Matrix monotone and convex functions

Otte Heinävaara

December 2, 2016

Contents

Chapter 1

Introduction

1.1 Foreword

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

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

we should also have

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

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

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

monotonic	monotonicity
increasing	?

How very inconvenient.

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

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

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

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

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

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

TODO (Brief explanations)

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

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

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

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

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

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

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

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

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

1.2 Plan of attack

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

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

1. Introduction
    - Introduction to the problem, motivation
    - Brief definition of the matrix monotonicity and convexity
    - Past and present
        - Loewner's original work, Loewner-Heinz -inequality
        - Students: Dobsch' and Krauss'
        - Subsequent simplifications and further results: Bendat-Sherman, Wigner-Neumann, Koranyi, etc.
        - Donoghue's work
        - Later proofs: Krein-Milman, general spectral theorem, interpolation spaces, short proofs etc.
        - Development of the convex case
        - Recent simplifications, integral representations
        - Operator inequalities
        - Multivariate case, other variants
        - Further open problems?
    - Scope of the thesis
2. Preliminaries (partially to be dumped to appendix?)
    - Positive matrices
        - Setup: finite (vs infinite) dimensional inner product spaces over $\mathbb{C}$ (vs $\mathbb{R}$), basic facts
        - Linear maps, adjoint, congruence, self-adjoint maps, spectral theorem: finite and infinite dimensional
        - Good properties of spectrum
        - Positive maps: basic properties (cone structure, Sylvester's criterion etc.)
    - Matrix functions
        - Several definitions

- Derivatives of matrix functions

- Divided differences: basic definition, properties, representations and smoothness
- Pick functions
  - Basic definitions and properties
  - Pick-Nevanlinna representations theorem
  - Pick matrices/ determinants
  - Compactness
  - Pick-Nevanlinna interpolation theorem
- Regularizations
  - Basic properties
  - Lemmas needed for some of the proofs

3. Monotonic and convex matrix functions

- Basics
  - Basic definitions and properties (cone structure, pointwise limits, compositions etc.)
  - Classes P_n, K_n and their properties
  - $-1/x$
  - One directions of Loewner's theorem
  - Examples and non-examples
- Pick matrices/determinants vs matrix monotone and convex functions
  - Proofs for (sufficiently) smooth functions
- Smoothness properties
  - Ideas, simple cases
  - General case by induction and regularizations
- Global characterizations
  - Putting everything together: we get original characterization of Loewner and determinant characterization

4. Local characterizations

- Dobsch (Hankel) matrix: basic properties, easy direction (original and new proof)

- Integral representations
  - Introducing the general weight functions for monotonicity and convexity (and beyond?)
  - Non-negativity of the weights
  - Proof of integral representations
- Proof of local characterizations

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

6. Loewner's theorem
   - Preliminary discussion, relation to operator monotone functions
   - Loewner's original proof
   - Pick-Nevanlinna proof
   - Bendat-Sherman proof
   - Krein-Milman proof
   - Koranyi proof
   - Discussion of the proofs
   - Convex case

7. Alternative characterizations (?)
   - Some discussion, maybe proofs

8. Bounded variations (?)
   - Dobsch' definition, basic properties
   - Decomposition, Dobsch' theorems

1.3 Some random ideas

1. It's easy to see that [Something]. Actually, it's so so easy that we have no excuse for not doing it.

2. TODO: Maximum of two matrices (at least as big), (a + b)/2 + abs(a - b)/2

Chapter 2

Preliminaries

As mentioned in the introduction, in order to understand matrix monotone and convex functions we first have to undestand positive matrices and matrix functions (and some other things).

2.1 Positive matrices

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

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

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

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

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

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

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

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

1'. 0 (zero matrix) is positive.

