Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Inequalities for Markov operators, majorization and the direction of time

Holger Stephan

submitted: December 12, 2013

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: holger.stephan@wias-berlin.de

No. 1896
Berlin 2013

2010 *Mathematics Subject Classification.* 54C45,60J05,15B51,82C03.

Key words and phrases. function-measure-duality, general classical physical system, Jensen's inequality, rearrangements of functions, stochastic matrix, Robin-Hood method .

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

In this paper, we connect the following partial orders: majorization of vectors in linear algebra, majorization of functions in integration theory and the order of states of a physical system due to their temporal-causal connection.

Each of these partial orders is based on two general inequalities for Markov operators and their adjoints. The first inequality compares pairs composed of a continuous function (observables) and a probability measure (statistical states), the second inequality compares pairs of probability measure. We propose two new definitions of majorization, related to these two inequalities. We derive several identities and inequalities illustrating these new definitions. They can be useful for the comparison of two measures if the Radon-Nikodym Theorem is not applicable.

The problem is considered in a general setting, where probability measures are defined as convex combinations of the images of the points of a topological space (the physical state space) under the canonical embedding into its bidual. This approach allows to limit the necessary assumptions to functions and measures.

In two appendices, the finite dimensional non-uniform distributed case is described, in detail. Here, majorization is connected with the comparison of general piecewise affine convex functions. Moreover, the existence of a Markov matrix, connecting two given majorizing pairs, is shown.

Contents

1 Introduction

Time is a natural partial order of physical states. According to the second law of thermodynamics, a generic physical process increases entropy and is therefore time irreversible. This means there is an inequality between the entropy of a physical system before and after a change of state.

From a mathematical point of view this is reflected by inequalities for Markov operators, describing the physical process. These inequalities define an partial order between the objects before and after the action of such operators.

A further partial order well known in mathematics is majorization. Classical majorization theory has at least two origins, majorization of two finite dimensional vectors in linear algebra and majorization of the rearrangements of two functions in integration theory in Euclidian spaces. In linear algebra, this partial order is generated by the action of a doubly stochastic matrix.

The similarity of these partial orders is well known. A vector or function is majorized by another if it is less spread out. The same goes for the example of states during a diffusion process in a homogeneous random matter – a typical irreversible physical process. The solution operators of the corresponding diffusion equation are Markov operators or their adjoints.

Clausius defined a physical process as irreversible if it is impossible to invert this process. However, it is found that a general physical process can be more complicated. If a state is attainable from another by a physical process, in general, the inverse process is possible, too. Generally, irreversibility holds for a pair of objects, only.

More precisely: We have two natural order relations. The first compares pairs composed of an observable and a state, the second compares pairs of states. (Note that these are orderings on the spaces of pairs, not between two elements of a pair.)

If, for example, a physical process transforms a pair of two states into another pair of two states, then there is no physical process transforming the states into the reverse direction (unless it is a deterministic process).

This is reflected by two general inequalities for Markov operators and their adjoints. The first inequality compares pairs composed of a continuous function and a measures, the second inequality compares pairs of two measures.

In [15] it was shown that the state change of a general classical physical system (GCPS) can be described by the action of Markov operators and their adjoints. This allows to describe irreversibility of GCPS by means of such inequalities.

Classical majorization theory is an partial order between two objects (two vectors or two functions), not between two pairs. This corresponds to the case in which we consider two pairs of states, implicitly assuming one of the states to be the equilibrium state of the process.

In this paper, we propose a generalization of the concept of majorization, corresponding to the order of physical processes by time based on the fundamental asymmetry – before and after the action of a Markov operator. The mathematical background for this asymmetry of Markov operators are inequalities of some entropy functionals which are consequences of Jensen's inequality satisfied by any Markov operator. There are two fundamental entropy functionals, each depending on a pair of objects. One depends on a pair composed of a continuous function and a probability measure, the other depends on a pair of probability measures.

In [15] it was shown that the modeling of GCPS leads to a special mathematical framework. This framework has been developed by Kaplan in his two books [7, 8]. It is a general approach to functional analysis, based on the concept of duality of ordered spaces. The main connection

to physics is the definition of probability measures (statistical states) as convex combinations of the images of the states under the canonical embedding of a topological space (state space) into its bidual. This setting describes intensive (observables, continuous functions) and extensive (states, probability measures) values together in one picture, but as different sides of the underlying duality.
The two entropy functionals generalize the classical notion of majorization. Moreover, the functional of two measures can be treated as some "extension" of the Radon-Nikodym Theorem. It turns out that "the rearrangement of the density" of one measure with respect to another exists even if there is no density, i.e. if the Radon-Nikodym Theorem fails. Furthermore, the explicitly given expression allows to use this "density" in applications, whereas the Radon-Nikodym Theorem is, in general, not constructive unless the density is a continuous function.
Due to the fact that Markov operators are solution operators to a wide class of evolution equations, the proved theorems can be applied to show contractivity of its solutions and global solvability. This can also help to understand their large time behavior. Here, the functionals can be treated as Lyapunov functions and the corresponding inequalities are an expression of their monotonicity.

The paper is organized as follows: In subsection 1.1, we collect the main definitions of the used mathematical objects like function spaces, probability measures and Markov operators.
In subsection 1.2, we briefly describe the physical meaning of the introduced objects and the time direction of the action of Markov operators and their adjoints.
In section 2, we define the main functionals, depending on an arbitrary convex function, and show their monotone behavior under the action of Markov operators.
Section 3 is devoted to majorization theory. In subsection 3.1 and 3.2, we collect some facts on the classical majorization theory. In subsection 3.3 we propose our new definition of majorization. In the following subsections we consider the main functionals for a special family of convex functions. This allows to transform Lebesgue integrals into Riemann integrals, which are often easier to handle. Moreover, we derive several helpful identities and inequalities. They them be useful for the comparison of two measures if the Radon-Nikodym Theorem is not applicable.
In Appendix A, we explain the main objects and theorems in the important case of a finite number of states. Here, majorization is connected with the comparison of general piecewise affine convex functions. Explicitely, we consider the non-uniform case. This is a generalization of the case well known in finite dimensional linear algebra.
In Appendix B, we construct a Markov matrix from given majorizated pairs. Here, measures are restricted to vectors of rational numbers.

1.1 Notations

We follow here the concept introduced by S. Kaplan in his two books [7, 8]. In particular, we consider the connection between a topological space and its bidual as essential link between points (state) and probability measures (statistical states). Kaplan's concept turns out to be the natural way of modeling GCPS [15].

1.1.1 The function spaces

The following setting defines functions and measures on a general topological space.

- $\mathcal{Z}$ is a compact topological Hausdorff space (e.g., a polish space).
- $\mathcal{C}(\mathcal{Z})$ is the Banach lattice of real valued continuous functions on $\mathcal{Z}$. It is the topological dual $\mathcal{Z}^*$ of $\mathcal{Z}$. A function $g \in \mathcal{C}$ is called positive if $g(z) \geq 0$ for $z \in \mathcal{Z}$.
- $\mathcal{C}^*(\mathcal{Z})$ is the dual space of $\mathcal{C}(\mathcal{Z})$, the space of radon measures on Borel sets $\mathcal{B}(\mathcal{Z})$ of $\mathcal{Z}$. It is the topological bidual $\mathcal{Z}^{**}$ of $\mathcal{Z}$. A measure $p \in \mathcal{C}^*$ is called positive if $p(B) \geq 0$ for $B \in \mathcal{B}$.
- $\langle \cdot, \cdot \rangle$ denote the dual product in $\mathcal{C}(\mathcal{Z}) \times \mathcal{C}^*(\mathcal{Z})$, $\langle g, p \rangle = \int_{\mathcal{Z}} g(z)p(dz)$, $g \in \mathcal{C}$, $p \in \mathcal{C}^*$. $p \geq 0$ is equivalent to $\langle g, p \rangle \geq 0$ for $g \geq 0$.
- $\mathcal{S}^*(\mathcal{Z}) \subset \mathcal{C}^*(\mathcal{Z})$ is the convex set of positive and normalized radon measures, i.e., probability measures $\mathcal{S}^*(\mathcal{Z}) = \{p \in \mathcal{C}^*(\mathcal{Z}) | p \geq 0, \|p\| = 1\}$. $\mathcal{S}^*(\mathcal{Z})$ is the unit sphere (simplex) of $\mathcal{C}^*(\mathcal{Z})$.
- The set of extremal elements, $\mathcal{S}_e^*(\mathcal{Z}) = \mathrm{extr}\mathcal{S}^*(\mathcal{Z})$, is the set of point measures $\mathcal{S}_e^*(\mathcal{Z}) = \{\delta_z \in \mathcal{C}^*(\mathcal{Z}) | z \in \mathcal{Z}\}$.
- $\mathcal{C}^{**}(\mathcal{Z})$ is the bidual, of $\mathcal{C}(\mathcal{Z})$. It contains the set of all bounded measurable functions $\mathcal{C}_{\mathcal{B}}(\mathcal{Z})$ the space of continuous functions on $\mathcal{Z}$ equipped with the topology generated by all Borel sets of $\mathcal{Z}$.
- $\langle \cdot, \cdot \rangle_*$ denote the dual product in $\mathcal{C}^{**}(\mathcal{Z}) \times \mathcal{C}^*(\mathcal{Z})$. $\langle \xi, p \rangle_* = \int_{\mathcal{Z}} \xi(z)p(dz)$, $\xi \in \mathcal{C}^{**}$, $p \in \mathcal{C}^*$.

The image of the canonical embedding i of $\mathcal{Z}$ in its bidual $\mathcal{Z}^{**} = \mathcal{C}^*(\mathcal{Z})$ is the set of point measures $\mathcal{S}_e^*(\mathcal{Z})$. We have $iz = \delta_z$. Thus, probability measures are the elements of the weak* closure of the convex hull of the image of the canonical embedding of some topological space in its bidual $\mathcal{S}^*(\mathcal{Z}) = \overline{\mathrm{conv}\{\delta_z | z \in \mathcal{Z}\}}^*$.
In this sense we can understood $\mathcal{C}(\mathcal{Z})$ as the Banach space of continuous functions on the extremal elements of the unit sphere of its own dual space $\mathcal{C}(\mathcal{Z}) = \mathcal{C}(\mathcal{S}_e^*(\mathcal{Z}))$. This connection

$$\mathcal{Z} \stackrel{\text{dual}}{\longrightarrow} \mathcal{C}(\mathcal{Z}) \stackrel{\text{dual}}{\longrightarrow} \mathcal{C}^*(\mathcal{Z}) \stackrel{\text{sphere}}{\longrightarrow} \mathcal{S}^*(\mathcal{Z}) \stackrel{\text{extr}}{\longrightarrow} \mathcal{S}_e^*(\mathcal{Z}) = \mathcal{Z}$$

gives the construction a solid structure and simplifies the technical problems.
The elements of the spaces are of two types: functions of points like the elements of $\mathcal{C}(\mathcal{Z})$ or $\mathcal{C}^{**}(\mathcal{Z})$ and functions of sets like the elements of $\mathcal{C}^*(\mathcal{Z})$. Inspired by the concepts in physics we call these objects intensive and extensive, resp.
We introduce some more notations:

- $\mathbb{1} \in \mathcal{C}(\mathcal{Z})$ is the constant function $\mathbb{1}(z) = 1,\ z \in \mathcal{Z}$

- $\mathbb{1}_B \in \mathcal{C}_{\mathcal{B}}(\mathcal{Z})$ is the indicator function $\mathbb{1}_B(z) = 1,\ z \in B, = 0$ otherwise. $\mathbb{1}_B(z) = \delta_z(B)$, $z \in \mathcal{Z}$, $B \in \mathcal{B}$.
- $z_{\min} \in \text{Argmin}_{z\in\mathcal{Z}} g(z)$, $z_{\max} \in \text{Argmax}_{z\in\mathcal{Z}} g(z)$
- $g_{\min} = g(z_{\min})$, $g_{\max} = g(z_{\max})$.
- $\overline{g}_p = \langle g, p\rangle$, the mean value of g with respect to p
- $g_+ = \sup\{g, 0\}$, $g_- = \sup\{-g, 0\}$.
- $\mathcal{C}_{[a,b]}$ is the band of such functions g with $[g_{\min}, g_{\max}] \subset [a, b]$.

