, w\rangle \end{aligned}$$

is antilinear bijection and the map

$$\begin{aligned} P_{n-1}(\mathbb{C}) &\to \{\text{Complex polynomials of degree at most } (2n-2) \text{ non-negative on } \mathbb{R}\} \\ p(z) &\mapsto p(z)\overline{p(\overline{z})} \end{aligned}$$

is surjection: composition of these maps is the correspondence between w and h. □

What's the moral of the story? If one unwraps all the definitions, matrix monotonicity is about positivity of some linear combinations of function values. Which linear combinations exactly? That is (more or less) explained in the main theorem.

6.9 A bit of history

Theorem 6.27 is usually stated in somewhat different terms. Functions of the form fh being $(2n-1)$-tone for some polynomials h can be also understood as certain matrix being positive, *Dobsch matrix*. Dobsch matrix (of order n) of $f : (a,b) \to \mathbb{R}$ at point $t \in (a,b)$ is the matrix

$$\begin{bmatrix} \frac{f'(t)}{1!} & \frac{f^{(2)}(t)}{2!} & \cdots & \frac{f^{(n)}(t)}{n!} \\ \frac{f^{(2)}(t)}{2!} & \frac{f^{(3)}(t)}{3!} & \cdots & \frac{f^{(n+1)}(t)}{(n+1)!} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{f^{(n)}(t)}{n!} & \frac{f^{(n+1)}(t)}{(n+1)!} & \cdots & \frac{f^{(2n-1)}(t)}{(2n-1)!} \end{bmatrix} \tag{6.31}$$

$$= \begin{bmatrix} [t,t]_f & [t,t,t]_f & \cdots & [t,\ldots,t]_f \\ [t,t,t]_f & [t,t,t,t]_f & \cdots & [t,t,\ldots,t]_f \\ \vdots & \vdots & \ddots & \vdots \\ [t,\ldots,t]_f & [t,t,\ldots,t]_f & \cdots & [t,t,t,\ldots,t]_f. \end{bmatrix}$$

Now an alternative version of 6.27 reads as follows.

Theorem 6.32. *Let $n \geq 1$. Then $f \in P_n(a,b) \cap C^{2n-1}(a,b)$, if and only if all Dobsch matrices of f of order n are positive for every $t \in (a,b)$.*

Proof. By simple computation

$$\frac{1}{(2n-1)!} \frac{d^{2n-1} \left(f(t) \left(\sum_{i=1}^n c_i t^{n-i}\right) \left(\sum_{i=1}^n \overline{c_i} t^{n-i}\right)\right)}{dt^{2n-1}} \Bigg|_{t=0} = \sum_{i,j=1}^n c_i \overline{c_j} \frac{f^{(i+j-1)}(0)}{(i+j-1)!}.$$

□

One can again get rid of the smoothness assumption by some careful considerations. TODO

6.10 Loewner's theorem

In addition to characterizing n-monotone functions, by theorem 6.19, the classes $P_n(a,b)$, Loewner characterized the classes $P_\infty(a,b)$.

Theorem 6.33. *$f \in P_\infty(a,b)$, if and only if there exist Pick function φ extending over the interval (a,b) such that $\varphi\big|_{(a,b)} = f$.*

Proof. The "if" direction is not too hard: the Loewner matrices are essentially limits of Pick matrices so the result follows rather immediately from 6.19.

The "only if" is the tricky part. Theorem 6.19 tells us that the Dobsch matrices are positive on (a, b). If we can somehow show that $f \in C^{\omega}(a, b)$, then we see that all points of (a, b) are Pick points of f, and we can extend it to weakly Pick function on some open set of upper half-plane, from which it extends to unique Pick function by Pick–Nevanlinna theorem 5.12.

It suffices to proof the following result.

Lemma 6.34. *Let $f \in C^{\infty}(a, b)$ such that $f^{(2n-1)}(t) \geq 0$ for every $t \in (a, b)$. Then $f \in C^{\omega}(a, b)$.*

Proof. We shall verify the conditions of the theorem 3.33.

The trick is first show that we have bound of the form $|f^{(n)}(t)| \leq n! C^{n+1}$ for odd n, and then use the following result.

