Matrix monotone and convex functions

Otte Heinävaara

January 20, 2018

Contents

Chapter 1

Matrix monotone functions – part 0

1.1 Disclaimer

Let's be honest: this master's thesis is really not about matrix monotone functions. What is it about, then? Well, unfortunately the only way I know how to answer that question is to explain what the matrix monotone functions are.[1] Hence the title.

1.2 What are matrix monotone functions?

Definition 1.1. 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 on (a, b), if for any two $n \times n$ Hermitian matrices A and B with spectra in (a, b), such that $B - A$ is positive semidefinite, also $f(B) - f(A)$ positive semidefinite. Here $f(A)$ and $f(B)$ are defined via functional calculus.

Now, it might not be too big of a surprise that, on the surface level at least, the main question of this thesis is the following.

Question 1.2. Fix positive integer n and an open interval (a, b). Which functions are n-monotone on (a, b)?

If all this makes sense to you, great! Feel free to skip this section. If not, what follows is an attempt to give some kind of handwavy picture of the setup. Alternatively, if you don't like handwaving, you may feel free to visit chapter 3 (TODO, which chapter) for rigorous foundations, although be warned: this might not be the book for you.

[1] Worry not: one need not read beyond this chapter to get some kind of answer to the question.

Matrix monotonicity is generalization of standard monotonicity 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

$$A \leq B \tag{1.3}$$

we should also have

$$f(A) \leq f(B). \tag{1.4}$$

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 the choice of words.
- Somehow I can't satisfactorily fill the following table:

monotonic	monotonicity
increasing	?

How very inconvenient.

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

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

Both of these questions can be (of course) answered in many ways, but for both of them, there is very natural, in fact tensorial answer. Instead of comparing matrices we can compare bilinear forms, $(0, 2)$-tensors. Similarly we can naturally apply function to linear mappings, $(1, 1)$-tensors.

For matrix (bilinear form) ordering we should first understand which matrices are *positive*, which here, a bit confusingly maybe, means "at least zero". We say that a form is positive if its diagonal is non-negative. This gives a partial order on the space of all bilinear forms.

For matrix functions, i.e. "how to apply function to matrix" the idea is to take a real 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, given the ring structure of linear maps themselves. If the argument (a linear map) is diagonalizable, this extension merely applies the function to the eigenvalues. This motivates to define $f(A)$ for linear map A to be linear map with same eigenspace structure as A but the eigenvalues changed from $\lambda \to f(\lambda)$ respectively. All this works for diagonalizable maps with real eigenvalues, so the domain isn't quite $\mathbb{R}^{n\times n}$ but that's okay. This extension idea is called **functional calculus**.

All this is kind of enough to make sense of matrix monotonicity, but to drastically simplify the setup it is customary to restrict the attention to a special set of diagonalizable matrices, which in this text are called **real maps**. They are exactly the symmetric matrices and they hold special place amidst the set of all matrices.

- They exactly correspond to symmetric bilinear forms.
- They correspond to diagonalizable linear maps with real eigenvalues and orthogonal eigenbasis.

In the second point we are thinking about everything in terms of standard inner product of $\mathbb{R}^n$. So the statement should be corrected to

- If considered as matrix of a linear map with respect to the standard orthonormal basis of $\mathbb{R}^n$ (with the standard inner product), then the linear map is diagonalizable with real eigenvalues and has orthogonal eigenbasis.

Real maps are usually called **Hermitian or self-adjoint matrices** and positive matrices **positive semidefinite matrices**. Now the definition of matrix monotinicity 1.1 should make sense. We will primarily call positive matrices **positive maps**.

Whether one should think about real maps as matrices, bilinear forms or linear maps depends on the context. If one does calculations, one might think about matrices. If one thinks about additive structure, bilinear forms are better suited. And of course functional calculus makes only sense with linear maps. We use the (linear) map terminology throughout mainly because it short. Also, it is a constant reminder that there is something tensorial going on.

1.3 . . . And why should we care?

It's easy to come up with one matrix monotone function: $x \mapsto \alpha x + \beta$ for $\alpha, \beta \in \mathbb{R}$ with $\alpha \geq 0$. It is n-monotone for every $n \geq 1$ on $(-\infty, \infty)$. This is the only easy example.

But there are lot more.

Matrix monotone functions are truly horrible. All matrix monotone functions are increasing (in the usual sense) but not vice versa. They have some obscure regularity properties. Constructing non-trivial matrix monotone functions is a pain. Although usual increasing real functions and matrix monotone functions should be very much interlinked, hardly any of the properties of increasing functions pass on to matrix monotonicity. Generally, if one attacks matrix monotone functions, especially of order $n > 2$, and doesn't use sophisticated weaponry, one will perish. The reader is encouraged to try.

All this is exactly what makes them so interesting. One is driven to ask the question:

Question 1.5. How should one think about matrix monotone functions?

If this sounds like the same question to you, think about increasing functions on $\mathbb{R}^n$. Function $f : \mathbb{R}^n \to \mathbb{R}$ is called n-increasing (this termonology lasts only for next couple of paragraphs) if $f(x_1, x_2, \ldots, x_n) \leq f(y_1, y_2, \ldots, y_n)$ whenever $x_i \leq y_i$ for every $1 \leq i \leq n$. Which functions are n-increasing? I would argue that n-increasing functions are awful, much more awful than usual (1-)increasing functions. The reason is that they don't have good additive structure.

TODO: pictures of n-increasing functions

One might say that "non-negative derivative" property (let's ignore smoothness issues for a while) makes increasing functions easy to understand, and while there is certain truth to that, I would argue that what makes them so simple is really the dual property: "increasing functions are sums of increasing step functions". This roughly implies that in order to understand increasing functions, it is enough to understand step functions, or just step functions with one jump upwards.

Note that we are heavily using the fact that increasing functions (of all types introduced before) form a convex cone:

Definition 1.6. Subset C of a vector space V over $\mathbb{R}$ is a convex cone if whenever $v, w \in C$ and $\alpha, \beta \geq 0$, also $\alpha v + \beta w \in C$.

Also, applicability of the "only needing to understand step functions" is somewhat limited: it doesn't really explain smoothness phenomena all too well, for instance. But it is always nice to know that some objects are really sums of other much simpler objects.

There's no such nice dual property for n-increasing functions (for $n > 1$). One can understand them locally with derivatives, but there are no simple decompositions. Same thing could be said about convex functions on $\mathbb{R}^n$.

Much more importantly for us, these is no such nice additive structure for n-monotone functions. This is by no means trivial (as it is not even with n-increasing functions). It is also not even clear what one means by "nice" and wheter even increasing functions are that "nice" in the end. These ideas shall however merely work as our guideline, so one should not be troubled.

All these issues can be, in a way, avoided by change of perspective: instead of trying to characterize matrix monotone functions by expressing them as sums of something simple, we express the definition itself as a sum of somethings simple. In particular we try to understand the "dual" (or a "predual" to be exact) of matrix monotone functions.

1.4 Dual cones

Let in the following V be a vector space over $\mathbb{R}$ and denote its dual by V^*.

Definition 1.7. For every subset C^* of V^* we define its **dual cone** to be

$$C = \{v \in V | w^* v \geq 0 \text{ for every } w^* \in V^*\} \subset V.$$

One immediately makes the following observation justifying the terminology.

Theorem 1.8. *Let $C^* \subset V^*$. Then the dual cone of C^* is a convex cone.*

Definition 1.9. Let $C \subset V$. Then C^* is a **predual** of C if C is the dual cone of C^*.

Of course only convex cones have preduals. Easy examples show that preduals are not unique in general (in fact never).

As an example for open interval (a, b) consider the set

$$P_1(a, b) := \{\text{Increasing functions } f : (a, b) \to \mathbb{R}\}.$$

This set is a convex cone. If one denote the evaluation functional or measure at x by δ_x, i.e. $\delta_x(f) = f(x_0)$, then one possible predual of $P_1(a, b)$ if given by

$$\{\delta_y - \delta_x | a < x < y < b\}.$$

I hope the reader agrees that this predual is in many ways much simpler than the set of increasing functions and yet it carries the information thereof. As we will see, if chosen suitably, preduals can offer convenient and clean way language to talk about cone itself. This is what this thesis is really about.

Chapter 2

Positive maps

2.1 Motivation

2.1.1 The right definition

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

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

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

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

Note that, contrary to the previous chapter, we snuck in the complex numbers and general vector space to the definition. It doesn't make much difference whether we talk about real or complex numbers but the author thinks that some of the arguments are more natural in the complex world. Also, having the general vector space V is mostly just reminder of the fact that there is something tensorial going on.

Theorem 1.8 immediately implies

Theorem 2.2. *The set*

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

is a convex cone.

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

In general one should think that the convex cones are models of positive real numbers. Such model need not be very good however: the whole vector space is always a convex cone. To fix this problem one introduces the concept of salient cone.

Definition 2.3. A convex cone $C \subset V$ is **salient cone**, or simply **salient**, if whenever both $v \in C$ and $-v \in C$, then necessarily $v = 0$.

Conveniently enough $\mathcal{H}_+(V)$ is a salient cone, but this is by no means trivial property.

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

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

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

It is also customary to give vector space a topology (and get a topological vector space in return). This leads to concept of **closed convex cone**, which is defined as one would expect. Note that as subset of dual lead to convex cones, subsets of continuous dual lead to closed convex cones.

A closed convex cone that is also salient is, as is somewhat customary, called **proper cone**. We have now

Theorem 2.5. *$\mathcal{H}_+(V)$ is a proper cone (with usual topology).*

Previous arguments carry directly to a much more general setting:

Theorem 2.6. *Let V be a topological vector space (over $\mathbb{R}$ or $\mathbb{C}$) and C^* a subset of its continuous dual. Assume that*

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

Then

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

is a proper convex cone of V.

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

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

Proposition 2.7. *If $A \in \mathcal{L}(V)$ is a orthogonal projection, then $A \geq 0$.*

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

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

so A is positive. □

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

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

2.1.2 Real maps and adjoint

Dual cone thinking lets us also lift other important notions.

Definition 2.8. We say that a map $A \in \mathcal{L}(V)$ is *real*, if

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

for any $v \in V$.

Definition 2.9. We say that a map $A \in \mathcal{L}(V)$ is *imaginary*, if

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

for any $v \in V$.

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

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

Theorem 2.10. *For any $A \in \mathcal{L}(V)$ there exists unique map $A^* \in \mathcal{L}(V)$, called the adjoint of A, for which for any $v \in V$ we have*

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

Proof. The uniqueness of adjoint is immediate from the lemma 2.4. 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.

The quadratic form for such map is given by

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

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

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

$\square$

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

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

for any $v, w \in V$.

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

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

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

2.1.3 More convincing

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

Definition 2.12. Let $W \subset V$ be a subspace and $A \in \mathcal{L}(V)$. Then the **compression** of A to W, denoted by A_W is the linear map

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

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

Lemma 2.13. *Let $W \subset V$ and $A \geq 0$. Then also $A_W \geq 0$. In particular all the eigenvalues of A are non-negative.*

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

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

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

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

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

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

Lemma 2.14. *Let $A_1 \in \mathcal{L}(V_1)$ and $A_2 \in \mathcal{L}(V_2)$. Then $A_1 \oplus A_2 \in \mathcal{H}_+(V_1 \oplus V_2)$, if and only if $A_1 \in \mathcal{H}_+(V_1)$ and $A_2 \in \mathcal{H}_+(V_2)$.*

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

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

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

2.2 Spectral theorem

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

Theorem 2.15. *Let $n = \dim(V)$. Then $A \in \mathcal{L}(V)$ is real if and only there exists real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ and for pairwise orthogonal vectors $v_1, v_2, \ldots, v_n \in V$ such that*

$$A = \sum_{i=1}^{n} \lambda_i P_{v_i}. \tag{2.16}$$

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

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

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

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

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

Lemma 2.17. *If $A \in \mathcal{H}_+(V)$ and $\langle Av, v \rangle = 0$ for some $v \in V$, then $Av = 0$ and $Aw \perp v$ for any $w \in v$.*

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

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

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

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

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

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

Proof of lemma 2.17. Take any $w \in V$. Now by assumption for any $c \in \mathbb{C}$ we have

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

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

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

In the representation 2.16 the numbers λ_i are evidently the eigenvalues of A and vectors v_i the corresponding eigenvectors; this is why we call it the *Spectral representation.* As remarked, the representation is of course not unique, but there is a way to make the Spectral representation unique, however. For this we have to change v_i to corresponding eigenspaces.

Theorem 2.18 (Spectral theorem)**.** *Let $A \in \mathcal{H}(V)$. Then there exists unique non-negative integer m, distinct real numbers $\lambda_1, \lambda_2, \ldots, \lambda_m$ and non-trivial orthogonal subspaces of V, $E_{\lambda_1}, E_{\lambda_2}, \ldots E_{\lambda_m}$ with $E_{\lambda_1} + E_{\lambda_2} + \ldots + E_{\lambda_m} = V$, such that*

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

Moreover, this representation is unique.

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

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

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

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

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

$$p(A) = c_n A^n + c_{n-1}A^{n-1} + \dots c_1 A + c_0 = \sum_{1\le i\le n} p(\lambda_i) P_{v_i}. \tag{2.19}$$

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

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

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

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

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

Finally, one should note that the lemma 2.17 enjoys following natural generalization.

Proposition 2.20. *If $A \ge 0$ and $A_W = 0$ for some subspace $W \subset V$ then we may decompose $A = A_{W^\perp} \oplus 0_W$.*

Proof. We prove the statement by induction on the dimension of W. Lemma 2.17 took care of the case $\dim(W) = 1$. When $\dim(W) > 1$ we may decompose $W = W' \oplus W''$ where $\dim(W') = 1$. Now $A_{W'} = (A_W)_{W'} = 0_{W'} = 0$, so we may decompose $A = A_{W'^\perp} \oplus 0_{W'}$. But $(A_{W'^\perp})_{W''} = A_{W''} = (A_W)_{W''} = 0_{W''} = 0$, so by the induction hypothesis $A_{W'^\perp} = (A_{W'^\perp})_{W''^\perp} \oplus 0_{W''} = A_{W^\perp} \oplus 0_{W''}$. Consequently $A = A_{W'^\perp} \oplus 0_{W'} = A_{W^\perp} \oplus 0_{W''} \oplus 0_{W'} = A_{W^\perp} \oplus 0_W$, as desired. $\square$

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

The previous proposition has an useful corollary.

Corollary 2.21. *If $W, W' \subset V$, then $(P_W)_{W'} = 0$ if and only if $W \perp W'$.*

Proof. Assume first that $(P_W)_{W'} = 0$. Then by the lemma 2.20 we have $W = \mathrm{im}(P_W) \subset W'^{\perp}$. The other direction is evident. □

2.2.1 Commuting real maps

Warning! Composition of positive maps need not be positive!

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

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

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

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

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

Lemma 2.22. *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*

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

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

Theorem 2.24. *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, $E_1, E_2, \dots E_m$ with $E_1+E_2+\dots+E_m = V$ 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 $E_1, E_2, \dots E_m$. 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 2.22.

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

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

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

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

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

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

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

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

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

The previous theorem sends a very important message.

Philosophy 2.25. 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 2.24 we have

Corollary 2.26. *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 2.27. *If $A, B \geq 0$, then AB is diagonalizable and has non-negative eigenvalues. Conversely, if C is diagonalizable and has non-negative eigenvalues, then it's of the form AB for some positive A and B.*

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

TODO: independence of random variables.

2.2.2 Symmetric product

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

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

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

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

Proposition 2.28. *If $v, w \in V \setminus \{0\}$, then*

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

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

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

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

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

Lemma 2.29. *If $A = P_v + P_w = I$, then v and w are orthogonal.*

Proof. Note that $I_{(v)} = A_{(v)} = (P_v)_{(v)} + (P_w)_{(v)} = I_{(v)} + (P_w)_{(v)}$, so $(P_w)_{(v)} = 0$. By lemma 2.20 we have $(w) \perp (v)$. □

Lemma 2.30. *If $A = P_v + P_w = 2P_u$ for some $u \in V$, then v, w and u are all parallel.*

Proof. Since $0 = A_{(u)^\perp} = (P_v + P_w)_{(u)^\perp} = (P_v)_{(u)^\perp} + (P_w)_{(u)^\perp} \geq 0$, we have $(P_v)_{(u)^\perp} = (P_w)_{(u)^\perp}$. Now by the lemma 2.20 we have $(v), (w) \subset (u)$: hence the claim. □

□

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

Proposition 2.31. *Let $\alpha, \beta, \gamma \in \mathbb{C}$ and $n \geq 2$. Then the expression $\alpha A^2 + \beta AB + \overline{\beta} BA + \gamma B^2$ is positive for any $A, B \geq 0$ if and only if $\alpha, \gamma \in [0, \infty)$ and $|\beta|^2 \leq \alpha\gamma$.*

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

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

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

TODO (is this true? probably) □

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

2.3 Congruence

2.3.1 $*$-conjugation

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

Definition 2.32. Let $A, B \in \mathcal{H}$. We say that B is $*$-*conjugate* of A if for some $C \in \mathcal{L}(V)$ we have $B = C^*AC$.

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

Definition 2.33. Let $A, B \in \mathcal{H}$. We say that A and B are *congruent* if they are $*$-conjugates of each other.

It is easily verified that congruence is a equivalence relation.

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

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

Theorem 2.34 (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 $\mathrm{span}\{v_i | 1 \leq i \leq k\}$, 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 2.35. *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. □

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 2.36. *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 2.37 (General Sylvester's Law of Inertia)**.** *For $A, B \in \mathcal{N}(V)$ and A is $*$-conjugate of B, if and only if $n_{\pm}(A) \leq n_{\pm}(B)$.*

Proof. The proof is essentially the same. □

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

2.3.2 Block decomposition

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

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

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

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

to get

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

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

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

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

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

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

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

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

then

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

and

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

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 2.39 (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 2.38 to arrive at so-called *LDL-decomposition* of M:

$$M = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}. \tag{2.40}$$

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

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

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

$$\det(A_W) = \det(A)\langle A^{-1}v, v\rangle. \tag{2.42}$$

By comparing determinants from two LDL-decompositions we arrive at

$$\det(A)\det(C - B^*A^{-1}B) = \det(C)\det(A - BC^{-1}B^*), \tag{2.43}$$

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

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

determinant of rank 1 update can be also easily calculated.

