Matrix monotone and convex functions

Otte Heinävaara

October 3, 2017

Contents

Chapter 1

Introduction

1.1 Foreword

This master's thesis is about matrix monotone and convex functions. Matrix monotonicity and convexity are generalizations of standard monotonicity and convexity of 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.

Matrix convexity, as you might have guessed by now, is defined as follows. A function f is *matrix convex* if for any two matrices A and B and $0 \leq t \leq 1$ we have

$$f(tA + (1-t)B) \leq tf(A) + (1-t)f(B). \tag{1.3}$$

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's in a way 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 say that matrix is positive if all it's eigenvalues are non-negative. Having done this, we immediately restrict ourselves to (symmetric) diagonalizable matrices with real eigenvalues, but we will later see that we can't do much "better". Also, since sum of positive matrices should be positive, we should further restrict ourselves to even stricter class of matrices, called Hermitian matrices, which correspond self-adjoint linear maps. Now everything works nicely but we still preserve non-trivial non-commutative structure.

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 monotone and convex 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 Plan of attack

This master's thesis is a comprehensive review of the rich theory of matrix monotone and convex 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

1.3 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.

1.4 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.

Chapter 2

Positive matrices

This chapter is titled "positive matrices", although "positive maps" might be more appropriate title. We are mostly going to deal with finite-dimensional objects, but many of the ideas could be generalized infinite-dimensional settings, where matrices lose their edge. Also, one should always ask whether it really clarifies the situation to introduce concrete matrices: matrices are good at hiding the truly important properties of linear mappings. The words "matrix" and "linear map" are used somewhat synonymously, although one should always remember that the former are just special representations for the latter.

2.1 Motivation

How should one order matrices? What should we require from ordering anyway?

We would definitely like to have natural total order on the space of matrices, but it turns out there are no natural choices for that. Partial order is the next best thing. Recall that a partial order on a set X is a binary relation $\leq$ on such that

1. $x \leq x$ for any $x \in X$.
2. For any $x, y \in X$ for which $x \leq y$ and $y \leq x$, necessarily $x = y$.
3. If for some $x, y, z \in X$ we have both $x \leq y$ and $y \leq z$, also $x \leq z$.

The third point is the main point, the first two are just there preventing us from doing something crazy. But we can do better: this partial order on matrices should also respect addition.

4. For any $x, y, z \in X$ such that $x \leq y$, we should also have $x + z \leq y + z$.

There's another way to think about this last point. Instead of specifying order among all the pairs, we just say which matrices are positive: matrix is positive if and only it's at least 0.

If we know all the positive matrices, we know all the "orderings". To figure out whether $x \leq y$, we just check whether $0 = x - x \leq y - x$, i.e. whether $y - x$ is positive. Also, positive matrices are just differences of the form $y - x$ where $x \leq y$. Now, conditions on the partial order are reflected to the set of positive matrices.

1'. 0 (zero matrix) is positive.

2'. If both x and $-x$ are positive, then $x = 0$.

3'. If both x and y are positive, so is their sum $x + y$.

Here $3'$ is kind of combination of 3 and 4.

The terminology here is rather unfortunate. Natural ordering of the reals satisfies all of the above with obvious interpretation of positive numbers, which however differs from the standard definition: 0 is itself positive in our above definition. This is undoubtedly confusing, but what can you do? For real numbers we have total order, so every number is either zero, strictly positive or strictly negative, so when we say non-negative, it literally means "not negative": we get all the positive numbers and zero. But with partial orders we might get more. So the main reasons why we are using this terminology are

1. It's short.

Also, now that we have decided to preserve the word "positive" for "at least zero" one might be tempted to preserve "strictly positive" for "at least zero, but not zero". We won't do that, we save that phrase for something more important.

To figure out a correct notion for positive maps, let's start simple. If we are in a 1-dimensional vector space V over $\mathbb{R}$ there's rather canonical choice for positivity. Any linear map is of the form $v \mapsto av$ for some $a \in \mathbb{R}$ and we should obviously say that a map is positive if $a \geq 0$ (note our non-standard terminology concerning positivity). More generally, if a map if scalar multiple of identity, map should be positive if and only if the corresponding scalar is non-negative.

Natural extension of this idea could be try the following: map is positive if all of its eigenvalues are non-negative. Of course, this doesn't quite work: not every map has real eigenvalues. But even if we restrict to maps with real eigenvalues, this property is not preserved in addition. Consider for example the pair

$$\begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -1 \\ -4 & 0 \end{bmatrix}$$

The two matrices have both distinct eigenvalues -2 and 2 and are hence diagonlizable, but their sum has characteristic polynomial x^2+9, which most definitely has no real zeros. In general one should not except summation and eigenvalues go very well together.

2.2 Real maps

2.2.1 Restricting linear maps

There's however quite clever way to go around this. Instead of requiring non-negativity of eigenvalues, we require that map "restricts" to positive map. The idea is: we already know which maps should be positive in one-dimensional spaces, or more generally, which scalar multiples of identity should be positive. Now we should require that when we restrict our look to one-dimensional subspaces, we should get a positive map.

Of course, one should first understand what restricting linear maps means. Usually if we have a linear map $A : V \to V$, we could take subspace $W \subset V$ and consider the usual restriction map $A\big|_W : W \to V$ given by $A\big|_W(w) = Aw$ for any $w \in W \subset V$. In other words $A\big|_W = A \circ J_W$, where J_W denotes the natural inclusion from W to V. But this map is going to wrong space. Instead we would like to define something satisfying

- Restriction is a linear map $(\cdot)_{V,W} = (\cdot)_W : \mathcal{L}(V) \to \mathcal{L}(W)$.
- If $A \in \mathcal{L}(V)$ and $A(W) \subset W$, restriction should coincide with the original map, in the sense that $A = J_W \circ A_W$.
- If $W' \subset W \subset V$, we should have $(\cdot)_{W'} = ((\cdot)_W)_{W'}$.

These properties don't uniquely define a linear map but they say that A_W should be of the form $P_{V,W} \circ A \circ J_W$ where $P_{V,W}$ is a projection, i.e. a map for which $P_{V,W} \circ J_W = I_W$. Moreover, these projections should satisfy $P_{V,W'} = P_{W,W'} \circ P_{V,W}$.

If we are working in a inner-product space, there's rather natural choice for the map $P_{V,W}$: orthogonal projections. Orthogonal projections are projections with $\ker(P) = \operatorname{im}(P)^{\perp}$. Such maps are easily seen to satisfy all the requirements. Finally, we will call our new concept *compression* instead of restriction, to distinguish between the two.

Definition 2.1. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and $W \subset V$ a subset. We define the map A_W, *compression* of A to W to be the linear map given by $P_W \circ A \circ J_W$.

