Operator monotone and operator convex functions: a survey
Technical note 0403, version 1

Michael A. Nielsen[1, *]

[1] School of Physical Sciences and School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Queensland 4072, Australia

(Dated: June 29, 2004)

We survey the important definitions and results in the mathematical theory of operator monotone and operator convex functions. Several simple applications of these results to quantum information science are given as illustrations.

I. OVERVIEW

This technical note is different in nature from my earlier technical notes. Instead of describing in detail some result or set of results, it surveys the important definitions and results in a subfield of science, together with a few applications of those concepts to quantum information science. For the most part, mathematical proofs or detailed chains of physical reasoning are omitted. As a result, it's a bit of an experiment for me, and I'm not sure I'll use the same approach again in future. The subject of the note is the mathematical theory of operator monotone and operator convex functions, and my notes follow the treatment in Chapter 5 of Bhatia [1].

II. BASIC DEFINITIONS AND EXAMPLES

Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a real-valued function. We will explain how to define a corresponding map $f : M_n \rightarrow M_n$, where M_n is the space of $n \times n$ Hermitian matrices. To see how to define such a map, suppose D is an $n \times n$ diagonal matrix with real diagonal entries $d_1, \ldots, d_n$. We define $f(D)$ to be the $n \times n$ diagonal matrix with diagonal entries $f(d_1), \ldots, f(d_n)$. Generalizing this definition, if A is any element of M_n then we can write $A = UDU^\dagger$ for some unitary U and diagonal matrix D. We define the induced map $f : M_n \rightarrow M_n$ by $f(UDU^\dagger) \equiv Uf(D)U^\dagger$. A more informal way of putting this definition is that we work in a basis in which A is diagonal, and simply apply f to each of the diagonal entries. In cases where A can be decomposed in many different ways as $A = UDU^\dagger$ it is an easy exercise to show that $f(A)$ does not depend upon the decomposition chosen. Note that with this definition we obtain $f(UAU^\dagger) = Uf(A)U^\dagger$ for any matrix A and any unitary U, i.e., f commutes with unitary conjugation.

For the rest of these notes we'll use the notation f to denote both real-valued functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and the corresponding functions $f : M_n \rightarrow M_n$. Sometimes it will be useful to restrict the domain of f to some subset S of $\mathbb{R}$. When this is done the induced map on Hermitian matrices is restricted in the obvious way to those Hermitian matrices whose spectrum lies in S. We'll state many of our definitions for the case $S = \mathbb{R}$, but extensions to more general S will often be obvious, especially when S is an interval in the real line.

To define operator monotonicity and convexity, we first need to introduce a partial order on Hermitian matrices. Given two Hermitian matrices $A, B \in M_n$ we define $A \leq B$ if $B - A$ is a positive matrix. We say a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is *operator monotone* if for all n and for all $A, B \in M_n$ satisfying $A \leq B$, we also have $f(A) \leq f(B)$.

Similarly, we say f is *operator convex* if for all n, for all $A, B \in M_n$, and for all $p \in [0, 1]$ we have

$$f(pA + (1-p)B) \leq pf(A) + (1-p)f(B). \tag{1}$$

With a little thought one can show that any convex function $f : \mathbb{R} \rightarrow \mathbb{R}$ is also continuous[1]. With this fact in hand, it follows that operator-convexity is equivalent to *operator mid-point convexity*, $f(\frac{A+B}{2}) \leq \frac{1}{2}f(A) + \frac{1}{2}f(B)$. An *operator concave* function f is a function such that $-f$ is operator convex.

Operator monotonicity and operator convexity are obviously generalizations of the familiar notions of monotone and convex functions on the real line. But there are some important differences in the operator case, arising from the fact that different matrices are diagonal in different bases. We now describe without proof a few examples of functions that are and are not operator monotone and operator convex, to give some insight into the new structure that arises when we generalize monotonicity and convexity to matrices.

Examples of operator monotonicity: The following functions are operator monotone: $\alpha + \beta t$ ($\beta \geq 0$); t^r on $(0, \infty)$ for $0 \leq r \leq 1$; $-1/t$ on $(0, \infty)$. The function t^2 is not operator monotone.