As usual, if there is no confusion, we omit $\mathcal{Z}$ and write $\mathcal{C}$, $\mathcal{S}^*$, ... instead of $\mathcal{C}(\mathcal{Z})$, $\mathcal{S}^*(\mathcal{Z})$, ...

1.1.2 Markov operators

- A Markov operator $\mathbf{M} : \mathcal{C}(\mathcal{Z}) \longrightarrow \mathcal{C}(\mathcal{Z})$ is a bounded linear, positivity and $\mathbb{1}$ conserving operator. We denote the set of Markov operator by $\mathcal{M}$

$$\mathcal{M} = \{\mathbf{M} \in \mathcal{L}(\mathcal{C}) \mid \mathbf{M} \geq 0,\ \mathbf{M}\mathbb{1} = \mathbb{1}\}$$

- The adjoint of a Markov operator $\mathbf{M}^* : \mathcal{C}^*(\mathcal{Z}) \rightarrow \mathcal{C}^*(\mathcal{Z})$ is defined by $\langle \mathbf{M}g, p\rangle = \langle g, \mathbf{M}^*p\rangle$ and is a bounded linear and probability measure conserving operator: $\mathbf{M}^*\mathcal{S}^*(\mathcal{Z}) \subset \mathcal{S}^*(\mathcal{Z})$. The adjoint of a bounded linear operator maps probability measure into itself if and only if it is the adjoint of a Markov operator. Note, that a bounded linear operator in $\mathcal{C}^*(\mathcal{Z})$ must not have a pre-adjoint one.
- The adjoint of a Markov operator has a adjoint, defined by $\langle \xi, \mathbf{M}^*p\rangle_* = \langle \mathbf{M}^{**}\xi, p\rangle_*$. $\mathbf{M}$ is the restriction of $\mathbf{M}^{**}$ on $\mathcal{C}$.
- A Markov operator can be described by a weak* continuous family $p_z \in \mathcal{S}^*(\mathcal{Z})$. We have $(\mathbf{M}g)(z) = \langle g, p_z\rangle$. Thus, Markov operators provide convex combinations.

  Markov operators and its adjoints have an integral representation

$$\begin{aligned}(\mathbf{M}g)(z) &= \int_{\mathcal{Z}} g(z')\omega(z, dz') = \langle g, \omega_z\rangle \\ (\mathbf{M}^*p)(B) &= \int_{\mathcal{Z}} \omega(z', B)p(dz') = \langle \omega_B, p\rangle\end{aligned}$$

  with a kernel $\omega : \mathcal{Z} \times \mathcal{B}(\mathcal{Z}) \rightarrow [0, 1]$.
- Every adjoint of a Markov operator has a fixed-point (or invariant measure) $\mu \in \mathcal{S}^*$ with $\mathbf{M}^*\mu = \mu$ (Theorem of Frobenius-Perron-Krein-Rutman).
- A Markov operator is band conserving: $\mathbf{M}\mathcal{C}_{[a,b]} \subset \mathcal{C}_{[a,b]}$. Thus, they satisfy the following maximum principle:

$$g_- \leq g \leq g_+ \quad \Longrightarrow \quad g_- \leq \mathbf{M}g \leq g_+$$

  The maximum principle is a characterization of Markov operators.

- Given a continuous function $\gamma : \mathcal{Z} \to \mathcal{Z}$, the composition operator $\mathbf{M}_\gamma g = g \circ \gamma$ is a Markov operator and is called deterministic Markov operator. Its adjoint maps point measures into itself: $\mathbf{M}_\gamma^* \mathbf{S}_e^* \subset \mathbf{S}_e^*$. More precisely, $\mathbf{M}_\gamma^* \delta_z = \delta_{\gamma(z)}$. In this sense, such an operator is deterministic. It is the image of γ by means of the canonical embedding of $\mathcal{Z}$ into $\mathcal{Z}^{**}$.

  For general measures $p \in \mathcal{S}$ we have $(\mathbf{M}_\gamma^* p)(B) = p(\gamma^{-1}(B))$.

  In the finite, n-dimensional case, the set of deterministic Markov operators (matrices) consists of n^n matrices with one 1 in each row. This are the extremal points of the convex set of Markov operators.

  In the general case, the set of deterministic Markov operators are not the extremal points of the set of Markov operators. There can be only one extremal points of this set, the identity $\mathbf{I}$.

- If a Markov operator $\mathbf{M}$ has an inverse $\mathbf{M}^{-1}$, is has to be a deterministic one. We have $\mathbf{M}_\gamma^{-1} = \mathbf{M}_{\gamma^{-1}}$ if the continuous function $\gamma : \mathcal{Z} \to \mathcal{Z}$ has a continuous inverse γ^{-1}.

  In the n-dimensional case, the set of invertible deterministic Markov operators consists of $n!$ matrices with one 1 in each row and each column – the representation of the permutation group on $\mathcal{Z}$.

In general, Markov operators can act from one Banach lattice into another. The state spaces before and after a state change must not be the same. If they are different, say $\mathcal{Z}_1$ (before) and $\mathcal{Z}_2$ (after), the corresponding spaces of observables are $\mathcal{C}(\mathcal{Z}_1)$ and $\mathcal{C}(\mathcal{Z}_2)$ and a Markov operator $\mathbf{M}$ is a linear bounded operator acting form $\mathcal{C}(\mathcal{Z}_1)$ into $\mathcal{C}(\mathcal{Z}_2)$. In this case positivity conserving means $g \geq_1 0 \implies (\mathbf{M}g) \geq_2 0$ and 1-preserving means $\mathbf{M}\mathbb{1}_1 = \mathbb{1}_2$, where $\geq_{1,2}$ are the corresponding partial orders and $\mathbb{1}_{1,2}$ are the corresponding unities. We will not unnecessarily complicate the notation and assume in the following Markov operators, acting in the same space. Nevertheless, what follows is true in the general case except, obviously, the existence of an invariant measure and the case, when the Markov operator is the solution operator to an evolution equations. This general situation covers the important case of rectangular matrices. In Appendix B we note, how this general situation is to be handled.

1.1.3 Operators, connecting objects of different types

There are some important operations, connecting extensive and intensive values.
For a probability measure $p \in \mathcal{S}^*$, we define the multiplication operator $\mathbf{Q}_p : \mathcal{C}(\mathcal{Z}) \longrightarrow \mathcal{C}^*(\mathcal{Z})$ acting as

$$(\mathbf{Q}_p g)(B) = \int_B g(z)p(dz), \;\; g \in \mathcal{C}, \;\; B \in \mathcal{B}(\mathcal{Z})$$

$\mathbf{Q}_p$ can be represented as an integral operator with kernel $p(B \cap B')$. (Note that this operator cannot be defined between different spaces $\mathcal{C}(\mathcal{Z}_1)$ and $\mathcal{C}^*(\mathcal{Z}_2)$.)
The inverse of $\mathbf{Q}_p$, the operator $\mathbf{Q}_p^{-1} : \mathcal{C}^* \to \mathcal{C}$, is in general an unbounded operator and is the restriction of the Radon-Nikodym operator to $\mathcal{C}$, $\mathbf{Q}_p^{-1} q = \frac{dq}{dp}$.
The operator $\mathbf{Q}_p^{-1}$ acting on q defines an intensive value $\mathbf{Q}_p^{-1} q$ from a pair (q, p) of extensive values. This is the typical way in physics to construct a value, defined at states from two values that can be measured by counting (e.g., velocity from covered distance and elapsed time).

For a functional $\Phi : \mathcal{C} \to \bar{\mathbb{R}}$ we define its convex conjugate $\Phi^* : \mathcal{C}^* \to \bar{\mathbb{R}}$ via the Legendre-Fenchel transform (see [3])

$$\Phi^*(p) = \sup_{g \in \mathcal{C}} \big(\langle g, p\rangle - \Phi(g)\big)$$

We have Youngs inequality

$$\Phi^*(p) + \Phi(g) \geq \langle g, p\rangle$$

with equality if p is in the subdifferential of Φ at g: $p \in \partial\Phi(g)$ or reverse, $g \in \partial\Phi^*(p)$.
The Legendre-Fenchel transform is the typical way in physics to construct a value of dual type from a given value (e.g., momentum from velocity).

1.2 Physical background of the objects

The elements of the main spaces $\mathcal{Z}$, $\mathcal{C}$ and $\mathcal{S}^*$ have important meanings for GCPS (see [15]). $\mathcal{Z}$ is the set of states, the state space. $\mathcal{C}$ is the set of observables and $\mathcal{S}^*$ is the set of statistical states.
Moreover, the elements of $\mathcal{C}$, functions of points can be understood as intensive values, whereas the elements of $\mathcal{S}^*$, positive and additive functions of sets can be understood as extensive values. $\mathcal{S}^*$ describes the real physical objects, whereas the elements of $\mathcal{C}$ are auxiliary objects and the names differ (e.g., potentials, densities, information, test functions), depending on the area of knowledge where they are used.
We use observables to decide, whether the system is in some state or not. $\mathbb{1}_B$ observes, whether the system is in a state from B or not. If an observable g coincides on two states, say $g(z_1) = g(z_2)$ for some $z_1, z_2 \in \mathcal{Z}$, the observable g is not able to distingwish them. For a fiven g, its range consists of all information about the observation. Detailed information about the state space is lost.
A probability measure is a mixed state, the weak* closure of convex combinations of pure states. Probabilities come into the play if the same experiment can give different results. The dual product $\langle g, p\rangle$ is the expected value of the observable g.
Using densities (intensive values) instead of measures (extensive values), the real type of the object changes. As a result, it can behave in a different way as expected (see the next subsubsections).
We have to consider Markov operators instead of deterministic Markov operators, because of the uncertainty of a state change. We treat here uncertainty in an epistemic sense – uncertainty due to incomplete knowledge. A Markov operator provides a sensible physical process. The adjoint of a non-markovian operator maps some probability measures out of the set $\mathcal{S}^*$. The result of such an action can not be understood in a physical sense.
The kernel $\omega(z, B)$ of a Markov operator is the probability to find the system after the state change in a state from B if it was before in the state z with certainty. The invariant measure of a Markov operator is called equillibrium state.