Theorem 2.2. *Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and $W \subset V$ subspace of V. Then the compression to W is unique linear contraction from $\mathcal{L}(V)$ to $\mathcal{L}(W)$, such that for any $A \in L(V, W)$ we have $(J_W \circ A)_W = A \circ J_W$. Moreover, if $W' \subset W$, we have $(\cdot)_{W'} = ((\cdot)_W)_{W'}$.*

Proof. Trust me, it's true. Also, even if it's not, this theorem is clearly here just to convince the reader that orthogonal projections are the only sensible choice; but does that really need convincing?

Okay, we'll come back to the proof. □

For one-dimensional compressions we have convenient representation. As one easily checks, one dimensional projection onto subspace spanned by vector v is given by

$$P_{(v)} = \frac{\langle \cdot, v\rangle}{\langle v, v\rangle} v,$$

as long as $v \neq 0$, and thus

$$A_{(v)} = \frac{\langle A\cdot, v\rangle}{\langle v, v\rangle} v.$$

If $w \in (v) \setminus \{0\}$, we could rewrite the previous in the form

$$A_{(v)}(w) = \frac{\langle Av, v\rangle}{\langle v, v\rangle} w = \frac{\langle Aw, w\rangle}{\langle w, w\rangle} w.$$

This gives rise to so called Rayleigh quotient $R(A, \cdot) : V \setminus \{0\} \to \mathbb{C}$, given by

$$R(A, v) = \frac{\langle Av, v\rangle}{\langle v, v\rangle}.$$

Compression in the direction of v is given by scaling by the corresponding Rayleigh quotient.

We will call $\langle Av, v\rangle$ the quadratic form of A, and denote it by $Q_A(v)$.

There's one more important property of compression we need. When map is compressed to a subspace, we naturally lose some information about the map. Knowing about all of the compressions, however, we can get our map back.

Lemma 2.3 (Injectivity of compression)**.** *If $A, B \in \mathcal{L}(V)$, $A = B$ if and only if $A_W = B_W$ for any one-dimensional subspace $W \subset V$.*

Proof. By linearity, it is sufficient to prove that if $Q_A(v) = 0$ for any $v \in V$, then $A = 0$. TODO polarization identity. □

2.2.2 Positive maps

Now that we have defined the compression we are ready define positive maps.

Definition 2.4. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. We say that a map $A \in \mathcal{L}(V)$ is positive, and write $A \geq 0$, if for any one-dimensional subspace $W \subset V$ the map A_W is positive, i.e. is induced by a non-negative real.

We denote the space of positive maps by $\mathcal{H}_+(V)$. Positive maps have the following useful properties.

Proposition 2.5. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space over $\mathbb{C}$. Then*

(i) $A \in \mathcal{L}(V)$ is positive if and only if A_W is positive for every subspace $W \subset V$.

(ii) If $A, B \in \mathcal{L}(V)$ are positive and $\alpha, \beta \geq 0$, also $\alpha A + \beta B$ is positive.

(iii) If $(A_i)_{i=1}^{\infty}$ are positive and $\lim_{i\to\infty} A_i = A$, also A is positive.

(iv) $A \in \mathcal{L}(V)$, A is positive if and only for any $v \in V$, or for any $v \in V$ with $|v| = 1$ we have $\langle Av, v\rangle \geq 0$, or still equivalently, for any $v \in V \setminus \{0\}$ the Rayleigh quotient $R(A, v)$ is non-negative.

(v) If both A and $-A$ are positive, then $A = 0$.

(vi) If A is positive, all of its eigenvalues are non-negative.

Proof. (i) Other direction is immediate. Also if for any subspace $W \subset V$ take any one-dimensional $W' \subset W$. Now $(A_W)_{W'} = A_{W'}$, is positive by assumption, and so is A_W.

(ii) The claim evidently holds for one-dimensional spaces. Now for any one-dimesional $W \subset V$ we have $(\alpha A + \beta B)_W = \alpha A_W + \beta B_W \geq 0$, by the one-dimensional case, so $\alpha A + \beta B \geq 0$.

(iii) Again, the claim evidently holds for one-dimensional spaces. Now for any one-dimesional $W \subset V$ we have $(\lim_{i\to\infty} A_i)_W = \lim_{i\to\infty}(A_i)_W \geq 0$, by the one-dimensional case, so $A \geq 0$.

(iv) These claims are immediate from our representation for one-dimensional compressions.

(v) If $A, -A \geq 0$, all the one-dimensional compressions of A are both non-negative and non-positive, so zero. But by the injectivity of compression, it follows that $A = 0$.

(vi) Note that if v is any eigenvector of v, $A_{(v)} w = \frac{\langle Av, v\rangle}{\langle v, v\rangle} w = \lambda w$, so by assumption $\lambda \geq 0$.

$\square$

The map $v \mapsto \langle Av, v\rangle$ is called the quadratic form of A, and is denoted by Q_A.

We can also lift other important notions.

Definition 2.6. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. We say that a map $A \in \mathcal{L}(V)$ is real, if for any one dimensional subspace $W \subset V$ the map A_W is real, i.e. is induced by real number.

Definition 2.7. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. We say that a map $A \in \mathcal{L}(V)$ is imaginary, if for any one dimensional subspace $W \subset V$ the map A_W is imaginary, i.e. is induced by imaginary number.

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. Hermitian 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.

2.2.3 Adjoint

We can also lift the notion of complex conjugate. If V is one-dimensional, $\overline{A}$, conjugate of A should be a linear map which is induced by the complex conjugate of the scalar inducing A.

Theorem 2.8. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Then for any $A \in \mathcal{L}(V)$ there exists unique map $A^* \in \mathcal{L}(V)$, which we will call adjoint of A, for which for any one-dimensional subspace W we have $(A^*)_W = \overline{A_W}$.*

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.

Note that the Rayleigh quotient for such map is given by

$$R(A, x) = \frac{\langle x, v\rangle\langle w, x\rangle}{\langle x, x\rangle}.$$

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

$$R(A^*, x) = \frac{\langle x, w\rangle\langle v, x\rangle}{\langle x, x\rangle} = \overline{\frac{\langle w, x\rangle\langle x, w\rangle}{\langle x, x\rangle}} = \overline{R(A, x)}.$$

□

Real maps are their own adjoints, and that is why they are often called *self-adjoint*.

There is also a meaningful way to extend the notion of adjoint for general (non-endomorphism) linear maps. In general setting, we don't have a notion of compression of linear map: there's no canonical way to restrict the codomain. We can however interpret a map in a bigger space. Indeed, any map $A \in \mathcal{L}(V, W)$ can be canonically interpreted as a map $\tilde{A} \in \mathcal{L}(V \oplus W)$: define $\tilde{A}(v, w) = (0, Av)$. We call this map the *symmetrization* of A. Now it makes sense to consider $(\tilde{A})^*$, the symmetrization has a unique adjoint. This adjoint does not in general live in $\mathcal{L}(V, W)$ anymore: but it turns out that it does live in $\mathcal{L}(W, V)$!