Examples of operator convexity: The functions t^2 and $1/t$ are operator convex. The functions t^3 and $|t|$ are not operator convex.

Example application: As an example of operator convexity in action, suppose $\{E_j\}$ is a set of measurement

* nielsen@physics.uq.edu.au and www.qinfo.org/people/nielsen

[1] We'll eventually see that operator convex functions are actually *analytic*. This remarkable fact follows without making any explicit smoothness assumptions.

operators satisfying $\sum_j E_j^\dagger E_j = I$. Suppose ρ is a density matrix and define $p_j \equiv \mathrm{tr}(E_j \rho E_j^\dagger), \rho_j' \equiv E_j \rho E_j^\dagger / p_j$. Using $\sum_j E_j^\dagger E_j = I$ we have $\rho = \sum_j p_j \frac{\sqrt{\rho} E_j^\dagger E_j \sqrt{\rho}}{p_j}$. Applying the polar decomposition we see that there exist unitary operators U_j such that $\sqrt{\rho} E_j^\dagger E_j \sqrt{\rho} / p_j = U_j \rho_j' U_j^\dagger$. Applying this observation, operator convexity, and the fact $f(UAU^\dagger) = Uf(A)U^\dagger$, we obtain

$$f(\rho) \leq \sum_j p_j U_j f(\rho_j') U_j^\dagger. \tag{2}$$

This can be applied to conclude, for example, that $\rho^2 \leq \sum_j p_j U_j \rho_j'^2 U_j^\dagger$. As a rather trivial special case of the result, by choosing $f(t) = t$, we obtain[2] $\rho \leq \sum_j p_j U_j \rho_j' U_j^\dagger$, which in turn implies the well-known majorization result $\lambda(\rho) \prec \sum_j p_j \lambda(\rho_j')$ [2, 3], where $\lambda(X)$ denotes the vector of eigenvalues of X, ordered into decreasing order. This result enables one to deduce, for example, majorization constraints on entanglement transformation, like those found in [4, 5, 6].

III. CHARACTERIZATIONS

Let P_j be a complete set of orthogonal projectors on some Hilbert space. We define the linear operation on matrices $\mathcal{E}(X) \equiv \sum_j P_j X P_j$. A linear operation that can be written in such a form is called a *pinching*. Such pinchings arise naturally as a description of the evolution of a density matrix under a quantum measurement or decoherence process.

Theorem 1 (Operator convexity and pinchings). *Let f be a real-valued function on an interval, I. Then f is operator convex if and only if $f(\mathcal{E}(X)) \leq \mathcal{E}(f(X))$ for all pinchings $\mathcal{E}$, and for all operators X whose spectrum lies within I.*

The backward implication in this theorem is not particularly useful — there are better ways of testing whether a function is operator convex. But the forward implication is very interesting. For example, if we can prove that the function $f(t) = t\ln(t)$ is operator convex on $[0, 1]$, then it follows that for a pinching $\mathcal{E}$ we have $\mathcal{E}(\rho)\ln[\mathcal{E}(\rho)] \leq \mathcal{E}(\rho \ln \rho)$, and taking the trace gives the entropy inequality $S(\rho) \leq S(\mathcal{E}(\rho))$. A similar line of thought based on the operator convex function $f(t) = t^2$ gives the inequality $\mathrm{tr}(\mathcal{E}(\rho)^2) \geq \mathrm{tr}(\rho^2)$. Inequalities such as these can, of course, be proven by other techniques, but it is interesting to see them drop out so simply from the theory of operator convex functions.

[2] By taking the trace of both sides we can see that the following inequality must be an equality.

The next theorem is essentially a generalization of the previous theorem, although now we consider only a restricted subset of the operator convex functions. To state the theorem, recall that a *contraction* is a linear operator K such that $\|K\| \leq 1$.