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

2.4 Loewner order

Definition 2.44. 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 2.45. *(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 $A \le B$ and $B \le A$ then $A = B$.

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

Example 2.46. If $W_1, W_2 \subset V$ are two subspaces of V we have $P_{W_1} \le P_{W_2}$ 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 $P_{W_2} = P_{W_1} + P_{W_3} \ge P_{W_1}$. Conversely if $P_{W_1} \le P_{W_2}$, then $0 \le (P_{W_1})_{W_2^\perp} \le (P_{W_2})_{W_2^\perp} = 0$, so $(P_{W_1})_{W_2^\perp} = 0$. But now lemma 2.21 implies that $W_1 \subset W_2$.

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

For example if $0 < a \le b$, with real numbers one could multiply the inequalities by the positive number $(ab)^{-1}$ to get $0 < b^{-1} \le a^{-1}$. 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 2.47. *If $A \le B$, then for any C we have $C^*AC \le C^*BC$.*

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

Theorem 2.48. *If $0 < A \le B$, then $B^{-1} \le A^{-1}$.*

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

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

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 $I \le X := A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ we have $X^{-1} \le I$, that is

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

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

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

Remark 2.49. Alternatively, we could conjugate both sides by $X^{-\frac{1}{2}}$ to arrive at the conclusion. Note that by doing this we have only used $*$-conjugation in the proof: actually we have $*$-conjugated altogether with

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

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

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

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

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

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

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

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

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

To prove the matrix version, namely

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

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

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

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

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

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

2.5 Eigenvalue inequalities

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

Proposition 2.50. *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 2.36 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 2.51 (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 2.36 it's compression to some k-dimensional subspace is positive. But then also compression of A to this same subspace is positive and hence it has k non-negative eigenvalues. $\square$

TODO picture of eigenvalues changing when compressed

Again one can prove that this result is strongest possible.

Proposition 2.52. *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 2.53. *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 2.50 $\lambda_k(A) \leq \lambda_k(B)$, 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 2.51 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 2.54. *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 2.55. *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 2.43. ☐

Proof of propostion 2.54. 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 2.52 is similar: the aim to first connect the characteristic polynomials of A and its compression and then do similar observations.

Lemma 2.56. *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 2.42. □

Proof of proposition 2.52. Proof is just an easier version of the proof of 2.54 □

These eigenvalue inequalities have interesting corollaries.

Corollary 2.57. *If $A, B \in \mathcal{H}^n(V)$, then $|\lambda_i(A) - \lambda_i(B)| \leq \sum_{i=1}^{n} |\lambda_i(A - B)| \leq n\|A - B\|$ for any $1 \leq i \leq n$.*

Proof. If $B - A = \sum_{i=1}^{n} \lambda_i(B - A)P_{v_i}$, write $A_j = A + \sum_{i=1}^{j} \lambda_i(B - A)P_{v_i}$. By using lemma 2.54 we may trace how the eigenvalues of A_j change when j increases. We conclude the given bound ... almost: this implies bound for terms $|\lambda_i(A) - \lambda_{\sigma(i)}(B)|$ for some permutation σ of $\{1, 2, \ldots, n\}$. But $\max_{1 \leq i \leq n} |\lambda_i(A) - \lambda_i(B)| \leq \max_{1 \leq i \leq n} |\lambda_i(A) - \lambda_{\sigma(i)}(B)|$ for any permutation σ (as can be seen by simple exchange argument, for instance). The last inequality is trivial, so we are done. □

TODO: change order of compression and projection eigenvalues converses.

2.6 Notes and references

2.7 Ideas

- Normal maps
- Square root of a matrix
- Ellipses map to ellipses
- adjoints of vectors
- Moore-Penrose pseudoinverse
- (canonical, löwdin) orthogonalization, polar decomposition and orthogonal Procrustes problem
- projection matrices
- Hilbert-Schmidt norm ($\rightarrow$ 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

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

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

- Lorenz order?

- BMV-conjecture (theorem)

- Proof difficulties

- Proof “sketch” (as in joke)

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

Chapter 3

Matrix functions

3.1 Functional calculus

Definition 3.1. For any $-\infty \leq a < b \leq \infty$ $f : (a,b) \to \mathbb{R}$ the associated matrix function on V is the map $f_V : \mathcal{H}_{(a,b)}(V) \to \mathcal{H}(V)$ given by

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

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 (2.19) and TODO.

Matrix functions enjoy many natural and useful properties.

Proposition 3.2. *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.

3.2 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 (3.2).

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.

3.3 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 3.3. *If $f\in C^1(a,b)$, $A\in\mathcal{H}_{(a,b)}$ such that $f(\lambda_i)=0=f'(\lambda_i)$ for $1\le i\le n$, then*

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

Proof. By lemma 2.57 we see that $|\lambda_i(A+H)-\lambda_i(A)|\le n\|H\|$. Now by Taylor expansion $|f(\lambda_i(A+H))|=n\|H\|f'(\xi_i)$ for some $\xi_i=\xi_{i,A,H}$ on $(\lambda_i(A)-n\|H\|,\lambda_i(A)+n\|H\|)$. Now

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

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

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

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

we have

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

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

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

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

3.4 Higher derivatives

TODO

TODO:

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

Chapter 4

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.

4.1 Basic properties

We first state the definition.

Definition 4.1. Let $(a,b) \subset \mathbb{R}$ be an open, possibly unbounded interval and n positive integer. We say that $f : (a,b) \to \mathbb{R}$ is n-monotone or matrix monotone of order n, if for any $A, B \in \mathcal{H}^n_{(a,b)}$, such that $A \leq B$ we have $f(A) \leq f(B)$.

We will denote the space of n-monotone functions on open interval (a,b) by $P_n(a,b)$. Note that in the notation we don't specify the space V; it doesn't really matter.

Proposition 4.2. *If* $\dim(V) = \dim(V')$*, then f is n-monotone in V if and only if it is n-monotone in V'.*

Proof. The reason is rather clear: inner product spaces of same dimension are isometric. □

One immediately sees that that all the matrix monotone functions are monotone as real functions.

Proposition 4.3. *If $f \in P_n(a,b)$, then f is increasing.*

Proof. Take any $a < x \leq y < b$. Now for $xI, yI \in \mathcal{H}^n_{(a,b)}$ we have $xI \leq yI$ so by definition

$$f(x)I = f(xI) \leq f(yI) = f(y)I,$$

from which it follows that $f(x) \leq f(y)$. This is what we wanted. $\square$

Actually, increasing functions have simple and expected role in n-monotone matrices.

Proposition 4.4. *Let (a,b) be an open interval and $f : (a,b) \to \mathbb{R}$. Then the following are equivalent:*

(i) *f is increasing.*

(ii) *$f \in P_1(a,b)$.*

(iii) *For any positive integer n and commuting $A, B \in \mathcal{H}^n_{(a,b)}$ such that $A \leq B$ we have $f(A) \leq f(B)$.*

Proof. Since 1×1 matrices are for our purposes just reals, $(i) \Leftrightarrow (ii)$ is clear. Also if (iii) holds, since in particular xI and yI commute for every x, y, if $x \leq y$, then $xI \leq yI$, and by assumption hence $f(x)I = f(xI) \leq f(yI) = f(y)I$ so $f(x) \leq f(y)$, which is to say that f is increasing.

Let us then prove that $(i) \Rightarrow (iii)$. If $A \leq B$ and A and B commute, by theorem 2.24 we may write $A = \sum_{i=1}^n a_i P_{v_i}$ and $B = \sum_{i=1}^n b_i P_{v_i}$ for some $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ and $v_1, v_2, \ldots, v_n$, orthonormal basis of V, with $a_i \leq b_i$. But now $f(A) = \sum_{i=1}^n f(a_i) P_{v_i}$ and $\sum_{i=1}^n f(b_i) P_{v_i}$ so

$$f(B) - f(A) = \sum_{i=1}^n (f(b_i) - f(a_i)) P_{v_i}$$

is positive, as f is increasing. $\square$

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

Let us then have some examples.

Proposition 4.5. *For any positive integer n, open interval (a,b) and $\alpha, \beta \in \mathbb{R}$ such that $\alpha \geq 0$ we have that $(x \mapsto \alpha x + \beta) \in P_n(a,b)$.*

Proof. Assume that for $A, B \in \mathcal{H}_{(a,b)}$ we have $A \leq B$. Now

$$f(B) - f(A) = (\alpha B + \beta I) - (\alpha A + \beta I) = \alpha(B - A).$$

Since by assumption $B - A \geq$ and $\alpha \geq 0$, also $\alpha(B - A) \geq 0$, so by definition $f(B) \geq f(A)$. This is exactly what we wanted. $\square$

That was easy. It's not very easy to come up with other examples, though.

One of the main properties of the classes of matrix monotone functions has still avoided our discussion, namely the relationship between classes of different orders. We already noticed that matrix monotone functions of all orders all monotonic, or $P_n(a, b) \subset P_1(a, b)$ for any $n \geq 1$. It should not be very surprising that we can make much more precise inclusions.

Proposition 4.6. *For any open interval (a, b) and positive integer n we have $P_{n+1}(a, b) \subset P_n(a, b)$.*

Proof. The idea is that if $\dim(V) \leq \dim(V')$, we can essentially find copy of V in V'. If $A, B \in \mathcal{H}^n(V)$, we can augment A and B to $V' = V \oplus \mathbb{C}$ by setting $A' = A \oplus c$ for any $c \in \mathbb{R}$.

Now if $A \leq B$, by picking any $c \in \mathbb{R}$ we see that $(A \oplus c) \leq (B \oplus c)$. Consequently if $f \in P_{n+1}(a, b)$, we have

$$f(A) \oplus f(c) = f(A \oplus c) \leq f(B \oplus c) = f(B) \oplus f(c),$$

which implies that $f(A) \leq f(B)$. $\square$

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.

4.2 Successes

As expected, certain basic properties of monotone functions carry to matrix monotone.

Proposition 4.7. *For any positive integer n and open interval (a, b) the set $P_n(a, b)$ is a convex cone, i.e. it is closed under taking summation and multiplication by non-negative scalars.*

Proof. This is easy: closedness under summation and scalar multiplication with non-negative scalars correspond exaclty to the same property of positive matrices. $\square$

We shall be using especially the previous result a lot.

Similarly we have composition and pointwise limits.

Proposition 4.8. *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 4.9. *If n-monotone functions $f_i : (a,b) \to \mathbb{R}$ converge pointwise to $f : (a,b) \to \mathbb{R}$ as $i \to \infty$, also f is n-monotone.*

Proof. As always, fix $A, B \in \mathcal{H}^n_{(a,b)}$ with $A \leq B$. Now by assumption

$$f(B) - f(A) = \lim_{i\to\infty} f_i(B) - \lim_{i\to\infty} f_i(A) = \lim_{i\to\infty} (f_i(B) - f_i(A)) \geq 0,$$

so also $f \in P_n(a,b)$. $\square$

4.3 Failures

Most of the common monotone functions fail to be matrix monotone. Let's try some non-examples.

Proposition 4.10. *Function $(x \mapsto x^2)$ is not n-monotone for any $n \geq 2$ and any open interval $(a,b) \subset \mathbb{R}$.*

Proof. Let us first think what goes wrong with the standard proof for the case $n = 1$.

Note that if $A \leq B$,

$$B^2 - A^2 = (B - A)(B + A)$$

is positive as a product of two positive matrices (real numbers).

There are two fatal flaws here when $n > 1$.

- $(B - A)(B + A) = B^2 - A^2 + (BA - AB)$, not $B^2 - A^2$.
- Product of two positive matrices need not be positive.

Note that both of these objections result from the non-commutativity and indeed, both would be fixed should A and B commute.

Let's write $B = A + H$ ($H \geq 0$). Now we are to investigate

$$(A + H)^2 - A^2 = AH + HA + H^2.$$

Note that $H^2 \geq 0$, but as we have seen in proposition 2.31, $AH + HA$ need not be positive! Also, if H is small enough, H^2 is negligible compared to $AH + HA$. We are ready to formulate our proof strategy: find $A \in \mathcal{H}^n_{(a,b)}$ and $\mathbb{H}^n_+$ such that $AH + HA \not\geq 0$. Then choose parameter $t > 0$ so small that $A + tH \in \mathcal{H}^n(a, b)$ and

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

and set the pair $(A, A + tH)$ as the counterexample. TODO (arbitrary intervals) $\square$

As a corollary with get

Corollary 4.11. *The function $\chi_{(0,\infty)}$ is not n-monotone for any $n \geq 2$.*

Proof. If $\chi_{x>0}$ were n-monotone so would be

$$x^2 = \int_0^\infty 2t\chi_{(t,\infty)}(x)dt.$$

$\square$

The function $\chi_{(0,\infty)}$ is in some sense canonical counterexample: every increasing function is more or less positive linear combination of its tranlates, so if monotone functions are not all matrix monotone, the reason is that it is not matrix monotone. For this reason we should really understand why it is not n-monotone for any $n > 1$.

The idea is the following: we are going to construct $A, B \in \mathcal{H}^2$ with the following properties:

1. $A \leq B$
2. A and B have both exactly one positive eigenvalue
3. A and B don't commute

If we can do this, A and B work as counterexamples. Indeed then $\chi_{(0,\infty)}(A) = P_{v_1}$ and $\chi_{(0,\infty)}(B) = P_{w_1}$ where eigenvectors v_1 and w_1 are eigenvectors of A and B corresponding to positive eigenvalues. But $\chi_{(0,\infty)}(A) \not\leq \chi_{(0,\infty)}(B)$ by 2.46.

Constructing such pair is very easy: just take A with eigenvalues -1 and 1 and consider B of the form $A + tH$ for some $H \geq 0$, $t > 0$ and such that A and H do not commute. For small enough H all of the conditions are easily satisfied.

As many properties of real numbers break with real maps, similarly many properties of monotone functions break when $n > 1$. As we saw with the square function example, product of two n-monotone functions need not be n-monotone in general, even if they are both positive functions. Similarly, taking maximums doesn't preserve monotonicity.

Proposition 4.12. *Maximum of two n-monotone functions need not be n-monotone for $n \geq 2$.*

Proof. Again, let's think what goes wrong with the standard proof for $n = 1$.

Fix open interval (a, b), positive integer $n \geq 2$ and two functions $f, g \in P^n(a, b)$. Take any two $A, B \in \mathcal{H}^n_{(a,b)}$ with $A \leq B$. Now $f(A) \leq f(B) \leq \max(f, g)(B)$ and $f(A) \leq f(B) \leq \max(f, g)(B)$. It follows that

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

as we wanted.

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

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

But now it is rather easy to see that we can take same pair as with $\chi_{(0,\infty)}$ as our counterexample. $\square$

The maximum problem is not too bad; similar statement doesn't hold for k-tone functions in general either and maybe it's more of a pleasent surprise that it holds for usual monotone functions, anyway. But there is very fundamental problem hidden in the square example.

Proposition 4.13. *There exists no $\alpha > 0$, and an open interval $(a, b) \subset \mathbb{R}$ such that $\alpha x + x^2 \in P_n(a, b)$.*

Proof. Adding linear term means just translating domain and codomain, which is not going to help: $x^2 + \alpha x = (x + \frac{\alpha}{2})^2 - \frac{\alpha^2}{4}$. $\square$

Why is this bad? If $f : (a, b) \to \mathbb{R}$ is not too bad (say Lipschitz), for large enough α the function defined by $g(x) = f(x) + \alpha x$ is increasing. But we can't do necessarily do the same thing in the matrix setting even for smooth or analytic functions. Although this might not be such a big surprise or a bad thing in the first place, it is worthwhile to investigate the underlying reason.

Consider the case of entire f and take $A, H \in \mathcal{H}^n_{(a,b)}$ with $H \geq 0$. As observed earlier, we have

$$\begin{aligned}\lim_{t \to 0} \frac{f(A + tH) - f(A)}{t} &= a_1 H \\ &+ a_2(HA + AH) \\ &+ a_3(HA^2 + AHA + A^2H) \\ &+ \ldots .\end{aligned}$$

In the real setting we could just increase a_1 to make the previous expression positive. In the matrix setting there is a problem: note that if H is of rank 1, increasing a_1 means "increasing the right-hand side only to one direction". The point is that if the right-hand side is not positive (map) in the first place, it might be (a priori) non-positive in a big subspace, so rank 1 machinery is not going to save the day.

On the other hand if $n = 2$, for instance, there is not too much room for things to go south. We still, a priori, can't guarantee positivity with a_1, but buffing first few Taylor coefficients starts to somehow affect the whole space.

When n gets larger we have more and more space to worry about, so we should start worrying about more and more Taylor coefficients.

This is all just heuristics, but it leads us to expect two things:

1. Something about things getting worse when n get larger
2. Something about things being not too bad when n is fixed as there is not too much space.

We will later see that both of these phenomena occur.

4.4 Pick functions are matrix monotone

Fortunately the affine functions are not the only matrix monotone functions. We have already discussed a second example.

Proposition 4.14. *We have $(x \mapsto -x^{-1}) \in P_n(a,b)$ for any $(a,b) \not\ni 0$ and $n \geq 1$.*

Proof. The result follows immediately from the theorem 2.48. □

Now also the tranlations $(x \mapsto \frac{1}{\lambda - x}) \in P_n(a,b)$ for any $n \geq 1$ and $(a,b) \not\ni \lambda$ so by the cone property also all the functions of the form

$$x \mapsto \alpha x + \beta + \sum_{i=1}^{m} \frac{t_i}{\lambda_i - x} \tag{4.15}$$

are n-monotone on (a,b) for any $n \geq 1$, $\alpha, t_1, t_2, \ldots, t_m \geq 0$ and $\beta, \lambda_1, \lambda_2, \ldots, \lambda_m$ where $\lambda_1, \lambda_2, \ldots, \lambda_m \notin (a,b)$.