Theorem 2.9. *For any $A \in \mathcal{L}(V, W)$ there exists a unique map $A^* \in \mathcal{L}(W, V)$ which we call the adjoint of A, such that $(\tilde{A})^* = (\tilde{A^*})$. Moreover, if $V = W$, the new notion of adjoint coincides with the old one.*

Proof. Of course, strictly speaking $\tilde{A}^*$ would be map in $\mathcal{L}(W \oplus V)$, not in $\mathcal{L}(V \oplus W)$, but the two spaces are canonically isomorphic.

The uniqueness follows from the already known uniqueness for the old notion. The map $(\cdot)^* : \mathcal{L}(V, W) \to \mathcal{L}(W, V)$ should again evidently be conjugate linear. Also, the same construction for basis elements of the form $A = (x \mapsto \langle x, v\rangle_V w)$ (where $v \in V$ and $w \in W$) works again. Indeed, for any $(x, y) \in V \oplus W$ the corresponding Rayleigh quotient is given by

$$R(\tilde{A}, (x, y)) = \frac{\langle (0, Ax), (x, y)\rangle_{V\oplus W}}{\langle (x, y), (x, y)\rangle_{V\oplus W}} = \frac{\langle Ax, y\rangle_W}{\langle x, x\rangle_V + \langle y, y\rangle_W} = \frac{\langle x, v\rangle_V \langle w, y\rangle_W}{\langle x, x\rangle_V + \langle y, y\rangle_W},$$

and it's clear that we may set $A^* = (x \mapsto \langle x, w\rangle_W v)$. Our construction also makes it clear that this notion coincides with the old one. $\square$

The previous proof also gives a convenient corollary, which is the most common definition for adjoint.

Corollary 2.10. *For any $A \in \mathcal{L}(V, W)$, the adjoint $A^* \in \mathcal{L}(W, V)$ is unique linear map such that for any $v \in V$ and $w \in W$ we have*

$$\langle Av, w\rangle_W = \langle v, A^* w\rangle_V.$$

In particular, map $A \in \mathcal{L}(V)$ is real for any $v, w \in V$ we have

$$\langle Av, w\rangle = \langle v, Aw\rangle.$$

and imaginary if for any $v, w \in V$

$$\langle Av, w\rangle = -\langle v, Aw\rangle.$$

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

Theorem 2.11. *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) $(AB)^* = B^*A^*$.

vi) $\ker(A) = \operatorname{im}(A)^\perp$

2.2.4 Examples

It's high time to have some examples.

Most obvious, although not very interesting, representatives of real/imaginary/positive maps are real/imaginary/positive scalar multiples of the identity. Projections are stereotypical examples of positive and hence real maps. Projections, proper definition. Indeed, one-dimensional projections are given by $A = (x \mapsto \langle x, v\rangle v)$ for some $v \in V$ with $|v| = 1$. For such maps $\langle Ax, x\rangle = \langle x, v\rangle\langle v, x\rangle = |\langle x, v\rangle|^2 \geq 0$. Higher dimensional projections are simply sums of one-dimensional ones, so they are also positive and real. More generally one could take any positive linear combination of projections to get much more positive maps, and real linear combination of projections to get real maps.

As we earlier noticed, however, not every map with real eigenvalues is real, and not every map non-negative eigenvalues is positive. It turns out that the basis elements of the form $A = (x \mapsto \langle x, v\rangle w)$ are real if and only if v and w are real multiples of each other, or to be precise, if there exists $\alpha, \beta \in \mathbb{R}$, not both 0, such that $\alpha v + \beta w = 0$. Indeed, by the corollary, the map is real if for any $x, y \in V$ we have

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

Now if v and w are not parallel, we can find x such that $\langle x, v\rangle = 0 \neq \langle x, w\rangle$, which contradicts the previous. The case of parallel v and w is easy to check.

While hunting for examples, it's worthwhile to note that in some sense $\mathcal{H}(V)$ is not essentially bigger than $\mathcal{H}_+(V)$: if $A \in \mathcal{H}(V)$ we can always find a positive real number

λ such that $A + \lambda I \in \mathcal{H}_+(V)$. To this end, note that the quadratic form of $A + \lambda I$ at $v \in V$ is $\langle (A + \lambda I)v, v\rangle = \langle Av, v\rangle + \lambda\langle v, v\rangle$. But if $\lambda \geq \|A\|$, the operator norm of A, the previous quantity is non-negative for any $v \in V$.

TODO: 2×2 case.

2.3 Spectral theorem

One might wonder if there are other examples of positive maps than positive linear combination of projections. Rather surprisingly, there are none.

Theorem 2.12. *$A \in \mathcal{L}(V)$ is positive if and only for some $m \geq 0$, $\lambda_i > 0$ and $v_i \in V$ for $1 \leq i \leq m$ we have*

$$A = \sum_{1 \leq i \leq m} \lambda_i P_{v_i}.$$

Proof. 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 2.13. *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. $\square$

TODO: image of the proof process in $\mathbb{R}^3$.

Proof of lemma 2.13. Take any $w \in V$. Now by assumption for any $c \in \mathbb{C}$ we have

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

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 $w \in V$. $\square$

As discussed, there should be obvious generalization for real maps: they real linear combinations for projections. We can however still improve the statement: we can take v_i's to be orthogonal, and hence $m \leq n$. We have thus arrived at Spectral theorem.

Theorem 2.14 (Spectral theorem)**.** *Let $A \in \mathcal{H}(V)$. Then there exists real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ and orthonormal vectors $v_1, v_2, \ldots, v_n$ such that*

$$A = \sum_{i=1}^{n} \lambda_i P_{v_i}. \tag{2.15}$$

Proof. Let's first check case of positive A. There's not very many things to change from the proof of theorem 2.12. Indeed, we again argue by induction. The case $n = 0$ is again clear. In the induction step we found that induction assumption applies to $A - \lambda I$ compressed to suitable $(n-1)$-subspace. There we can cook up required representation, and bring back the representation for $A - \lambda I$ itself. That is we have

$$A - \lambda I = \sum_{i=1}^{n-1} \lambda_i P_{v_i}$$

for orthonormal v_i's and non-negative λ_i's. But then if v_n is a missing orthonormal vector, we find that

$$A = \sum_{i=1}^{n} (\lambda_i + \lambda) P_{v_i},$$

where $\lambda_n = 0$. But this is what we wanted.

For non-positive A, simply add suitable multiple of identity to get $B := A + \lambda I \geq 0$ and apply what we have proved to B. If we have representation for B, we can easily cook up one for A: just subtract λ for λ_i's in the representation of B. $\square$

In the representation 2.15 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.* Such representation is of course not unique: if $A = I$, we could again choose v_i's to be any orthonormal basis of V.