Lemma 6.35. *Let $f \in C^2(a, b)$ such that $|f(x)| \leq M_0$ and $|f^{(2)}(x)| \leq M_2$ for any $x \in (a, b)$. Then*

$$|f'(x)| \leq \max\left(2\sqrt{M_0 M_2}, \frac{8M_0}{b-a}\right)$$

for any $x \in (a, b)$.

Proof. Take any $x_0 \in (a, b)$ and set $f'(x_0) = c$: we shall prove the given bound of c. Without loss of generality we may assume that $c \geq 0$ and $x_0 \leq \frac{a+b}{2}$. The idea is that as $f^{(2)}$ is not too big, f' has to be positive and reasonably big interval around the point x_0 which means that f has to increase a lot around x_0. By the assumption it can't increase more than $2M_0$, however.

To make this argument precise and effective, we split into too cases.

1. $M_2(b - x_0) > c$: this means that we have

$$f'(x) \geq c - M_2(x - x_0)$$

for $x_0 \leq x \leq \frac{c}{M_2} + x_0$ and hence

$$2M_0 \geq f\left(\frac{c}{M_2} + x_0\right) - f(x_0) \geq \int_{x_0}^{\frac{c}{M_2}+x_0} (c - M_2(x - x_0))\, dx \geq \frac{c^2}{2M_2},$$

which yields the first inequality.

2. $M_2(b-x_0) \leq c$: now we have

$$f'(x) \geq c\frac{b-x}{b-x_0},$$

for every $x_0 \leq x < b$

$$2M_0 \geq f(x) - f(x_0) \geq \int_{x_0}^{x} c\frac{b-x}{b-x_0}dx \geq \frac{c}{2(b-x_0)}\left((b-x_0)^2 - (b-x)^2\right).$$

Letting $x \to b$ and using $(b-x_0) \geq \frac{b-a}{2}$ we get the second inequality.

TODO: pictures of function and it's derivatives TODO: better proof □

To prove the bound for odd n, we would like to play the same game as in the proof of lemma 3.31, but the unfortunate thing is that the even order terms are breaking the inequality. We can salvage the situation by getting rid of them. Assume first that $0 \in (a,b)$. Trick is to consider the Taylor expansion for $f(x) - f(-x)$, centered at 0, instead:

$$f(x) - f(-x) = 2\left(\sum_{i=1}^{n} \frac{f^{(2i-1)}(0)}{(2i-1)!}x^{2i-1}\right) + \int_0^x \frac{f^{(2n+1)}(t) + f^{(2n+1)}(-t)}{(2n)!}(x-t)^{2n}dt.$$

But now we can simply follow the same argument. □

□

TODO:

- Why smoothness
- Examples
- Pick functions are monotone
- Heaviside function
- Trace inequalities: if f is monotone/convex then $\mathrm{tr}f$ is monotone/convex. Proof idea: we may write $\mathrm{tr}f$ as a limit of finite sum of translations of Heaviside functions (monotone case) or absolute values (convex case), so its sufficient to prove the claim for these functions. For monotone case it hence suffices to prove that if $A \leq B$, B has at least as many non-negative eigenvalues as A. But this is clear by subspace characterization of non-negative eigenvalues. For convex case, it suffices to prove

that $\mathrm{tr}|A| + \mathrm{tr}|B| \geq \mathrm{tr}|A+B|$ for any $A, B \in \mathcal{H}^n(a,b)$. For this, note that if $(e_i)_{i=1}^n$ is eigenbasis of $A+B$, we have

$$\begin{aligned} \mathrm{tr}|A+B| &= \sum_{i=1}^n \langle |A+B|e_i, e_i\rangle \\ = \sum_{i=1}^n |\langle (A+B)e_i, e_i\rangle| &\leq \sum_{i=1}^n |\langle Ae_i, e_i\rangle| + \sum_{i=1}^n |\langle Be_i, e_i\rangle| \\ \leq \sum_{i=1}^n \langle |A|e_i, e_i\rangle + \sum_{i=1}^n \langle |B|e_i, e_i\rangle &= \mathrm{tr}|A| + \mathrm{tr}|B| \end{aligned}$$