Taking pointwise limits we arrive at

Theorem 4.16. *If $f \in P(a,b)$, then $f \in P_n(a,b)$ for every $n \geq 1$.*

Proof. The reason is of course that every element of $P(a,b)$ is pointwise limit of the functions of the form 4.15. Why is that exactly? One should be a bit careful as things aren't compactly supported. TODO (explain why everything is compact afterall) □

4.5 Derivative and Loewner's characterization

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

Theorem 4.17. *Let $f \in C^1(a,b)$ and $n \geq 1$. Then the following are equivalent:*

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

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

$$D^1_n f_A(H) \geq 0.$$

(iii) For any $A \in \mathcal{H}^n_{(a,b)}$ and P one dimensional (orthogonal) projection we have

$$D^1_n f_A(P) \geq 0.$$

(iv) For any $A \in \mathcal{H}^n_{(a,b)}$, $H \geq 0$ and $v \in V$ the map

$$t \mapsto \langle f(A + tH)v, v\rangle$$

is increasing.

(v) For any $A \in \mathcal{H}^n_{(a,b)}$, P one dimensional (orthogonal) projection and $v \in V$ the map

$$t \mapsto \langle f(A + tP)v, v\rangle$$

is increasing.

Proof. $(i) \Leftrightarrow (ii)$ follows from Spectral theorem linearity of D^1_n, and $(iii) \Leftrightarrow (iv)$ from Spectral theorem. One also easily sees that the derivative of the map

$$t \mapsto \langle f(A + tH)v, v\rangle$$

is the map

$$t \mapsto \langle D^1_n f_A(H)v, v\rangle$$

so also $(i) \Leftrightarrow (iii)$ is clear. □

We already noticed that we can express $D^1_n f_A(H) = ([\lambda_i, \lambda_j]_f)_{1 \leq i,j \leq n} \circ H$, where Hadamard product is taken along eigenbasis of A. We can however make the following simple observation:

Lemma 4.18. *Let $A \in \mathcal{H}$. Then $A \geq 0$, if and only if $A \circ B$ for every $B \geq 0$.*

Proof. If the Hadamard product is along $(e_i)_{1\leq i\leq n}$, we have $A = A \circ \left(\sqrt{n}P_{\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} e_i}\right)$, and hence have the "if". Note that for only if we only need to verify the inequality for A and B both one dimensional projections. But now one easily sees that $A \circ B$ is non-negative multiple of projection. □

We hence have the following characterization.

Theorem 4.19. *$f \in P_n(a,b) \cap C^1(a,b)$, if and only if the matrix*

$$([\lambda_i, \lambda_j]_f)_{i,j} \geq 0 \tag{4.20}$$

for any $\lambda_1, \lambda_2, \ldots, \lambda_n \in (a,b)$.

This is the original characterization by Loewner, and it is pretty much just saying that function is matrix monotone if its (matrix) derivative is positive. The matrix 4.20 is called, appropriately, Loewner matrix (of function f on points $\lambda_1, \lambda_2, \ldots, \lambda_n$). Using the characterization it is in general not very easy to check that the function is n-monotone: we would have to check positivity of the matrix for any tuple on the interval. Also the characterization is not local one: in order to check monotonicity we need to know the behaviour on the whole interval. This is just a reflection of the fact that the space in which we are working on, space of real maps, is itself in a way spread around the interval.

4.6 Local characterization

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

Proposition 4.21. *For any $n \geq 1$, P_n is a local property meaning that whenever $f \in P_n(a,b)$ and $P_n(c,d)$ for some $a < c < b < d$, then also $f \in P_n(a,d)$.*

The reason for this is hidden in the Loewner matrix.

Note that Loewner matrix is essentially something we saw before: it is just a Pick matrix when all the points are on the real line. We observed before that positivity of the Pick matrix was some kind of manifestation of the strength of the Cauchy's integral formula. Namely, if f happens to analytic, in some suitable set, we can write

$$([\lambda_i, \lambda_j]_f)_{i,j} = \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{(z-\lambda_i)(z-\lambda_j)} dz.$$

Now the positivity the matrix means that for any $c_1, c_2, \ldots, c_n \in \mathbb{C}$ the quantity

$$\sum_{1\leq i,j\leq n} c_i\overline{c_j}[\lambda_i,\lambda_j]_f = \frac{1}{2\pi i}\int_\gamma f(z)\left(\sum_{1\leq i\leq n}\frac{c_i}{z-\lambda_i}\right)\left(\sum_{1\leq i\leq n}\frac{\overline{c_i}}{z-\lambda_i}\right) dz.$$

But here's the trick: we may write

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

for some polynomial of degree less than n, and indeed, if the λ's are distinct there's a one-to-one correspondence between polynomials q and the λ's. It follows that we may rewrite the integral as

$$\frac{1}{2\pi i}\int_\gamma f(z)\frac{q(z)\overline{q(\overline{z})}}{p_\Lambda(z)^2}dz.$$

Note that $z \mapsto q(z)\overline{q(\overline{z})})$ is a polynomial of degree at most $(2n-2)$ non-negative on the real line. Easy application of the Fundamental theorem of algebra reveals that all such polynomials are actually of the previous form.

Lemma 4.22. *h is polynomial of degree at most $(2n-2)$ non-negative on $\mathbb{R}$, if and only if it is of the form $p(z)\overline{p(\overline{z})}$ for some complex polynomial of degree of at most $(n-1)$.*

Proof. It is easy to see that all of the polynomials of the specific form fit the bill. Conversely, if h is non-negative on real axis, it's roots all appear in pairs: either with strict complex conjugate pairs, of pairs of double real roots. We may take p to be $\sqrt{a_n}\prod(z-z_i)$ where z_i range over representatives of all the pairs and a_n is the leading coefficient of h. $\square$

Write $h(z) = q(z)\overline{q(\overline{z})})$.

Finally note that resulting expression,

$$\frac{1}{2\pi i}\int_\gamma f(z)\frac{h(z)}{p_\Lambda(z)^2}dz$$

is nothing but the divided difference of the function fh at points $\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n, \lambda_n$. By 5.24 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 5.41.

Lemma 4.23. *Let k and n be positive integers and $d_1, d_2, \ldots, d_m$ be positive integers with $d_1 + d_2 + \ldots + d_m = k+1$. Assume that $n \ge (\max_{1\le i\le m} d_i) - 1$. Let $f \in C^n(a,b)$. Then the following are equivalent.*

(i) f is k-tone.

(ii) For any $a < x_1 < x_2 < \ldots < x_m < b$ we have

$$[x_1, x_1, \ldots, x_1, x_2, \ldots, x_2, \ldots, x_m, \ldots, x_m]_f \geq 0$$

where x_i appears d_i times.

Proof. $(i) \Rightarrow (ii)$ is clear.

For the other direction let's take any $a < y_1 < y_2 < \ldots < y_n < y_{n+1} < b$. We need to prove that

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f \geq 0.$$

The idea is to bunch the variables together using the mean value theorem.

Let's consider the function $g_1(x) = [x, y_{d_1+1}, y_{d_1+2}, \ldots, y_{k+1}]_f$. In terms of g_1 we need to prove that

$$[y_1, y_2, \ldots, y_{d+1}]_{g_1} \geq 0.$$

Since $f \in C^{d_1-1}(a,b)$, $g_1 \in C^{d_1-1}(a, y_{d_1+1})$ and hence by the mean value theorem we have

$$[y_1, y_2, \ldots, y_{d_1}]_g = [x_1, x_1, \ldots, x_1]_{g_1}$$

for some $a < x_1 < y_{d+1}$. Consequently,

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f = [x_1, x_1, \ldots, x_1, y_{d_1+1}, \ldots, y_{k+1}]_f.$$

Next step is to bunch together the next d_2 terms: consider now the map $g_2(x) = [x_1, x_1, \ldots, x_1, x, y_{d_1+d_2+1,\ldots,y_{k+1}}]_f$ and observe that we are to verify that

$$[y_{d_1+1}, \ldots, y_{d_1+d_2}]_{g_2} \geq 0.$$

Again use mean value theorem to replace $y_{d_1+1}, \ldots, y_{d_1+d_2}$ by x_2's.

One should be bit careful here: the number x_1 certainly depends on all the y's, so once we have fixed it we can't say that

$$[y'_{d_1+1}, y'_{d_1+2}, \ldots, y'_{d_1+d_2}]_{g_2} = [y_1, \ldots, y_{d_1}, y'_{d_1+1}, \ldots, y'_{d_1+d_2}, y_{d_1+d_2+1}, \ldots, y_{k+1}]_f,$$

for instance, anymore. This is of course not a problem.

Making m steps of the previous form we finally find numbers $x_1, x_2, \ldots, x_m$ such that

$$[y_1, y_2, \ldots, y_k, y_{k+1}]_f = [x_1, x_1, \ldots, x_1, x_2, \ldots, x_2, \ldots, x_m, \ldots, x_m]_f \geq 0$$

and we are done. $\square$

We have hence the following.

Theorem 4.24. *Let $n \geq 1$ and $f \in C^1(a,b)$. Then $f \in P_n(a,b)$, if and only if fh is $(2n-1)$-tone whenever h is polynomial of degree at most $(2n-2)$, non-negative on real line.*

This result has a curious corollary.

Corollary 4.25. *If $n \geq 1$ and fh is $(2n+1)$:tone for every polynomial h of degree at most $2n$, then fh is $(2n-1)$-tone for every polynomial h of degree at most $(2n-2)$. In particular f is k-tone for every $k = 1, 3, 5, \ldots, 2n-3, 2n-1, 2n+1$.*

Although we strictly speaking only proved this corollary for $f \in C^1$, it holds true without extra assumptions, as can be seen with the following alternate proof.

Proof of corollary 4.25. Take any h, a polynomial of at most $(2n-2)$ non-negative on real axis and points $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n} < b$. We should prove that

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \geq 0.$$

The idea is the following: if f is C^1, we have

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} = [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, t, t]_{fh(\cdot - t)^2} \geq 0,$$

Now, actually $fh(\cdot - t)^2$ is always differentiable at t, so the previous at least should hold without the smoothness TODO, but one can take safer route. For any $a < t, s < b$ we have

$$[\lambda_1, \lambda_2, \ldots, \lambda_{2n}, t, s]_{fh(\cdot - t)^2} \geq 0.$$

Expanding this Leibniz rule leads to

$$\begin{aligned} 0 \leq\, & [\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \\ & + (s-t)[\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s]_{fh} \end{aligned}$$

But by choosing t and s suitably we can definitely make the second term non-positive, so the first term is non-negative, as we wanted. Indeed choose first arbitrary s and then choose t on (a, s) or on (s, b), depending on the sign of the divided difference. Or if one so prefers, we have

$$\begin{aligned} & 2[\lambda_1, \lambda_2, \ldots, \lambda_{2n}]_{fh} \\ = & [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s+r, s]_{fh(\cdot - s - r)^2} \\ + & [\lambda_1, \lambda_2, \ldots, \lambda_{2n}, s-r, s]_{fh(\cdot - s + r)^2} \end{aligned}$$

for small enough r.

Final claim follows by setting $h \equiv 1$. $\square$

4.7 Regularity

Finally, one might ask whether we can get rid of C^1-assumption, and it turns out that we can. This leads to question: why should matrix monotone functions be C^1? Why should they be even continuous?

The continuity question we actually almost answered previously. There are not very many ways increasing function can be discontinuous. For every increasing function we have $\lim_{t\to 0^-} f(t) \leq f(0) \leq \lim_{t\to 0^+} f(t)$, and discontinuity means that at least one of the inequalities is strict. Thus every discontinuous increasing function looks like step function at discontinuity.

Lemma 4.26. *Let $f \in P_n(a,b)$ for some $n > 1$. Then f is continuous.*

Proof. Without loss of generality it suffices to check the continuity at 0. We may decompose $f(t) = c(t) + c_1\chi_{\{x>0\}}(t) + c_2\chi_{\{x=0\}}(t)$, where c is continuous at 0, and $0 \leq c_2 \leq c_1$. We aim to prove that $c_1 = 0$, in which case f is also continuous at 0.

As in the proof of corollary 4.11 we $A \leq B$ (with non-zero eigenvalues) such that $\chi_{x>0}(A) \not\leq \chi_{x>0}(B)$. Now for small enough $t > 0$, we have $tA, tB \in \mathcal{H}_{(a,b)}$ and we have

$$\begin{aligned} f(tB) - f(tA) = c(tB) - c(tA) & \\ + c_1\left(\chi_{\{x>0\}}(tB) - \chi_{\{x>0\}}(tA)\right) & \\ + c_2\left(\chi_{\{x=0\}}(tB) - \chi_{\{x=0\}}(tA)\right) & \end{aligned}$$

The last difference vanishes as tA and tB have non-zero eigenvalues. By continuity of c the first difference tends to zero with decreasing t. The second difference is independent of t, so as $f \in P_n(a,b)$, we conclude that

$$c_1\left(\chi_{\{x>0\}}(B) - \chi_{\{x>0\}}(A)\right) \geq 0,$$

which is only possible if $c_1 = 0$. □

Differentiablity is much more subtle however. Using similar ideas one could prove the matrix monotone function cannot have corners, i.e. cannot look locally like multiple of absolute value + affine function (up to first order), but there are far worse ways in which continuous function can be non-differentiable.

This is the point where divided differences come in. If one interprets matrix monotonicity solely with formulas, function being matrix monotone just means that some special set of linear combinations of the function values are positive. But this is precisely the type of condition we had with k-tone functions, and regularity started appearing. It is hence not such a big surprise that there should be some kind smoothness going on.

4.8 Main Theorem

Theorem 4.27. *Let $n \geq 1$. Then $f \in P_n(a,b)$, if and only if fh is $(2n-1)$-tone whenever h is polynomial of degree at most $(2n-2)$, non-negative on the real line.*

Proof. For the version with extra assumption, the starting point was to take derivative of the matrix function. Although we now cannot do that, we can try to replicate the proof otherwise.

Instead of proving that

$$[\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n, \lambda_n]_{fh} \geq 0$$

for any $a < \lambda_1, \lambda_2, \ldots, \lambda_{n-1}, \lambda_n < b$, we should prove that

$$[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh} \geq 0.$$

λ's should be eigenvalues of some map, but now there are $2n$ of them. Natural guess would be that they are eigenvalues of two maps, A and B.

But now everything starts to make sense: whenever A, B with $A \leq B$ and $w \in V$ the quantity

$$\langle (f(B) - f(A))w, w\rangle$$

is non-negative. On the other hand this can be expanded as some kind of linear combination of values of f at eigenvalues of A and B. Same is true for the divided differences, so there might be a chance to choose A, B and w such that

$$\langle (f(B) - f(A))w, w\rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}.$$

Moreover, should we find some kind of correspondence between triplets (A, B, w) and pairs $((\lambda_i)_{i=1}^{2n}, h)$, we would be done. This is the content of the main lemma.

Lemma 4.28. *If $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ and h is polynomial of degree at most $(2n-2)$ non-negative on the real line, we may find a strict projection pair (A, B) such that*

$$\langle (f(B) - f(A))w, w\rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}$$

for any $f : (a,b) \to \mathbb{R}$.

Conversely, if (A, B) is a strict projection pair and $w \in V$, then there exists $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ and polynomial h of degree at most $(2n-2)$, non-negative on the real line such that for any $f : (a,b) \to \mathbb{R}$ we have

$$\langle (f(B) - f(A))w, w\rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh}.$$

Before proving the lemma we show how it implies the theorem.

Assume first that $f \in P_n(a,b)$. We need to prove that fh is $(2n-1)$-tone for any h polynomial of degree at most $(2n-2)$ non-negative on real line. But any divided difference of such fh can be expressed by the main lemma 4.28 as $\langle (f(B)-f(A))w, w\rangle$ for some projection pair (A,B), and the previous is non-negative by the assumption.

Conversely, assume that fh is $(2n-1)$-tone for any suitable h and take any $A \leq B$. Write $B - A = \sum_{i=1}^n c_i P_{v_i}$ for some $c_i \geq 0$. To prove that $f(B) - f(A) \geq 0$ we simply need to prove that $f(A + \sum_{i=1}^k c_i P_{v_i}) - f(A + \sum_{i=1}^{k-1} c_i P_{v_i}) \geq 0$ for any $1 \leq k \leq n$, as $f(B) - f(A)$ is sum of such terms. We may hence assume that (A,B) projection pair.

We may also assume that (A,B) is strict. Indeed, if this would not be the case, we could decompose $V = \text{span}\{v_1\} \oplus V'$, where v_1 is the eigenvector, and factorize $A = A_{\text{span}\{v_1\}} \oplus A_{V'}$ and $P_w = 0 \oplus (P_w)_{V'}$. But now checking that $f(B) - f(A) \geq 0$ boils down to checking that $f(B_{V'}) - f(A_{V'}) \geq 0$, which would follow if we could prove that $f \in P_{n-1}(a,b)$. But this follows if we add the sentence "We induct on n." as the first sentence of this proof and use lemma 4.25.

Finally in this case, by the lemma 4.28 we may find $a < \lambda_1 < \lambda_2 < \ldots < \lambda_{2n-1} < \lambda_{2n} < b$ such that

$$\langle (f(B)-f(A))w, w\rangle = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{2n-1}, \lambda_{2n}]_{fh} \geq 0$$

and we are finally done.

In the "if"-direction we could alternatively make use of the continuity of f, which is guaranteed by the lemma 5.31

□

Let us then complete proof by proving the lemma 4.28.

Proof of lemma 4.28. The proof is based on lemmas 2.53 and 2.54. To find the connection we first assume f is entire. Then if and (A,B) is a strict projetion pair with $B - A = vv^*$ for some $v \in V$ and $w \in V$ we have

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

The integrand equals

$$\frac{h(z)}{\prod_{i=1}^n (z-\lambda_i(A)) \prod_{i=1}^n (z-\lambda_i(B))} f(z),$$

where $h(z) = \det(zI - B)\langle (zI - B)^{-1}v, w\rangle \det(zI - A)\langle (zI - A)^{-1}w, v\rangle$ and hence

$$\langle (f(B) - f(A))w, w\rangle = [\lambda_1(A), \ldots, \lambda_n(A), \lambda_1(B), \ldots, \lambda_n(B)]_{fh}.$$

Note that this identity evidently holds without any extra smoothness assumptions.