There is 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: 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 $p(x) = c_n x^n + c_{n-1} x^{n-1} + \ldots c_1 x + c_0$, with $c_i \in \mathbb{R}$, we should write

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

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 $v = \sum_{1\leq i\leq n} x_i v_i$, where $(v_i)_{i=1}^n$ is a eigenbasis for A and $x_i = \langle x, v_i\rangle$. Now $Ax = \sum_{1\leq i\leq n} \lambda_i x_i v_i$, so for instance

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

Similarly, if $A \geq 0$, 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 $A = \sum_{1\leq i\leq n} \lambda_i P_{v_i}$, we can simply set $A^{\frac{1}{2}} = \sum_{1\leq i\leq n} \lambda_i^{\frac{1}{2}} P_{v_i}$. As for the uniqueness, note that if B is a positive square root for A and $B = \sum_{1\leq i\leq n} \lambda_i' P_{v_i'}$, then $B^2 = \sum_{1\leq i\leq n} \lambda_i'^2 P_{v_i'}$. 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.

Spectral theorem also makes it clear that positiveness and realness of a map is preserved under direct sums.

Corollary 2.18. *Let $A \in \mathcal{L}(V)$ and $B \in \mathcal{L}(W)$ and consider the map $A\oplus B \in \mathcal{L}(V\oplus W)$. Then $A \oplus B$ is real (positive), if and only if A and B are.*

2.3.1 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 2.19. *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 $P_{W_1} P_{W_2} = 0 = P_{W_2} P_{W_1}$. □

Definition 2.20. We say that two $W_1, W_2 \subset V$ subspaces commute if the respective projections commute.

Theorem 2.21. *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 2.19.

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. $\square$

The previous theorem sends a very important message.

Philosophy 2.22. 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 2.21 we have

Corollary 2.23. *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 2.24. *If $A, B \geq 0$, then AB has non-negative eigenvalues. Conversely, if C has non-negative eigenvalues, then it's of the form AB for some positive A and B. TODO: details*

Proof. Postponed. □

TODO: independence of random variables.

2.3.2 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.

Proposition 2.25. *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. TODO □

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

2.4 Congruence

2.4.1 $*$-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.

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. If this is the case, we say that A and C are congruent. 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 2.26 (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 2.27. *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 2.28. *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 2.29 (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. TODO □

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.

2.4.2 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

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

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

$$D^*MD = \begin{bmatrix} I & 0 \\ -B^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix}. \tag{2.30}$$

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 2.31 (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 2.30 to arrive at so-called *LDL-decomposition* of M:

$$M = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}. \tag{2.32}$$

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

$$A^{-1} + A^{-1}B(C - B^*A^{-1}B)^{-1}B^*A^{-1} = (A - BC^{-1}B^*)^{-1}, \tag{2.33}$$

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

$$\det(A_W) = \det(A)\langle A^{-1}v, v\rangle. \tag{2.34}$$

By comparing determinants from two LDL-decompositions we arrive at

$$\det(A)\det(C - B^*A^{-1}B) = \det(C)\det(A - BC^{-1}B^*), \tag{2.35}$$

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.

2.5 Loewner order

Definition 2.36. 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 2.5 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 2.37. *(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 2.38. 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 $P_{W_1} \leq P_{W_2}$, for any $v \in W_1$ we have $Q_{P_{W_1}}(v) = \|v\|^2 \leq \langle P_{W_2} v, v \rangle = Q_{P_{W_2}}(v)$, where the inequality can hold if and only if $v \in W_2$.

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 2.39. *If $A \leq B$, then for any C we have $C^* A C \leq C^* B C$.*

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

Theorem 2.40. *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 $I \leq X := A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ 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 2.41. 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

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

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

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

If one writes $X = A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$, 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.

2.6 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 2.42. *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 2.28 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 2.43 (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 2.28 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. □

TODO picture of eigenvalues changing when compressed

Again one can prove that this result is strongest possible.

Proposition 2.44. *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 2.45. *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 2.42 $\lambda_k(A) \leq \lambda_k(B)$, 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 2.43 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 2.46. *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 2.47. *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 2.35. □

Proof of propostion 2.46. 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 2.44 is similar: the aim to first connect the characteristic polynomials of A and its compression and then do similar observations.

Lemma 2.48. *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 2.34. ☐

Proof of proposition 2.44. Proof is just an easier version of the proof of 2.46 ☐

TODO: change order of compression and projection eigenvalues converses.

2.7 Notes and references

2.8 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 ($\rightarrow$ 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

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

  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

Divided differences

3.1 Motivation

Divided differences are derivatives without limits.

Consider a function $f : \mathbb{R} \to \mathbb{R}$. Its (first) divided difference is defined as

$$\begin{aligned}[\cdot,\cdot]_f : \mathbb{R}^2 \setminus \{x \neq y\} &\to \mathbb{R} \\ [x, y]_f &= \frac{f(x) - f(y)}{x - y}.\end{aligned}$$

If f is sufficiently smooth, we should also define $[x, x]_f = f'(x)$: if $f \in C^1(\mathbb{R})$, this gives continuous extension to $[\cdot,\cdot]_f$. Much of the power of divided differences comes however from the fact that they conveniently carry same information even if we do not do such extension.

Consider the case of increasing f. This information is exactly carried by the inequality $[x, y]_f \geq 0$. Again, if f is differentiable, this is equivalent to $f'(x) = [x, x]_f \geq 0$. There are many ways to see this fact, one of the more standard being the mean value theorem: If $x \neq y$, for some ξ between x and y we have

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

Now if the derivative is everywhere non-negative, so are divided differences. Also divided differences are sort of approximations for derivative.

We can push these ideas to higher orders. Second order divided differences should be something that captures second order behaviour of a function. In particular, if $f \in C^2(\mathbb{R})$ has non-negative second derivative everywhere, i.e. it is convex, its second divided difference should be non-negative, and vice versa. Standard definition of convexity is almost what we are looking for: f is convex if for any $x, y \in \mathbb{R}$ and $0 \leq t \leq 1$ we have $tf(x) + (1-t)f(y) \geq f(tx + (1-t)y)$. So if we define the mapping $[\cdot, \cdot, \cdot]_f : \mathbb{R}^2 \times [0,1]$ by $tf(x) + (1-t)f(y) - f(tx + (1-t)y)$, we have $[x, y, t]_f \geq 0$ for any $(x, y, t) \in \mathbb{R}^2 \times [0,1]$ if and only if $f^{(2)}(x) \geq 0$ for any $x \in \mathbb{R}$.

There is however much better version for the function. If we write $z = tx + (1-t)y$, we can solve that $t = \frac{z-y}{x-y}$ and express

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

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 is 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 crap the previous definition and set instead

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

One would naturally except 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 Basic properties

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 3.1. 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

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

As with the first order divided differences, if 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.

Albeit a rather direct generalization for the cases $n = 1$ and $n = 2$, We defined divided differences only for real valued functions, but codomain could just as well any real or complex vector space. it's not very clear why such definition should correspond to anything useful. We have however the following important properties.