- What about trace inequalities for k-tone functions? Eigen-package seems to find a counterexample for 6-tone functions and $n = 2$, but it's hard to see if there's enough numerical stability. At divided differences of polynomials vanish. First non-trivial question would be: If $A_j = A + jH$ for $0 \leq j \leq 3$ and $H \geq 0$. Then is it necessarily the case that
$$\mathrm{tr}\left(A_3|A_3| - 3A_2|A_2| + 3A_1|A_1| - A_0|A_0|\right) \geq 0?$$
This would imply that 3-tone functions would lift to trace 3-tone functions. Maybe expressing this as a contour integral from $-i\infty \to i\infty$ a same tricks as in the paper. First projection case: H is projection. Or: approximate by integrals of heat kernels. It should be sufficient to proof things for k-fold integrals or heat kernel, or by scaling just for gaussian function.

- How is the previous related to the $|\cdot|$ not being operator-convex: quadratic form inequality for eigenvectors is not enough.

- The previous also implies that
$$f(Q_A(v)) \leq Q_{f(A)}(v)$$
for any convex f. Using this and Minkowski one sees that p-schatten norms are indeed norms.

- For f, g generalization (Look at $h(X) = g(\mathrm{tr} f(X))$) we need that f is convex. What else? h is convex if it is convex for diagonalizable matrices and f is convex and g increasing. For the diagonalizable maps it is sufficient that f is increasing and $g = f^{-1}$ and $\log \circ f \circ \exp$ is convex.

- Von Neumann trace inequality, more trace inequalities.

- On Generalizations of Minkowski's Inequality in the Form of a Triangle Inequality, Mulholland
- There should nice proof for Loewner theorem, like the blog post for Bernstein's big theorem.

Chapter 7

Matrix k-tone functions

7.1 Matrix convex functions

Having charaterized matrix monotone functions it is natural to ask what happens with convex functions.

Definition 7.1. We say that $f : (a, b) \to \mathbb{R}$ is matrix convex of order n if for every $A, B \in \mathcal{H}^n(V)$ and $t \in [0, 1]$ we have

$$f(tA + (1-1)B) \leq tf(A) + (1-t)f(B).$$

We denote class of matrix convex functions of order n on interval (a, b) by $P_n^2(a, b)$. Many of the properties of matrix monotone functions translate immediately to matrix convexity.

Proposition 7.2. *Let $(a, b) \subset \mathbb{R}$ be an open interval $n \geq 1$. Then*

1. *$P_n^2(a, b)$ is a convex closed cone.*
2. *$P_{n+1}^2(a, b) \subset P_n^2(a, b)$.*
3. *$(x \mapsto \alpha_2 x^2 + \alpha_1 x + \alpha_0) \in P_n^2(a, b)$ if $\alpha_n \geq 0$.*
4. *$(x \mapsto |x|^{-1}) \in P_n^2(a, b)$ whenever $0 \notin (a, b)$.*

Proof. 1. This is clear.

2. The proof is essentially the same as in the monotone case.

3. We have
$$\begin{aligned}
&t(\alpha_2 A^2+\alpha_1 A+\alpha_0 I)+(1-t)(\alpha_2 B^2+\alpha_1 B+\alpha_0 I)\\
&-(\alpha_2(tA+(1-t)B)^2+\alpha_1(tA+(1-t)B)+\alpha_0 I)\\
=&\alpha_2 t(1-t)(A-B)^2.
\end{aligned}$$

4. It's is clearly sufficient to prove that $x \mapsto x^{-1}$ is convex on $(0,\infty)$. For this we should prove that for any $A, B > 0$ and $t \in [0,1]$ we have
$$tA^{-1}+(1-t)B^{-1} \geq (tA+(1-t)B)^{-1}.$$
By doing the congruence trick, i.e. $*$-conjugating by $A^{\frac{1}{2}}$ and setting $X = A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ we are left with
$$t+(1-t)X^{-1} \geq (t+(1-t)X)^{-1},$$
which would follow if we can prove the respective scalar inequality. But
$$t+(1-t)x^{-1}-(t+(1-t)x)^{-1} = \frac{t(1-t)}{x(t+(1-t)x)}(x-1)^2,$$
so the scalar inequality is true.