1.2.1 Markov operators and direction of time

We understand time as a parameter of order, defining the causal order of successive states. In this sense, the notion of time can be replaced by any other ordering parameter like the step in an iteration scheme or a control parameter of a device.

This treatment is in contrast to the understanding of time in classical mechanics as the continuous parameter, parameterizing the trajectory of a particle or body.
In what follows, we consider two time points t and $t' > t$, calling the states at these points of time “before” and “after”. We define the order of physical processes – the action of the adjoint of a Markov operator – as “time running forward” as it is intuitively understood.
Then, the action of a Markov operator is a step backwards in time, $g = \mathbf{M}g'$; whereas the action of the adjoint of a Markov operator is a step forwards in time, $p' = \mathbf{M}^* p$. The dual product taken at different points of time is conserved:

$$\langle g', p' \rangle = \langle g', \mathbf{M}^* p \rangle = \langle \mathbf{M}g', p \rangle = \langle g, p \rangle$$

The backward or forward directed action of Markov operators and their adjoints is well known from evolution equations in probability theory. The solution operators of the Kolmogorov–Chapman backward equation are Markov operators, whereas the solution operators of the Kolmogorov–Chapman forward equation are adjoints of Markov operators (see, e.g., [11]).
Since a Markov operator is band conserving, $\mathbf{M}\mathcal{C}_{[a,b]} \subset \mathcal{C}_{[a,b]}$, an observation before is “less spread out” or “less detailed” than afterwards. An observation becomes “clearer” in time.
A deterministic Markov operator $\mathbf{M}_\gamma$ is an abstract pullback operator. Its adjoint $\mathbf{M}^*_\gamma$ is an abstract pushforward operator, well known, for example, in optimal transport theory in which it is written as $\mathbf{M}^*_\gamma p = \gamma_{\#} p$.

1.2.2 Reverse physical processes and densities

The density of one probability measure with respect to another is an intensive value. It is interesting how it is mapped, if the measure is mapped forward in time by the adjoint of a Markov operator $\mathbf{M}^*$. Let $p, q \in \mathcal{S}^*$, $p' = \mathbf{M}^* p$ and $q' = \mathbf{M}^* q$. We define densities $h = \mathbf{Q}_p^{-1} q$ and $h' = \mathbf{Q}_{p'}^{-1} q'$, assuming them to exist and to be continuous.
We define an operator $\mathbf{X} = \mathbf{Q}_{p'}^{-1} \mathbf{M}^* \mathbf{Q}_p$ and assume that this is a bounded linear operator in $\mathcal{C}$. We do not analyze the assumptions on $\mathbf{M}$ and p to be sure that $\mathbf{X}$ is a well defined bounded linear operator in $\mathcal{C}$ (see subsection 3.6). In simple situations this is the case (e.g. in case of a finite $\mathcal{Z}$ with strict positive p and mixing $\mathbf{M}$, see Appendix A and B).
Obviously, $\mathbf{X} \geq 0$ and we have

$$\mathbf{X}\mathbb{1} = \mathbf{Q}_{p'}^{-1} \mathbf{M}^* \mathbf{Q}_p \mathbb{1} = \mathbf{Q}_{p'}^{-1} \mathbf{M}^* p = \mathbf{Q}_{p'}^{-1} p' = \mathbb{1} \ .$$

Hence, $\mathbf{X}$ is a Markov operator. Moreover, we have

$$\begin{aligned} \mathbf{X}h &= \mathbf{Q}_{p'}^{-1} \mathbf{M}^* \mathbf{Q}_p h = \mathbf{Q}_{p'}^{-1} \mathbf{M}^* q = \mathbf{Q}_{p'}^{-1} q' = h' \ , \\ \mathbf{X}^* p' &= \mathbf{Q}_p \mathbf{M}^{**} \mathbf{Q}_{p'}^{-1} p' = \mathbf{Q}_p \mathbf{M}^{**} \mathbb{1} = \mathbf{Q}_p \mathbb{1} = p \ . \end{aligned}$$

Thus, $\mathbf{X}$ maps h to h', at first glance forward in time, but being a Markov operator and not the adjoint, this is the inverse time direction. Undoubtedly, this behavior does not invert the time, but the physical process. $\mathbf{X}^*$ is an action forward in time, but it models the inverse physical process. Hence, for $\mathbf{X}$ p is the state after and p' before.
If one wants to describe the evolution using densities, one has to take the inverse physical process and backwards running time (see subsubsection 2.1.3).
The case $\mathbf{X} = \mathbf{M}$, i.e. if a Markov operator $\mathbf{M}$ and a state p satisfy $\mathbf{Q}_{p'} \mathbf{M} = \mathbf{M}^* \mathbf{Q}_p$, describes an invertible physical process. The special case, when $p = \mu$ is an invariant measure of $\mathbf{M}^*$, the operator $\mathbf{M}$ is called a detailed balance operator and the underlying process is called reversible. Note, this does not mean time reversiblity. A reversible process is time irreversible as well in the sense that corresponding entropy functionals are monotone in time.

2 Inequalities for Markov operators

Any Markov operator satisfies a wide class of simple inequalities. Some of them are consequences of Jensen's inequality.

2.1 The main functionals

We define two real valued functionals. Each of them depends on a pair of objects, on $\mathcal{C} \times \mathcal{S}^*$ and $\mathcal{S}^* \times \mathcal{S}^*$.

$$
\begin{aligned}
H[g,p] &= \langle F(g), p\rangle, \ g \in \mathcal{C}, \ p \in \mathcal{S}^* && (1)\\
H^*[q,p] &= \sup_{g\in\mathcal{C}} \big(\langle g, q\rangle - \langle F(g), p\rangle\big), \ p, q \in \mathcal{S}^* && (2)
\end{aligned}
$$

It is easy to show that $H^*[q,p] \geq H^*[p,p] = F^*(1)$ (see [14]).
The importance of these functionals is in their asymmetry with respect to the action of a Markov operator shown in Theorem 2. This asymmetry is the key for the new definitions of majorization.

2.1.1 Convexity of the main functionals

Obviously, $H[g,\cdot]$ is linear on $\mathcal{S}^*$. Moreover, we have

Theorem 1

(i) $H[\cdot,p]$ *is convex on* $\mathcal{C}$

(ii) $H^*[\cdot,p]$ *is convex on* $\mathcal{S}^*$

(iii) $H^*[q,\cdot]$ *is convex on* $\mathcal{S}^*$

Proof:

(i) This is a simple consequence of the convexity of F and the positivity of p.

(ii) We set $\alpha_{1,2} \in [0,1]$ with $\alpha_1 + \alpha_2 = 1$. Then, we have

$$\langle g, \alpha_1 q_1 + \alpha_2 q_2\rangle - \langle F(g), p\rangle = \alpha_1\big(\langle g, q_1\rangle - \langle F(g), p\rangle\big) + \alpha_2\big(\langle g, q_2\rangle - \langle F(g), p\rangle\big)$$

Applying $\sup_{g\in\mathcal{C}}$, it follows

$$
\begin{aligned}
\sup_{g\in\mathcal{C}} \big(\langle g, \alpha_1 q_1 + \alpha_2 q_2\rangle - \langle F(g), p\rangle\big) &\leq \alpha_1 \sup_{g\in\mathcal{C}} \big(\langle g, q_1\rangle - \langle F(g), p\rangle\big) + \\
&+ \alpha_2 \sup_{g\in\mathcal{C}} \big(\langle g, q_2\rangle - \langle F(g), p\rangle\big)
\end{aligned}
$$

This shows $H^*[\alpha_1 q_1 + \alpha_2 q_2, p] \leq \alpha_1 H^*[q_1, p] + \alpha_2 H^*[q_2, p]$.

(iii) This can be shown as just. □

2.1.2 Monotonicity of the main functionals

The following theorem shows the monotonicity of the defined functionals under the action of Markov operators.

Theorem 2 *For any convex function F, any Markov operator $\mathbf{M}$, any $g \in \mathcal{C}$ and any $p, q \in \mathcal{S}^*$ the following inequalities hold:*

$$\begin{aligned} H[\mathbf{M}g, p] &\leq H[g, \mathbf{M}^* p] & (3) \\ H^*[\mathbf{M}^* q, \mathbf{M}^* p] &\leq H^*[q, p] & (4) \end{aligned}$$

In (3) equality holds for deterministic Markov operators. In (4) equality holds for deterministic Markov operators with dense range.

Proof: The inequalities (3) and (4) together with boundedness and convexity were already proved in [14]. □

For completeness we repeat the easy proofs of the monotonicity, based on Jensen's inequality. The classical Jensen inequality for probability measures p reads

$$\big\langle F(g), p \big\rangle \geq F\big(\langle g, p\rangle\big)$$

with equality if $p = \delta_z$ is a point measure. Since a Markov $\mathbf{M}$ operator can be represented as a dual product with a family of probability measures, say $\omega_z \subset \mathcal{S}^*$, it follows

$$\big\langle F(g), \omega_z \big\rangle \geq F\big(\langle g, \omega_z\rangle\big)$$

equivalent to the inequality

$$\mathbf{M}F(g) \geq F(\mathbf{M}g) \tag{5}$$

This is an inequality between continuous functions in $\mathcal{C}$. Equality holds for strict convex functions F for deterministic Markov operators (families of point measures) and reads

$$(\mathbf{M}_\gamma F)(g) = (F \circ g) \circ \gamma = F \circ (g \circ \gamma) = F(\mathbf{M}_\gamma g)$$

Taking the dual product of (5) with some $p \in \mathcal{S}^*$ it follows, since $p \geq 0$,

$$\langle \mathbf{M}F(g), p \rangle \geq \langle F(\mathbf{M}g), p \rangle \tag{6}$$

This is a general variant of Karamata's inequality and is equivalent to (3).
Finally, (4) follows from

$$\begin{aligned} H^*[\mathbf{M}^* q, \mathbf{M}^* p] &= \sup_{g \in \mathcal{C}} \big(\langle g, \mathbf{M}^* q\rangle - \langle F(g), \mathbf{M}^* p\rangle\big) \\ &= \sup_{g \in \mathcal{C}} \big(\langle \mathbf{M}g, q\rangle - \langle \mathbf{M}F(g), p\rangle\big) \\ &\leq \sup_{g \in \mathcal{C}} \big(\langle \mathbf{M}g, q\rangle - \langle F(\mathbf{M}g), p\rangle\big) \\ &= \sup_{f \in R(\mathbf{M})} \big(\langle f, q\rangle - \langle F(f), p\rangle\big) \\ &\leq \sup_{g \in \mathcal{C}} \big(\langle g, q\rangle - \langle F(g), p\rangle\big) = H^*[q, p] \end{aligned}$$

Denote as usual $g = \mathbf{M}g'$, $p' = \mathbf{M}^* p$ and $q' = \mathbf{M}^* q$, inequalities (3) and (4) read

$$\begin{aligned} H[g', p'] &\geq H[g, p] & (7) \\ H^*[q', p'] &\leq H^*[q, p] & (8) \end{aligned}$$

Note, the functionals H and H^* have different monotonicities in time.