Now when (A, B) ranges over all strict projection pairs, the permutations of tuples

$$(\lambda_1(A), \ldots, \lambda_n(A), \lambda_1(B), \ldots, \lambda_n(B)) \tag{4.29}$$

range over all tuples of distinct numbers on (a, b). Hence to prove the lemma, we should prove that for fixed strict projection pair (A, B), as w ranges over V, h ranges over all polynomials of degree at most $(2n-2)$, non-negative on $\mathbb{R}$. This follows from lemma 4.22 and the following observation.

Lemma 4.30. *If (A, B) is a projection pair with $B - A = vv^*$ then*

$$\det(zI - A)(zI - A)^{-1}v = \det(zI - B)(zI - B)^{-1}v$$

Proof. As $zI - A = zI - B + vv^*$, multiplying both sides from left by $(zI - A)$ leads to the equivalent

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

which follows from 2.55. □

It follows that if $p(z) = \det(zI - B)\langle (zI - B)^{-1}v, w\rangle$, $h(z) = p(z)\overline{p(\overline{z})}$, so to finish the proof, we need only need to observe that when w ranges over V, $\det(zI - B)\langle (zI - B)^{-1}v, w\rangle$'s range over all complex polynomials of degree at most $(n-1)$. But this is clear as components of $\det(zI - A)(zI - A)^{-1}v$ with respect to eigenbasis of A, $(e_i)_{i=1}^n$ are $p_j(z) = \prod_{i\neq j}(z - \lambda_i(B))\langle v, e_i\rangle$, which are clearly linearly independent polynomials over $\mathbb{C}$.

To recap, the map

$$\begin{aligned} V &\to P_{n-1}(\mathbb{C}) = \{\text{Complex polynomials of degree at most } (n-1)\} \\ w &\mapsto \det(zI - A)\langle (zI - A)^{-1}v, w\rangle \end{aligned}$$

is antilinear bijection and the map

$$\begin{aligned} P_{n-1}(\mathbb{C}) &\to \{\text{Complex polynomials of degree at most } (2n-2) \text{ non-negative on } \mathbb{R}\} \\ p(z) &\mapsto p(z)\overline{p(\overline{z})} \end{aligned}$$

is surjection: composition of these maps is the correspondence between w and h. □

What's the moral of the story? If one unwraps all the definitions, matrix monotonicity is about positivity of some linear combinations of function values. Which linear combinations exactly? That is (more or less) explained in the main theorem.

4.9 A bit of history

Theorem 4.27 is usually stated in somewhat different terms. Functions of the form fh being $(2n-1)$-tone for some polynomials h can be also understood as certain matrix being positive, *Dobsch matrix*. Dobsch matrix (of order n) of $f : (a, b) \to \mathbb{R}$ at point $t \in (a, b)$ is the matrix

$$\begin{bmatrix} \frac{f'(t)}{1!} & \frac{f^{(2)}(t)}{2!} & \cdots & \frac{f^{(n)}(t)}{n!} \\ \frac{f^{(2)}(t)}{2!} & \frac{f^{(3)}(t)}{3!} & \cdots & \frac{f^{(n+1)}(t)}{(n+1)!} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{f^{(n)}(t)}{n!} & \frac{f^{(n+1)}(t)}{(n+1)!} & \cdots & \frac{f^{(2n-1)}(t)}{(2n-1)!} \end{bmatrix} \tag{4.31}$$

$$= \begin{bmatrix} [t,t]_f & [t,t,t]_f & \cdots & [t,\ldots,t]_f \\ [t,t,t]_f & [t,t,t,t]_f & \cdots & [t,t,\ldots,t]_f \\ \vdots & \vdots & \ddots & \vdots \\ [t,\ldots,t]_f & [t,t,\ldots,t]_f & \cdots & [t,t,t,\ldots,t]_f. \end{bmatrix}$$

Now an alternative version of 4.27 reads as follows.

Theorem 4.32. *Let $n \geq 1$. Then $f \in P_n(a, b) \cap C^{2n-1}(a, b)$, if and only if all Dobsch matrices of f of order n are positive for every $t \in (a, b)$.*

Proof. By simple computation

$$\frac{1}{(2n-1)!} \frac{d^{2n-1} \left(f(t) \left(\sum_{i=1}^{n} c_i t^{n-i}\right) \left(\sum_{i=1}^{n} \overline{c_i} t^{n-i}\right)\right)}{dt^{2n-1}}\Bigg|_{t=0} = \sum_{i,j=1}^{n} c_i \overline{c_j} \frac{f^{(i+j-1)}(0)}{(i+j-1)!}.$$

□

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

4.10 Loewner's theorem

In addition to characterizing n-monotone functions, by theorem 4.19, the classes $P_n(a, b)$, Loewner characterized the classes $P_\infty(a, b)$.

Theorem 4.33. *$f \in P_\infty(a, b)$, if and only if there exist Pick function φ extending over the interval (a, b) such that $\varphi\big|_{(a,b)} = f$.*

Proof. The "if" direction is not too hard: the Loewner matrices are essentially limits of Pick matrices so the result follows rather immediately from 4.19.

TODO: Fix the following (no more Pick points)

The "only if" is the tricky part. Theorem 4.19 tells us that the Dobsch matrices are positive on (a, b). If we can somehow show that $f \in C^{\omega}(a, b)$, then we see that all points of (a, b) are Pick points of f, and we can extend it to weakly Pick function on some open set of upper half-plane, from which it extends to unique Pick function by Pick–Nevanlinna theorem 6.19.

It suffices to proof the following result.

Lemma 4.34. *Let $f \in C^{\infty}(a, b)$ such that $f^{(2n-1)}(t) \geq 0$ for every $t \in (a, b)$. Then $f \in C^{\omega}(a, b)$.*

Proof. We shall verify the conditions of the theorem 5.43.

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 4.35. *Let $f \in C^{2}(a, b)$ such that $|f(x)| \leq M_0$ and $|f^{(2)}(x)| \leq M_2$ for any $x \in (a, b)$. Then*

$$|f'(x)| \leq \max\left(2\sqrt{M_0 M_2}, \frac{8M_0}{b-a}\right)$$

for any $x \in (a, b)$.

Proof. Take any $x_0 \in (a, b)$ and set $f'(x_0) = c$: we shall prove the given bound of c. Without loss of generality we may assume that $c \geq 0$ and $x_0 \leq \frac{a+b}{2}$. The idea is that as $f^{(2)}$ is not too big, f' has to be positive and reasonably big interval around the point x_0 which means that f has to increase a lot around x_0. By the assumption it can't increase more than $2M_0$, however.

To make this argument precise and effective, we split into too cases.

1. $M_2(b - x_0) > c$: this means that we have

   $$f'(x) \geq c - M_2(x - x_0)$$

   for $x_0 \leq x \leq \frac{c}{M_2} + x_0$ and hence

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

   which yields the first inequality.

2. $M_2(b - x_0) \leq c$: now we have

$$f'(x) \geq c\frac{b-x}{b-x_0},$$

for every $x_0 \leq x < b$

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

Letting $x \to b$ and using $(b - x_0) \geq \frac{b-a}{2}$ we get the second inequality.

TODO: pictures of function and it's derivatives TODO: better proof □

To prove the bound for odd n, we would like to play the same game as in the proof of lemma 5.41, but the unfortunate thing is that the even order terms are breaking the inequality. We can salvage the situation by getting rid of them. Assume first that $0 \in (a, b)$. Trick is to consider the Taylor expansion for $f(x) - f(-x)$, centered at 0, instead:

$$f(x) - f(-x) = 2\left(\sum_{i=1}^{n} \frac{f^{(2i-1)}(0)}{(2i-1)!}x^{2i-1}\right) + \int_{0}^{x} \frac{f^{(2n+1)}(t) + f^{(2n+1)}(-t)}{(2n)!}(x-t)^{2n}dt.$$

But now we can simply follow the same argument. □

□

TODO:

- Why smoothness
- Examples
- Pick functions are monotone
- Heaviside function
- Trace inequalities: if f is monotone/convex then $\mathrm{tr}f$ is monotone/convex. Proof idea: we may write $\mathrm{tr}f$ as a limit of finite sum of translations of Heaviside functions (monotone case) or absolute values (convex case), so its sufficient to prove the claim for these functions. For monotone case it hence suffices to prove that if $A \leq B$, B has at least as many non-negative eigenvalues as A. But this is clear by subspace characterization of non-negative eigenvalues. For convex case, it suffices to prove

that $\mathrm{tr}|A| + \mathrm{tr}|B| \geq \mathrm{tr}|A+B|$ for any $A, B \in \mathcal{H}^n(a,b)$. For this, note that if $(e_i)_{i=1}^n$ is eigenbasis of $A+B$, we have

$$\begin{aligned} \mathrm{tr}|A+B| &= \sum_{i=1}^n \langle |A+B| e_i, e_i \rangle \\ = \sum_{i=1}^n |\langle (A+B)e_i, e_i\rangle| &\leq \sum_{i=1}^n |\langle Ae_i, e_i\rangle| + \sum_{i=1}^n |\langle Be_i, e_i\rangle| \\ \leq \sum_{i=1}^n \langle |A|e_i, e_i\rangle + \sum_{i=1}^n \langle |B|e_i, e_i\rangle &= \mathrm{tr}|A| + \mathrm{tr}|B| \end{aligned}$$

- What about trace inequalities for k-tone functions? Eigen-package seems to find a counterexample for 6-tone functions and $n = 2$, but it's hard to see if there's enough numerical stability. At divided differences of polynomials vanish. First non-trivial question would be: If $A_j = A + jH$ for $0 \leq j \leq 3$ and $H \geq 0$. Then is it necessarily the case that
$$\mathrm{tr}\left(A_3|A_3| - 3A_2|A_2| + 3A_1|A_1| - A_0|A_0|\right) \geq 0?$$
This would imply that 3-tone functions would lift to trace 3-tone functions. Maybe expressing this as a contour integral from $-i\infty \to i\infty$ a same tricks as in the paper. First projection case: H is projection. Or: approximate by integrals of heat kernels. It should be sufficient to proof things for k-fold integrals or heat kernel, or by scaling just for gaussian function.

- How is the previous related to the $|\cdot|$ not being operator-convex: quadratic form inequality for eigenvectors is not enough.

- The previous also implies that
$$f(Q_A(v)) \leq Q_{f(A)}(v)$$
for any convex f. Using this and Minkowski one sees that p-schatten norms are indeed norms.

- For f, g generalization (Look at $h(X) = g(\mathrm{tr} f(X))$) we need that f is convex. What else? h is convex if it is convex for diagonalizable matrices and f is convex and g increasing. For the diagonalizable maps it is sufficient that f is increasing and $g = f^{-1}$ and $\log \circ f \circ \exp$ is convex.

- Von Neumann trace inequality, more trace inequalities.

- On Generalizations of Minkowski's Inequality in the Form of a Triangle Inequality, Mulholland
- There should nice proof for Loewner theorem, like the blog post for Bernstein's big theorem.

Chapter 5

k-tone functions

5.1 Motivation

As mentioned in the introduction, k-tone functions correspond to the functions with non-negative k'th derivative[1]. What should this mean?

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

Theorem 5.1. *Let $f : (a, b) \to \mathbb{R}$ be differentiable. Then f is increasing, if and only if $f'(x) \geq 0$ for every $x \in (a, b)$.*

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

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

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

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

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

[1] The terminology is not very established, and such functions are also occasionally called k-monotone or k-convex.

While this proof by the mean value theorem works in more general setting, if $f \in C^1$, one has more instructive proof.[2]

Alternate proof for the theorem 5.1 (in the case $f \in C^1(a,b)$). Note that if $f \in C^1(a,b)$, we may write

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

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

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

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

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

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

Aside from capturing the first derivative, divided difference has two rather convenient properties.

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

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

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

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

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

Theorem 5.2. *Let $f : (a, b) \to \mathbb{R}$ be twice differentiable. Then f is convex, if and only if $f^{(2)}(x) \geq 0$ for every $x \in (a, b)$.*

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

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

This suggest that we may write

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

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

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

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

TODO: picture

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

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

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

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

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

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

the second divided difference of f.

One would hope 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.

5.2 Divided differences

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

Definition 5.3. Let $n \geq 0$. For any real function $f : (a, b) \to \mathbb{R}$ we define the corresponding n'th divided difference $[\cdots]_f : (a, b)^{n+1} \setminus D_{n+1}$ by setting

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

We will soon prove that divided differences (of order n) are simply weighted averages of the n'th derivative.

5.2.1 Basic properties

Divided differences have the following important properties.

Proposition 5.4. *Divided differences are symmetric in the variable, i.e. for any* $f : (a,b) \to \mathbb{R}$ *and pairwise distinct* $a < x_0, x_1, \ldots, x_n < b$ *permutation* $\sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ *we have*

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

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

Proof. Easy to check. □

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

Proposition 5.5. *For any* $f : (a,b) \to \mathbb{R}$ *and pairwise distinct* $x_0, x_1, \ldots, x_n \in (a,b)$ *we have*

$$[x_0, x_1, \ldots, x_n]_f = \frac{[x_0, x_1, \ldots, x_{n-1}]_f - [x_1, x_2, \ldots, x_n]_f}{x_0 - x_n} = [x_0, x_n]_{[\cdot, x_1, \ldots, x_{n-1}]_f} \tag{5.6}$$

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

$$[y_1, y_2, \ldots, y_m]_{[\cdot, x_1, x_2, \ldots, x_n]_f} = [y_1, y_2, \ldots, y_m, x_1, x_2, \ldots, x_n]_f. \tag{5.7}$$

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

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

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

We call 5.7 the *nesting property* of divided differences. Although the analogy isn't perfect, one could think that this identity says that m'th derivative of the n'th derivative is the $(n+m)$'th derivative.

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

Proposition 5.8. *We have* $[x_0, x_1, \ldots, x_n]_{(x \mapsto x^n)} = 1$ *and* $[x_0, x_1, \ldots, x_n]_p = 0$ *for any polynomial of degree at most* $n-1$*. In other words,* $[x_0, x_1, \ldots, x_n]_f$ *is the leading coefficient of the Lagrange interpolation polynomial on pairs* $(x_0, f(x_0)), (x_1, f(x_1), \ldots, (x_n, f(x_n))$*.*

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

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

and the leading coefficient of this polynomial is exactly the divided difference. $\square$

5.2.2 Peano representation

Coming back to the original motivation, divided differences enjoy an integral representation also for larger n, albeit somewhat more complicated.

Theorem 5.9. *If $f \in C^n(a, b)$, then for any pairwise distinct $a < x_0, x_1, x_2, \dots, x_n < b$ we have*

$$[x_0, x_1, \dots, x_n]_f = \int_{\mathbb{R}} f^{(n)}(t) w(t) dt, \tag{5.10}$$

where

$$w(t) := w_{x_0,x_1,\dots,x_n}(t) = \frac{1}{(n-1)!} \sum_{i=0}^{n} \frac{((x_i - t)_+)^{n-1}}{\prod_{j\neq i}(x_i - x_j)}. \tag{5.11}$$

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

Proof. Note that the weight is simply the n'th divided difference of the map $g_{t,n} : x \mapsto \frac{1}{(n-1)!}((x - t)_+)^{n-1}$. This is not very surprising: one should think that $g_{t,n}$ is the function whose n'th derivative is δ_t. Now if we plug in $f = g_{t,n}$, (as in the proof of 5.2), we, at least morally, get the claim. While the previous argument could be pushed through, we take safer route. To prove that the formula even makes sense, we should prove the claim on the support. It is clear that w is zero whenever $t \geq \max(x_i)$. If on the other hand $t \leq \min(x_i)$, $w(t)$ is simply a n'th divided difference of the map $x \mapsto \frac{1}{(n-1)!}(x - t)^{n-1}$, which is zero by the proposition 5.8.

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

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

where

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

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

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

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

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

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

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

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

we have

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

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

The weight 5.11 is called *Peano kernel* (of order n, whenever there are $(n+1)$ points). The points $x_0, x_1, \ldots, x_n$ are called the nodes of w.

TODO: pictures of Peano kernels

As an very important corollary we get the following.

Theorem 5.12 (Mean value theorem for divided differences)**.** *Let $n \geq 1$ and $f \in C^n(a,b)$. Then for any pairwise distinct $x_0, x_1, \ldots, x_n \in (a,b)$ we have*

$$\min_{0 \leq i \leq n} (x_i) < \xi < \max_{0 \leq i \leq n} (x_i)$$

such that

$$[x_0, x_1, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}. \tag{5.13}$$

Proof. This follows immediately from 5.9.

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

By linearity and proposition 5.8 it suffices to verify the claim for the case $f(x_i) = 0$ for $0 \leq i \leq n$, or to verify the following claim.

Lemma 5.14. *If f is n times differentiable, and has $n+1$ roots, then $f^{(n)}$ has a root (in the interior of the convex hull of the roots).*

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

□

The mean value theorem could be also used to prove the non-negativity of the weight w: if w were somewhere negative, one could construct function with non-negative derivative and negative divided difference, which would contradict 5.13.

As in the case $n = 1$, if for $n > 1$ we can continuously extend divided differences to the set D_{n+1}, we should do that, and we identify the resulting function with the original one. We will later proof that, as expected, this can be done, if and only $f \in C^n(a,b)$. In this case by 5.13 the extesion satisfies

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

which together with 5.6 is enough to describe the divided differences with values of the function and its derivative.

5.2.3 Basis elements

There's very instructive alternate way to think about theorem 5.9.

Theorem 5.15. *Let $f \in C^k(a,b)$. Then for any $a < c \leq x \leq d < b$ we have*

$$f(x) = f(c) + (x-c)f'(c) + \ldots + (x-c)^{k-1}\frac{f^{(k-1)}(c)}{(k-1)!} + \int_c^d g_{t,k}(x)dt.$$

Proof. This is just a restatement of the usual Taylor expansion. □

The previous observation could have also be used to prove the identity 5.10 itself. This is a kind of result elegance of which would benefit from the quotient point of view: should we consider k-tone functions only up to polynomials of degree less than k, would we get rid of the first k summands.