$\square$

With linear combinations of the previous examples we can again build large number of n-convex functions. TODO

By now it ought to be no surprise that not every convex function is matrix convex. Canonical counterexample is absolute value. In turns out that we have

Proposition 7.3. *Let $v, w \in V \setminus \{0\}$. Then*
$$P_v + P_w \geq |P_v - P_w|,$$
if and only if v and w are parallel or orthogonal, i.e. if and only if P_v and P_w commute.

Proof. As everything is happening in (at most) two dimensional subspace of V, we may assume that V is two dimensinal in the first place. Note that in this case $P_v - P_w$ has 0 trace, so its eigenvalues are additive inverses of each other. Consequently the absolute value is multiple of identity.

It follows that both sides of the inequality commute, and as they are both positive, it suffices to check when the squared inequality holds. This leads to an equivalent inequality
$$P_v P_w + P_w P_v \geq 0,$$
which was already discussed in 2.26. $\square$

7.2 Convexity and the second derivative

So far everything has worked pretty much the same way as with the monotone functions, but with derivative things to start to look very different. Theorem 6.17 has natural analog in the convex case.

Theorem 7.4. *Let $n \geq 1$ and $f \in C^2(a,b)$. Then the following are equivalent:*

(i) $f \in P_n^2(a,b)$.

(ii) For any $A, H \in \mathcal{H}_{(a,b)}^n$ we have

$$D_n^2 f_A(H) \geq 0.$$

(iii) For any $A, H \in \mathcal{H}_{(a,b)}^n$ and $v \in V$ the map

$$t \mapsto \langle f(A+tH)v, v\rangle$$

is convex.

Now, there is a big problem here: there are only three conditions in this theorem. In the monotone case we could take H to be projection, but with the convex case this is not (a priori) possible anymore. Recall that working with projections was rather easy (we could even have very explicit formulas for everything) but with general positive maps similar arguments are hopeless. TODO

TODO

7.3 Matrix k-tone functions

After having defined the notion of k-tone function in the real setting, it is natural to ponder what happens with matrix setting. Defining the notion itself is already a bit cumbersome: with monotone and convex functions the usual definitions make immediately sense but divided differences cause some problems. One cannot simply say that f is matrix monotone if

$$[A,B]_f = \frac{f(B) - f(A)}{B - A},$$

since the right-hand side doesn't make much sense. We can however use an equivalent definition from the theorem 6.17.

7.4 Basic properties

Definition 7.5. We say that $f : (a, b) \to \mathbb{R}$ is matrix k-tone of order n if for every $A \in \mathcal{H}^n(V)$ and $B \in \mathcal{H}^n_+(V)$ and $v \in V$ the function

$$t \mapsto \langle f(A + tB)v, v\rangle$$

is k-tone.

Denote the class of matrix k-tone functions of order n on interval (a, b) as $P^k_n(a, b)$ (so $P^1_n(a, b) = P_n(a, b)$).

This definition doesn't exactly coincide with our definition for matrix convex functions, where we needed no assumption on the "sign" of B. As we will later see, however, this alternate definition leads to same set of functions.

As in the monotone case, we can list many natural properties of classes $P^k_n(a, b)$, proofs of which are very similar to the monotone case.

Proposition 7.6. *Let $(a, b) \subset \mathbb{R}$ be an open interval $n \geq 1$, and $k \geq 1$. Now*

1. *$P^k_n(a, b)$ is a convex closed cone.*
2. *$P^k_{n+1}(a, b) \subset P^k_n(a, b)$.*
3. *$\big(x \mapsto \alpha_k x^k + \ldots + \alpha_1 x + \alpha_0\big) \in P^k_n(a, b)$ if $\alpha_n \geq 0$.*
4. *$\big(x \mapsto (-1)^k x^{-1}\big) \in P^k_n(a, b)$.*

Proof. TODO □

Not surprisingly, we have also the following derivative characterization.