Proposition 3.2. *Divided differences are symmetric in the variable, i.e. for any $f : (a, b) \to \mathbb{R}$, $a < x_0, x_1, \ldots, x_n < b$ permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ we have*

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

and linear in the function, i.e. for any $\alpha, \beta \in \mathbb{R}$, $f, g : (a, b) \to \mathbb{R}$ and $a < x_0, x_1, \ldots, x_n < b$ we have

$$[x_0, x_1, \ldots, x_n]_{\alpha f + \beta g} = \alpha [x_0, x_1, \ldots, x_n]_f + \beta [x_0, x_1, \ldots, x_n]_g.$$

In addition, divided differences can be calculated recursively as

$$[x_0, x_1, \ldots, x_n]_f = \frac{[x_0, x_1, \ldots, x_{n-1}]_f - [x_1, \ldots, x_n]_f}{x_0 - x_n}. \tag{3.3}$$

Proof. Easy to check. □

Also, we have following classic characterization.

Proposition 3.4. *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. Claims are easily derived from each other since if f is itself a polynomial of degree at most n, its Lagrange interpolation polynomial is f itself. Recall that the Lagrange interpolation polynomial of a dataset $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ 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 in our case the leading coefficient of this polynomial leads exactly to our definition for the divided differences. □

These observation already partially justify the terminology: as higher order derivatives are defined recursively using (first order) derivatives, higher order divided differences can be calculated recursively using (the usual) divided differences.

The most important property of the divided differences is the following.

Theorem 3.5 (Mean value theorem for divided differences)**.** *Let $n \geq 1$ and $f \in C^n(a,b)$. Then for any $x_0, x_1, \ldots, x_n$ we have $\min_{0 \leq i \leq n}(x_i) \leq \xi \leq \max_{0 \leq i \leq n}(x_i)$ such that*

$$[x_0, x_1, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}. \tag{3.6}$$

Proof. We prove the statement assuming additionally that the divided differences define a continuous function on the whole set $(a,b)^{n+1}$: this will proven later. Note that if one manages to prove the statement for distinct points, one may take sequence of tuples of distinct points, $((x_i^{(j)})_{i=0}^n)_{j=1}^\infty$ converging to $(x_i)_{i=0}^\infty$. Now the left-hand side will converge to the respective divided difference (assuming the continuity), and by moving to a convergent subsequence, so will the ξ_n's on the right-hand side. By the continuity of $f^{(n)}$ we are done.

For the case of distinct x_i's note that we have already proven the statement for polynomials of order at most n. By linearity it hence suffices to prove the statement for $C^n(a,b)$ functions vanishing on the set $\{x_i | 0 \leq i \leq n\}$. This we know already for $n = 1$; this is the mean value theorem. Let us prove the statement by induction on n. To simplify notation we may assume that $x_0 < x_1 < \ldots < x_n$. Note that by the mean value theorem, given that $f(x_i) = 0$ for any $0 \leq i \leq n$, we also have $f'(y_i) = 0$ for some $x_i < y_i < y_{i+1}$, for $0 \leq i \leq n-1$. By the induction hypothesis the $(n-1)$:th derivative of f', $f^{(n)}$ has a zero ξ with $x_0 \leq \xi \leq x_n$. But this is exactly what we wanted.

TODO: figure of recursive procedure. □

We'll get back to smoothness in a minute. This is already a very precise sense in which divided differences work like derivatives, up to a constant. In some sense though $\frac{f^{(n)}}{n!}$, the Taylor coefficients are even more natural objects than the pure derivatives. They are the coefficients in the Taylor expansion, and they satisfy very natural Leibniz rule

$$\frac{(fg)^{(n)}(x)}{n!} = \sum_{k=0}^{n} \left(\frac{f^{(k)}(x)}{k!}\right)\left(\frac{g^{(n-k)}(x)}{(n-k)!}\right),$$

which is of course just a formula for polynomial convolution.

Divided differences enjoy similar Leibniz formula, which is related to a generalization of Taylor expansion, called Newton expansion. In Newton expansion we first fix a sequence of points $x_0, x_1, \ldots x_n \in (a,b)$, say pairwise distinct for starters. For $f : (a,b) \to \mathbb{R}$ and

$x \in (a, b)$ we may start a process of rewriting

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

By taking first $1, 2, \dots$ terms of the sum, one obtains Newton form of interpolating polynomial for the first $1, 2, \dots,$ points of the sequence $x_0, x_1, \dots$. If the points x_i coincide, we get

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} + [x, x_0, x_0, \dots, x_0]_f (x - x_0)^n, \tag{3.7}$$

the usual Taylor expansion. Consequently

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

and as is well known, this error term can be also written in the form

$$[x, x_0, x_0, \dots, x_0]_f = \frac{1}{(x - x_0)^n} \int_{x_0}^{x} \frac{f^{(n)}(t)(x - t)^{n-1}}{(n-1)!} dt.$$

Remark 3.8. 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 $z_0, z_1, \dots, z_n, \dots \in \mathbb{C}$ and $z \in \mathbb{C}$ the following converges

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

If the points coincide we recover the usual Taylor expansion and from the 3.7 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.

Divided differences enjoy also the following nesting property.

Proposition 3.9. *For any $f : (a, b) \to \mathbb{R}$ and pairwise distinct $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m$ we have*

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

In particular

$$\frac{d^m}{dx^m}\left([x, x_1, x_2, \ldots, x_n]_f\right) = m![x, x, \ldots, x, x_1, x_2, \ldots, x_m]$$

Proof. Note that both $[y_1, y_2, \ldots, y_m]_{[\cdot, x_1, x_2, \ldots, x_n]_f}$ and $[y_1, y_2, \ldots, y_m, x_1, x_2, \ldots, x_n]_f$ satisfy 3.3 and they agree when $m = 1$. ☐

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 3.10 (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, 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

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

As $z \mapsto [z,x_0,x_1,x_2,\dots,x_n]_f$ 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. □

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 $[X]_f = [x_0,x_1,\dots,x_n]_f$, 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 3.11. For instance 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

Proposition 3.12 (Leibniz formula for divided differences)**.** *We have*

$$\begin{aligned}[x_0, x_1, \dots, x_n]_{fg} &= [x_0]_f [x_0, x_1, \dots, x_n]_g \\ &+ [x_0, x_1]_f [x_1, \dots, x_n]_g \\ &+ [x_0, x_1, x_2]_f [x_2, \dots, x_n]_g \\ &+ \dots \\ &+ [x_0, x_1, x_2, \dots, x_n]_f [x_n]_g.\end{aligned}$$

Proof. Write Newton expansion for f and g with points reversed for g. The rest follows as in the final proof of Theorem (3.10). □

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 3.13. *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. □

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 proof of 3.12. Note that if we expand the divided differences, we are almost in the situation of the lemma 3.13; now we just have product of two functions instead. Story is the same. □