2.1.3 Reduced functionals for invariant measures and Lyapunov functions

We fix a Markov operator $\mathbf{M}$ and consider an arbitrary fixed point (invariant measures) $\mu \in \mathcal{S}^*$ of $\mathbf{M}^*$. For this measure we define

$$
\begin{aligned}
L[g] &= H[g,\mu] = \langle F(g), \mu\rangle, \; g \in \mathcal{C} && (9)\\
L^*[q] &= H^*[q,\mu] = \sup_{g\in\mathcal{C}} \big(\langle g, q\rangle - \langle F(g),\mu\rangle\big), \; q \in \mathcal{S}^* && (10)
\end{aligned}
$$

Then, the inequalities (3) and (4) reduce to

$$
\begin{aligned}
L[\mathbf{M}g] &\le L[g], && (11)\\
L^*[\mathbf{M}^*q] &\le L^*[q] \; . && (12)
\end{aligned}
$$

These functionals are well known as Lyapunov function, for example in the theory of evolution equations like (see, e.g. [1])

$$
\begin{aligned}
\partial_s g(s) &= \mathbf{A}g(s), \; g(0) = g_0, \;\; \text{in } \mathcal{C} && (13)\\
\partial_t p(t) &= \mathbf{A}^* p(t), \; p(0) = p_0, \;\; \text{weak* in } \mathcal{C}^* && (14)
\end{aligned}
$$

If $\mathbf{T}(s_2 - s_1)$ and $\mathbf{T}^*(t_2 - t_1)$ are the solution operators (semigroups) of these equations, from inequalities (11) and (12) follows with $t_2 \ge t_1 \ge 0$, $g(t_1) = \mathbf{T}(s_1)g_0$, $g(t_2) = \mathbf{T}(s_2)g_0$, $p(t_1) = \mathbf{T}^*(t_1)p_0$, $p(t_2) = \mathbf{T}^*(t_2)p_0$, $\mathbf{M} = \mathbf{T}(s_2 - s_1)$ it follows

$$L[g_2] \le L[g_1], \; L^*[p_2] \le L^*[p_1] \; .$$

This allows the investigation of the asymptotic behavior of the solutions to the equations. The existence of a stationary solution, i.e., an invariant measure of the semigroup $\mathbf{T}^*(t)$ is a consequence of the Kakutani-Markov Theorem.
The equations (13) and (14) is the usual writing of the Kolmogorov-Chapman equations. However s and t are different times. This is the reason, why for this equations time monotonicity holds with respect to an invariant measure. Writing both equations with the same forward running time t with $t_a \le t \le t_b$ we have

$$
\begin{aligned}
\partial_t g(t) &= -\mathbf{A}g(t), \; g(t_b) = g_b, \;\; \text{in } \mathcal{C}\\
\partial_t p(t) &= \mathbf{A}^* p(t), \; p(t_a) = p_a, \;\; \text{weak* in } \mathcal{C}^*
\end{aligned}
$$

For this equations – even for time depending operators $\mathbf{A} = \mathbf{A}(t)$ – the functionals H and H^* can be used as Lyapunov functions with the inequalities (7) and (8). (see [16]).

2.1.4 The connection with the common entropy functional

As usual, instead of the weak* equation (14) a weak equation, e.g., in $L_2(\mu)$ for the density h of p with respect to the invariant measure μ is considered. For such an equation the entropy functional

$$L[h] = \int_{\mathcal{Z}} G\big(h(z)\big)\mu(dz) \qquad (15)$$

with some convex function $G : \mathbb{R} \to \mathbb{R}$ is defined. This is a special case of the functional H^*:

Assuming, there is a density h of q with respect to p, i.e., $q = \mathbf{Q}_p h$. Then, functional $H^*[q,p]$ reads

$$\begin{aligned} H^*[q,p] &= \sup_{g\in\mathcal{C}} \big(\langle g,q\rangle - \langle F(g),p\rangle\big) = \sup_{g\in\mathcal{C}} \big(\langle g,\mathbf{Q}_p h\rangle - \langle F(g),p\rangle\big) = \\ &= \sup_{g\in\mathcal{C}} \big(\langle g\cdot h,p\rangle - \langle F(g),p\rangle\big) = \sup_{g\in\mathcal{C}}\langle g\cdot h - F(g),p\rangle \end{aligned}$$

Since $p \geq 0$, we can take the supremum pointwise and can conclude

$$\tilde{H}[h,p] := H^*[q,p]\Big|_{q=\mathbf{Q}_p h} = \int_{\mathcal{Z}} F^*\big(h(z)\big)p(dz) \tag{16}$$

where F^* is the conjugate convex function of F. This is precisely functional H since F can be an arbitrary convex function and F and F^* constitute the same set.
This functional can not be used as a Lyapunov function, because h and p are mapped in different time directions. Therefore, this functional can be used only in the case, when h or p do not change in time, i.e., if one of them is the stationary solution. Thus, this functional is used with $p = \mu$ (the case $h = \mathbb{1}$ is trivial and uninteresting), consequently it is (15).
Clearly, the right hand side of (16) makes no sense, if the radon-Nikodym derivative h does not exist. As usual, it is defined to be equal to $+\infty$ in this case. However, $H^*[q,p]$ can be finite even in this case, if F^* behaves well enough (see [14]).

2.2 Time irreversibility and the second law of thermodynamics

While everybody knows that time is running forward and has a fieling, whether a physical process can be inverted or not, there is not unique definition of time irreversibility of physical processes. Clausius defined irreversibility mutatis mutandis as: A states p' is reachable from the state p by an action, but the opposite is impossible. Allready Planck countered that may be a completely different process can lead to a reversal of the event.
Obviously, a periodic or almoust periodic physical processes is reversible. The process itsself take back the system to any former state. But if two states – the states before and after a state change – are given, we have to find out the conditions for a physical state change to be undone. Actually this means, is it possible to determine the direction of time from two given states.
Clearly, we don't know all possible physical processes which are known today or which will be later discovered. Therefore, we are not interested to understand whether there is such a physical process, but whether we can preclude its existence for mathematical reasons, i.e., whether the non-existence of such a physical process is a logical consequence.
Knowing that a physical state change is provided by the action of a Markov operator, we will say that a process is irreversible, if there cannot exist a Markov operator doing the reverse action. If the reverse action is provided by some non-Markov operator, this means that it maps some probability measure out of the set of probability measures (e.g., to a non-positive measure). This is a logical contradiction because such a result is physically non understandable.
In this sense we try to give to one of the monotonicity inequalities namely to the inequality (8) a physical sense. This inequality is connected with probability measures that are widely accepted to understood as statistical states.
The inequality (7) is connected with a continuous function and a probability measures. The interpretation of the continuous function depends on the context of the problem of consideration. It can be interpretated as observable, test function, potential, information and much

more. Thus, the interpretation of inequality (7) can not be general. Moreover, inequality (8) corresponds to the typical understanding of the second law of thermodynamic. That's why, we concentrate here on inequality (8).

At first we state that there is no talk about a complete reversal of all actions of a general Markov operator. Given a Markov operator $\mathbf{M}$. We consider its range, the set $R(\mathbf{M}^*) = \{p' \in \mathcal{S}^* : \ p' = \mathbf{M}^* p, \ p \in \mathcal{S}^*\}$. Is there a Markov operator $\mathbf{X}$ with $p = \mathbf{X}p'$ for all $p' \in R(\mathbf{M}^*)$? Obviously we have $\mathbf{X} = \mathbf{M}^{-1}$ on $R(\mathbf{M}^*)$. This is possible only if $\mathbf{M}$ is a deterministic Markov operator. This is, for example, a simple consequence of the fact, that the spectrum of a Markov operator lies in the closed unit disc. Thus, a invertible Markov operator must have a spectrum in the closed unit circle, i.e., it must be a deterministic Markov operator.

To demand all actions to be invertible is one extreme case. The other is to demand that one actions can be inverted. Given a Markov operator $\mathbf{M}$ and two states p and p' with $p' = \mathbf{M}^* p$. Is there a Markov operator $\mathbf{X}$ with $p = \mathbf{X}^* p'$? This problem was allready solved in subsubsection 1.2.2. $\mathbf{X} = \mathbf{Q}_{p'}^{-1} \mathbf{M}^* \mathbf{Q}_p$ is such an operator. We claim that such an operator allways exists (see subsection 3.6). In simple situations this is true (see Appendix B).

This is in contrast to the usual interpretation of irreversibility with respect to one state like the definition of irreversibility by Clausius: "A states p' is reachable from the state p by an action, but the opposite is not possible." This interpretation is connected with the reduced functional (9) and implicitly considers states with respect to an equilibrium state. Then, the statement is a consequence of inequality (14).

The next step is the question, whether the action of two states is invertible. Given a Markov operator $\mathbf{M}$ and two pairs of states (p, q) and (p', q') with $p' = \mathbf{M}^* p$ and $q' = \mathbf{M}^* q$. Is there a Markov operator $\mathbf{X}$ with $p = \mathbf{X}^* p'$ and $q = \mathbf{X}^* q'$? We show that such a Markov operator cannot exist if $H^*[q', p'] < H^*[q, p]$ for some functional H^* (i.e., for some convex function, generating H^*). We show the non-existence of the mentioned operator $\mathbf{X}$ by contradiction. Assume that there exists a $\mathbf{X}$ with $p = \mathbf{X}^* p'$ and $q = \mathbf{X}^* q'$. Then, we can conclude from (8) $H^*[\mathbf{X}^* q', \mathbf{X}^* p'] \leq H^*[q', p']$. Hence, $H^*[q, p] \leq H^*[q', p']$, a contradiction to $H^*[q', p'] < H^*[q, p]$. In general, in (8) equality $H^*[q', p'] = H^*[q, p]$ is possible. This means, $\mathbf{M}$ acts on the states p and q like a deterministic Markov operator. Since we consider the action of $\mathbf{M}^*$ only on p and q, we don't know how $\mathbf{M}^*$ is acting on other states. On two states we cannot distingwish Markov operator from deterministic ones. A Markov operator $\mathbf{M}$ is called mixing, if $\mathbf{M}^*$ maps extremal points of $\mathcal{S}$ to inner points. For such operators the strong inequality $H^*[\mathbf{M}^* q, \mathbf{M}^* p] < H^*[q, p]$ holds for any $p, q \in \mathcal{S}^*$ if the convex function, generating H^* is strict convex (see [14]).

3 Majorization

Classical majorization theory has at least two origins, majorization of two finite dimensional vectors in linear algebra and majorization of the rearrangements of two functions in integration theory in Euclidian spaces. In both cases this is an partial order between two objects (two vectors or two integrable functions). In both cases functionals (sums or integrals) of the objects are compared. Thus, implicitely can underlying measure is assumed.
We recall here the main statements of the mentioned theories.