Theorem 7.7. *Let $n, k \geq 1$ and $f \in C^k(a, b)$. Then the following are equivalent:*

(i) *$f \in P^k_n(a, b)$.*

(ii) *For any $A \in \mathcal{H}^n_{(a,b)}$ and $H \geq 0$ we have*

$$D^k_n f_A(H) \geq 0.$$

Proof. TODO □

But there's a problem: there's no obvious way to change the H in the statement of the previous theorem to one-dimensional projection, when $k > 1$. The issue is that when $k > 1$, the map

$$H \mapsto D_n^k f_A(H)$$

is not linear anymore! It's a horrible mess instead.

TODO

- Is this section really needed?
- How to deal with smoothness issues cleanly?

Chapter 8

Trace functions

8.1 Absolute Value

As adjoint behaves as conjugate, it would be natural to guess that

$$|A| := (A^*A)^{\frac{1}{2}},$$

absolute value of a map, would have many similar properties as the standard absolute value.

The following list of properties of the absolute value make it clear that this is indeed good definition.

- $|A| \geq 0$ for any $A \in \mathcal{L}(V)$ and $|A| = A$, if and only if $A \geq 0$.
- For any $A \in \mathcal{H}(V)$ we have $-|A| \leq A \leq |A|$, or equivalently $|Q_A(v)| \leq Q_{|A|}(v)$ for any $v \in V$
- For any $v \in V$ we have $\|Av\| = \||A|v\|$.

Note that in general we have $|A| \neq |A^*|$, and maps need not even go between the same spaces.

One might be tempted to think that we have triangle inequality, i.e.

$$|A + B| \leq |A| + |B|,$$

for any $A, B \in \mathcal{L}(V)$, or at least $A, B \in \mathcal{H}$. Such inequality doesn't hold, but it's not that far from being true. Like in the real case, one would like to add

$$-|A| \leq A \leq |A| \quad \text{and} \quad -|B| \leq B \leq |B|,$$

to get

$$-(|A| + |B|) \leq A + B \leq |A| + |B|.$$

The problem is that we can't make any further conclusions: just because $-Y \leq X \leq Y$, it is not necessarily the case that $|X| \leq Y$. Thinking in quadratic forms we get the inequality

$$|Q_{A+B}(v)| \leq Q_{|A|+|B|}(v), \tag{8.1}$$

for any $v \in V$, but this does not imply that $Q_{|A+B|}(v) \leq Q_{|A|+|B|}(v)$. Indeed $|Q_{A+B}(v)| \leq Q_{|A+B|}(v)$, as we noticed, so the inequality is going to the wrong direction. If however v is an eigenvector of $A + B$, we have $|Q_{A+B}(v)| = Q_{|A+B|}(v)$, and it follows that

$$Q_{|A+B|}(v) \leq Q_{|A|+|B|}(v)$$

holds for eigenvectors v of $|A + B|$. Summing over the eigenvector we see that

$$\mathrm{tr}|A + B| \leq \mathrm{tr}|A| + \mathrm{tr}|B|,$$

so instead of the full inequality, we get inequality for traces. There is a nice generalization for the previous we'll get back to.

Chapter 9

Representations

Over the course of this thesis we have mentioned various representations results of the following form:

$$f(x) = \int h_t(x) d\mu(t),$$

or

$$f = \int h_t d\mu(t).$$

Here μ is Borel measure on some set and f and h_t's are functions of some kind. Functions h_t should be thought of some kind of basis functions. Although such results have been hardly used, one cannot just leave them unmentioned.

Most of the representation results in the thesis can be understood in terms of Choquet theory. The idea is the following: the sets we are concerned with are convex and the functions h_t the extreme points of these convex set. Extreme points are the points that can't be expressed as a non-trivial linear combination of points in the set.

Now if one has say compact convex set in $K \subset \mathbb{R}^n$, K should be roughly given by its boundary: compact convex sets are equal if and only they have same boundary. But more is true: every point in K can be expressed as a convex combination of extreme points of K (actually of at most $n + 1$ extreme points).

Same holds true much more generally, in infinite dimensional spaces.

We have already discussed this kind of phenomenon: we noted that every Pick function is essentially sum of functions of the form $z \mapsto \frac{1}{\lambda - z}$. Such functions (together with affine ones) are the extreme points of the cone of Pick functions. Similarly we found basis functions for k-tone functions: they are exactly the extreme points of the k-tone functions. TODO splitting Pick measures.