Theorem 2 (Operator convexity and contractions). *Let I be an interval containing 0, and let f be a real function defined on I. Then f is operator convex on I and $f(0) \leq 0$ if and only if $f(KXK^\dagger) \leq Kf(X)K^\dagger$ for all contractions K.*

This theorem has the following consequence. Consider the operator convex function $f(t) = t^2$. Then for all density matrices ρ and contractions E (which can be interpreted as measurement operators), the theorem tells us that $(E\rho E^\dagger)^2 \leq E\rho^2 E^\dagger$. Defining $p \equiv \mathrm{tr}(E\rho E^\dagger)$ and $\rho' \equiv E\rho E^\dagger / p$, we obtain $p^2 \rho'^2 \leq E\rho^2 E^\dagger$. This allows us to deduce that

$$p^2 \mathrm{tr}(\rho'^2) \leq \mathrm{tr}(\rho^2), \tag{3}$$

which relates the probability of a measurement event occuring to the trace squared before and after the event. A little thought shows that the result $(E\rho E^\dagger)^2 \leq E\rho^2 E^\dagger$ could have been reached easily by other means, but this is still an interesting result.

We'll see below that another natural place contractions arise in quantum mechanics is as unital quantum operations, $\mathcal{E}$. However, I can't think of any place where such unital operations act naturally in the adjoint fashion required by the theorem.

A result similar in flavour to Theorem 2 is the following theorem, in which a special type of contraction, the projection, plays the role of the contractions in Theorem 2.

Theorem 3 (Operator convexity and projections). *Let I be an interval containing 0, and let f be a real function defined on I. Then f is operator convex on I and $f(0) \leq 0$ if and only if $f(PXP) \leq Pf(X)P$ for all projectors P.*

The forward implications of Theorems 2 and 3 may be viewed as special cases of the following elegant proposition, which we generalize even further below.

Proposition 1. *Let I be an interval containing 0, and let f be a real function defined on I. Then f is operator convex on I and $f(0) \leq 0$ if and only if $f(K_1XK_1^\dagger + K_2YK_2^\dagger) \leq K_1f(X)K_1^\dagger + K_2f(Y)K_2^\dagger$ for all Hermitian X and Y whose spectrum is contained within I, and for all K_1, K_2 such that $K_1K_1^\dagger + K_2K_2^\dagger \leq I$.*

A natural generalization of this result suggests itself. Let $\mathcal{K}$ be a completely positive map with operation elements $\{K_j\}$, so that $\mathcal{K}(X) = \sum_j K_j X K_j^\dagger$. We say $\mathcal{K}$ is *subunital* if $\mathcal{K}(I) \leq I$, i.e., $\sum_j K_j K_j^\dagger \leq I$. The forward implications of Theorems 2, 3, and Proposition 1 have the following elegant generalization to subunital operations.

Theorem 4 (Operator convexity and operation elements). *Let I be an interval containing 0, and let f be a real function defined on I. Then f is operator convex on I and $f(0) \leq 0$ if and only if $f(\sum_j K_j X_j K_j^\dagger) \leq \sum_j K_j f(X_j) K_j^\dagger$ for all Hermitian X_j whose spectrum is contained within I, and for all sets of $\{K_j\}$ such that $\sum_j K_j K_j^\dagger \leq I$.*

Proof: This result goes beyond Bhatia, who only proves Proposition 1, and so we outline a proof. The idea is to induct on the number of operation elements. The case where there is only one operation element is the case of Theorem 2. To do the inductive step, we define $\tilde{K}_1 \equiv K_1$, $\tilde{K}_2 \equiv \sqrt{I - \sum_{j\neq 1} K_j K_j^\dagger}$, $\tilde{X}_1 \equiv X_1$ and $\tilde{X}_2 \equiv \tilde{K}_2^{-1} \left(\sum_{j\neq 1} K_j X_j K_j^\dagger\right) \tilde{K}_2^{-1}$. Provided $\tilde{K}_2$ is invertible, the result now follows by applying Proposition 1 to $\tilde{K}_1, \tilde{K}_2, \tilde{X}_1, \tilde{X}_2$, and using the inductive hypothesis. When $\tilde{K}_2$ is not invertible a standard continuity argument can be used to establish the result.

