Matrix monotone and convex functions

Otte Heinävaara

March 1, 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
        - 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. Preliminaries (partially to be dumped to appendix?)

- Divided differences: basic definition, properties, representations and smoothness
- Positive matrices
  - Setup: finite (vs infinite) dimensional inner product spaces over $\mathbb{C}$ (vs $\mathbb{R}$), basic facts
  - Linear maps, adjoint, congruence, self-adjoint maps, spectral theorem: finite and infinite dimensional
  - Good properties of spectrum
  - Positive maps: basic properties (cone structure, Sylvester's criterion etc.)
- Matrix functions
  - Several definitions
  - Derivatives of matrix functions
- Divided differences: basic definition, properties, representations and smoothness
- Pick functions
  - Basic definitions and properties
  - Pick-Nevanlinna representations theorem
  - Pick matrices/ determinants
  - Compactness
  - Pick-Nevanlinna interpolation theorem
- Regularizations
  - Basic properties
  - Lemmas needed for some of the proofs

3. 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

4. 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

5. Structure of the classes P_n and K_n, interpolating properties (?)
   - Strict inclusions, strict smoothness conditions
   - Strictly increasing functions
   - Extreme values
   - Interpolating properties

6. 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

7. Alternative characterizations (?)

   - Some discussion, maybe proofs

8. Bounded variations (?)

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

1.3 Some random ideas

1. It's easy to see that [Something]. Actually, it's so so easy that we have no excuse for not doing it.
2. When is matrix of the form $f(a_i + a_j)$ positive: f is completely monotone (?).
3. Polynomial regression...
4. TODO: Maximum of two matrices (at least as big), (a + b)/2 + abs(a - b)/2
5. If $\langle Ax, y\rangle = 0$ implies $\langle x, Ay\rangle = 0$, then A is constant times hermitian.
6. Angularity preserving functions
7. If subspace of linear maps are diagonalizable with real eigenvalues, is there a inner product such that subspace consists of only Hermitian maps

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.

Ideas how to rewrite this section:

- Map is positive, if all of its restrictions are. One dimensional maps are positive if and only the scalar is positive.
- Map is real, if all of its restrictions are. One dimensional maps are real if and only the scalar is real.
- Adjoint commutes with restriction. Adjoint of one dimensional map is its conjugate.
- How should one restrict linear mappings? (Use inclusions and finally project)
- Central notion is adjoint
- $*$-conjugation is natural notion with adjoint
- Increasing sequence of subspaces and restriction with positive determinant $\Rightarrow$ the whole map is positive. Equivalently, map is positive if it has positive determinant and has a subspace where it is positive.

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 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 $\ker(P) = \mathrm{im}(P)^{\perp}$. Such maps are easily seen to satisfy all the requirements.

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, $*$-restriction 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 $*$-restriction 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. TODO □

For one-dimensional restrictions 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

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

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

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

Restriction 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 restricions we need. When map is restricted to a subspace, we naturally lose some imformation about the map. Knowing about all of the restrictions, however, we can get our map back.

Lemma 2.3. *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. □

2.3 Positive maps

Now that we have defined the restrictions 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.

The following properties are evident.

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* $v \in V$ *we have* $\langle Ax, x\rangle \geq 0$, *or 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.*

TODO: more/better explanation This is where we see the usefulness of the restriction: it allows us to lift all the useful properties of reals to linear mappings. 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, and write $A \geq 0$, 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, and write $A \geq 0$, 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 called Hermitian and Skew-Hermitian and as with positive maps, many of their properties are lifted form usual complex numbers.

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 is immediate from the injectivity of restrictions. TODO $\square$

2.4 Ideas

- Square root of a matrix
- Sylvester's criterion
- adjoints of vectors
- projections with inclusions
- order of projections
- Loewner order
- Schur complements
- Moore-Penrose pseudoinverse
- (canonical, löwdin) orthogonalization, polar decomposition and orthogonal Procrustes problem
- projection matrices
- Hilbert-Schmidt norm ($\to$ matrix functions?) and inner product
- Hilbert spaces
- Real vs. complex
- Positive definite kernels
- Weakly positive matrices
- Hlawka inequality for determinant

Ideas how to rewrite this section:

- Map is positive, if all of its restrictions are. One dimensional maps are positive if and only the scalar is positive.
- Map is real, if all of its restrictions are. One dimensional maps are real if and only the scalar is real.
- Adjoint commutes with restriction. Adjoint of one dimensional map is its conjugate.
- How should one restrict linear mappings? (Use inclusions and finally project)

- Central notion is adjoint
- $*$-conjugation is natural notion with adjoint
- Increasing sequence of subspaces and restriction with positive determinant $\Rightarrow$ the whole map is positive. Equivalently, map is positive if it has positive determinant and has a subspace where it is positive.