9.1 Minkowski–Carathéodory Theorem

Definition 9.1. Let E be a vector space C its convex subset. A point $x \in C$ is *extreme point* of C whenever $x = ty + (1-t)z$ for $y, z \in C$ and $t \in [0,1]$ then $x = y$ or $x = z$.

Theorem 9.2 ((Weak) Minkowski–Carathéodory Theorem)**.** *Let $C \subset \mathbb{R}^n$ be convex and compact. Then it is convex hull of its extreme points, i.e. for any $x \in C$ we may find $m \geq 1$, $x_1, x_2, \ldots, x_m$ and $t_1, t_2, \ldots, t_m \geq 0$ with $t_1 + t_2 + \ldots + t_m = 1$ such that*

$$x = \sum_{i=1}^{m} t_i x_i.$$

Proof. TODO □

9.2 Basis k-tone functions

We noticed that k-tone functions correspond more or less to functions with non-negative k'th derivative. In other words, k-tone functions should be k-fold integrals of positive functions, at least in sufficiently smooth setting. For instance $f : (a,b) \to \mathbb{R}$ is increasing and smooth if and only if it's of the form

$$f(x) = \int_{x_0}^{x} \rho(t) dt \tag{9.3}$$

for some positive $\rho \in C^{\infty}(a,b)$ and $x_0 \in (a,b)$, up to a constant at least. For non-smooth case, we could require ρ only to be a positive L^1-function: this gives us absolutely continuous increasing functions. If we further drop ρ but replace the Lebesgue measure by an arbitrary Radon measure μ, we get every right-continuous increasing function. Measuretheoretically these are already all the increasing functions, although we miss some functions like $\chi_{(0,\infty)}$.

If $\mu = \delta_0$, for instance, $f = \chi_{[0,\infty)}$. One could think that δ_0 gives a jump for f at 0. More generally, if μ is positive linear combination if m (distinct) Dirac deltas, f is a function with m jumps. Now every Radon measure is a weak limit of positive linear combination Dirac deltas, so every increasing function is limit of finite sums of jump functions, at least in some weak sense.

This fact is actually contained in 9.3: we may rewrite

$$f(x) = \int_{a}^{b} \chi_{[t,\infty)}(x) d\mu(t),$$

f is essentially sum of functions of the form $\chi_{[t,\infty)}$, again up to a constant. We will call those the basis functions for

The point is: whenever something holds for any step function, it should hold for any increasing function. In this context by "something" I mean linear inequalities: if ν is a signed Radon measure such that for any step function $\chi_{[t,\infty)}$ we have

$$\int \chi_{[x,\infty)}(t)d\nu(t),$$

then also

$$\int f(t)d\nu(t)$$

for any increasing function. Actually we should also require that $\int d\nu(t) = 0$. I'm being deliberately vague about the domains, they don't really matter too much.

Things get much more interesting when we move to k-tone functions of higher order. For k-tone functions, i.e. convex functions we can make similar statements.

We can write any (smooth enough) convex function in the form

$$f(x) = \int_{x_0}^{x} \int_{x_0}^{x_1} \rho(t)dtdx_1,$$

at least up to a constant and linear term. By simple partial integration this can be rewritten as

$$f(x) = \int_{x_0}^{x} (x-t)\rho(t)dt,$$

or even, better, as

$$f(x) = \int_{a}^{b} (x-t)_+\rho(t)dt,$$

where $(x-t)_+$ denotes $\max(0, x-t)$. What this means is that the functions $(\cdot - t)_+$ work as a basis functions for convex functions, up to a affine term. By affine transformation we could equivalently take the functions of the form $|\cdot - t|$ as a basis functions.

Now if a linear equality holds for functions of the form $|x-t|$, it holds for any convex function. So since for any $x_1, x_2, \ldots, x_m \in \mathbb{R}$ we have

$$\sum_{1\le i\le m} |x_i - t| \ge m \left| \frac{\sum_{1\le i\le m} x_i}{m} - t \right|,$$

also for any convex function

$$\sum_{1\le i\le m} f(x_i) \ge mf\left(\frac{\sum_{1\le i\le m} x_i}{m} \right),$$

Jensen's inequaltity.