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

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

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

The terminology here is rather unfortunate. Natural ordering of the reals satisfies all of the above with obvius interpretation of positive numbers, which however differs from the standard definition: 0 is itself positive in our above definition. This is undoubtedly confusing, but what can you do? For real numbers we total order, so every number is either zero, strictly positive or strictly negative, so when we say non-negative, it literally means "not negative": we get all the positive numbers and zero. But with partial orders we might get more. For a cheap fix, we could say that matrix is strictly positive if it's positive but not zero. But won't do that for reasons to be explained later.

TODO (jotain vaan väliin, ehkä)

It of course depends on the ground field. It hardly makes any sense to order matrices over $\mathbb{F}_p$: even 1×1 matrices, namely (canonically) the elements of $\mathbb{F}_p$ defy reasonable ordering. But real numbers, for instance, have ordering, so there's a serious change that all real matrices could be ordered.

We will first try to order all real square matrices. (Actually, we won't even try to order non-square matrices.) 1×1 matrices are easy to order, but as soon one moves to larger matrices, one faces difficult decisions:

$$\text{Is } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \leq \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \geq \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}?$$

That's okay, we don't necessarily have to order all pairs of matrices. But there are other problems. We would like the ordering of the matrices to be independent of the choice matrix representation.

2.1.1 Hermitian matrices

Chapter 3

Pick-Nevanlinna functions

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

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

3.1 Basic properties and examples

Most obvious examples of Pick functions are of course affine functions of the form $\alpha x + \beta$ where $\alpha, \beta \in \mathbb{R}$ and $\alpha \geq 0$. Of course one could also take $\beta \in \overline{\mathbb{H}}_{+}$.

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.

Chapter 4

Monotone and Convex matrix functions

We already introduced monotone and convex matrix functions in the introduction, but now that we have properly defined and discussed underlying structures we should take a deeper look. As mentioned, monotone and convex matrix functions are sort of generalizations for the standard properties of reals, and this is why we should undestand which of the phenomena for the real functions carry to matrix functions and which do not.

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

4.1 Basic properties of the matrix monotone functions

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 $A, B \in \mathcal{H}_{(a,b)}$, such that $A \leq B$ we have $f(A) \leq f(B)$.

We will denote the space of n-monotone functions on open interval (a, b) by $P_n(a, b)$. One immediately sees that that all the matrix monotone functions are monotone as real functions.

Proposition 4.2. *If $f \in P_n(a, b)$, f is increasing.*

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

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

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

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

Proposition 4.3. *Let (a,b) be an open interval and $f : (a,b) \to \mathbb{R}$. Then the following are equivalent:*

(i) f is increasing.

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

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

Proof. TODO $\square$

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

Let us then have some examples.

Proposition 4.4. *For any positive integer n, open interval (a,b) and $\alpha, \beta \in \mathbb{R}$ such that $\alpha \geq 0$ we have that $(x \mapsto \alpha x + \beta) \in P_n(a,b)$.*

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

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

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

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

Proposition 4.5. *Function $(x \mapsto x^2)$ is not n-monotone for any $n \geq 2$ and any open interval $(a,b) \subset \mathbb{R}$.*

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

Note that if $A \leq B$,

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

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

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

- $(B - A)(B + A) = B^2 - A^2 + (BA - AB)$, not $B^2 - A^2$.

- Product of two positive matrices need not be positive.

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

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

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

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

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

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

TODO □

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

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

Proposition 4.6. *For any positive integer n and open interval (a,b) the set $P_n(a,b)$ is a convex cone, i.e. it is closed under taking summation and multiplication by non-negative scalars.*

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

We should be a bit careful though. As we saw with the square function example, product of two n-monotone functions need not be n-monotone in general, even if they are both positive functions; similar statement holds for increasing functions. Similarly, taking maximums doesn't preserve monotonicity.

Proposition 4.7. *Maximum of two n-monotone functions need not be n-monotone for $n \geq 2$.*

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

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

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

as we wanted.

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

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

TODO $\square$

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

We shall be using especially the previous result a lot.

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

Proposition 4.10. *For any open interval (a, b) and positive integer n we have $P_{n+1}(a, b) \subset P_n(a, b)$.*

Proof. TODO $\square$

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.