QED

Choosing $f(t) = t^2$ as our operator convex function, we see that the Hilbert-Schmidt norm is monotonic with respect to any doubly stochastic (that is, completely positive, trace-preserving, and unital) map, i.e., $\mathrm{tr}(\mathcal{E}(X)^2) \leq \mathrm{tr}(X^2)$ for all doubly stochastic $\mathcal{E}$. I should mention that just prior to making this observation, a different proof of this result was communicated to me by Tobias Osborne (2004), who sent me a proof based on the Lieb-Ruskai [7] operator Schwarz inequality, $\mathcal{E}(X)^\dagger \mathcal{E}(X) \leq \mathcal{E}(X^\dagger X)$, which is obviously closely related to Theorem 4. Osborne later pointed me to a paper of Raginsky [8] (which I have not yet read), which apparently gives the same proof, attributing the operator Schwarz inequality to Kadison.

On a related note, in [9] I conjectured that the Hilbert-Schmidt norm is monotonic with respect to any completely positive trace-preserving map $\mathcal{E}$, i.e., $\mathrm{tr}(\mathcal{E}(X)^2) \leq \mathrm{tr}(X^2)$ for all Hermitian X. This conjecture is not correct, as may be seen by choosing $\mathcal{E}$ to be the partial trace operation taking a state of two qubits, A and B, to a state of qubit A alone, and choosing X to be the density matrix for a maximally entangled state of the two qubits. It would be interesting to determine if some weaker monotonicity result can be proved in this context. For example, with a little work one can show that the Hilbert-Schmidt norm is contractive on the space of single-qubit density matrices, with respect to any completely positive trace-preserving map, i.e.,

$$\mathrm{tr}\left(\mathcal{E}(\rho - \sigma)^2\right) \leq \mathrm{tr}\left((\rho - \sigma)^2\right). \tag{4}$$

A natural generalization of this result is that the Hilbert-Schmit norm is contractive with respect to any completely positive trace-preserving map $\mathcal{E}$ having the same domain and range. It would be interesting to prove this more general result correct, or to find a counterexample.

IV. CONNECTIONS BETWEEN OPERATOR MONOTONICITY AND OPERATOR CONVEXITY

The next two theorems connect operator monotonicity with operator convexity and concavity. We will see later that operator monotonicity can be characterized very simply in terms of analytic continuations, so these relationships also greatly simplify the study of operator convexity and concavity.

Theorem 5. *Let $f : [0, \infty) \to [0, \infty)$ be continuous. Then f is operator monotone if and only if f is operator concave.*

As an example of this theorem in action, it follows that t^2 cannot be operator monotone, since t^2 is not concave, and thus cannot be operator concave. Of course, it's difficult to get all that much information out of this theorem at this stage, since we don't have a good understanding of either operator monotone or operator concave functions. Later on, when we understand operator monotone functions better, the utility of this theorem will become more apparent.

Theorem 6. *Let f be a continuous real function on $[0, \alpha)$. Then f is operator convex and $f(0) \leq 0$ if and only if $f(t)/t$ is operator monotone on $(0, \alpha)$.*

As an example of this theorem in action, we see that t^2 must be operator convex, since t is operator monotone. More generally, for any r in the range $1 \leq r \leq 2$ we see that t^r must be operator convex, since t^{r-1} is operator monotone. As for Theorem 5, the true utility of this theorem will not become apparent until later, when we have a powerful characterization of when a function is operator monotone.

V. OPERATOR MONOTONICITY AND ANALYTICITY

In this section we give a general characterization of the operator monotone functions in the language of complex analysis. Key to this is the remarkable fact that operator monotone functions are smooth to all orders; in fact, we'll see that they are analytic.

Theorem 7. *Let $f : I \to \mathbb{R}$ be an operator monotone function on an interval I. Then f is C^∞, that is, the derivatives of f exist to all orders.*

This fact is used to establish the following integral representation for operator monotone functions.

Theorem 8 (Integral representation for operator monotone functions). *Let f be a nonconstant operator monotone function on $(-1, 1)$. Then there exists a unique probability measure μ on $[-1, 1]$ such that*

$$f(t) = f(0) + f'(0) \int_{-1}^{1} \frac{t}{1 - \lambda t} d\mu(\lambda). \tag{5}$$

As a simple application of the this theorem we obtain the following representation theorem for nonlinear operator convex functions.