5.2.4 Identities

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

Proposition 5.16. *Let n, k, f, g, $f_1, f_2, \ldots, f_k$ and $x_0, x_1, \ldots, x_n$ be such that the folliowing identities make sense.*

(i) (Newton expansion)

$$\begin{aligned} f(x) = {} & [x_0]_f + [x_0, x_1]_f(x-x_0) + [x_0,x_1,x_2]_f(x-x_0)(x-x_1) + \ldots \\ & + [x_0, x_1, \ldots, x_n]_f(x-x_0)(x-x_1)\cdots(x-x_{n-1}) \\ & + [x, x_0, x_1, \ldots, x_n]_f(x-x_0)(x-x_1)\cdots(x-x_n), \end{aligned} \tag{5.17}$$

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

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} + [x, x_0, x_0, \ldots, x_0]_f(x-x_0)^n, \tag{5.18}$$

(ii) (Product rule)

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

(iii) (Leibniz rule)

$$[x_0,x_1,\dots,x_n]_{fg} = [x_0]_f[x_0,\dots,x_n]_g + [x_0,x_1]_f[x_1,\dots,x_n]_g + \dots \\ + [x_0,x_1,\dots,x_{n-1}]_f[x_{n-1},x_n]_g + [x_0,x_1,\dots,x_n]_f[x_n]_g. \tag{5.19}$$

More generally

$$[x_0,x_1,\dots,x_n]_{f_1f_2\cdots f_k} = \sum_{0=i_0<i_1<i_2<\dots<i_{k-1}<i_k=n} \prod_{j=1}^{k}[x_{i_{j-1},\dots,x_{i_j}}]_{f_j}$$

(iv) (Chain rule)

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

(v) (Faà di Bruno formula)

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

Proof sketches. (i) Easy induction using 5.6. Notice that this formula makes it also clear that the divided difference agrees with the degree n coefficient of the interpolating polynomial.

(ii) Easy to check.

(iii) Induction using the product rule (i.e. the case $n = 1$) and the nesting rule 5.7. Alternatively one could write Newton expansions of both f and g with sequences $(x_0,x_1,\dots,x_n)$ and $(x_n,x_{n-1},\dots,x_0)$ and notice that the given sum gives exactly the leading term of the interpolating polynomial of fg. The more general case follows from the case of two functions by induction.

(iv) Easy to check.

(v) A bit tedious induction using the Leibniz rule and 5.6.

□

5.2.5 k-tone functions

All these observations are more than enough to verify that our definition of divided differences gives us suitable notion k-tone functions.

Definition 5.20. $f : (a, b) \to \mathbb{R}$ is called k-tone if for any $x_0, x_1, \ldots, x_n \in (a, b)$ of distinct points we have

$$[x_0, x_1, \ldots, x_n]_f \geq 0,$$

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

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

Theorem 5.21. *Let k be an non-negative integer and (a, b) an open interval. Then $P^{(k)}(a, b) \subset \mathbb{R}^{(a,b)}$ is (almost) a proper cone.*

Proof. Since the divided differences are continuous linear functional result follows if we can prove that

$$[\cdot, \cdot, \ldots, \cdot]_f = 0 \Leftrightarrow f = 0.$$

This isn't quite true, instead we have

$$[\cdot, \cdot, \ldots, \cdot]_f = 0 \Leftrightarrow \ f \text{ is a polynomial of degree less } n.$$

To see why this is true, note that we already proved "$\Leftarrow$" -direction. The other direction follows immediately from the Newton expansion 5.17. □

Mean value theorem tells us that C^k k-tone functions are exactly the functions with non-negative k'th derivative.

TODO: quotient topological vector space

5.2.6 Cauchy's integral formula

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

Lemma 5.22 (Cauchy's integral formula for divided differences)**.** *If γ is a closed counter-clockwise curve enclosing the numbers $x_0, x_1, \ldots, x_n$, we have*

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

Proof. Easy induction, verifying 5.6, by taking Cauchy's integral formula as a base case. Alternatively, the claim is a direct consequence of the Residue theorem.

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

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

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

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

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

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

Example 5.23. We may express the Lagrange interpolation polynomial of a analytic function f and sequence $X = (x_i)_{i=0}^n$ by

$$P_X(x) = \frac{1}{2\pi i}\int_\gamma \frac{p_X(x)-p_X(z)}{x-z}\frac{f(z)}{p_X(z)}dz = [x_0,x_1,\dots,x_n]_{f[x,\cdot]_{p_X}}.$$

More generally, if some of the points coincide, we get so-called Hermite interpolation polynomial.

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

Alternate proof for the Leibniz rule for the divided differences. Write Newton expansion for f and g with points reversed for g. The rest follows as in the final proof of Theorem (5.22). $\square$

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

Lemma 5.24. *Let $n, m \geq 0$. Assume that for some constants $c_{i,j}$ and $a_{i,j} \in (a, b)$ we have*

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

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

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

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

Rest of the alternate proof. Note that if we expand the divided differences, we are almost in the situation of the lemma 5.24; now we just have product of two functions instead. Story is the same. $\square$

5.2.7 Locality

One of the properties of the divided differences, which might not be clear from the definition, is that they can also be used to model local phenomena. One of the important properties of the k-tone functions is that if a function is k-tone on two overlapping intervals, then the function is k-tone on their union. While this definitely holds for C^k functions, it's not really clear how to change this argument for the general case.

If one thinks that k-tone functions have k'th derivative as a positive measure, the locality property should be a special case of the general property of distributions.

Proposition 5.25. *Let $a < c < b < d$ and μ distribution on (a, d), restriction of which to (a, b) and (c, d) is a positive measure. Then μ is a positive measure.*

Proof. We should prove that $\mu(f)$ is non-negative for every non-negative test function f on (a, d). But every such function can be written as sum of two non-negative test functions, f_1 and f_2, f_1 supported on (a, b) and f_2 on (c, d), so $\mu(f) = \mu(f_1) + \mu(f_2) \geq 0$ by the assumption. □

They key idea in the proof was to split the test functions to two parts, one supported on (a, b) and one on (c, d). It turns out that we can do the same thing with Peano kernels.

Lemma 5.26. *Let $a < c < b < d$ be reals and w a Peano kernel supported on (a, d). Then w can written as a (finite) weighted average of Peano kernels, all of which are supported either on (a, b) or on (c, d).*

Proof. Let n be the order of the Peano kernel and let $a < x_0 < x_1 < \ldots < x_n < d$ be the nodes or w.

The case $n = 1$ is rather clear: we simply split characteristic function of an interval to characteristic function of two intervals. TODO: picture. In terms of the kernels, if $a < x_0 < c < b < x_1 < d$, we can pick $c < y_0 < b$ and write

$$w_{x_0,x_1} = \frac{y_0 - x_0}{x_1 - x_0} w_{x_0,y_0} + \frac{x_1 - y_0}{x_1 - x_0} w_{y_0,x_1} :$$

this is a sought decomposition.

When $n = 2$ the goal is not much harder: we have a triangle whose corners have (distinct) x-coordinates on the interval (a, d) and we should split it to smaller triangles in such a way that

- No triangle has two equal x-coordinates (this is the property of the Peano kernels)
- All triangles have all their corners' x-coordinates either on (a, b) and (c, d).

TODO: picture We call such triangles good. While the above picture should be rather convincing already, one can write an general algorithm generating such decomposition.

Step 1. If the triangle is good already, we are done.

Step 2. Pick $y_0 \in (c, b)$, which does not coincide any of the x_0, x_1, x_2.

Step 3. Divide the triangle to two triagles with x-coordinates (x_0, x_1, y_0) and (x_1, y_0, x_2), as illustrated.

Step 4. Run this algorithm recursively for these two triangles.

Why does this algorithm terminate? Note that if any of the x_i's are in (c,b), the triangle is either good or $x_1 \in (c,b)$ (or maybe both). In the former case we are done, and in the latter the two parts of the split are both good. If none of x_i's are in (c,b), both of the parts of the split has a coordinate in (c,b), so also this case leads to a good split. In other words we can keep splitting triangles in such a way that they either become good or they have more nodes on (c,b).

It's easy to verify that splitting the triangle corresponds to the identity

$$w_{x_0,x_1,x_2} = \frac{y_0 - x_0}{x_2 - x_0} w_{x_0,x_1,y_0} + \frac{x_1 - y_0}{x_2 - x_0} w_{x_1,y_0,x_2}.$$

When $n > 2$ geometric picture is largely lost (at least by the author), but the algebra generalizes perfectly: we can still split Peano kernels using the following identity:

$$w_{x_0,x_1,\ldots,x_n} = \frac{y_0 - x_0}{x_n - x_0} w_{x_0,x_1,\ldots,x_{n-1},y_0} + \frac{x_n - y_0}{x_n - x_0} w_{x_1,\ldots,x_n,y_0}. \tag{5.27}$$

Where does this come from? Recall that the Peano kernels are nothing more than the divided differences of the functions $g_{t,n} = \frac{1}{(n-1)!}((\cdot - t)_+)^{n-1}$,. The first identity immediately generalizes

$$[x_0, x_1]_f = \frac{y_0 - x_0}{x_1 - x_0}[x_0, y_0]_f + \frac{x_1 - y_0}{x_1 - x_0}[y_0, x_1]_f,$$

where f is now any function. By the nesting property the identity 5.27 is nothing more than the previous identity applied to $f = [\cdot, x_1, \ldots, x_{n-1}]_{g_{t,n}}$. Note that we need $x_0 < y_0 < x_n$, so that the weighted average is really a convex combination.

Now we are ready to generalize the algorighm to higher n:

Step 1. If the kernel is good already, we are done.

Step 2. Pick $y_0 \in (c,b)$, which does not coincide any of the $x_0, x_1, \ldots, x_n$.

Step 3. Divide the kernel to two kernels with nodes $(x_0, x_1, \ldots, x_n, y_0)$ and $(y_0, x_1, \ldots, x_n)$ as in the 5.27.

Step 4. Run this algorithm recursively for these two kernels.

This algorithm terminates basically because of the same reason: if the kernel isn't good already, the two splits have more nodes on (c,b), and this quantity cannot increase forever. TODO: figure □

While this property is of independent interest, the real use of it is its generalization to divided differences.

Lemma 5.28. *Let $a < c < b < d$ be reals and $x_0, x_1, \dots, x_n \in (a, d)$. Then we may find $N, M \in \mathbb{N}$, sequences $(y_{0,i}, \dots, y_{n,i})$ and $(z_{0,j}, \dots, z_{n,j})$ and numbers p_i and q_j for $1 \le i \le N$ and $1 \le j \le M$, such that*

- $\sum_{i=1}^{N} p_i + \sum_{j=1}^{M} q_j = 0$ *and $p_i, q_j \ge 0$ for every $1 \le i \le N$ and $1 \le j \le M$.*
- $y_{0,i}, y_{1,i}, \dots, y_{n,i}$ *are pairwise distinct elements of (a, b) for every $1 \le i \le N$.*
- $z_{0,j}, z_{1,j}, \dots, z_{n,j}$ *are pairwise distinct elements of (c, d) for every $1 \le i \le M$.*
- *For every $f : (a, d) \to \mathbb{R}$ we have*

$$[x_0, x_1, \dots, x_n]_f = \sum_{i=1}^{N} p_i [y_{0,i}, y_{1,i}, \dots, y_{n,i}]_f + \sum_{j=1}^{M} q_j [z_{0,j}, z_{1,j}, \dots, z_{n,j}]_f.$$

Proof. Proof is almost identical to that of the lemma 5.26: we more-or-less just replace the identity 5.27 by

$$[x_0, x_1, \dots, x_n]_f = \frac{y_0 - x_0}{x_n - x_0}[x_0, x_1, \dots, x_{n-1}, y_0]_f + \frac{x_n - y_0}{x_n - x_0}[x_1, \dots, x_n, y_0]_f, \tag{5.29}$$

which is valid because of essentially the same reasoning, and replace the word "kernel" with word "tuple". □

We are now ready to prove the locality property of the k-tone functions.

Proposition 5.30. *$P^{(k)}$ is a local property i.e. $P^{(k)}(a, b) \cap P^{(k)}(c, d) \subset P^{(k)}(a, d)$ for any $-\infty \le a \le c < b \le d \le \infty$. 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. This follows immediately from lemma 5.28. □

Note that we could have also used the splitting property 5.29 to slightly simplify the proof of theorem 5.9. In the induction step we managed to prove that we have

$$[x_0, x_1, \dots, x_{n-1}, t]_{g_{t,n}} = \frac{1}{n-1}[x_0, x_1, \dots, x_{n-1}]_{g_{t,n-1}} = \frac{1}{n-1} w_{x_0, \dots, x_{n-1}}(t) \ge 0$$

for any $x_0, x_1, \ldots, x_{n-1}$. But this readily implies that the divided differences are non-negative on all tuples as we have

$$\begin{aligned}
w_{x_0,x_1,\ldots,x_{n-1},x_n} &= [x_0, x_1, \ldots, x_{n-1}, x_n]_{g_{t,n}} \\
&= \frac{t - x_0}{x_n - x_0}[x_0, x_1, \ldots, x_{n-1}, t]_{g_{t,n}} + \frac{x_n - t}{x_n - x_0}[x_1, \ldots, x_n, t]_{g_{t,n}} \\
&= \frac{1}{n-1}\frac{t - x_0}{x_n - x_0}[x_0, x_1, \ldots, x_{n-1}]_{g_{t,n-1}} + \frac{1}{n-1}\frac{x_n - t}{x_n - x_0}[x_1, \ldots, x_n]_{g_{t,n-1}} \\
&= \frac{1}{n-1}\frac{t - x_0}{x_n - x_0}w_{x_0,x_1,\ldots,x_{n-1}} + \frac{1}{n-1}\frac{x_n - t}{x_n - x_0}w_{x_1,\ldots,x_n} \\
&\geq 0.
\end{aligned}$$

Of course, this approach only works if $\min(x_i) \leq t \leq \max(x_i)$, but if this is not the case, as observed, the divided differences are zero anyway. The previous identity can be also used to recursively compute Peano kernels.

5.3 Regularity

The real power of the divided differences comes in when are used to carry regularity information.

Theorem 5.31. *Let $k \geq 2$. Then $f \in P^{(k)}(a, b)$, if and only if $f \in C^{k-2}(a, b)$ and $f^{(k-2)}$ is convex.*

"Proof". Let $f \in P^{(k)}(a, b)$. Since $f^{(k)}$ is a positive measure, $f^{(k-1)}$ is increasing and $f^{(k-2)}$ is convex. As convex functions are continuous, we are done with $\Rightarrow$. Conversely, if $f \in C^{k-2}(a, b)$ and $f^{(k-2)}$ is convex, then $f^{(k-2)}$ has second derivative as a positive measure. But this measure is also the k'th derivative of f, so $f \in P^{(k)}(a, b)$. □

Even though the previous argument isn't exactly sound (at least given our current machinery), the result is true. In this section we will translate the proof to the language of the divided differences.

The first step is to connect the divided differences of a function to the divided differences (of one lower order) of the derivative.

Lemma 5.32. *Let $f \in C^1(a, b)$. Then for any (pairwise distinct) $x_0, x_1, \ldots, x_n \in (a, b)$ we have*

$$[x_0, x_1, \ldots, x_{n-1}]_{f'} = \sum_{i=0}^{n-1}[x_0, x_1, \ldots, x_{i-1}, x_i, x_i, x_{i+1}, \ldots, x_{n-1}]_f \tag{5.33}$$

and

$$(5.34) \qquad [x_0, x_1, \ldots, x_n]_f = \int_0^1 [x_0, x_1, \ldots, x_{n-1}]_{f'(s\cdot+(1-s)x_n)} ds$$
$$= \int_0^1 [sx_0 + (1-s)x_n, \ldots, sx_{n-1} + (1-s)x_n]_{f'} s^{n-1} ds.$$

Proof. Note that divided differences of f have repeated entries in the first identity. As mentioned, these values of the divided difference are defined as a continuous extension. We will take the existence of this extension given for now.

We have

$$\begin{aligned}
[x_0, x_1, \ldots, x_{n-1}]_{f'} &= \lim_{h\to 0} [x_0, x_1, \ldots, x_{n-1}]_{\frac{f(\cdot+h)-f(\cdot)}{h}} \\
&= \lim_{h\to 0} \frac{[x_0, x_1, \ldots, x_{n-1}]_{f(\cdot+h)} - [x_0, x_1, \ldots, x_{n-1}]_f}{h} \\
&= \lim_{h\to 0} \frac{[x_0 + h, x_1 + h, \ldots, x_{n-1} + h]_f - [x_0, x_1, \ldots, x_{n-1}]_f}{h}.
\end{aligned}$$

Now the approach is basically the same as with differentiation of multivariate functions: we write the difference as sum of n differences: the difference can be expressed as sum of differences where only one of the entries are changed at time.

$$\begin{aligned}
&\lim_{h\to 0} \frac{[x_0 + h, x_1 + h, \ldots, x_{n-1} + h]_f - [x_0, x_1, \ldots, x_{n-1}]_f}{h} \\
&= \lim_{h\to 0} \Bigg(\frac{[x_0 + h, x_1 + h, \ldots, x_{n-1} + h]_f - [x_0 + h, x_1 + h, \ldots, x_{n-2} + h, x_{n-1}]_f}{h} \\
&+ \frac{[x_0 + h, x_1 + h, \ldots, x_{n-2} + h, x_{n-1}]_f - [x_0 + h, x_1 + h, \ldots, x_{n-2}, x_{n-1}]_f}{h} \\
&+ \ldots \\
&+ \frac{[x_0 + h, x_1, \ldots, x_{n-1}]_f - [x_0, x_1, \ldots, x_{n-1}]_f}{h} \Bigg) \\
&= \lim_{h\to 0} \left(\sum_{i=0}^{n} [x_0 + h, \ldots, x_{i-1} + h, x_i + h, x_i, x_{i+1}, \ldots, x_{n-1}]_f \right).
\end{aligned}$$

Now assuming the claim on the continuity, the limit is exactly what we wanted.

First equality of second claim was already essentially proved in the proof of theorem 5.9; the second is a simple computation. $\square$

Note that the proof essentially gives also the following identity.