3.4 Peano kernels

We already noticed that the n:th divided differences of the form $[x, x, \dots, x, y]_f$ can be written as an integral of n:th derivative of f with some positive polynomial weight. This is no anomaly: if $f \in C^n(a, b)$ and $a < x_0, x_1, \dots, x_n < b$, there exits a function $w = w_{x_0, x_1, \dots, x_n}$ such that

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

It turns out that w, *Peano kernel* with nodes $x_0, x_1, \ldots, x_n$ is piecewise polynomial compactly supported $C^{n-2}(\mathbb{R})$-function. TODO

3.5 k-tone functions

Definition 3.14. $f : (a, b) \to \mathbb{R}$ 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 3.15. *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 3.16. 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 3.17. *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 3.18. *$P^{(k)}$ 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. □

3.6 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

$$f(x) = \int_{x_0}^{x} \rho(t) dt \tag{3.19}$$

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 is fact is actually contained in 3.19: 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

$$f(x) = \int_{x_0}^{x} \int_{x_0}^{x_1} \rho(t) dt dx_1,$$

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

$$\sum_{1\leq i\leq m} |x_i - t| \geq m|\frac{\sum_{1\leq i\leq m} x_i}{m} - t|,$$

also for any convex function

$$\sum_{1\leq i\leq m} f(x_i) \geq mf\left(\frac{\sum_{1\leq i\leq m} x_i}{m}\right),$$

Jensen's inequaltity.

3.7 Majorization

Of course there should be a larger family of inequalities which hold for functions of the form $|x-t|$: it turns out that there is a rather simple characterization for such inequalities, by *majorization.*

Definition 3.20. Let $x = (x_i)_{i=1}^n$ and $y = (y_i)_{i=1}^n$ be two sequences of reals. We say that y majorizes x, and write $x \prec y$, if TODO

Theorem 3.21 (Polya-Hardy-Littlewood-Karamata-Inequality)**.** *Let (a, b) be an open interval, n positive integer, and $x = (x_i)_{i=1}^n \in (a, b)^n$ and $y = (y_i)_{i=1}^n \in (a, b)^n$. Then the following are equivalent.*

1. $x \prec y$
2. *For any real number t we have*

$$\sum_{1\leq i\leq n} |x_i - t| \leq \sum_{1\leq i\leq n} |y_i - t|.$$

3. *For any convex $f : (a, b) \to \mathbb{R}$ we have*

$$\sum_{1\leq i\leq n} f(x_i) \leq \sum_{1\leq i\leq n} f(y_i).$$

Proof. TODO □

3.8 Spectral majorization

In addition to being convenient notion to discuss k-tone functions, majorization explains many phenomena related to spectra of real maps. Basic fact is the following.

Proposition 3.22. *If $A \leq B$, the spec$(A) \prec_1$ spec(B).*

Proof. By Theorem (?) it suffices to check that for any $t \in \mathbb{R}$ we have $\#([t, \infty) \cap \text{spec}(A)) \leq \#([t, \infty) \cap \text{spec}(B))$. Translating by tI this amounts to proving that if $A \leq B$, B has at least as many non-negative eigenvalues as A. But this follows from Lemma 2.28. $\square$

TODO: Higher orders, majorization

3.9 Smoothness

Theorem 3.23. *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 3.24. *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 3.3 and 3.6.*

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 3.3. By construction, our extension satisfies 3.3 and 3.6 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 3.25. *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 3.26. *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 3.3 we have

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

□

Let us then take $2 \leq m = n$. By the lemma 3.25 and the case $m = 1$ we see that $f \in C^1(a, b)$. But by the lemma 3.26 $(n-1)$: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 3.25 immediately implies the remaining cases $\leq m < n$, by induction on n. $\square$

There is however more interesting equivalence to be made.

Theorem 3.27. *Let $f : (a, b) \to \mathbb{R}$ and $n \geq 1$. 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. $\square$

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

Proof of the theorem 3.15. We start with a lemma.

Lemma 3.28. *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. $\square$

Combining the lemma with theorem 3.27 gives the right smoothness. Convexity condition is implied by the lemma 3.26 combined with induction on k. $\square$

3.10 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 3.29. *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 $\left|\frac{f^{(n)}(x_0)}{n!}\right| \le C^{n+1}$ 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

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

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). $\square$

Of course, we could just as well replace the closed inteval by any compact compact subset of (a,b). The previous result is some kind of relative of 3.27. Also theorem 3.15 has rather interesting relative.

Theorem 3.30 (Bernstein's little theorem)**.** *If $f:(a,b)\to\mathbb{R}$ is k-tone for every $k\ge 0$, then f is real-analytic on (a,b).*

Proof. We prove that the conditions of the theorem 3.29 are satisfied. Pick any $a<x_0<x<b$. Now for any $n\ge 0$ 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

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

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. $\square$

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 $(-1)^n f^{(n)}(t)\ge 0$ 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 3.31 (Bernstein's big theorem)**.** *If $f : (0, \infty) : \mathbb{R}$ 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 3.27?

Chapter 4

Matrix functions

4.1 Functional calculus

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

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

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

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. TODO $\square$

4.4 Slick proof

We give slick proof for the fact $f \in C^\omega(a,b) \cap P_\infty(a,b)$ then the Pick matrices are positive. Pick matrices being positive is equavalent to

$$\frac{1}{2\pi i}\int_\gamma f(z)\left|\sum_{1\leq i\leq n}\frac{c_i}{z-z_i}\right| dz \geq 0$$

for any $c_i \in \mathbb{C}$ and $z_i \in \mathbb{H}_+$ such that γ encloses z_i's. Now what we know is that this holds if z_i's are real and we can let some points coincide. But then we consider $g_k(z) = \frac{1}{z^k}\sum_{i=1}^n c_i \frac{z^k - z_i^k}{z-z_i}$, this converges pointwise to what we want, at least for suitable z_i's, and it is of the right form, so we have the claim, essentially. To elaborate a little, we integrate

$$\frac{1}{2\pi i}\int_\gamma f(z)|g_k(z)|^2 dz.$$

As $k \to \infty$, this approaches what we want. Absolute values might a bit off, but something like this.

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.*

We will not prove this quite surprising result yet; the message is 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

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

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 estimate.

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. $\square$

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. TODO □

Curiosly 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 f is analytic.

Proof. TODO □

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

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

In the language of matrices this really says that

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

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

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

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 $z_1, z_2, \ldots, z_n \in \mathbb{D}(z_0, r) \cap U$ 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

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

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

Matrix 5.15 is called *extended Pick matrix*.

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 say that the Taylor coefficients of φ at z_0 don't grow too fast. Second one tells us that if 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 $c_0 =?$.*

Proof. TODO □

Lemma 5.17. *Let $U \subset \mathbb{H}_+$ be open and $\varphi : U \to \overline{\mathbb{H}}_+$. Let $z_0 \in U$ and $r > 0$. 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. TODO. □

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. □

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. TODO □

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 extensions of φ to open subsets of upper half-plane for which all the Pick matrices are positive.