3.1 Classical majorization theory in linear algebra

There is a huge amount of literature on majorization theory in linear algebra. We refer here only to the classical book of Marshal and Olkin [10]. But we can not resist to mention the introduction to this topic in Steele's excellent book [13].
In linear algebra, the classical definition of majorization reads as follows (see, e.g., [10]):
Let $x = (x_i)_{i=1}^n$ and $y = (y_i)_{i=1}^n$ be two real valued sequences and $x^\downarrow$ and $y^\downarrow$ their decreasing rearrangements (the vector with the same components, sorted in decreasing order). We say x majorize y, denoted by $y \prec_{\mathbb{1}} x$ (we explain the reason for the index $\mathbb{1}$ later), if

$$\sum_{i=1}^k y_i^\downarrow \leq \sum_{i=1}^k x_i^\downarrow, \; k = 1, ..., n-1 \tag{17}$$

$$\sum_{i=1}^n y_i^\downarrow = \sum_{i=1}^n x_i^\downarrow = \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \tag{18}$$

A matrix $\mathbf{D}$ is called stochastic, if the entries are nonnegative and the rows sum up to one, i.e., if $\mathbf{D}\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the vector with all components equal to 1.
Note that the ordering of the vectors is a special operation. We have to take two permutation matrices $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ such that $\mathbf{\Pi}_x x = x^\downarrow$ and $\mathbf{\Pi}_y y = y^\downarrow$.
A matrix $\mathbf{D}$ is called doubly stochastic, if both $\mathbf{D}$ and its transposed $\mathbf{D}^*$ are stochastic matrices.
The main theorem in the classical majorization theory in linear algebra states that the following statements are equivalent:

(*i*) **Majorization:** $y \prec_{\mathbb{1}} x$

(*ii*) **Doubly stochastic:** $\exists$ doubly stochastic $\mathbf{D}$ with $y = \mathbf{D}x$

(*iii*) **Karamata's inequality:** $\sum_{i=1}^n F(y_i) \leq \sum_{i=1}^n F(x_i)$ for all convex $F : \mathbb{R} \to \mathbb{R}$

(*iv*) **Hardy-Littlewood inequality:** For all $z \in \mathbb{R}_n$ we have $\sum_{i=1}^n z_i^\downarrow y_i^\downarrow \leq \sum_{i=1}^n z_i^\downarrow x_i^\downarrow$.

Due to the equivalence, any of the above statements can be taken as the definition of majorization. Note that Karamata's inequality does not require the ordering of the vectors, since the sum is invariant with respect to permutations.
Given two majorizing vectors x and y, the doubly stochastic matrix $\mathbf{D}$ connecting them can be found explicitely by the so-called Robin-Hood method (see Appendix B). Moreover, this method is a proof of Birkhoff's famos Theorem, stating that any doubly stochastic matrix is a convex combination of permutation matrices or, in other words, the permutation matrices are the extremal elements of the convex set of doubly stochastic matrices.

3.2 Classical majorization in integration theory

Majorization theory for functions and the connection to integration theory in Euclidian spaces is a powerful tool, e.g., to estimate norms of functions in various function spaces. We refer here to the books of Lieb and Loss [9] and Bennett and Sharpley [2].
As usual majorization of two functions g and f in a functions space on a given nonnegative measure p is defined using a rearrangement of the functions (usually denoted by $g^*(t)$ and $f^*(t)$), similar to the ordering of vectors in linear algebra.
Some authors start the discussion of the rearrangement of functions with a "rearrangement of sets". From a physical point of view, this implies a rearrangement of the underlying state space $\mathcal{Z}$. Moreover, the measure p is fixed and the same for g and f. As usual, the explanation is restricted to non-atomic or uniform and purely atomic measures.
Because the rearrangement of the state space is unphysical and due to the desired larger generality, we try to avoid the notion of rearrangement and define majorization in another way. Previously, we sketch briefly the common way to make the differences clearer. We use here the same notation as it will be used in subsection 3.4.
We denote the level set of the function $g \in \mathcal{C}$ at a point x by

$$L_g(x) = g^{-1}\big((x, \infty)\big) = \big\{z \in \mathcal{Z} | g(z) > x\big\}$$

Due to the continuity of g, this is an open set. From $g_{\min} \le g(z) \le g_{\max}$ it follows

$$\begin{aligned} L_g(x) &= \mathcal{Z}, \; x < g_{\min} \\ L_g(x) &= \emptyset, \; x \ge g_{\max} \\ L_g(x) &\subset L_g(x'), \; x \ge x' \end{aligned}$$

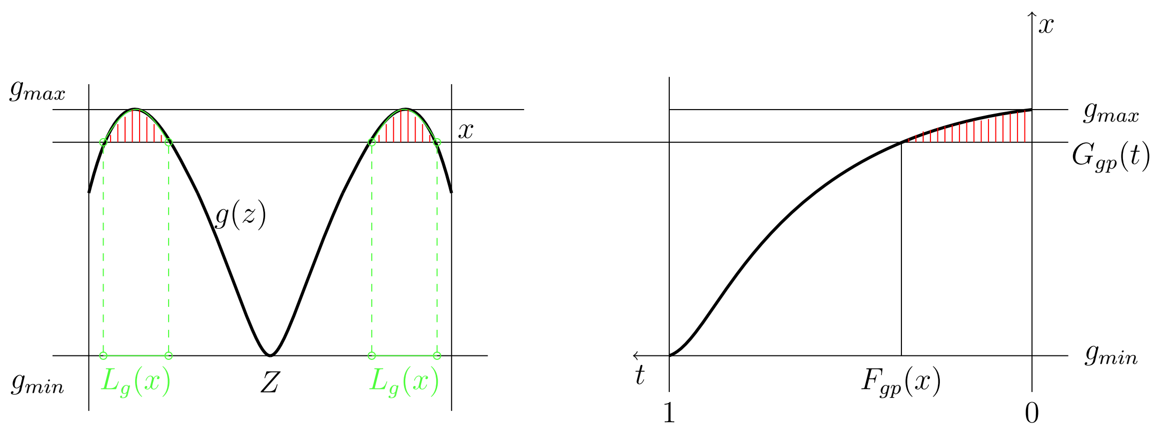
For a given probability measure $p \in \mathcal{S}^*(\mathcal{Z})$ we define by

$$F_{gp}(x) = p\big(L_g(x)\big) = \langle \mathbb{1}_{L_g(x)}, p\rangle = p\Big(g^{-1}\big((x, \infty)\big)\Big) = \big(p \circ g^{-1}\big)(x, \infty) \tag{19}$$

the distribution function of g with respect to p and by

$$G_{gp}(t) = \inf\{x | F_{gp}(x) \le t\}$$

its invers. The picture below illustrate the situation.



The distribution function and its invers have the well known properties

$$\begin{aligned} F_{gp}(x) &= 1,\ x < g_{\min} \\ F_{gp}(x) &= 0,\ x \geq g_{\max} \\ F_{gp}(x) &\leq F_{gp}(x'),\ x \geq x' \\ G_{gp}(t) &\leq G_{gp}(t'),\ t' \geq t \\ F_{gp} &: [g_{\min}, g_{\max}] \rightarrow [0,1] \\ G_{gp} &: [0,1] \rightarrow [g_{\min}, g_{\max}] \end{aligned}$$

In the usual notation we have $G_{gp}(t) = g^*(t)$ and $G_{fp}(t) = f^*(t)$. The measure p is fixed from the begining and must be non-atomic or uniform and purely atomic.
Now, g majorizes f – written $f \prec_p g$ (the index point out the fixed measure) – if

$$\int_0^t f^*(t')dt' \leq \int_0^t g^*(t')dt' \tag{20}$$

for all $t \in [0,1)$ and

$$\int_0^1 f^*(t')dt' = \int_0^1 g^*(t')dt' \ .$$

In our notation this means $\int_0^t G_{gp}(t')dt' \leq \int_0^t G_{fp}(t')dt'$ for $t \in [0,1)$ and $\langle g, p\rangle = \langle f, p\rangle$.
In [2] it is shown that there exists a measure preserving operator in $L_1(p)$, mapping f to g. Clearly, this requires a fixed measure. It is not inverstigated how this operator acts on functions from $\mathcal{C}(\mathcal{Z})$.

3.3 A new definition of majorization

Time irreversibility corresponds to a Markov operator acting on a pair of objects, an observable-measure pair or a pair of two measures. We define majorization of a pair of objects in such a way that one pair in majorized by another pair, if the first is a pair "before" and the second is a pair "after". As in subsection 1.1 we denote by a prime an object after the action of a Markov operator.

Definition 1: Majorization of a pair composed of a continuous function and a probability measure
A pair (g', p') is majorized by (g, p), written $(g', p') \prec (g, p)$ if there exists a Markov operator $\mathbf{M}$ with $g = \mathbf{M}g'$ and $p' = \mathbf{M}^*p$.

Definition 2: Majorization of a pair of measures
A pair (p', q') is majorized by (p, q), written $(p', q') \prec^* (p, q)$ if there exists a Markov operator $\mathbf{M}$ with $p' = \mathbf{M}^*p$ and $q' = \mathbf{M}^*q$.

In the classical theory, majorization is defined by some inequalities, (17) or (20). Then, the existence of an operator mapping the majorized objects to each other is a consequence. This assumes implicitely that a measure is fixed. This is a special case of the general situation. We see no reason to restrict ourself to this special case. Once a measure is not fixed from the

beginning, we concentrate on the main reason of the connection of two majorizated objects – the action of a Markov operator.
Naturally, it would be nice, to have a simple criterion, whether one pair is majorize by another one. In the next subsection we derive inequalities – special cases of Karamata's inequality and similar to (17) and (20) – that such pairs of objects have to satisfy. Unfortunately, we don't know a proof for the reverse statement: If such inequalities are satisfied, does a corresponding Markov operator exists. In special cases this can be shown (see Appendix B, a proof based on the Robin-Hood method). A general theorem we left as a conjecture (subsection 3.6).