Proposition 5.35. *Let $x_0, x_1, \ldots, x_n$ and $y_0, y_1, \ldots, y_n$ be pairwise distinct points on (a, b). Then for any $f : (a, b) \to \mathbb{R}$ we have*

$$[y_0, y_1, \ldots, y_{n-1}]_f - [x_0, x_1, \ldots, x_{n-1}]_f = \sum_{i=0}^{n-1} [x_0, \ldots, x_{i-1}, x_i, y_i, y_{i+1}, \ldots, y_{n-1}]_f (y_i - x_i).$$

Next step is to connect the regularity of divided differences to regularity of divided differences of the derivative. Denote

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

Note that $D_{n+1,2}$ is exactly the set where the divided differences aren't defined. Still, if f is smooth enough, we should be able to continuosly extend the divided differences to this set, or at least to some subset set of it. This thinking leads to the following notion of the regularity of a function.

Definition 5.36. Let $f : (a, b) \to \mathbb{R}$ and $k \geq 0$. We call f *weakly C^k* (on (a, b)), or write $f \in C^k_w(a, b)$, if its order k divided differences can be continuosly extended to $(a, b)^{k+1}$.

Our aim is to prove that function is weakly C^k, if and only if it's C^k. Note that this trivially holds for $k = 0$.

Lemma 5.37. *Let $n \geq k$. Then $f \in C^k_w(a, b)$, if and only if the order n divided differences of f extend continuosly to $(a, b)^{n+1} \setminus D_{n+1,k+2}$.*

Proof. We prove the statement by induction on n, taking $n = k$ as a base case.

Note that the case $n = k$ is exactly the definition.

In the induction step fix $n > k$ take any $C^k_w(a, b)$. By the induction hypothesis its divided differences of order $n - 1$ extend continuously to $(a, b)^n \setminus D_{n,k+2}$. Now take any $(x_0, x_1, \ldots, x_n) \in (a, b)^{n+1} \setminus D_{n+1,k+2}$. Consider any sequence of tuples $(y_{0,j}, y_{1,j}, \ldots, y_{n,j})_{j=1}^{\mathbb{N}}$ such that $(y_{i,j}) \to x_i$ as $j \to \infty$ for every $0 \leq i \leq n$ and for every fixed $j \in \mathbb{N}$. We should prove that the sequence

$$([y_{0,j}, y_{1,j}, \ldots, y_{n,j}]_f)_{j=1}^{\mathbb{N}}$$

converges. By permutation we may assume that $x_0 \neq x_n$. But since

$$[y_{0,j}, y_{1,j}, \ldots, y_{n,j}]_f = \frac{[y_{0,j}, y_{1,j}, \ldots, y_{n-1,j}]_f - [y_{1,j}, y_{2,j}, \ldots, y_{n,j}]_f}{y_{0,j} - y_{n,j}},$$

$(y_{0,j}, \ldots, y_{n-1,j}), (y_{1,j}, \ldots, y_{n,j}) \in (a, b)^n \setminus D_{n,k+2}$ for every $j \in \mathbb{N}$, and these sequences converge, we see that the divided differences in the numerator converge. As also the denominator converges to non-zero number, the whole expression converges, and we are done with the first direction.

Assume then that the order n divided differences extend continuously to $(a,b)^{n+1} \setminus D_{n+1,k+2}$. Our aim is to prove that $f \in C^k_w(a,b)$. To this end we prove that we can extend $(n-1)$'th divided differences continously to $(a,b)^n \setminus D_{n,k+2}$, as then the induction hypothesis finishes the claim. So take any sequnce $(y_{0,j}, y_{1,j}, \ldots, y_{n-1,j})_{j=1}^{\mathbb{N}}$ converging to $(x_0, x_1, \ldots, x_{n-1}) \in (a,b)^n \setminus D_{n,k+2}$ and choose additional sequence $(z_0, z_1, \ldots, z_{n-1})$ of pairwise distinct points distinct from all the x_i's and $y_{i,j}$'s. Now we can write

$$[y_{0,j}, \ldots, y_{n-1,j}]_f = [z_0, z_1, \ldots, z_{n-1}]_f + \sum_{i=0}^{n-1} [z_0, \ldots, z_{i-1} z_i, y_{i,j}, y_{i+1,j}, \ldots, y_{n-1,j}]_f (y_{i,j} - z_i).$$

As by the induction hypothesis the right-hand side converges, so does the left-hand side, and we are done. □

Theorem 5.38. *Let $f : (a,b) \to \mathbb{R}$, $k \geq 1$. Then $f \in C^k_w(a,b)$, if and only if $f \in C^1(a,b)$ and $f' \in C^{k-1}_w(a,b)$.*

Proof. We start with the "$\Rightarrow$"-direction.

Let's start by proving that f is continuouly differentiable. Lemma 5.37 easily implies that it is sufficient prove this for the case $k = 1$. But in this case we now that the limits $\lim_{x \to x_0} [x, x_0]_f = [x_0, x_0]_f$ exist and f is hence differentiable with $f'(x) = [x,x]_f$. Also, $x \mapsto [x,x]_f = f'(x)$ is continuous.

Now the identity 5.33 easily implies the claim.

For the "$\Leftarrow$"-direction take any sequence $(y_{0,j}, \ldots, y_{k,j})_{j=1}^{\mathbb{N}}$ of elements of $(a,b)^{k+1}$ converging to $(x_0, \ldots, x_k) \in (a,b)^{k+1}$. Now by 5.34

$$[y_{0,j}, y_{1,j}, \ldots, y_{k,j}]_f = \int_0^1 [s y_{0,j} + (1-s) y_{k,j}, \ldots, s y_{k-1,j} + (1-s) y_{k,j}]_{f'} s^{k-1} ds$$

As $j \to \infty$,, we have $(y_{0,j}, \ldots, y_{n,j}) \to (x_0, x_1, \ldots, x_k)$ and hence also $(s y_{0,j} + (1 - s) y_{k,j}, \ldots, s y_{k-1,j} + (1-s) y_{k,j}) \to (s x_0 + (1-s) x_n, \ldots, s x_{n-1} + (1-s) x_k)$ uniformly (over s). As $(n-1)$'th divided differences of f' extend continuosly, they are uniformly continuous over all the compact sets, so in particular the integrand converges uniformly to

$$[s x_0 + (1-s) x_n, \ldots, s x_{n-1} + (1-s) x_k]_{f'} s^{n-1},$$

and hence also the integral converges, which was to be shown. □

Corollary 5.39. *$f \in C^k_w(a,b)$ if and only if $f \in C^k(a,b)$.*

Proof. Simply apply lemma 5.38 inductively. □

Just like one can carry regularity information, one can carry boundedness information.

Lemma 5.40. *Let $f : (a,b) \to \mathbb{R}$ and $n \geq 2$. Then the n'th order divided differences of f are bounded, if and only if $f \in C^1$ and the order $(n-1)$ divided differences of f' are bounded. Moreover, the bounds satisfy*

$$\sup_{a<x_0<x_1<\ldots<x_{n-1}<b} |[x_0, x_1, \ldots, x_{n-1}]_{f'}| = n \sup_{a<x_0<x_1<\ldots<x_n<b} |[x_0, x_1, \ldots, x_n]_f|$$

Proof. The bounds follow rather immediately from the identities 5.33 and 5.34, so it only remains to verify that $f \in C^1$ given the conditions. Since the n'th divided difference corresponds to n'th derivative, if it is bounded, $(n-1)$'th derivative should be continuous. Thus we should prove that this is indeed the case by proving that $(n-1)$'th divided differences of f extend continuously to the whole of $(a,b)^n$.

Note that lemma 5.35 immediately implies that $(n-1)$'th divided difference of f is Lipschitz. But Lipschitz functions can be always extended as Lipschitz functions, so we are done by lemma 5.38. □

Theorem 5.41. *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. Again, simply apply lemma 5.40 inductively. □

Finally, one can carry positivity.

Lemma 5.42. *Let $f : (a,b) \to \mathbb{R}$ and $k \geq 3$. Then f is k-tone, if and only if $f' \in C^1(a,b)$ and f' is $(k-1)$-tone.*

Proof. Again, only the claim on the regularity is non-trivial as the k-tone claim follows easily from 5.33 and 5.34. As with the bounded case the idea is that if f is k-tone $f^{(k)}$ is positive and hence $f^{(k-1)}$ is increasing, and consequently locally bounded. We should hence prove that the $(n-1)$'th divided differences are bounded, as then 5.40 would imply the claim. But this follow easily from 5.35. □

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

Proof of the theorem 5.31. Yet again, simply apply lemma 5.42 inductively. □

5.4 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 5.43. *Let $f : (a,b) \to \mathbb{R}$. Then f is real analytic, if and only if for every closed subinteval $[c,d]$ of (a,b) there exists constant C such that for any $n \geq 1$*

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

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

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

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

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

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

which may be estimated by

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

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

Of course, we could just as well replace the closed inteval by any compact subset of (a,b). The previous result is some kind of relative of 5.41. Also theorem 5.31 has rather interesting relative.

Theorem 5.44 (Bernstein's little theorem)**.** *If $f : (a,b) \to \mathbb{R}$ is k-tone for every $k \geq 0$, then f is real-analytic on (a,b).*

Proof. We prove that the conditions of the theorem 5.43 are satisfied. Pick any $a < x_0 < x < b$. Now for any $n \geq 0$ we have

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

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

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

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

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 $(-1)^n f^{(n)}(t) \geq 0$ for any $t > 0$, then f extends to analytic function to right half-plane. This is readily implied by the previous observation. Latter version has however a considerable strenghtening.

Theorem 5.45 (Bernstein's big theorem). *If $f : (0, \infty) \to \mathbb{R}$ is smooth such that $(-1)^n f^{(n)}(t) \geq 0$ for any $t > 0$, then f is Laplace transform of a radon measure μ on $[0, \infty)$, that is we have*

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

for every $x > 0$.

We will postpone the proof.
TODO:

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

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

- Harnack-type inequalities for derivatives of Pick functions?
- Smooth function is in $P(0, \infty)$ if it's negative of Laplace transform of Laplace transform of a measure on $[0, \infty)$?
- Are there better bounds for theorem 5.41?

Chapter 6

Pick-Nevanlinna functions

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

6.1 Examples and basic properties

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.

As one might have guessed by now, Pick functions are our set of "positive analytic functions".

Theorem 6.1. *$\mathcal{P} \subset \{$analytic maps on $\mathbb{H}_+\}$ is a proper cone.*

Proof. Again, since the evaluation functionals are continuous, we are essentially left to check that the fourth condition, i.e. we should prove that

$$\Im(\varphi) = 0 \Leftrightarrow \varphi = 0.$$

And again, this isn't quite true, even for analytic functions. Imaginary part being constant merely implies that function is a real constant. Good enough. $\square$

Again, one could instead consider the quotient space, analytic maps up to a constant, and we would have a proper cone, but this kind of thinking isn't bringing much to the discussion.

So far we have made no mention on the topology, as it's usually taken to be the topology of locally uniform convergence. This definitely works (as it makes the evaluation functionals continuous), but we can do much better. It namely turns out that we can consider the set of Pick functions as a proper cone of $\mathbb{C}^{\mathbb{H}_+}$, set of all functions, with product topology.

Proposition 6.2. *If $(\varphi_i)_{i=1}^{\infty}$ is a sequence of Pick functions converging pointwise, the limit function is also a Pick function.*

This result far from clear: pointwise limits of analytic functions need not in general be analytic. We will not prove the result yet, but it strongly suggests that there is something more going on; Pick functions are very rigid. Note also that if Pick functions are thought of a subset of all functions, the definition of the cone doesn't really fit the general framework of theorem 2.6. This suggests that we are missing some better functionals, or better predual.

6.2 Boundary

To understand the rigidity phenomena we take look at a close relative to Pick functions, *Schur functions*. Schur funtions are analytic maps from open unit disc to closed unit disc. Classic fact about these functions is the Schwarz lemma.

Theorem 6.3 (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 $\omega \in \mathbb{S}$.*

The textbook proof is based on two observations about analytic functions.

- If φ is analytic at a with $\varphi(a) = 0$, then $\varphi/(\cdot - a)$ is also analytic.
- If φ is analytic on closed unit disc and $|\varphi| \leq 1$ on the boundary of the disc, then $|\varphi| \leq 1$ inside the disc.

The first observation might not be very surprising, and it holds for smooth functions also. The second, on the other hand, is a true manifestation of the nature of the analytic maps: we can bound analytic functions simply by bounding them on the boundary of the domain. More generally, one knows everything about an analytic function on a domain simply by knowing it on a boundary, by Cauchy's integral formula.

This suggests that we should be able to recognize also Pick functions looking only at their boundary values. Actually even more is true: it suffices to look at the imaginary parts.

Proposition 6.4. *Let $\varphi : U \to \mathbb{C}$ be analytic, such that $\overline{\mathbb{H}_+} \subset U$, and φ is continuous at ∞. Then if the imaginary part of φ is non-negative on the real axis, φ is Pick function.*

Proof. This follows immediately from the minimum principle applied to the harmonic function $\Im(\varphi)$. □

6.3 Integral representations

Recall that imaginary part of an analytic function determines also its real part, up to a constant, so we can also recover the function itself. This can be also done explicitly.

Theorem 6.5. *Let $\varphi : U \to \mathbb{C}$ be analytic, such that $\overline{\mathbb{H}_+} \subset U$, and $\varphi(z) = O(|z|^{-\varepsilon})$ for some $\varepsilon > 0$ at infinity. Then for any $z \in \mathbb{H}_+$ we have*

$$\varphi(z) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\Im(\varphi)(\lambda)}{\lambda - z} d\lambda$$

Proof. Note that the integral defines an analytic function, imaginary part of which equals

$$\frac{\Im(z)}{\pi}\int_{\mathbb{R}}\frac{\Im(\varphi)(\lambda)}{(\lambda-z)(\lambda-\overline{z})}d\lambda.$$

This expression however equals $\Im(\varphi(z))$ by Poisson integral formula. By letting $z\to\infty$ one sees that also the real constants match.

Alternatively one could observe that for a closed counter clockwise oriented curves γon the upper half-plane, enclosing z, we have

$$\varphi(z)=\frac{1}{2\pi i}\int_{\gamma}\frac{\varphi(\lambda)}{\lambda-z}d\lambda.$$

Now given the bound, we may deform the contour to real axis. By comparing this identity and our goal, we are left to prove that

$$\frac{1}{2\pi i}\int_{\gamma}\frac{\overline{\varphi(\lambda)}}{\lambda-z}d\lambda=\frac{1}{2\pi i}\overline{\int_{\gamma}\frac{\varphi(\lambda)}{\lambda-\overline{z}}d\lambda}.$$

But this is clear as $\varphi/(\cdot-\overline{z})$ is analytic in the upper half-plane. □

Compare this with theorem 5.15: in both cases we can express a (regular enough) positive element (k-tone functions and Pick functions) as a linear combination of some kind of basis elements.

There's of course nothing really special about the decay assumption $\varphi(z)=O(|z|^{-\varepsilon})$; it's there just to make everything converge.

One can guarantee the convergence also by other means. Note that the integrand behaves like $\frac{1}{\lambda-z}$, if we subtract something from it something behaving the same way at the infinity (something not depending on z), we ought to improve convergence, but only change the value of the function by a constant. As an example, consider the integral

$$\frac{1}{\pi}\int_{\mathbb{R}}\left(\frac{1}{\lambda-z}-\frac{[|\lambda|>1]}{\lambda}\right)\Im(\varphi)(\lambda)d\lambda. \tag{6.6}$$

It converges to an analytic function as long as, say, $\Im(\varphi)$ is bounded. As before, its imaginary part coincides with φ's so the functions equal up to a real constant. Now it's not clear however that the functions should equal and indeed they need not: the right-hand side doesn't see real constants.

Note that the previous idea could be used to construct Pick functions. Everything still makes sense if we replace $\Im(\varphi)$ by some other positive function, as long as the integral converges. Heck, we could replace it by any positive measure for which $\mu((\lambda^2+1)^{-1})<\infty$.

(Almost) all the examples given before are actually just special cases of this construction. The rational functions $\frac{1}{\lambda - z}$, where $\lambda \in \mathbb{R}$ are obtained by setting $\mu = \delta_\lambda$. The power functions are obtained as

$$\begin{aligned} z^p &= c_0 + \frac{1}{\pi}\int_{-\infty}^{0}\left(\frac{1}{\lambda - z} - \frac{1}{\lambda - 1}\right)\Im(\lambda^p)d\lambda \\ &= c_0 + \frac{1}{\pi}\int_{-\infty}^{0}\left(\frac{1}{\lambda - z} - \frac{1}{\lambda - 1}\right)|\lambda|^p \sin(\pi p)d\lambda, \end{aligned}$$

for some constant c_0 (which can be seen to be 1 by setting $z = 1$). Logarithm is even simpler:

$$\log(z) = \int_{-\infty}^{0}\left(\frac{1}{\lambda - z} - \frac{1}{\lambda - 1}\right)d\lambda.$$

Tangent function could be obtained by putting δ-measures to the points of the form $\frac{\pi}{2}+n\pi$, where $n \in \mathbb{Z}$, the singularities of tangent.

The only exception is the function $z \mapsto \alpha z$ – it can't be expressed as such integral. But even this failure is really more about poor point of view, as we will see in a minute. With these observations in mind it ought to be not too surprising that we have the following.

Theorem 6.7. *$\varphi \in \mathcal{P}$, if and only*

$$\varphi(z) = \alpha z + \beta + \int_{-\infty}^{\infty}\left(\frac{1}{\lambda - z} - \frac{\lambda}{\lambda^2 + 1}\right)d\mu(\lambda) \tag{6.8}$$

for some $\alpha \geq 0$ and $\beta \in \mathbb{R}$ and a Borel measure μ with $\int_{-\infty}^{\infty}(\lambda^2+1)^{-1}d\mu(\lambda) < \infty$.

Choosing $\lambda \mapsto \frac{\lambda}{\lambda^2+1}$ is common choice in the literature and is convenient as

- It's real. This is really not too big of a deal though, but it aids in keeping track with the real and imaginary parts.
- We may recover the constant β as $\Re(\varphi(i))$.