Corollary 1 (Integral representation for operator convex functions). *Let f be a nonlinear operator convex function on $(-1,1)$. Then there exists a unique probability measure μ on $[-1,1]$ such that*

$$f(t)=f(0)+f'(0)t+\frac{f''(0)}{2}\int_{-1}^{1}\frac{t^2}{1-\lambda t}d\mu(\lambda). \tag{6}$$

The representation of Theorem 8 represents a real breakthrough that allows us to connect the theory of operator monotonicity to complex analysis. In particular, we see that we can define an analytic continuation of f to the entire complex plane except for $z\in(-\infty,1]\cup[1,\infty)$:

$$f(z)\equiv f(0)+f'(0)\int_{-1}^{1}\frac{z}{1-\lambda z}d\mu(\lambda). \tag{7}$$

Next, observe that

$$\mathrm{Im}\left(\frac{z}{1-\lambda z}\right)=\mathrm{Im}\left(\frac{z}{|1-\lambda z|^2}\right). \tag{8}$$

Defining the upper half plane $H_+\equiv\{z\in\mathbb{C}:\mathrm{Im}(z)>0\}$, we see that f maps H_+ into itself. These observations can be used to prove the forward implication in the following theorem.

Theorem 9. *f is an operator monotone function on (a,b) if and only if f has an analytic continuation to the upper half plane H_+ that maps H_+ into itself.*

The reverse implication in this theorem requires results from the theory of complex analysis that we have not proved. The reverse implication is, however, remarkably powerful, giving us a simple recipe to determine when a function is operator monotone.

Indeed, there is a remarkably powerful general theory of such functions. We define a *Pick function* to be all those complex analytic functions defined on H_+, and such that their range is in the closed upper half plane $\{z\in\mathbb{C}:\mathrm{Im}(z)\geq 0\}$. We denote the set of Pick functions by P. Elementary results in complex analysis can be used to show that the non-constant Pick functions take H_+ into H_+, and thus the non-constant operator monotone functions correspond precisely to the non-constant Pick functions. Some examples of Pick functions are as follows: z^r, for $0\leq r\leq 1$; $\log(z)$; $\tan(z)$; $-1/z$. Furthermore, if $f,g\in P$, then so is $f\circ g$. This lets us prove, for example, that if f is in P, then so is $-1/f$.

VI. WHAT TO REMEMBER

Extending real functions to functions on Hermitian matrices; the induced function commutes with unitary conjugation; definition of operator monotone functions; definition of operator convex functions; definition of midpoint operator convexity, and equivalence to operator convexity.

Examples: $-1/t$ and t^r $(0\leq r\leq 1)$ are operator monotone on $(0,\infty)$. t^2 and $1/t$ are operator convex, the latter on $(0,\infty)$.

Definition of a pinching; characterization of operator convex functions in terms of pinchings; definition of a contraction; characterization of operator convexity in terms of contractions; characterization of operator convexity in terms of operation elements.

Equivalence between operator monotonicity and operator concavity for maps from the positive half-line to the positive half-line; equivalence between operator monotonicity and operator convexity of a related function on an interval.

Analyticity of operator monotone functions; integral representation for operator monotone functions; correspondence between operator monotone functions and Pick functions; fact that the Pick functions are closed under composition.

Acknowledgments

Thanks to Tobias Osborne for enlightening correspondence on the contractivity of the Hilbert-Schmidt norm.

[1] R. Bhatia, *Matrix analysis* (Springer-Verlag, New York, 1997).
[2] M. A. Nielsen, Phys. Rev. A **63**, 022114 (2001), arXiv:quant-ph/0008073.
[3] C. Fuchs and K. Jacobs, arXiv:quant-ph/0009101 (2000).
[4] M. A. Nielsen, Phys. Rev. Lett. **83**, 436 (1999), arXiv:quant-ph/9811053.
[5] G. Vidal, Phys. Rev. Lett. **83**, 1046 (1999).
[6] D. Jonathan and M. B. Plenio, Phys. Rev. Lett. **83**, 1455 (1999), arXiv:quant-ph/9903054.
[7] E. H. Lieb and M. B. Ruskai, Advances in Mathematics **12**, 269 (1974).
[8] M. Raginsky, J. Phys. A: Math. Gen. **35**, L585 (2002), arXiv:math-ph/0207041.
[9] M. A. Nielsen (2004), technical note 0401, available at http://www.qinfo.org/people/nielsen/blog/.