3.4 The main functionals for special convex functions

We denote by $x \in \mathbb{R}$ a point of the range of functions from $\mathcal{C}$ and by $t \in [0,1]$ a point of the range of probablility measures from $\mathcal{S}^*$. A general integral $\langle g,p\rangle = \int_{\mathcal{Z}} g(z)p(dz)$ is a Lebesgue integral on the topological space $\mathcal{Z}$. Often, it is much easier to show properties of functions by analizing Riemann integrals instead of Lebesgue integrals. Therefore, a typical approach in integration theory is to consider Riemann integrals over x or t of some other functions, say a and b, resp., depending on g and p. In other words, there are functions $a_{gp}(x)$ and $b_{gp}(t)$ sought, satisfying

$$\langle g,p\rangle = \int_{\mathcal{Z}} g(z)p(dz) = \int_{g_{\min}}^{g_{\max}} a_{gp}(x)dx = \int_0^1 b_{gp}(t)dt \tag{21}$$

In such a transition, only the information is included that actually is contained in g and p. The information of the structure of $\mathcal{Z}$ is lost.
Given, for example, a function g with $g(z_1) = g(z_2)$ for some $z_1 \neq z_2$, this means, the observable g cannot distingwish such points. Fixing such an observable, we decompose the state space in distingwishable states. The function $a_{gp}(x)$ has to collect both, indistinguishably states with respect to a given observable and the measure of the set of such states. This is the physical background of the transition from the Lebesgue integral over the state space to a Riemann integral over the observation results.
The function $a_{gp}(x)$ and $b_{gp}(t)$ will be the subgradients of a pair of special convex function. We comment the classical way of this transition, based on a rearrangements of the function g in subsubsection 3.7.
The proposed method allows to construct Riemann integrals of the mentioned type also for a pair of measures according to the transition from H to H^*. This aproach allows to construct "rearrangements of densities of two measure" even if such a density does not exist because the Radon-Nikodym theorem fails.
We consider the main functionals for a family of special convex functions

$$F_x(\xi) = (\xi - x)_+$$

and their conjugate

$$F_x^*(\theta) = x\theta,\ \theta \in [0,1] \ \text{ and } \ = +\infty \text{ otherwise} .$$

We define the following functionals for a pair of a continuous function and a probability measure and a pair of two measures

$$M_{gp}(x) \;=\; H[g,p]\Big|_{F=F_x} = \langle (g-x)_+, p\rangle,\ x \in \mathbb{R} \tag{22}$$

$$S_{qp}(x) \;=\; H^*[q,p]\Big|_{F=F_x^*} = \sup_{g\in\mathcal{C}_{[0,1]}} \big(\langle g,q\rangle - \langle xg,p\rangle\big),\ x \in \mathbb{R} \tag{23}$$

Since $F_x^{**} = F_x$, we have for a measure $q = \mathbf{Q}_p g$ with continuous density g with respect to p the equality $S_{qp}(x) = M_{gp}(x)$. Moreover, we have $\langle g, p\rangle = \langle \mathbf{Q}_p^{-1} q, p\rangle = \langle \mathbf{Q}_p^{-1} p, q\rangle = \langle \mathbb{1}, q\rangle = 1$.

Lemma 1 *M_{gp} has the following properties:*

(i) $M_{gp}(x) = \langle g, p\rangle - x$, $x < g_{min}$

(ii) $M_{gp}(x) = 0$, $x \geq g_{max}$

(iii) $M_{gp}(\cdot)$ *is monotone non-increasing.*

(iv) $M_{gp}(\cdot)$ *is convex.*

(v) $M_{gp}(x)$ *is continuous for* $x \in \mathbb{R}$.

Proof:

(i) This follows from $(g - x)_+ = g - x$ for $x < g_{\min}$.

(ii) This follows from $(g - x)_+ = 0$ for $x \geq g_{\max}$.

(iii) This follows from the fact that $(g - x)_+$ is monotone non-increasing in x.

(iv) We set $\alpha_{1,2} \in [0, 1]$ with $\alpha_1 + \alpha_2 = 1$. Since $(\xi - x)_+$ is convex as a function of ξ, we have

$$(\xi - \alpha_1 x_1 + \alpha_2 x_2)_+ \leq \alpha_1(\xi - x_1)_+ + \alpha_2(\xi - x_2)_+$$

and therefore from definition (22) and $p \geq 0$

$$M_{gp}(\alpha_1 x_1 + \alpha_2 x_2) \quad \leq \quad \alpha_1 M_{gp}(x_1) + \alpha_2 M_{gp}(x_2)$$

(v) This follows from the convexity of M_{gp}. □

To prove similiar properties for S_{qp} we define

$$X_{qp} = \inf_{B\in\mathcal{B}} \frac{q(B)}{p(B)},\ Y_{qp} = \sup_{B\in\mathcal{B}} \frac{q(B)}{p(B)},\ Z_{qp} = \sup_{B:\ p(B)=0} q(B)$$

Since $p, q \geq 0$ and $\mathcal{Z} \in \mathcal{B}$, we have $0 \leq X_{qp} \leq 1 \leq Y_{qp} \leq +\infty$ and $Z_{qp} \geq 0$.
With a continuous density $g = \mathbf{Q}_p^{-1} q$ we have $X_{qp} = g_{\min}$, $Y_{qp} = g_{\max}$ and $Z_{qp} = 0$.

Lemma 2 *S_{qp} has the following properties:*

(i) $S_{qp}(x) = 1 - x$, $x \leq X_{qp}$.

(ii) $S_{qp}(x) = Z_{qp}$, $x \geq Y_{qp}$.

(iii) $S_{qp}(\cdot)$ *is monotone non-increasing.*

(iv) $S_{qp}(\cdot)$ *is convex.*

(v) $S_{qp}(x)$ *is continuous for* $x \in \mathbb{R}$.

Proof:

(*i*) Since $p, q \in \mathcal{S}^*$, for $x \leq 0$ the expression under the supremum is non-negative and therefore the supremum is attained at the larges function $g = \mathbb{1}$. Hence $S_{qp}(x) = \langle \mathbb{1}, q\rangle - \langle x\mathbb{1}, p\rangle = 1 - x$.

(*ii*) Taking in definition (23) $g \equiv 1$ and $g \equiv 0$, we conclude $S_{qp}(x) \geq 0$ for $x \geq 1$ and $S_{qp}(x) \geq 1 - x$ for $x \leq 0$.

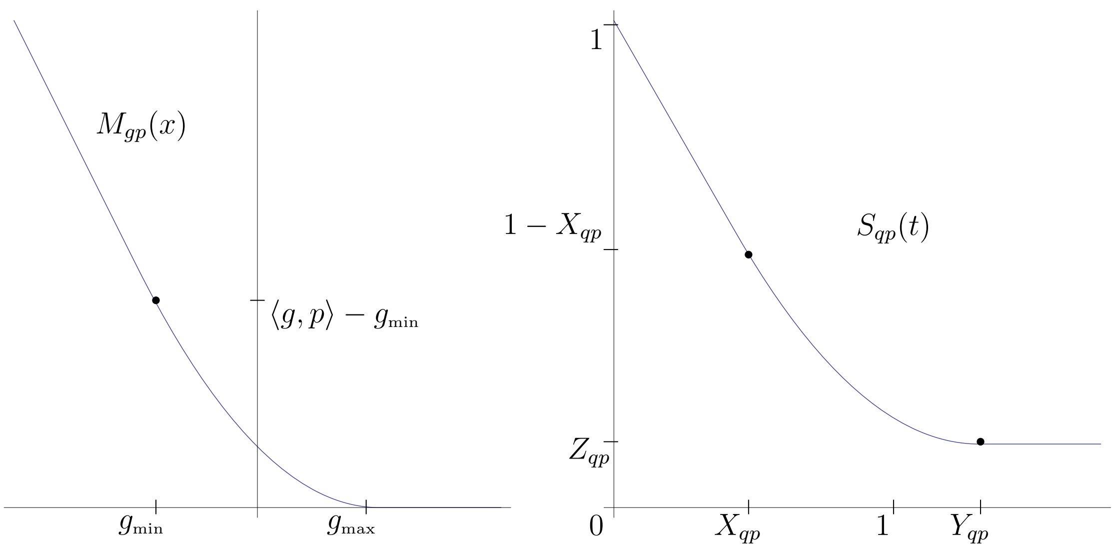
(*iii*) This follows because $S_{qp}(x)$ is a supremum over a function monotone non-increasing in x.

(*iv*) We set $\alpha_{1,2} \in [0,1]$ with $\alpha_1 + \alpha_2 = 1$ and use $\sup(a+b) \leq \sup(a) + \sup(b)$ to show

$$\begin{aligned}
\langle g, q\rangle - \langle(\alpha_1 x_1 + \alpha_2 x_2)g, p\rangle &= \langle(\alpha_1 + \alpha_2)g, q\rangle - \alpha_1\langle x_1 g, p\rangle - \alpha_2\langle x_2 g, p\rangle \\
\langle g, q\rangle - \langle(\alpha_1 x_1 + \alpha_2 x_2)g, p\rangle &= \alpha_1\big(\langle g, q\rangle - \langle x_1 g, p\rangle\big) + \alpha_2\big(\langle g, q\rangle - \langle x_2 g, p\rangle\big) \\
\sup_{g \in \mathcal{C}_{[0,1]}} \big(\langle g, q\rangle - \langle(\alpha_1 x_1 + \alpha_2 x_2)g, p\rangle\big) &\leq \alpha_1 \sup_{g \in \mathcal{C}_{[0,1]}} \big(\langle g, q\rangle - \langle x_1 g, p\rangle\big) + \\
&+ \alpha_2 \sup_{g \in \mathcal{C}_{[0,1]}} \big(\langle g, q\rangle - \langle x_2 g, p\rangle\big) \\
S_{qp}(\alpha_1 x_1 + \alpha_2 x_2) &\leq \alpha_1 S_{qp}(x_1) + \alpha_2 S_{qp}(x_2)
\end{aligned}$$

(*v*) This follows from the convexity of S_{qp}. □

The following pictures illustrate the functionals M_{gp} and S_{qp}.



Beside the functionals M_{gp} and S_{qp} we define some Fenchel-Legendre transforms of them:

$$A_{gp}(t) = \inf_{x \in \mathbb{R}} (xt + M_{gp}(x)), \; t \in [0,1] \tag{24}$$

$$T_{qp}(t) = \inf_{x \in \mathbb{R}_+} (xt + S_{qp}(x)), \; t \in [0,1] \tag{25}$$

As usual, we define $\inf = -\infty$ and $\sup = \infty$ outside of the domain of definition.

The inf-expression is connected with the usual Fenchel-Legendre transform $F^*(t)$ of a function $F(x)$ defined by $F^*(x) = \sup_{x\in\mathbb{R}}(xt - F(x))$ by the relation

$$\inf_{x\in\mathbb{R}}(xt + F(x)) = -\sup_{x\in\mathbb{R}}(-xt - F(x)) = -F^*(-t)$$

Thus, we have $A_{gp}(t) = -M^*_{gp}(-t)$ and $T_{qp}(t) = -S^*_{qp}(-t)$.