Chapter 3

Divided differences

Divided differences are derivatives without limits.

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

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

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.

TODO:

- Mean value theorem, coefficient of the interpolating polynomial
- Basic properties, product rule.
- k-tone functions, smoothness, and representation
- Jensen and Karamata inequalities, generalizations
- Cauchy's integral formula

- Polynomial interpolation
- Peano Kernels: Smoothness properties, Bernstein (?) polynomials as examples.
- Opitz formula
- Regularization techniques

Chapter 4

Matrix functions

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

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 often 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 $(\varphi_i)_{i=1}^n$ 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 actually a considerable strengthening of the previous result.

Proposition 5.2. *If $(\varphi_i)_{i=1}^n$ 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.

5.2 Schur functions

As we have noticed, Pick functions need not be injections or surjections. Some are both: simple examples are functions of the form $\alpha z + \beta$ and $\frac{\alpha}{\lambda - z} + \beta$ for $\alpha > 0$ and $\beta, \lambda \in \mathbb{R}$. And that's all.

Before trying to understand why is that, we have to change the point of view. All the previous functions are rational functions, but even more is true: they are all Möbius transformations. Möbius transformations are analytic bijections of extended complex plane i.e. Riemann sphere, to itself. Our examples all exactly those Möbius transformation which map the extended real axis to itself, and don't change the orientation, so the upper half-plane is mapped to itself and not to the lower half-plane. When viewed as a part of the Riemann sphere, upper half-plane is just a hemisphere. Of course it shouldn't matter too much which hemisphere we are looking at, so we could also consider mappings from unit disc to itself (or closed unit disc, to be precise). These mappings are called *Schur* *functions* and class of Schur functions is denoted by $\mathcal{S}$. It's then natural to conjecture that bijective Schur functions are exactly the Möbius transformations which map unit circle to unit circle, and don't change the orientation so that the inside is mapped to the inside.

These claims are easily derivable from each other as follows. Consider the pair of Möbius transformations

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

They are inverses of each other and map the (open) unit disc to upper half-plane and back, respectively. Now take any bijective Schur function $\psi : \mathbb{D} \to \mathbb{D}$. Then $\varphi = \xi \circ \psi \circ \eta$ is bijective Pick function. Similarly one could invert $\psi = \eta \circ \varphi \circ \xi$. This means that bijections can be paired: if all bijective Pick functions are Möbius transformations, so are all bijective Schur functions, since non-Möbiusness on one side would give rise to non-Möbiusness on the other side.

Still before proving anything we should think about this relation a bit further. We noticed that every bijective Pick function has a corresponding Schur function pair. This correspondence is however by no means unique, it was merely our choice to choose such ξ and η. Still, there is need to restrict ourselves to bijections anymore. If one takes **any** Schur function $\psi : \mathbb{D} \to \mathbb{D}$ we can form the corresponding Pick function by taking $\varphi = \xi \circ \psi \circ \eta$. This gives rise to bijection $\mathcal{S} \to \mathcal{P}$, and the inverse should be rather obvious by now. Of course, it's not a big surprise that there would be such bijection, that is to say that the sets are equal in size, but our bijection preserves composition of functions. All this is to say that in some sense these classes are almost the same.

One should be a bit more careful here though: we have included also real constant functions to our class $\mathcal{P}$ and we should also add unimodular constants to $\mathcal{S}$. For these the bijection doesn't quite work; we can mostly do a natural extension, but then one would be forced to map the constant function -1 to the constant ∞. This means that we should add the constant infinity function to our Pick functions. We will not do this, as it would change the whole business to Riemann sphere, since it will bring other technical problems, but we will try to indicate when you should think about this extension.