To better explain the appearance of the linear term, we can write the integral in a sligtly different form. Denoting $d\nu(\lambda) = \frac{d\mu(\lambda)}{\lambda^2+1}$, the formula reads

$$\varphi(z) = \alpha z + \beta + \int_{-\infty}^{\infty}\frac{\lambda z + 1}{\lambda - z}d\nu(\lambda).$$

Here ν is just a finite Borel measure. Now it kind of makes sense to extend the domain of this measure to infinity: the linear term merely corresponds to δ-measure at infinity point. Of course, should one formalize this line of thought, the question on the type of extended real line had to be asked and one should address the topology. The answer is that one should glue the real line into a circle. One shouldn't worry about such issues, though, as these thoughts are here merely for intuition, at least for know. The giveaway is that α should be really thought as a part of the measure μ, even though this might not make perfect sense.

We will not prove theorem 6.7 yet, but it shall work as a motivation. In order to understand Pick functions, we should understand their boundaries.

We will call the family

$$\{z \mapsto z\} \cup \{z \mapsto \frac{1}{\lambda - z} | \lambda \in \mathbb{R}\}$$

extreme Pick functions. Finite linear combinations of the extreme Pick functions are called **simple Pick functions**. Finally, we will call a Pick function extentable, if it is bounded and analytically extends over the real line. Such Pick functions enjoy representations of the form 6.6.

6.4 Pick functionals

Question 6.9. How can we recover the measure μ from the Pick function?

If μ is given by a continuous, bounded function, i.e. $d\mu(\lambda) = f(\lambda)d\lambda$, it's not very hard to see that

$$f(\lambda) = \frac{1}{\pi} \lim_{y \to 0^+} \Im(\varphi(\lambda + iy)).$$

This doesn't however work with rational functions with poles on the real line. One might try to salvage the situation by saying that poles should correspond to δ-measures, but even if that would be true in some sense, we are only scratching the surface. What to do if the measure is for instance the uniform measure on the Cantor set?

Beauty of the measure theory is of course that we don't even need to make sense of the measure pointwise; everything is hidden in the averages. Could we recover the measures of open intervals then? Is it true that

$$\mu((a, b)) = \lim_{y \to 0^+} \frac{1}{\pi} \int_a^b \Im(\varphi(\lambda + iy))d\lambda?$$

Even this isn't quite true: the problem is that if μ contains δ-measure at a or at b, the right-hand side doesn't see this properly. It turns out that this is only problem though.

We have however already encountered much better averages: the imaginary part of a Pick function φ is a weighted average of the imaginary parts of φ on the real line. We only proved this in the case of bounded φ, and indeed, the proper generalization should be

$$\Im(\varphi(z)) = \alpha\Im(z) + \frac{\Im(z)}{\pi}\int_{\mathbb{R}} \frac{d\mu(\lambda)}{(\lambda - z)(\lambda - \overline{z})}.$$

Bear in mind that we also consider the constant α be part of the measure. One can take this idea much further: if q is any rational function with simple poles, no poles on $\mathbb{R}$, and decay $O(|z|^{-2})$ at infinity, the expression

$$\int_{\mathbb{R}} q(\lambda) d\mu(\lambda)$$

makes sense. Even better, partial fraction expansion allows us to write this integral in terms of as a linear combination of values of φ and its conjugate. Indeed, partial fraction decomposition allows us to write q as

$$q(\lambda) = c_0 \frac{\lambda}{\lambda^2 + 1} + \sum_{i=1}^{M} c_i \left(\frac{1}{\lambda - a_i} - \frac{\lambda}{\lambda^2 + 1} \right),$$

where a_i's are the poles of q. The decay condition tells that $c_0 = 0$. If $\Im(a_i) > 0$, term corresponds to multiple of $\varphi(a_i)$, and if $\Im(a_i) < 0$, we get the conjugate. Explicitly, if we abuse notation a tad by writing $\varphi(a_i) = \overline{\varphi(\overline{a_i})}$ if $\Im(a_i) < 0$, we have

$$\begin{aligned}
\int_{\mathbb{R}} q(\lambda) d\mu(\lambda) &= \int_{\mathbb{R}} \sum_{i=1}^{M} c_i \left(\frac{1}{\lambda - a_i} - \frac{\lambda}{\lambda^2 + 1} \right) \mu(\lambda) \\
&= \sum_{i=1}^{M} c_i \left(\varphi(a_i) - \alpha a_i - \beta \right),
\end{aligned}$$

which rewrites to

$$\sum_{i=1}^{M} c_i \varphi(a_i) = \sum_{i=1}^{M} c_i \left(\alpha a_i + \beta \right) + \int_{\mathbb{R}} q(\lambda) d\mu(\lambda).$$

We can get some kind of glimpse of the measure just by looking at linear combinations of the values of Pick function.

These ideas get really useful when q is non-negative on the real line. It follows easily from the lemma 4.22 that such rational functions can be written in the form

$$q(\lambda) = \left(\sum_{i=1}^{n} \frac{c_i}{\lambda - \lambda_i} \right) \left(\sum_{i=1}^{n} \frac{\overline{c_i}}{\lambda - \overline{\lambda_i}} \right)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}\backslash\mathbb{R}$ and $c_1, c_2, \ldots, c_n \in \mathbb{C}$. Running through the same calculations we see that

$$\sum_{1\le i,j\le n} c_i\overline{c_j}\frac{\varphi(\lambda_i) - \overline{(\varphi(\lambda_j))}}{\lambda_i - \overline{\lambda_j}} = \alpha\left|\sum_{i=1}^{n} c_i\right|^2 + \int_{\mathbb{R}} q(\lambda)d\mu(\lambda) \ge 0.$$

Definition 6.10. Functional in $\left(\mathbb{C}^{\mathbb{H}_+}\right)^*$ of the form

$$\sum_{1\le i,j\le n} c_i c_j \frac{\delta_{\lambda_i} - \overline{\delta_{\lambda_j}}}{\lambda_i - \overline{\lambda_j}},$$

$\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}\setminus\mathbb{R}$ and $c_1, c_2, \ldots, c_n \in \mathbb{C}$, is called a **Pick functional**. This is all of course to say that

$$\varphi \mapsto \sum_{1\le i,j\le n} c_i c_j \frac{\varphi(\lambda_i) - \overline{\varphi(\lambda_j)}}{\lambda_i - \overline{\lambda_j}}.$$

Given $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}\setminus\mathbb{R}$, the matrix

$$\begin{bmatrix} [\lambda_1, \overline{\lambda_1}]_\varphi & [\lambda_1, \overline{\lambda_2}]_\varphi & \cdots & [\lambda_1, \overline{\lambda_n}]_\varphi \\ [\lambda_2, \overline{\lambda_1}]_\varphi & [\lambda_2, \overline{\lambda_2}]_\varphi & \cdots & [\lambda_2, \overline{\lambda_n}]_\varphi \\ \vdots & \vdots & \ddots & \vdots \\ [\lambda_n, \overline{\lambda_1}]_\varphi & [\lambda_n, \overline{\lambda_2}]_\varphi & \cdots & [\lambda_n, \overline{\lambda_n}]_\varphi \end{bmatrix}$$

is called a **Pick matrix**. Pick functionals are simply values of quadratic forms of Pick matrices.

We denote the set of Pick functionals by $\mathcal{P}^*$.

Theorem 6.11. *Let $p^* \in \left(\mathbb{C}^{\mathbb{H}_+}\right)^*$. Then the following are equivalent.*

(i) $p^* \in \mathcal{P}^*$

(ii) $p^*(\varphi) \ge 0$ *for any extreme Pick function* φ.

(iii) $p^*(\varphi) \ge 0$ *for any extentable Pick function* φ.

(iv) $p^*(\varphi) \ge 0$ *for any Pick function* φ.

Proof. The implications $(i) \Rightarrow (ii) \Rightarrow (iii)$ and $(iv) \Rightarrow (ii)$, up to one detail: if $p^*(\varphi) \geq 0$ for any extreme Pick function φ, we should prove that $p^*(1) = 0$. But follows as soon as one notes that

$$p^*\left(\frac{1}{\lambda - \cdot}\right) = \frac{p^*(1)}{\lambda} + O\left(\frac{1}{\lambda^2}\right).$$

$(iii) \Rightarrow (iv)$: It suffices to prove that the extentable Pick functions are dense (with respect to the topology of Pointwise convergence) in the set of all Pick function. But for this it is enough to find a sequence of Pick functions $(g_n)_{n=1}^{\infty}$ such that

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

as then we have $\varphi \circ g_n \to \varphi$ pointwise as $n \to \infty$ for every Pick function φ the functions $\varphi \circ g_n$ are evidently bounded and extend analytically over the real line.

It is not very hard to check that we may take

$$g_n(z) = \frac{z + \frac{i}{n}}{1 - \frac{iz}{n}}.$$

$(ii) \to (i)$: By the construction of the $\mathcal{P}^*$ it contains all the functionals **with finite support**, which give positive values for all extreme Pick functions. Thus it remains to be noted that all the continuous functionals in $\left(\mathbb{C}^{\mathbb{H}_+}\right)^*$ have finite support. But this follows from the general fact that the dual of a product equals direct sum of the duals. $\square$

It is useful to note that one does not even need to test every extreme Pick function to check that functional is Pick functional, dense subset suffices. This is clear for the functions $z \mapsto \frac{1}{\lambda - z}$ but also holds for $z \mapsto z$, when this is intepreted as the function with $\lambda = \infty$ (and dense subset refers to the circle topology). Indeed, we have

$$p^*\left(\frac{1}{\lambda - \cdot}\right) = p^*\left(\frac{1}{\lambda - \cdot} - \frac{1}{\lambda}\right) = \frac{p^*(\cdot)}{\lambda^2} + O\left(\frac{1}{\lambda^3}\right).$$

Finally, the following correspondence should be clear by now.

Lemma 6.12. *There is a natural linear bijective linear map between non-negative rational functions vanishing at infinity and $\mathcal{P}^*$.*

6.5 Weakly Pick functions

Theorem 6.11 implies that$\mathcal{P}^*$ is really just the dual cone of $\mathcal{P}$. It turns out that the "converse" is also true: $\mathcal{P}^*$ is also a predual of $\mathcal{P}$.

Definition 6.13. We will elements of $(\mathcal{P}^*)^*$ **weakly Pick functions**.

In layman's terms, weakly Pick functions are ones, which look like Pick functions if one only considers linear functionals. The aim of this section is to show that weakly Pick functions are exactly the Pick functions. We already proved one direction in the theorem 6.11.

The other direction is tricky. Note that weakly Pick maps map to the upper half-plane so the interesting part is to prove that weakly Pick maps are analytic. For this we are going to verify bounds for the divided differences of φ. Recall that by theorem 5.41 it suffices to verify that the order 2 divided differences are locally bounded to prove that φ is (continuously) differentiable. Strictly speaking we only proved the result on real line, but the prove would be almost identical in the complex case.

The idea is the following: we are going to formulate everything terms of the linear functionals. This idea is best illustrated with an example.

Lemma 6.14 (Harnack inequality)**.** *Let φ be a weakly Pick function. Then for every compact $K \subset \mathbb{H}_+$ there exists a constant C_K such that*

$$\frac{\Im(\varphi(z))}{\Im(z)} \leq C_K \frac{\Im(\varphi(w))}{\Im(w)}$$

for every $z, w \in K$.

Proof. Note that the sought inequality can be rephrased as positivity of the linear functional

$$\varphi \mapsto C_K \frac{\Im(\varphi(w))}{\Im(w)} - \frac{\Im(\varphi(z))}{\Im(z)}.$$

By theorem 6.11 it suffices to prove that there exists constant C_K such that the previous inequality holds for any extreme Pick function. This implies that we should have

$$\frac{1}{|\lambda - z|^2} \leq \frac{C_K}{|\lambda - w|^2}$$

for every $\lambda \in \mathbb{R}$. But

$$\left|\frac{\lambda - w}{\lambda - z}\right|^2 \leq \left|1 + \frac{z - w}{\lambda - z}\right|^2 \leq \left(1 + \frac{|z - w|}{\Im(z)}\right)^2,$$

so we can definitely find such constant. $\square$

Similarly, one can prove that weakly Pick functions are continuous.

Theorem 6.15. *Let φ be a weakly Pick function. Then φ is continuous.*

Proof. Our aim is to bound the divided difference $|[z, w]_\varphi|$. Now the problem is that this expression is not linear in the function anymore. There's a way to fix this problem however: we bound $\Re(\omega[z, w]_\varphi)$ for $\omega \in \mathbb{S}$. This expression is linear in the function, and we have

$$|z| \leq C \Leftrightarrow \Re(\omega z) \leq C \text{ for every } \omega \in \mathbb{S}.$$

Observe that

$$\Re\left(\frac{\omega}{(\lambda - z)(\lambda - w)}\right) \leq \frac{1}{|\lambda - z||\lambda - w|} \leq \frac{1}{2}\left(\frac{1}{|\lambda - z|^2} + \frac{1}{|\lambda - w|^2}\right)$$

for every $\omega \in \mathbb{S}$. It follows that

$$|[z, w]_\varphi| \leq \frac{1}{2}\left(\frac{\Im(\varphi(z))}{\Im(z)} + \frac{\Im(\varphi(w))}{\Im(w)}\right)$$

for any weakly Pick function. Combining this with Harnack inequality 6.14 yields that any weakly Pick function is locally Lipschitz, so in particular continuous. □

The previous argument can be easily extended to the following.

Theorem 6.16. *Let φ be a weakly Pick function. Then for any $n \geq 1$ and $z_0, z_1, \ldots, z_n$ we have*

$$|[z_0, z_1, \ldots, z_n]_\varphi| \leq \frac{1}{2\Im(z_2)\Im(z_3)\ldots\Im(z_n)}\left(\frac{\Im(\varphi(z_0))}{\Im(z_0)} + \frac{\Im(\varphi(z_1))}{\Im(z_1)}\right).$$

In particular any weakly Pick function is analytic and hence a Pick function.

Proof. Simply note that

$$\Re\left(\frac{\omega}{(\lambda - z_0)(\lambda - z_1)\cdots(\lambda - z_n)}\right) \leq \frac{1}{\Im(z_2)\Im(z_3)\ldots\Im(z_n)}\frac{1}{|\lambda - z_0||\lambda - z_1|}$$

and follow the argument in the proof of theorem 6.15. □

It is worthwhile to note that as one really only needs to get bound for order 2 divided differences in the proof of 6.16, one only needs to keep track of small family of pick functionals, in particular ones with support of at most 3 points. This observation is known as the Hindmarsh theorem.

Corollary 6.17. *$\varphi : \mathbb{H}_+ \to \mathbb{C}$ is weakly Pick, if and only if it Pick function.*

Proof of theorem 6.2. This follows immediately from 6.17. □

6.6 Pick-Nevanlinna extrapolation theorem

There's a remarkable generalization to the theorem 6.17.

Definition 6.18. Let $X \subset \mathbb{H}_+$. We say that $\varphi : X \to \mathbb{C}$ is weakly Pick on X if $p^*(\varphi) \geq 0$ for any Pick functional p^* supported on X.

Theorem 6.19. *Let $U \subset \mathbb{H}_+$ be ope and assume that φ is weakly Pick on U. Then there exists a unique pick function $\tilde{\varphi}$ such that $\tilde{\varphi}\big|_U = \varphi$.*

The proof is based on the following lemma.

Lemma 6.20. *Let $U \subset \mathbb{H}_+$ be open and assume that φ is weakly Pick on U. Let $z_0 \in U$. Then there exists unique weakly Pick $\tilde{\varphi} : U \cap \mathbb{D}(z_0, \Im(z_0))$ such that $\tilde{\varphi}\big|_U = \varphi$.*

Proof. Take any sequence $z_1, z_2, \ldots$ converging to z_0. By 6.16 the Newton series at converges in $\mathbb{D}(z_0, \Im(z_0))$. We claim that this series gives the (necessarily unique) extension for φ to $\mathbb{D}(z_0, \Im(z_0)) \setminus U$.

To this end take any Pick functional p^* supported on $\mathbb{D}(z_0, \Im(z_0)) \cup U$ and apply it to our $\tilde{\varphi}$. The functional correponds to some non-negative rational function r. Now if we replace all the evaluations of p^* at $\mathbb{D}(z_0, \Im(z_0)) \setminus U$ by truncation of Newton series, we can interpret the result as a new linear functional, say p_N^*. The corresponding rational function is also changed (say to r_N): all the terms of the form $\frac{1}{\lambda - w_0}$ for $w_0 \in \mathbb{D}(z_0, \Im(z_0)) \setminus U$ are replaced by

$$\frac{1}{\lambda - z_1} + \frac{(w_0 - z_1)}{(\lambda - z_1)(\lambda - z_2)} + \ldots + \frac{(w_0 - z_1) \cdots (w_0 - z_{N-1})}{(\lambda - z_1) \cdots (\lambda - z_N)},$$

where N is the point of truncation. Difference between these rational functions can be easily bounded by

$$\left(\frac{|w_0 - z_0|}{\Im(z_0)} \right)^N \frac{C}{|\lambda - z_0|^2},$$

where C is some constant not depending on N. But this means that r_N can't be too small, as r was non-negative to begin with. Indeed, by summing over all the evaluations of p^* at $\mathbb{D}(z_0, \Im(z_0)) \setminus U$, we see that

$$r_N \geq -\frac{C'}{|\lambda - z_0|^2} \rho^N$$

for some $\rho < 1$ and $C' > 0$ (again, not depending on N). It follows that

$$p^*(\tilde{\varphi}) = \lim_{N \to \infty} p_N^*(\tilde{\varphi}) \geq \lim_{N \to \infty} -C' \frac{\Im(\varphi(z_0))}{\Im(z_0)} \rho^N = 0,$$

hence the claim. $\square$

Proof of theorem 6.19. Consider all weakly Pick extensions of φ (to open supersets of U), ordered by restriction. These maps trivially satisfy conditions of Zorn's lemma so we may Pick maximal map, $\tilde{\varphi}$. It follows immediately from lemma 6.20 that the domain of $\tilde{\varphi}$ is the whole $\mathbb{H}_+$.

Of course, Zorn's lemma is not really necessary here: one could write explicit extension scheme (TODO: picture). □

6.7 Notes and references

TODO:

- Poincaré metric: discs are discs, Apollonius circle
- Examples of representing measures behind functions and functions behind representing measures
- Spectral commutant lifting theorem
- Use Morera's theorem to prove weak Hindmarsh's theorem
- Nice formula for finite Pick extension (rational function case)

How to rewrite majority of this section:

- Concentrate on Pick-integrals
- Proof open Pick interpolation by extending thinking about Pick-integrals and then showing that the extension is unique, thus showing that everything works.
- Split the Pick measure weakly with $1/(x^2+1)$ (How to do the same thing with k-tone functions?)