Bibliography

[1] T. Ando. Majorization, doubly stochastic matrices, and comparison of eigenvalues. *Linear Algebra and Its Applications*, 118:163–248, 1989.

[2] T. Ando. Parameterization of minimal points of some convex sets of matrices. *Acta Scientiarum Mathematicarum (Szeged)*, 57:3–10, 1993.

[3] T. Ando. Majorizations and inequalities in matrix theory. *Linear algebra and its Applications*, 199:17–67, 1994.

[4] J. Antezana, E. R. Pujals, and D. Stojanoff. Iterated aluthge transforms: a brief survey. *Revista de la Unión Matemática Argentina*, 49(1):29–41, 2008.

[5] B. C. Arnold. Majorization: Here, there and everywhere. *Statistical Science*, pages 407–413, 2007.

[6] M. F. Atiyah. Convexity and commuting Hamiltonians. *Bull. London Math. Soc.*, 14(1):1–15, 1982.

[7] D. Azagra. Global approximation of convex functions. *arXiv preprint arXiv:1112.1042*, 2011.

[8] J. Bendat and S. Sherman. Monotone and convex operator functions. *Trans. Amer. Math. Soc.*, 79:58–71, 1955.

[9] H. Bercovici, C. Foias, and A. Tannenbaum. A spectral commutant lifting theorem. *Transactions of the American Mathematical Society*, 325(2):741–763, 1991.

[10] R. Bhatia. *Positive definite matrices.* Princeton university press, 2009.

[11] R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[12] Z. Brady. Inequalities and higher order convexity. *arXiv preprint arXiv:1108.5249*, 2011.

[13] P. Bullen. A criterion for n-convexity. *Pacific Journal of Mathematics*, 36(1):81–98, 1971.

[14] S. Burgdorf. *Trace-positive polynomials, sums of hermitian squares and the tracial moment problem*. PhD thesis, 2011.

[15] K. Cafuta, I. Klep, and J. Povh. A note on the nonexistence of sum of squares certificates for the bessis–moussa–villani conjecture. *Journal of Mathematical Physics*, 51(8):083521, 2010.

[16] E. Carlen. Trace inequalities and quantum entropy: an introductory course. *Entropy and the quantum*, 529:73–140, 2010.

[17] P. Chansangiam. A survey on operator monotonicity, operator convexity, and operator means. *Int. J. Anal.*, pages Art. ID 649839, 8, 2015.

[18] F. Clivaz. *Stahl's Theorem: Insights and Intuition on its Proof and Physical Context*. PhD thesis, 2014.

[19] F. Clivaz. Stahl's theorem (aka bmv conjecture): Insights and intuition on its proof. In *Spectral Theory and Mathematical Physics*, pages 107–117. Springer, 2016.

[20] C. de boor. *A practical guide to splines*. Springer Verlag., 1978.

[21] C. de Boor. Divided differences. *Surv. Approx. Theory*, 1:46–69, 2005.

[22] O. Dobsch. Matrixfunktionen beschränkter Schwankung. *Math. Z.*, 43(1):353–388, 1938.

[23] W. F. Donoghue, Jr. *Monotone matrix functions and analytic continuation*. Springer-Verlag, New York-Heidelberg, 1974. Die Grundlehren der mathematischen Wissenschaften, Band 207.

[24] A. Eremenko. Herbert stahl's proof of the bmv conjecture. *arXiv preprint arXiv:1312.6003*, 2013.

[25] M. Floater and T. Lyche. Two chain rules for divided differences and faa di bruno's formula. *Mathematics of Computation*, 76(258):867–877, 2007.

[26] U. Franz, F. Hiai, and E. Ricard. Higher order extension of Löwner's theory: operator k-tone functions. *Trans. Amer. Math. Soc.*, 366(6):3043–3074, 2014.

[27] G. Frobenius. Ueber die Entwickelung analytischer Functionen in Reihen, die nach gegebenen Functionen fortschreiten. *J. Reine Angew. Math.*, 73:1–30, 1871.

[28] F. Hansen. Trace functions as laplace transforms. *Journal of mathematical physics*, 47(4):043504, 2006.

[29] F. Hansen. The fast track to Löwner's theorem. *Linear Algebra Appl.*, 438(11):4557–4571, 2013.