If one only thinks about composition one can of course do lot more. Take any simply connected domain in $U \subset \mathbb{C}$. By Riemann mapping theorem there's a analytic bijection $\xi_U \mathbb{D} \to U$. For the domain U we could define similar class of functions, and via ξ_U and it's inverse we could connect the classes. Again, one should be a bit careful with the boundary.

In many ways Pick and Schur functions are most natural of these classes: they are closed under addition and multiplication, respectively. Also, they both contain the identity of the respective operations, so these properties are barely true.

5.2.1 Automorphisms of the unit disc

As mentioned, the functions ξ and η are not unique. All such mappings are however of a very simple form. If $\rho : \mathbb{D} \to \mathbb{D}$ is analytic bijection, the function $\xi \circ \rho$ is an analytic bijection from unit disc to upper half-plane. Conversely, if ξ_1 is an analytic bijection from unit disc to upper half-plane, $\eta \circ \xi_1$ is an analytic bijection. Hence to understand the diversity of the analytic mappings from $\mathbb{D}$ to $\mathbb{H}_+$, we need to understand the analytic self-maps of the unit disc.

All analytic bijections from the unit disc to itself, called the automorhphisms of the unit disc, are given by

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

where $a \in \mathbb{D}$ and $\omega \in \mathbb{S}$. We will write $\rho_a = \rho_{a,1}$. As $|1 - \overline{a}z|^2 - |a - z|^2 = (1 + |a||z|^2 - \overline{a}z - \overline{z}a) - (|a|^2 + |z|^2 - \overline{a}z - \overline{z}a) = (1 - |z|^2)(1 - |a|^2)$, one readily sees that such mappings are indeed analytic bijections. Schwarz lemma explains why these are the all.

5.3 Schwarz lemma

At first sight one might not guess that Pick function have strong regularity properties. In some sense they however work like Schur functions, and they feel much more restrictive. If one considers a Schur function ψ mapping zero to itself very classic lemma of Schwarz states that $|\psi(z)| \leq z$ for any $z \in \mathbb{D}$. For completeness let's review a standard proof.

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 $\omega \in \mathbb{S}$.*

Proof. By the assumption ψ can be represented as a locally uniformly convergent power series in unit disc, of the form $\sum_{n=1}^{\infty} a_1 z^n$. Now $\psi(z)/z := \sum_{n=0}^{\infty} a_{n+1} z^n$ defines also an analytic function in unit disc. For any $0 < r < 1$, by the maximum modulus principle, we have

$$\sup_{z \in \mathbb{D}(0,r)} \left| \frac{\psi(z)}{z} \right| \leq \sup_{z \in \mathbb{S}(0,r)} \left| \frac{\psi(z)}{z} \right| \leq \frac{1}{r},$$

so by letting $r \to 1$ we get $\sup_{z \in \mathbb{D}} \left| \frac{\psi(z)}{z} \right| \leq 1$ and hence for any $z \in \mathbb{D}$

$$|\psi(z)| \leq \left| \frac{\psi(z)}{z} \right| |z| \leq |z|$$

TODO $\square$

One might argue that this proof hides all the mysteriousness in the maximum modulus principle, which is quite surprising itself too. Maximum modulus principle can be seen as a consequence of the Cauchy's integral formula.

There is somewhat better form of the Schwarz lemma, called the invariant form or 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 automorphism of the unit disc.

Note that one obtains the usual Schwarz lemma if $z_1 = 0 = \psi(z_1)$. One may check that if ψ is indeed automorphism, 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. $\square$

There are many ways to think about these results. One immediate interpretation is that Schur functions are in some precise sense very rigid. Knowing a Schur function in a point immediately restricts the values the functions might attain at some other points. We could make this more precise. If we consider the points z_1, z_2 and the value $\psi(z_1)$ how exactly is the value $\psi(z_2)$ restricted? If $z_1 = 0 = \psi(z_1)$, we are in the original Schwarz lemma and $\psi(z_2)$ is simply restricted in a closed disc of radius $|z_2|$ around 0. TODO