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. □

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 $\varphi \to \overline{\mathbb{H}}_+$ 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.

Lemma 5.21. *Let $A \subset \mathcal{H}_+$ and $\varphi \to \overline{\mathbb{H}}_+$ 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. TODO

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 ball. We also know that $W_{F_1 \cup F_2} \subset W_{F_1} \cap W_{F_2}$. But this means that the family

$$\{W_F \subset \mathbb{H}_+ | F \text{ is finite (non-empty) subset of } A\}$$

has finite intersection property, and since it's members are compact, we know that they have non-empty intersection. This intersection is the place where we find the w_0. □

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 5.21 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 $z_1, z_2, \ldots, z_n \in \mathcal{H}_+ \setminus \{i\}$. Theorem 5.7 for φ and points $i, z_1, \ldots, z_n$ 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 $z_1, z_2, \ldots, z_n$, 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. If $z_1 = i = w_1$, we just noted that this matrix being positive is equivalent to smaller matrix, namely

$$\left(\frac{\frac{w_i - z_i}{1 + z_i w_i} - \frac{\overline{w_i} - \overline{z_i}}{1 + \overline{z_i}\overline{w_i}}}{z_i - \overline{z_j}} \right)_{2 \le i,j \le n}$$

being positive. By inductive hypothesis this then means that there exists a Pick function $\tilde{\varphi}$ with $\tilde{\varphi}(z_i) = \frac{w_i - z_i}{1 + z_i w_i}$. But then $\varphi(z) = \frac{\tilde{\varphi}(z) + z}{1 - z\tilde{\varphi}(z)}$ fits the bill. TODO: general case (introduce Schur transform) □

5.6 Compactness

One can lift the previous argument of the Pick-Nevanlinna interpolation theorem to infinite sets by the following compactness result.

Theorem 5.22. *Let $F = \{\varphi_j | j \in J\}$ be a family of Pick functions uniformly bounded in a point (or upper half-plane). 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 cavaet. The problem is that Schur function constant -1 corresponds to constant ∞, which isn't proper Pick function. But by the boundedness condition this cannot happen. $\square$

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 domain 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 5.22. 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. By Arzelà-Ascoli theorem for every compact set K we can find a subsequence convergent in K and by taking exhaustin upper half-plane by nested compact sequence and taking the diagonal sequence we get the claim. By proposition 5.2 the limit is also Pick function. $\square$

Of course, one could use the previous approach 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.

TODO:

- Poincaré metric: discs are discs, Apollonius circle
- Pick-Nevanlinna-Herglotz representation theorem
- 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

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 $P_n(a, b)$. One immediately sees that that all the matrix monotone functions are monotone as real functions.

Proposition 6.2. *If $f \in P_n(a, b)$, 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 6.3. *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 2.21 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. □

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 6.4. *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. □

That was easy. It's not very easy to come up with other examples, though. Most of the common monotone functions fail to be matrix monotone. Let's try some non-examples.

Proposition 6.5. *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 2.25, $AH + HA$ 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 $AH + HA \ngeq 0$. Then choose parameter $t > 0$ so small that $A + tH \in \mathcal{H}^n(a, b)$ and

$$(A + tH)^2 - A^2 = t(AH + HA + tH^2) \ngeq 0$$

and set the pair $(A, A + tH)$ as the counterexample. □

At this point several other important properties of the matrix monotone functions should be clear.

Proposition 6.6. *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 should be a bit careful though. 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; similar statement holds for increasing functions. Similarly, taking maximums doesn't preserve monotonicity.

Proposition 6.7. *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 $g = \text{id}$. So far the only n-monotone functions we know are affine functions so that's essentially our only hope for counterexamples.

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 $f(A) = a_1 P_{v_1}$ and $f(B) = b_1 P_{w_1}$ where a_1 and b_1 are the positive eigenvalues of A and B, with respective eigenvectors v_1 and w_1. But $f(A) \not\leq f(B)$ as v_1 and w_1 are linearly independent.

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. $\square$

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 $f \in P_n(a, b)$. □

We shall be using especially the previous result a lot.

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 6.10. *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 Derivative and Loewner's characterization

As in the real case, also in the matrix world we may characterize monotonicity with derivatives.

Theorem 6.11. *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. TODO □

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 6.12. *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. □

We hence have the following characterization.

Theorem 6.13. $f \in P_n(a,b) \cap C^1(a,b)$, *if and only if the matrix*

$$([\lambda_i, \lambda_j]_f)_{i,j} \geq 0 \tag{6.14}$$

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 6.14 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.

6.3 Local characterization

It nevertheless turns out that n-monotonicity is a local property.

Proposition 6.15. *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

$$\sum_{1 \leq i \leq n} \frac{c_i}{z - \lambda_i} = \frac{q(z)}{\prod_{1 \leq i \leq n} (z - \lambda_i)} = \frac{q(z)}{p_\Lambda(z)}$$

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 6.16. *h is polynomial of degree at most $(2n-2)$ 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 real 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 3.13 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 3.27.

Lemma 6.17. *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 $n \geq (\max_{1\leq i\leq m} d_i) - 1$. 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. □

We have hence the following.

Theorem 6.18. *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 6.19. *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 6.19. 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 suitable 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 $h \equiv 1$. $\square$

6.4 Main Theorem

Finally, one might ask whether we can get rid of C^1-assumption, and it turns out that we can.

Theorem 6.20. *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 6.21. *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 6.21 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 6.19.

Finally in this case, by the lemma 6.21 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 3.15

□

Let us then complete proof by proving the lemma 6.21.

Proof of lemma 6.21. The proof is based on lemmas 2.45 and 2.46. 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

$$(\lambda_1(A), \dots, \lambda_n(A), \lambda_1(B), \dots, \lambda_n(B)) \tag{6.22}$$

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 6.16 and the following observation.

Lemma 6.23. *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 2.47. $\square$

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. $\square$

6.5 A bit of history

Theorem 6.20 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

$$
\begin{bmatrix}
\frac{f'(t)}{1!} & \frac{f^{(2)}(t)}{2!} & \cdots & \frac{f^{(n)}(t)}{n!} \\
\frac{f^{(2)}(t)}{2!} & \frac{f^{(3)}(t)}{3!} & \cdots & \frac{f^{(n+1)}(t)}{(n+1)!} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{f^{(n)}(t)}{n!} & \frac{f^{(n+1)}(t)}{(n+1)!} & \cdots & \frac{f^{(2n-1)}(t)}{(2n-1)!}
\end{bmatrix}
\tag{6.24}
$$
$$
= \begin{bmatrix}
[t,t]_f & [t,t,t]_f & \cdots & [t,\ldots,t]_f \\
[t,t,t]_f & [t,t,t,t]_f & \cdots & [t,t,\ldots,t]_f \\
\vdots & \vdots & \ddots & \vdots \\
[t,\ldots,t]_f & [t,t,\ldots,t]_f & \cdots & [t,t,t,\ldots,t]_f.
\end{bmatrix}
$$

Now an alternative version of 6.20 reads as follows.