Chapter 7

Matrix k-tone functions

7.1 Matrix convex functions

Having charaterized matrix monotone functions it is natural to ask what happens with convex functions.

Definition 7.1. We say that $f : (a, b) \to \mathbb{R}$ is matrix convex of order n if for every $A, B \in \mathcal{H}^n(V)$ and $t \in [0, 1]$ we have

$$f(tA + (1 - 1)B) \leq tf(A) + (1 - t)f(B).$$

We denote class of matrix convex functions of order n on interval (a, b) by $P_n^2(a, b)$. Many of the properties of matrix monotone functions translate immediately to matrix convexity.

Proposition 7.2. *Let $(a, b) \subset \mathbb{R}$ be an open interval $n \geq 1$. Then*

1. *$P_n^2(a, b)$ is a convex closed cone.*
2. *$P_{n+1}^2(a, b) \subset P_n^2(a, b)$.*
3. *$(x \mapsto \alpha_2 x^2 + \alpha_1 x + \alpha_0) \in P_n^2(a, b)$ if $\alpha_n \geq 0$.*
4. *$(x \mapsto |x|^{-1}) \in P_n^2(a, b)$ whenever $0 \notin (a, b)$.*

Proof. 1. This is clear.

2. The proof is essentially the same as in the monotone case.

3. We have

$$\begin{aligned} &t(\alpha_2 A^2 + \alpha_1 A + \alpha_0 I) + (1-t)(\alpha_2 B^2 + \alpha_1 B + \alpha_0 I) \\ &- (\alpha_2 (tA + (1-t)B)^2 + \alpha_1 (tA + (1-t)B) + \alpha_0 I) \\ =&\alpha_2 t(1-t)(A-B)^2. \end{aligned}$$

4. It's is clearly sufficient to prove that $x \mapsto x^{-1}$ is convex on $(0, \infty)$. For this we should prove that for any $A, B > 0$ and $t \in [0,1]$ we have

$$tA^{-1} + (1-t)B^{-1} \geq (tA + (1-t)B)^{-1}.$$

By doing the congruence trick, i.e. $*$-conjugating by $A^{\frac{1}{2}}$ and setting $X = A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ we are left with

$$t + (1-t)X^{-1} \geq (t + (1-t)X)^{-1},$$

which would follow if we can prove the respective scalar inequality. But

$$t + (1-t)x^{-1} - (t + (1-t)x)^{-1} = \frac{t(1-t)}{x(t + (1-t)x)}(x-1)^2,$$

so the scalar inequality is true.

□

With linear combinations of the previous examples we can again build large number of n-convex functions. TODO

By now it ought to be no surprise that not every convex function is matrix convex. Canonical counterexample is absolute value. In turns out that we have

Proposition 7.3. *Let $v, w \in V \setminus \{0\}$. Then*

$$P_v + P_w \geq |P_v - P_w|,$$

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

Proof. As everything is happening in (at most) two dimensional subspace of V, we may assume that V is two dimensinal in the first place. Note that in this case $P_v - P_w$ has 0 trace, so its eigenvalues are additive inverses of each other. Consequently the absolute value is multiple of identity.

It follows that both sides of the inequality commute, and as they are both positive, it suffices to check when the squared inequality holds. This leads to an equivalent inequality

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

which was already discussed in 2.28. □

7.2 Convexity and the second derivative

So far everything has worked pretty much the same way as with the monotone functions, but with derivative things to start to look very different. Theorem 4.17 has natural analog in the convex case.

Theorem 7.4. *Let $n \geq 1$ and $f \in C^2(a,b)$. Then the following are equivalent:*

(i) $f \in P_n^2(a,b)$.

(ii) For any $A, H \in \mathcal{H}^n_{(a,b)}$ *we have*

$$D_n^2 f_A(H) \geq 0.$$

(iii) For any $A, H \in \mathcal{H}^n_{(a,b)}$ *and* $v \in V$ *the map*

$$t \mapsto \langle f(A+tH)v, v\rangle$$

is convex.

Now, there is a big problem here: there are only three conditions in this theorem. In the monotone case we could take H to be projection, but with the convex case this is not (a priori) possible anymore. Recall that working with projections was rather easy (we could even have very explicit formulas for everything) but with general positive maps similar arguments are hopeless. TODO

TODO

7.3 Matrix k-tone functions

After having defined the notion of k-tone function in the real setting, it is natural to ponder what happens with matrix setting. Defining the notion itself is already a bit cumbersome: with monotone and convex functions the usual definitions make immediately sense but divided differences cause some problems. One cannot simply say that f is matrix monotone if

$$[A,B]_f = \frac{f(B)-f(A)}{B-A},$$

since the right-hand side doesn't make much sense. We can however use an equivalent definition from the theorem 4.17.

7.4 Basic properties

Definition 7.5. We say that $f : (a, b) \to \mathbb{R}$ is matrix k-tone of order n if for every $A \in \mathcal{H}^n(V)$ and $B \in \mathcal{H}^n_+(V)$ and $v \in V$ the function

$$t \mapsto \langle f(A + tB)v, v\rangle$$

is k-tone.

Denote the class of matrix k-tone functions of order n on interval (a, b) as $P^k_n(a, b)$ (so $P^1_n(a, b) = P_n(a, b)$).

This definition doesn't exactly coincide with our definition for matrix convex functions, where we needed no assumption on the "sign" of B. As we will later see, however, this alternate definition leads to same set of functions.

As in the monotone case, we can list many natural properties of classes $P^k_n(a, b)$, proofs of which are very similar to the monotone case.

Proposition 7.6. *Let $(a, b) \subset \mathbb{R}$ be an open interval $n \geq 1$, and $k \geq 1$. Now*

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

Proof. TODO □

Not surprisingly, we have also the following derivative characterization.

Theorem 7.7. *Let $n, k \geq 1$ and $f \in C^k(a, b)$. Then the following are equivalent:*

(i) *$f \in P^k_n(a, b)$.*

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

$$D^k_n f_A(H) \geq 0.$$

Proof. TODO □

But there's a problem: there's no obvious way to change the H in the statement of the previous theorem to one-dimensional projection, when $k > 1$. The issue is that when $k > 1$, the map

$$H \mapsto D_n^k f_A(H)$$

is not linear anymore! It's a horrible mess instead.

TODO

- Is this section really needed?
- How to deal with smoothness issues cleanly?

Chapter 8

Trace functions

8.1 Absolute Value

As adjoint behaves as conjugate, it would be natural to guess that

$$|A| := (A^*A)^{\frac{1}{2}} \, ,$$

absolute value of a map, would have many similar properties as the standard absolute value.

The following list of properties of the absolute value make it clear that this is indeed good definition.

- $|A| \geq 0$ for any $A \in \mathcal{L}(V)$ and $|A| = A$, if and only if $A \geq 0$.
- For any $A \in \mathcal{H}(V)$ we have $-|A| \leq A \leq |A|$, or equivalently $|Q_A(v)| \leq Q_{|A|}(v)$ for any $v \in V$
- For any $v \in V$ we have $\|Av\| = \||A|v\|$.

Note that in general we have $|A| \neq |A^*|$, and maps need not even go between the same spaces.

One might be tempted to think that we have triangle inequality, i.e.

$$|A + B| \leq |A| + |B|,$$

for any $A, B \in \mathcal{L}(V)$, or at least $A, B \in \mathcal{H}$. Such inequality doesn't hold, but it's not that far from being true. Like in the real case, one would like to add

$$-|A| \leq A \leq |A| \quad \text{and} \quad -|B| \leq B \leq |B|,$$

to get

$$-(|A| + |B|) \leq A + B \leq |A| + |B|.$$

The problem is that we can't make any further conclusions: just because $-Y \leq X \leq Y$, it is not necessarily the case that $|X| \leq Y$. Thinking in quadratic forms we get the inequality

$$|Q_{A+B}(v)| \leq Q_{|A|+|B|}(v), \tag{8.1}$$

for any $v \in V$, but this does not imply that $Q_{|A+B|}(v) \leq Q_{|A|+|B|}(v)$. Indeed $|Q_{A+B}(v)| \leq Q_{|A+B|}(v)$, as we noticed, so the inequality is going to the wrong direction. If however v is an eigenvector of $A + B$, we have $|Q_{A+B}(v)| = Q_{|A+B|}(v)$, and it follows that

$$Q_{|A+B|}(v) \leq Q_{|A|+|B|}(v)$$

holds for eigenvectors v of $|A + B|$. Summing over the eigenvector we see that

$$\mathrm{tr}|A + B| \leq \mathrm{tr}|A| + \mathrm{tr}|B|,$$

so instead of the full inequality, we get inequality for traces. There is a nice generalization for the previous we'll get back to.

Chapter 9

Representations

Over the course of this thesis we have mentioned various representations results of the following form:

$$f(x) = \int h_t(x) d\mu(t),$$

or

$$f = \int h_t d\mu(t).$$

Here μ is Borel measure on some set and f and h_t's are functions of some kind. Functions h_t should be thought of some kind of basis functions. Although such results have been hardly used, one cannot just leave them unmentioned.

Most of the representation results in the thesis can be understood in terms of Choquet theory. The idea is the following: the sets we are concerned with are convex and the functions h_t the extreme points of these convex set. Extreme points are the points that can't be expressed as a non-trivial linear combination of points in the set.

Now if one has say compact convex set in $K \subset \mathbb{R}^n$, K should be roughly given by its boundary: compact convex sets are equal if and only they have same boundary. But more is true: every point in K can be expressed as a convex combination of extreme points of K (actually of at most $n + 1$ extreme points).

Same holds true much more generally, in infinite dimensional spaces.

We have already discussed this kind of phenomenon: we noted that every Pick function is essentially sum of functions of the form $z \mapsto \frac{1}{\lambda - z}$. Such functions (together with affine ones) are the extreme points of the cone of Pick functions. Similarly we found basis functions for k-tone functions: they are exactly the extreme points of the k-tone functions. TODO splitting Pick measures.

9.1 Minkowski–Carathéodory Theorem

Definition 9.1. Let E be a vector space C its convex subset. A point $x \in C$ is *extreme point* of C whenever $x = ty + (1-t)z$ for $y, z \in C$ and $t \in [0,1]$ then $x = y$ or $x = z$.

Theorem 9.2 ((Weak) Minkowski–Carathéodory Theorem)**.** *Let $C \subset \mathbb{R}^n$ be convex and compact. Then it is convex hull of its extreme points, i.e. for any $x \in C$ we may find $m \geq 1$, $x_1, x_2, \ldots, x_m$ and $t_1, t_2, \ldots, t_m \geq 0$ with $t_1 + t_2 + \ldots + t_m = 1$ such that*

$$x = \sum_{i=1}^{m} t_i x_i.$$

Proof. TODO ☐

9.2 Basis k-tone functions

We noticed that k-tone functions correspond more or less to functions with non-negative k'th derivative. In other words, k-tone functions should be k-fold integrals of positive functions, at least in sufficiently smooth setting. For instance $f : (a,b) \to \mathbb{R}$ is increasing and smooth if and only if it's of the form

$$f(x) = \int_{x_0}^{x} \rho(t) dt \tag{9.3}$$

for some positive $\rho \in C^{\infty}(a,b)$ and $x_0 \in (a,b)$, up to a constant at least. For non-smooth case, we could require ρ only to be a positive L^1-function: this gives us absolutely continuous increasing functions. If we further drop ρ but replace the Lebesgue measure by an arbitrary Radon measure μ, we get every right-continuous increasing function. Measuretheoretically these are already all the increasing functions, although we miss some functions like $\chi_{(0,\infty)}$.

If $\mu = \delta_0$, for instance, $f = \chi_{[0,\infty)}$. One could think that δ_0 gives a jump for f at 0. More generally, if μ is positive linear combination if m (distinct) Dirac deltas, f is a function with m jumps. Now every Radon measure is a weak limit of positive linear combination Dirac deltas, so every increasing function is limit of finite sums of jump functions, at least in some weak sense.

This fact is actually contained in 9.3: we may rewrite

$$f(x) = \int_{a}^{b} \chi_{[t,\infty)}(x) d\mu(t),$$

f is essentially sum of functions of the form $\chi_{[t,\infty)}$, again up to a constant. We will call those the basis functions for

The point is: whenever something holds for any step function, it should hold for any increasing function. In this context by "something" I mean linear inequalities: if ν is a signed Radon measure such that for any step function $\chi_{[t,\infty)}$ we have

$$\int \chi_{[x,\infty)}(t)d\nu(t),$$

then also

$$\int f(t)d\nu(t)$$

for any increasing function. Actually we should also require that $\int d\nu(t) = 0$. I'm being deliberately vague about the domains, they don't really matter too much.

Things get much more interesting when we move to k-tone functions of higher order. For k-tone functions, i.e. convex functions we can make similar statements.

We can write any (smooth enough) convex function in the form

$$f(x) = \int_{x_0}^{x} \int_{x_0}^{x_1} \rho(t)dtdx_1,$$

at least up to a constant and linear term. By simple partial integration this can be rewritten as

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

or even, better, as

$$f(x) = \int_{a}^{b} (x-t)_+\rho(t)dt,$$

where $(x-t)_+$ denotes $\max(0, x-t)$. What this means is that the functions $(\cdot - t)_+$ work as a basis functions for convex functions, up to a affine term. By affine transformation we could equivalently take the functions of the form $|\cdot - t|$ as a basis functions.

Now if a linear equality holds for functions of the form $|x-t|$, it holds for any convex function. So since for any $x_1, x_2, \ldots, x_m \in \mathbb{R}$ we have

$$\sum_{1\le i\le m} |x_i - t| \ge m \left| \frac{\sum_{1\le i\le m} x_i}{m} - t \right|,$$

also for any convex function

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

Jensen's inequaltity.

Bibliography

[1] T. Ando. Majorization, doubly stochastic matrices, and comparison of eigenvalues. *Linear Algebra and Its Applications*, 118:163–248, 1989.

[2] T. Ando. Parameterization of minimal points of some convex sets of matrices. *Acta Scientiarum Mathematicarum (Szeged)*, 57:3–10, 1993.

[3] T. Ando. Majorizations and inequalities in matrix theory. *Linear algebra and its Applications*, 199:17–67, 1994.

[4] J. Antezana, E. R. Pujals, and D. Stojanoff. Iterated aluthge transforms: a brief survey. *Revista de la Unión Matemática Argentina*, 49(1):29–41, 2008.

[5] B. C. Arnold. Majorization: Here, there and everywhere. *Statistical Science*, pages 407–413, 2007.

[6] M. F. Atiyah. Convexity and commuting Hamiltonians. *Bull. London Math. Soc.*, 14(1):1–15, 1982.

[7] D. Azagra. Global approximation of convex functions. *arXiv preprint arXiv:1112.1042*, 2011.

[8] J. Bendat and S. Sherman. Monotone and convex operator functions. *Trans. Amer. Math. Soc.*, 79:58–71, 1955.

[9] H. Bercovici, C. Foias, and A. Tannenbaum. A spectral commutant lifting theorem. *Transactions of the American Mathematical Society*, 325(2):741–763, 1991.

[10] R. Bhatia. *Positive definite matrices.* Princeton university press, 2009.

[11] R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[12] Z. Brady. Inequalities and higher order convexity. *arXiv preprint arXiv:1108.5249*, 2011.

[13] P. Bullen. A criterion for n-convexity. *Pacific Journal of Mathematics*, 36(1):81–98, 1971.

[14] S. Burgdorf. *Trace-positive polynomials, sums of hermitian squares and the tracial moment problem*. PhD thesis, 2011.

[15] K. Cafuta, I. Klep, and J. Povh. A note on the nonexistence of sum of squares certificates for the bessis–moussa–villani conjecture. *Journal of Mathematical Physics*, 51(8):083521, 2010.

[16] E. Carlen. Trace inequalities and quantum entropy: an introductory course. *Entropy and the quantum*, 529:73–140, 2010.

[17] P. Chansangiam. A survey on operator monotonicity, operator convexity, and operator means. *Int. J. Anal.*, pages Art. ID 649839, 8, 2015.

[18] F. Clivaz. *Stahl's Theorem: Insights and Intuition on its Proof and Physical Context*. PhD thesis, 2014.

[19] F. Clivaz. Stahl's theorem (aka bmv conjecture): Insights and intuition on its proof. In *Spectral Theory and Mathematical Physics*, pages 107–117. Springer, 2016.

[20] C. de boor. *A practical guide to splines*. Springer Verlag., 1978.

[21] C. de Boor. Divided differences. *Surv. Approx. Theory*, 1:46–69, 2005.

[22] O. Dobsch. Matrixfunktionen beschränkter Schwankung. *Math. Z.*, 43(1):353–388, 1938.

[23] W. F. Donoghue, Jr. *Monotone matrix functions and analytic continuation*. Springer-Verlag, New York-Heidelberg, 1974. Die Grundlehren der mathematischen Wissenschaften, Band 207.

[24] A. Eremenko. Herbert stahl's proof of the bmv conjecture. *arXiv preprint arXiv:1312.6003*, 2013.

[25] M. Floater and T. Lyche. Two chain rules for divided differences and faa di bruno's formula. *Mathematics of Computation*, 76(258):867–877, 2007.

[26] U. Franz, F. Hiai, and E. Ricard. Higher order extension of Löwner's theory: operator k-tone functions. *Trans. Amer. Math. Soc.*, 366(6):3043–3074, 2014.

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

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

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

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

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

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

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

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

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

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

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

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

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

[40] C. R. Johnson and C. J. Hillar. Eigenvalues of words in two positive definite letters. *SIAM journal on matrix analysis and applications*, 23(4):916–928, 2002.

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

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

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

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

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

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

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

[48] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: theory of majorization and its applications*. Springer Series in Statistics. Springer, New York, second edition, 2011.

[49] R. F. Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902.

[50] H. Mulholland. On generalizations of minkowski's inequality in the form of a triangle inequality. *Proceedings of the London Mathematical Society*, 2(1):294–307, 1949.

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

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

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

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

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

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

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

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

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

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

[61] H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

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