[30] F. Hansen and G. Kjærgård Pedersen. Jensen's inequality for operators and löwner's theorem. *Mathematische Annalen*, 258(3):229–241, 1982.

[31] F. Hansen and J. Tomiyama. Differential analysis of matrix convex functions. *Linear Algebra Appl.*, 420(1):102–116, 2007.

[32] F. Hansen and J. Tomiyama. Differential analysis of matrix convex functions. II. *JIPAM. J. Inequal. Pure Appl. Math.*, 10(2):Article 32, 5, 2009.

[33] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge university press, 1952.

[34] U. Helmke and J. Rosenthal. Eigenvalue inequalities and schubert calculus. *Mathematische Nachrichten*, 171:207–226, 1995.

[35] F. Hiai and D. Petz. *Introduction to matrix analysis and applications*. Springer Science & Business Media, 2014.

[36] C. J. Hillar. Advances on the bessis–moussa–villani trace conjecture. *Linear algebra and its applications*, 426(1):130–142, 2007.

[37] A. Hindmarsh. Pick's conditions and analyticity. *Pacific Journal of Mathematics*, 27(3):527–531, 1968.

[38] J. A. Holbrook. Spectral variation of normal matrices. *Linear algebra and its applications*, 174:131–144, 1992.

[39] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[40] C. R. Johnson and C. J. Hillar. Eigenvalues of words in two positive definite letters. *SIAM journal on matrix analysis and applications*, 23(4):916–928, 2002.

[41] J. Karamata. Sur une inégalité relative aux fonctions convexes. *Publications de l'Institut mathematique*, 1(1):145–147, 1932.

[42] I. Klep and M. Schweighofer. Sums of hermitian squares and the bmv conjecture. *Journal of Statistical Physics*, 133(4):739–760, 2008.

[43] A. A. Klyachko. Stable bundles, representation theory and Hermitian operators. *Selecta Math. (N.S.)*, 4(3):419–445, 1998.

[44] A. Knutson and T. Tao. Honeycombs and sums of Hermitian matrices. *Notices Amer. Math. Soc.*, 48(2):175–186, 2001.

[45] G. Kowalewski. *Interpolation und genäherte Quadratur*. Teubner Leipzig und Berlin, 1932.

[46] F. Kraus. Über konvexe Matrixfunktionen. *Math. Z.*, 41(1):18–42, 1936.

[47] K. Löwner. Über monotone Matrixfunktionen. *Math. Z.*, 38(1):177–216, 1934.

[48] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: theory of majorization and its applications*. Springer Series in Statistics. Springer, New York, second edition, 2011.

[49] R. F. Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902.

[50] H. Mulholland. On generalizations of minkowski's inequality in the form of a triangle inequality. *Proceedings of the London Mathematical Society*, 2(1):294–307, 1949.

[51] R. Nevanlinna. *Über beschränkte funktionen die in gegebenen punkten vorgeschriebene werte annehmen...* Buchdruckerei-a.-g. Sana, 1919.

[52] G. Pick. Über die beschränkungen analytischer funktionen, welche durch vorgegebene funktionswerte bewirkt werden. *Mathematische Annalen*, 77(1):7–23, 1915.

[53] A. Pinkus and D. Wulbert. Extending n-convex functions. *Studia Math*, 171(2):125–152, 2005.

[54] T. Popoviciu. *Les fonctions convexes*, volume 992. Bussière, 1945.

[55] A. W. Roberts and D. E. Varberg. *Convex functions*, volume 57. Academic Press, 1974.

[56] D. Sarason. Generalized interpolation in H^∞. *Trans. Amer. Math. Soc.*, 127:179–203, 1967.

[57] D. Sarason. Nevanlinna-pick interpolation with boundary data. *Integral Equations and Operator Theory*, 30(2):231–250, 1998.

[58] H. R. Stahl et al. Proof of the bmv conjecture. *Acta Mathematica*, 211(2):255–290, 2013.

[59] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 7.1)*, 2016. `http://www.sagemath.org`.

[60] J. A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[61] H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

[62] F. Zhang. *Matrix theory: basic results and techniques.* Springer Science & Business Media, 2011.