Theorem 6.25. *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. TODO □

One can again get rid of the smootness assumption by some careful considerations. TODO

6.6 Loewner's theorem

In addition to characterizing n-monotone functions, by theorem 6.13, the classes $P_n(a,b)$, Loewner characterized the classes $P_\infty(a,b)$.

Theorem 6.26. *$f\in P_\infty(a,b)$, 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 6.13.

The "only if" is the tricky part. Theorem 6.13 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 6.27. *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 3.29.

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 6.28. *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

$$2M_0 \geq f\left(\frac{c}{M_2} + x_0\right) - f(x_0) \geq \int_{x_0}^{\frac{c}{M_2}+x_0} \left(c - M_2(x - x_0)\right) dx \geq \frac{c^2}{2M_2},$$

   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: two pictures. TODO: better proof $\square$

To prove the bound for odd n, we would like to play the same game as in the proof of lemma 3.27, 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 know we can simply follow the same argument. $\square$

TODO: another trick with $(a,b) = (0,\infty)$ $\square$

TODO:

- 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 $0 \le j \le 3$ and $H \ge 0$. Then is it necessarily the case that
$$\operatorname{tr}\left(A_3|A_3| - 3A_2|A_2| + 3A_1|A_1| - A_0|A_0|\right) \ge 0?$$
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
$$f(Q_A(v)) \le Q_{f(A)}(v)$$
for any convex f. Using this and Minkowski one sees that p-schatten norms are indeed norms.

- For f, g generalization (Look at $h(X) = g(\operatorname{tr} f(X))$) 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

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 6.11.

7.1 Basic properties

Definition 7.1. 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 7.2. *Let $(a, b) \subset \mathbb{R}$ be an open interval $n \geq 1$, and $k \geq 1$. Now*

1. *$P_n^k(a, b)$ is a convex closed cone.*
2. *$P_{n+1}^k(a, b) \subset P_n^k(a, b)$.*
3. *$\left(x \mapsto \alpha_k x^k + \ldots + \alpha_1 x + \alpha_0\right) \in P_n^k(a, b)$ if $\alpha_n \geq 0$.*
4. *$\left(x \mapsto (-1)^k x^{-1}\right) \in P_n^k(a, b)$.*

Proof. TODO □

Not surprisingly, we have also the following derivative characterization.

Theorem 7.3. *Let $n, k \geq 1$ and $f \in C^k(a, b)$. Then the following are equivalent:*

(i) $f \in P_n^k(a, b)$.

(ii) For any $A \in \mathcal{H}_{(a,b)}^n$ and $H \geq 0$ we have

$$D_n^k f_A(H) \geq 0.$$

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.

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] 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.

[26] G. Frobenius. Ueber die Entwickelung analytischer Functionen in Reihen, die nach gegebenen Functionen fortschreiten. *J. Reine Angew. Math.*, 73:1–30, 1871.

[27] F. Hansen. Trace functions as laplace transforms. *Journal of mathematical physics*, 47(4):043504, 2006.

[28] F. Hansen. The fast track to Löwner's theorem. *Linear Algebra Appl.*, 438(11):4557–4571, 2013.

[29] F. Hansen and G. Kjærgård Pedersen. Jensen's inequality for operators and löwner's theorem. *Mathematische Annalen*, 258(3):229–241, 1982.

[30] F. Hansen and J. Tomiyama. Differential analysis of matrix convex functions. *Linear Algebra Appl.*, 420(1):102–116, 2007.

[31] F. Hansen and J. Tomiyama. Differential analysis of matrix convex functions. II. *JIPAM. J. Inequal. Pure Appl. Math.*, 10(2):Article 32, 5, 2009.

[32] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge university press, 1952.

[33] U. Helmke and J. Rosenthal. Eigenvalue inequalities and schubert calculus. *Mathematische Nachrichten*, 171:207–226, 1995.

[34] F. Hiai and D. Petz. *Introduction to matrix analysis and applications*. Springer Science & Business Media, 2014.

[35] C. J. Hillar. Advances on the bessis–moussa–villani trace conjecture. *Linear algebra and its applications*, 426(1):130–142, 2007.

[36] A. Hindmarsh. Pick's conditions and analyticity. *Pacific Journal of Mathematics*, 27(3):527–531, 1968.

[37] J. A. Holbrook. Spectral variation of normal matrices. *Linear algebra and its applications*, 174:131–144, 1992.

[38] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[39] 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.

[40] J. Karamata. Sur une inégalité relative aux fonctions convexes. *Publications de l'Institut mathematique*, 1(1):145–147, 1932.

[41] I. Klep and M. Schweighofer. Sums of hermitian squares and the bmv conjecture. *Journal of Statistical Physics*, 133(4):739–760, 2008.

[42] A. A. Klyachko. Stable bundles, representation theory and Hermitian operators. *Selecta Math. (N.S.)*, 4(3):419–445, 1998.

[43] A. Knutson and T. Tao. Honeycombs and sums of Hermitian matrices. *Notices Amer. Math. Soc.*, 48(2):175–186, 2001.

[44] G. Kowalewski. *Interpolation und genäherte Quadratur.* Teubner Leipzig und Berlin, 1932.

[45] F. Kraus. Über konvexe Matrixfunktionen. *Math. Z.*, 41(1):18–42, 1936.

[46] K. Löwner. Über monotone Matrixfunktionen. *Math. Z.*, 38(1):177–216, 1934.

[47] 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.

[48] 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.

[49] 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.

[50] R. Nevanlinna. *Über beschränkte funktionen die in gegebenen punkten vorgeschriebene werte annehmen...* Buchdruckerei-a.-g. Sana, 1919.

[51] G. Pick. Über die beschränkungen analytischer funktionen, welche durch vorgegebene funktionswerte bewirkt werden. *Mathematische Annalen*, 77(1):7–23, 1915.

[52] A. Pinkus and D. Wulbert. Extending n-convex functions. *Studia Math*, 171(2):125–152, 2005.

[53] T. Popoviciu. *Les fonctions convexes*, volume 992. Bussière, 1945.

[54] A. W. Roberts and D. E. Varberg. *Convex functions*, volume 57. Academic Press, 1974.

[55] D. Sarason. Generalized interpolation in H^∞. *Trans. Amer. Math. Soc.*, 127:179–203, 1967.

[56] D. Sarason. Nevanlinna-pick interpolation with boundary data. *Integral Equations and Operator Theory*, 30(2):231–250, 1998.

[57] H. R. Stahl et al. Proof of the bmv conjecture. *Acta Mathematica*, 211(2):255–290, 2013.

[58] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 7.1)*, 2016. http://www.sagemath.org.

[59] J. A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[60] 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.

[61] F. Zhang. *Matrix theory: basic results and techniques*. Springer Science & Business Media, 2011.