Other interpretation is that we may factor Schur functions. If ψ is a Schur function with $\psi(z) = 0$, then also $\frac{\psi(z)}{z}$ is a Schur function; this is the main step in the proof of Schwarz lemma. More generally, for any Schur function ψ, the function

$$\psi_{z_0}(z) = \frac{1 - \overline{z_0} z}{z_0 - z} \frac{\psi(z_0) - \psi(z)}{1 - \overline{\psi(z_0)}\psi(z)}$$

is also a Schur function for any $z_0 \in \mathbb{D}$. In particular, if $\psi(z_0) = 0$ for some $z_0 \in \mathbb{D}$, we can write

$$\psi(z) = \frac{z - z_0}{1 - \overline{z_0} z} \psi_{z_0}(z).$$

There is a correponding variant of the Schwarz-Pick theorem for the upper half-plane, for the 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*

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

Proof. Apply the Schwarz-Pick theorem to the map $\xi \circ \varphi \circ \eta$. □

In a way there's no natural variant of the original Schwarz lemma for the upper half-plane. Direct analogue would be to consider Pick functions with $\varphi(i) = i$, but one might wonder if this is really any simpler. The problem is that there's no canonical center to the upper half-plane, although we've been implicitely taking it as i (by choosing ξ and η). One might however argue that there's no canonical center to the unit disc either. With automorphisms ρ_a we may map any point of the unit disc to any other point. For any Schur function ψ and $a \in \mathbb{D}$ we can consider the map $\rho_a \circ \psi \circ \rho_a$. TODO

5.3.1 Poincaré metric

There's really nice interpretation for the Schwarz-Pick theorem. The unit disc $\mathbb{D}$ can be equipped with a hyperbolic metric, which for any two points $z_1, z_2 \in \mathbb{D}$ is given by

$$2\tanh^{-1}\left|\frac{z_1 - z_2}{1 - \overline{z_1}z_2}\right|.$$

This is space is called Poincaré disc. One may check that the previous is indeed a metric, and automorhisms of the unit disc are exactly the isometries of this space. Now Schwarz-Pick theorem states that any Schur function decreases distances i.e. Schur functions are contractions on Poincaré disc.

Analogously one could interpret that the Pick functions are contractions in Poincaré upper half-plane, metric space in upper half-plane in which the metrix is given by

$$2\tanh^{-1}\left|\frac{z_1 - z_2}{z_1 - \overline{z_2}}\right|.$$

5.3.2 Pick Matrices

Schwarz-Pick theorem was about restricting Schur functions at one point. If we fix two points instead, things get more complicated. Let's say we have a Schur function ψ and we know that $\psi(0) = 0$ and $\psi(\frac{1}{2}) = \frac{1}{4}$. Surely there are such Schur functions, $z \mapsto \frac{1}{2}z$ for instance, but at least Schwarz-Pick theorem doesn't immediately fix such function, which it would do should we change $\frac{1}{4}$ to $\frac{1}{2}$. The question is: what kind of values could a Schur function attain at some other point, say in $i\frac{1}{2}$. There are two things we can say immediately.

TODO (Introduce Pick matrices earlier)

5.4 Pick-Nevanlinna interpolation theorem

We have seen that the Schur functions are contractions in Poincaré disc and Pick function are contractions in Poincaré upper half-plane. Not every contraction in these spaces is however analytic. Take disc for instance. It turns out that $z \mapsto \Re(z)$ is **not** contraction in the Poincaré disc, but for small enough positive ε, $z \mapsto \varepsilon\Re(z)$ definitely is. One could immediately push this example to upper half-plane, by using the maps ξ and η, or do something similar from the scratch and take, say, $z \to i + \varepsilon\Re(1 + 2i - \frac{2i}{z+i})$. In the both examples the philosophy is the same: the hyperbolic metric in both cases look locally like (a scaled version of) standard metric. So if we map everything to small neighbourhood everyhing behaves nicely. In the disc, this is very simple, just scale. In the upper half-plane, we can for instance first map everything to some compact set in the upper half-plane. Then break the analyticity and scale to get contractivity.