Corollary 1 *$M_{gp}(x)$ and $S_{qp}(x)$ can be caluclated from $A_{gp}(t)$ and $T_{qp}(t)$ by*

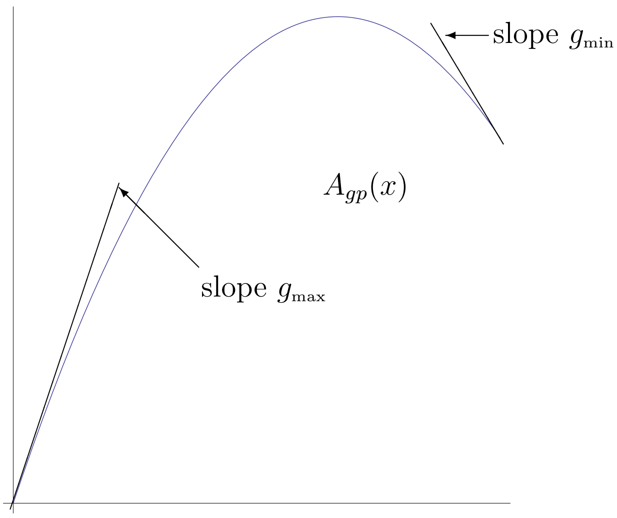
$$M_{gp}(x) = \sup_{t\in[0,1]}(A_{gp}(t) - xt) \quad (26)$$
$$S_{qp}(x) = \sup_{t\in[0,1]}(T_{qp}(t) - xt) \quad (27)$$

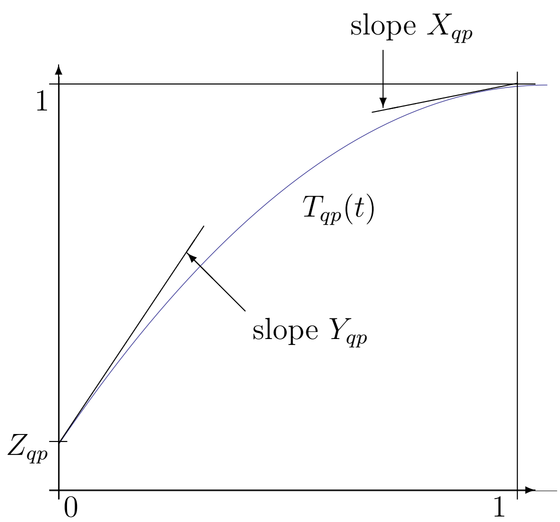
Proof: The right expressions in (26) and (27) are the second convex conjugate or bipolar of $M_{gp}(x)$ and $S_{qp}(x)$. They coincide with $M_{gp}(x)$ and $S_{qp}(x)$ due to convexity. We have

$$M_{gp}(x) = \sup_{t\in[-1,0]}(xt - M^*_{gp}(t)) = \sup_{t\in[-1,0]}(xt + A^*_{gp}(-t)) = \sup_{t\in[0,1]}(-xt + A^*_{gp}(t)) \; ,$$

i.e., (26). (27) is shown analogously. □

The following pictures illustrate the functionals A_{gp} and $T_{qp}(x)$.





Lemma 3 *A_{gp} has the following properties:*

(i) *$A_{gp}(\cdot)$ is concave.*

(ii) *$A_{gp}(0) = 0$.*

(iii) *$A_{gp}(1) = \langle g, p\rangle$.*

Proof:

(i) This follows from definition (24).

(*ii*) From definition (24) we have $A_{gp}(0) = \inf_{x\in\mathbb{R}} M_{gp}(x)$. Thus, $A_{gp}(0) \geq 0$ due to Lemma 1. Since $M_{gp}(g_{\max}) = 0$ we have equality.

(*iii*) Since M_{gp} is convex, it has the larges slope for $x < g_{\min}$. From definition (24) we have $A_{gp}(1) \geq \inf_{x<g_{\min}}(x + M_{gp}(x)) = \inf_{x<g_{\min}}(x + \langle g, p\rangle - x) = \langle g, p\rangle$. But for any $x < g_{\min}$ we have equality. □

A_{gp} has one-side derivatives: $A'_{gp}(+0) = g_{\max}$ and $A'_{gp}(1-0) = g_{\min}$.

Lemma 4 *T_{qp} has the following properties:*

(i) $T_{qp}(\cdot)$ is concave.

(ii) $T_{qp}(\cdot)$ is monotone non-decreasing.

(iii) $T_{qp}(0) = Z_{qp}$.

(iv) $T_{qp}(1) = 1$.

Proof:

(*i*) This follows from definition (25).

(*ii*) This is because T_{qp} has nonnegative slopes from Y_{qp} down to X_{qp}.

(*iii*) We have $T_{qp}(0) = \inf_{x\in\mathbb{R}_+} S_{qp}(x) = Z_{qp}$.

(*iv*) We have $x + S_{qp}(x) \geq x + (1-x) = 1$ with equality for $x \leq X_{qp}$. Hence, $T_{qp}(1) = \inf_{x\in\mathbb{R}_+}(x + S_{qp}(x)) = 1$. □

The connection with the Radon-Nikodym Theorem gives the following

Lemma 5 *The measure q is differentiable with respect to p, if and only if $Z_{qp} = 0$.*

Proof: If $Z_{qp} = 0$, then $p(B) = 0$ implies $q(B) = 0$. Thus, the Radon-Nikodym Theorem holds. Conversely, if $Z_{qp} > 0$, then there is a B with $p(B) = 0$ and $q(B) > 0$. Hence, the Radon-Nikodym Theorem fails.

3.4.1 The subdifferentials

$M_{gp}(\cdot)$ is a convex function. Thus, it has a monotone non-increasing subdifferential ∂M_{gp} satisfying

$$M_{gp}(y) - M_{gp}(x) \geq \partial M_{gp}(x)(y - x),\ x, y \in \mathbb{R}$$

$A_{gp}(\cdot)$ is a concav function. Thus, it has a monotone non-decreasing subdifferential ∂A_{gp} satisfying

$$A_{gp}(s) - A_{gp}(t) \leq \partial A_{gp}(t)(s - t),\ s, r \in [0, 1]$$

We have

$$x \in A_{gp}(t) \iff t \in M_{gp}(x),\ (x, t) \in \Phi_{gp}$$

By Φ_{gp}, we denote the maximal monotone (decreasing) map

$$\begin{aligned}\Phi_{gp} &= \big\{(x,t):\ x\in\mathbb{R},\ t\in\partial M_{gp}(x)\big\}\\ &= \big\{(x,t):\ t\in[0,1],\ x\in\partial A_{gp}(t)\big\}\end{aligned}$$

Analogously we have a monotone non-increasing subdifferential ∂S_{qp} and a monotone non-decreasing subdifferential ∂T_{qp}.
We choose one point from every set of the subdifferentials by setting

$$\begin{aligned}F_{gp}(x) &= -\sup\partial M_{gp}(x)\\ G_{gp}(t) &= \inf\partial A_{gp}(t)\\ U_{qp}(x) &= -\sup\partial S_{qp}(x)\\ V_{qp}(t) &= \inf\partial T_{qp}(t)\end{aligned}$$

F_{gp} and G_{gp} are inverse to each other in the sense of

$$\begin{aligned}G_{gp}(t) &= \inf\{x|F_{gp}(x)\le t\}\\ F_{gp}(x) &= \inf\{t|G_{gp}(t)\le x\}\ .\end{aligned}$$

Similar we have

$$\begin{aligned}V_{qp}(t) &= \inf\{x|U_{qp}(x)\le t\}\\ U_{qp}(x) &= \inf\{t|V_{qp}(t)\le x\}\ .\end{aligned}$$

From the general connections of convex functions and their subdifferentials it is well know that the functions $M_{gp}(x)$, $S_{qp}(x)$, $A_{gp}(t)$ and $T_{qp}(t)$ can be represented as Riemann integrals of the functions $F_{gp}(x)$, $U_{qp}(x)$, $G_{gp}(t)$ and $V_{qp}(t)$ resp. To regard all cases, we define the endpoints of the integration intervalls not depending on g,p,q, using the invariant values $M_{gp}(+\infty)=0$, $A_{gp}(0)=0$, $S_{qp}(0)=0$ and $T_{qp}(1)=1$. Hence, we set

$$M_{gp}(x) = \int_x^{g_{\max}} F_{gp}(x')dx' \tag{28}$$
$$A_{gp}(t) = \int_0^t G_{gp}(t')dt',\ t\in[0,1] \tag{29}$$

and

$$S_{qp}(x) = 1-\int_0^x U_{qp}(x')dx' \tag{30}$$
$$T_{qp}(t) = 1-\int_t^1 V_{qp}(t')dt',\ t\in[0,1] \tag{31}$$

The functionals $M_{gp}(x)$, $S_{qp}(x)$, $A_{gp}(t)$ and $T_{qp}(t)$ provide a number of other representations. This will be the main result of the subsubsections 3.4.5 and 3.4.6. Before presenting them, we prove a couple of lemmas.

3.4.2 The level sets

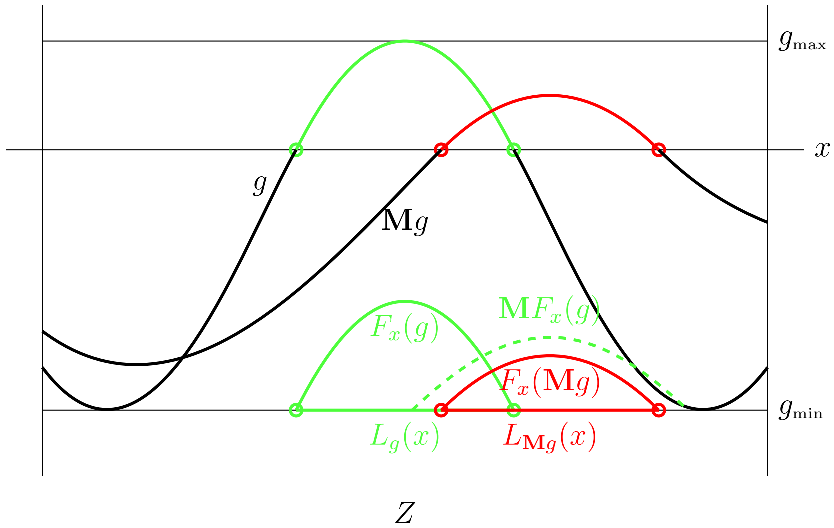
We recall the definition of the level set of the function $g \in \mathcal{C}$ at a point x by

$$L_g(x) = g^{-1}\big((x, \infty)\big) = \big\{z \in \mathcal{Z} | g(z) > x\big\}$$

Obviously,

$$\langle (g - x)_+, p\rangle = \langle g - x, p\rangle_{L_g(x)} = \langle g, p\rangle_{L_g(x)} - xp\big(L_g(x)\big) \ . \tag{32}$$

The following picture is an illustration of Jensen's inequality (5) for the special case $F = F_x$, i.e. $\mathbf{M}F_x(g) \geq F_x(\mathbf{M}g)$ and the connection of level sets and $F_x(g) = (g - x)_+$.



3.4.3 The range of a measure

In the sequel we compare these functionals for some measure p with the same functional for the measure $\mathbf{M}^*p$. We denote by

$$E_p = \{t \in [0, 1] : \exists B \in \mathcal{B},\ p(B) = t\},$$

the range of the measure p as a function of Borel sets. Since, we restrict ourselves to probability measures, $E_p \subset [0, 1]$. A main assumption of the classical majorization theory is $E_p = T_{\mathbf{M}^*p}$. This is clear, because the idea of rearrangement of a function or the rewriting of a vector in decreasing order is only possible, if the measures of the underlying sets before and after the rearrangement are "comparable". This is only possible, if 1) $p = \mu$ (starting with a fixed measure and considering all with respect to this measure); 2) $E_p = [0, 1]$ (the non atomic case) and 3) $E_p = \{0, 1/n, 2/n, ..., (n - 1)/n, 1\}$ (the uniform, pure atomic case).
Since the main idea of this paper is to consider arbitrary situations, we try to define majorization without using the idea of rearrangement. The consideration of general sets E_p leads to the most technical difficulties in the sequel. The pure atomic but non-uniform case is considered in Appendix A.