5.4.1 Pick Matrices

There is however rather simple characterization for Schur functions requiring a bit more than contractivity. It turns out, that the condition "$\psi : \mathbb{D} \to \mathbb{D}$ is contraction" is equivalent to the matrix

$$\begin{bmatrix} \frac{1-|\psi(z_1)|^2}{1-|z_1|^2} & \frac{1-\overline{\psi(z_1)}\psi(z_2)}{1-\overline{z_1}z_2} \\ \frac{1-\overline{\psi(z_2)}\psi(z_1)}{1-\overline{z_2}z_1} & \frac{1-|\psi(z_2)|^2}{1-|z_2|^2} \end{bmatrix}$$

being positive. Such matrix is called Pick matrix. Conveniently, even the condition "ψ maps $\mathbb{D}$ to itself" is built-in in the positivity. This is not entirely obvious but it may straighforwardly, although tediously verified by checking that the Schwarz-Pick inequality is equivalent to the determinant of the Pick matrix being non-negative.

Similarly, for any sequence of say n points in the unit disc we may form the matrix

$$\begin{bmatrix} \frac{1-|\psi(z_1)|^2}{1-|z_1|^2} & \frac{1-\overline{\psi(z_1)}\psi(z_2)}{1-\overline{z_1}z_2} & \cdots & \frac{1-\overline{\psi(z_1)}\psi(z_n)}{1-\overline{z_1}z_n} \\ \frac{1-\overline{\psi(z_2)}\psi(z_1)}{1-\overline{z_2}z_1} & \frac{1-|\psi(z_2)|^2}{1-|z_2|^2} & \cdots & \frac{1-\overline{\psi(z_2)}\psi(z_n)}{1-\overline{z_2}z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-\overline{\psi(z_n)}\psi(z_1)}{1-\overline{z_n}z_1} & \frac{1-\overline{\psi(z_n)}\psi(z_2)}{1-\overline{z_n}z_2} & \cdots & \frac{1-|\psi(z_n)|^2}{1-|z_n|^2} \end{bmatrix}$$

TODO:

- Poincaré metric
- Pick matrices
- 2 proofs of the Pick-Nevanlinna interpolation theorem
- Hindmarsh's theorem
- Pick-Nevanlinna-Herglotz representation theorem
- Compactness
- Examples of representing measures behind functions and functions behind representing measures
- "Pointed" Pick-Nevanlinna interpolation: two proofs (one with Pick-Nevanlinna, one with congretely extending)
- Spectral commutant lifting theorem

Chapter 6

Monotone and Convex matrix functions

We already introduced monotone and convex 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 and convex 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.

We will start with the matrix monotone functions; much of the discussion carries quite directly to the convex case.

6.1 Basic properties of the matrix monotone functions

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 $A, B \in \mathcal{H}_{(a,b)}$, 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. TODO $\square$

The equivalence of the first two is almost obvious and from this point on we shall identify 1-monotone and increasing functions. But the third point is very important: it is exactly the non-commutative nature which makes the classes of higher order interesting.

Let us then have some examples.

Proposition 6.4. *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. $\square$

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 TODO, $AH + HA$ need not be positive! Also, if H is small enough, H^2 is negligible compared to $AH + HA$. We are ready to formulate our proof strategy: find $A \in \mathcal{H}^n_{a,b}$ and $\mathbb{H}^n_+$ such that $AH + HA \not\geq 0$. Then choose parameter $t > 0$ so small that $A + tH \in \mathcal{H}^n(a, b)$ and

$$(A + tH)^2 - A^2 = t(AH + HA + tH^2) \not\geq 0$$

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

TODO □

In a similar manner one could show the similar statement for the functions $(x \mapsto x^k)$.

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 Now $f(A) \leq f(B) \leq \max(f, g)(B)$. It follows that

$$\max(f, g)(A) = \max(f(A), g(A)) \leq \max(f, g)(B),$$

as we wanted.

Here the flaw is in the expression $\max(f(A), g(A))$: what is maximum of two matrices? This is an interesting question and we will come back to it a bit later, but it turns out that however you try to define it, you can't satisfy the above inequality.

We still need proper counterexamples though. Let's try $f \equiv 0$ and $g = \mathrm{id}$. So far the only n-monotone functions we know are affine functions so that's essentially our only hope for counterexamples.

TODO □

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. □

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. TODO □

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.

TODO:

- Examples
- Pick functions are monotone
- Heaviside function
- Positive derivative
- Smoothness properties
- Characterizations