3.4.4 Two extremal properties of the level set

The following Lemma shows that under all Borel sets B, the level sets maximize the value of $\langle g, p\rangle_B - xp(B)$.

Lemma 6 *Let $g \in \mathcal{C}$, $B \in \mathcal{B}(\mathcal{Z})$ and $x \in \mathbb{R}$ be arbitrary. Then, the following inequality holds*

$$\langle g, p\rangle_B - xp(B) \leq \langle g, p\rangle_{L_g(x)} - xp(L_g(x)) \tag{33}$$

Proof: We have $g(z) > x$ for $z \in L_g(x)$ and $g(z) \leq x$ for $z \notin L_g(x)$. Hence, from

$$\begin{aligned} L_g(x) &= (L_g(x) \setminus B) \cup (B \cap L_g(x)) \\ B &= (B \setminus L_g(x)) \cup (B \cap L_g(x)) \end{aligned}$$

follows

$$\begin{aligned} \langle g, p\rangle_{L_g(x)} - \langle g, p\rangle_B &= \langle g, p\rangle_{L_g(x)\setminus B} + \langle g, p\rangle_{B\cap L_g(x)} - \langle g, p\rangle_{B\setminus L_g(x)} - \langle g, p\rangle_{B\cap L_g(x)} = \\ &= \langle g, p\rangle_{L_g(x)\setminus B} - \langle g, p\rangle_{B\setminus L_g(x)} \geq \\ &\geq xp(L_g(x) \setminus B) + xp(B \cap L_g(x)) - xp(B \setminus L_g(x)) - xp(B \cap L_g(x)) = \\ &= xp(L_g(x)) - xp(B) \end{aligned}$$

This proves the Lemma. □

Corollary 2 *Let $g \in \mathcal{C}$ and $x, y \in \mathbb{R}$ be arbitrary. Then, the following inequality holds*

$$\langle g, p\rangle_{L_g(x)} - \langle g, p\rangle_{L_g(y)} \geq x\Big(p\big(L_g(x)\big) - p\big(L_g(y)\big)\Big) \tag{34}$$

Proof: This is (33) with $B = L_g(y)$. □

Lemma 7 $F_{gp}(x) = p\big(L_g(x)\big)$.

Proof: From Corrolary 2 we have

$$\langle g, p\rangle_{L_g(y)} - \langle g, p\rangle_{L_g(x)} \geq y\Big(p\big(L_g(y)\big) - p\big(L_g(x)\big)\Big)$$

Adding $xp\big(L_g(x)\big)$ we conclude

$$\Big(\langle g, p\rangle_{L_g(y)} - yp\big(L_g(y)\big)\Big) - \Big(\langle g, p\rangle_{L_g(x)} - xp\big(L_g(x)\big)\Big) \geq -p\big(L_g(x)\big)(y - x)$$

Hence, $M_{gp}(y) - M_{gp}(x) \geq -p\big(L_g(x)\big)(y - x)$ for all $x, y \in \mathbb{R}$. Thus, $-p\big(L_g(x)\big) \in M_{gp}(x)$. □

3.4.5 Lebesgue and Riemann integral representations

We collect some integral representations starting with the so-called layer cake representation.

Lemma 8 *Let $F(x)$ be a function with a Riemann integrable derivative. Then*

$$\langle F(g), p\rangle = \int_{\mathcal{Z}} F\big(g(z)\big)p(dz) = \int_{g_{min}}^{g_{max}} F_{gp}(x)F'(x)dx + F(g_{min})$$

holds.

Proof: Beside the set $L_g(x)$ we consider the set

$$C_g = \{(z,x) \mid g(z) > x,\ x \ge g_{\min}\}$$

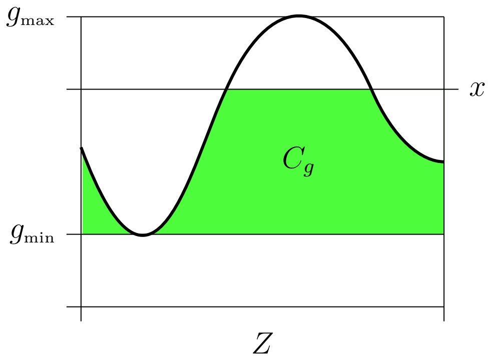
i.e., C_g is the set under the function g from x down to $g_{\min}$. For $\mathbb{1}_{C_g}$ – the characteristic function of C_g – we have

$$\int_{g_{\min}}^{g_{\max}} \mathbb{1}_{C_g}(z,x)h(x)dx = \int_0^{g(z)} h(x)dx$$

(this is a cut along the set C_g for fixed z) and

$$\int_{\mathcal{Z}} \mathbb{1}_{C_g}(z,x)p(dz) = \int_{L_g(x)} p(dz) = F_{gp}(x)$$

because for fixed x, the cut of C_g is the level set $L_g(x)$. Using this equalities we get from Fubini's Theorem



$$\begin{aligned}\langle F(g),p\rangle - F(g_{\min}) &= \langle F(g)-F(g_{\min}),p\rangle = \int_{\mathcal{Z}} \big(F(g(z)) - F(g_{\min})\big)p(dz) = \\ &= \int_{\mathcal{Z}} \left(\int_{g_{\min}}^{g(z)} F'(x)dx\right) p(dz) = \int_{\mathcal{Z}} \left(\int_{g_{\min}}^{g(z)} \mathbb{1}_{C_g}(z,x)F'(x)dx\right) p(dz) = \\ &= \int_{g_{\min}}^{g_{\max}} \left(\int_{\mathcal{Z}} \mathbb{1}_{C_g}(z,x)p(dz)\right) F'(x)dx = \int_{g_{\min}}^{g_{\max}} F_{gp}(x)F'(x)dx\ .\end{aligned}$$

Hence,

$$\langle F(g),p\rangle = \int_{g_{\min}}^{g_{\max}} F_{gp}(x)F'(x)dx + F(g_{\min}) \qquad \square \tag{35}$$

Since $F_{gp}(x)$ and $G_{gp}(t)$ are inverse to each other, for suitable functions F and G, we can write the integration by parts rule as an identity for Riemann integrals:

$$\int_0^1 F(G_{gp}(t))G'(t)dt + G(0)F(g_{\max}) = G(1)F(g_{\min}) + \int_{g_{\min}}^{g_{\max}} G\big(F_{gp}(x)\big)F'(x)dx \tag{36}$$

The special case $G(t) = t$ leads to the following identity:

$$\int_0^1 F(G_{gp}(t))dt = F(g_{\min}) + \int_{g_{\min}}^{g_{\max}} F_{gp}(x)F'(x)dx \tag{37}$$

Together with (35) we obtain

$$\langle F(g),p\rangle = \int_0^1 F(G_{gp}(t))dt = F(g_{\min}) + \int_{g_{\min}}^{g_{\max}} F_{gp}(x)F'(x)dx \tag{38}$$

This is the desired representation (21) of the Lebesgue integral as Riemann integrals over the range of the continuous function and over the range of the measure, resp.

In (36), The special case $F(x) = x$ leads to the identity

$$\int_0^1 G_{gp}(t)G'(t)dt + G(0)g_{\max} \;=\; G(1)g_{\min} + \int_{g_{\min}}^{g_{\max}} G\big(F_{gp}(x)\big)dx \qquad (39)$$

Integrating by parts ones more, from (38) and (36), we obtain identities containing M_{gp} and A_{gp}:

$$\int_{g_{\min}}^{g_{\max}} F_{gp}(x)F'(x)dx \;=\; F'(g_{\min})(\langle g,p\rangle - g_{\min}) + \int_{g_{\min}}^{g_{\max}} M_{gp}(x)F''(x)dx \qquad (40)$$

$$\int_0^1 G_{gp}(t)G'(t)dt \;=\; \langle g,p\rangle G'(1) - \int_0^1 A_{gp}(t)G''(t)dt \qquad (41)$$

Remark: All identities remain true, if we replace $g_{\min}$ by some $c_- \leq g_{\min}$ and $g_{\max}$ by some $c_+ \geq g_{\min}$.
The identities allow a representation of the functional H for an arbitrary, but smooth enought, convex function F in terms of piecewise affine convex functions F_x.

Corollary 3 *The following identity holds for $c \leq g_{min}$:*

$$\langle F(g),p\rangle \;=\; F(c) + F'(c)(\langle g,p\rangle - c) + \int_c^{g_{max}} \langle (g-x)_+,p\rangle F''(x)dx \qquad (42)$$

Proof: This is a consequence of (38) and (40) together with the definition of M_{gp}. □
Identity (42) can be regarded as a consequence of the Taylor expansion for $F\big(g(z)\big)$:

$$F\big(g(z)\big) \;=\; F(c) + \big(g(z)-c\big)F'(c) + \int_c^{g(z)} \big(g(z)-x\big)F''(x)dx$$

Analogously, we obtain the following identities containing S_{qp}, T_{qp}, U_{qp} and V_{qp}.

$$\int_0^1 F(V_{qp}(t))G'(t)dt + G(0)F(Y_{qp}) = G(1)F(X_{qp}) + \int_{X_{qp}}^{Y_{qp}} G\big(U_{qp}(x)\big)F'(x)dx$$

$$\int_0^1 F(V_{qp}(t))dt \;=\; F(X_{qp}) + \int_{X_{qp}}^{Y_{qp}} U_{qp}(x)F'(x)dx$$

$$\int_0^1 V_{qp}(t)G'(t)dt + G(0)Y_{qp} \;=\; G(1)X_{qp} + \int_{X_{qp}}^{Y_{qp}} G\big(U_{qp}(x)\big)dx$$

$$\int_{X_{qp}}^{Y_{qp}} U_{qp}(x)F'(x)dx \;=\; F'(X_{qp})(1-X_{qp}) - F'(Y_{qp})Z_{qp} + \int_{X_{qp}}^{Y_{qp}} S_{gp}(x)F''(x)dx$$

$$\int_0^1 V_{qp}(t)G'(t)dt \;=\; \langle g,p\rangle G'(1) - \int_0^1 T_{qp}(t)G''(t)dt$$

$$H^*[q,p] \;=\; \int_0^1 F^*(V_{qp}(t))dt = F^*(X_{qp}) + \int_{X_{qp}}^{Y_{qp}} U_{qp}(x)F^{*\prime}(x)dx$$

This identities can be used, if q has to Radon-Nikodym derivative with respect to p. In this case, $Y_{qp} = +\infty$ and $Z_{qp} > 0$ is possible. Nevertheless, for suitable functions F the integrals can be finite.

3.4.6 Equivalent representations for the special functionals

For practical reasons we show some other ways for the calculation of the functionals.

Theorem 3 *The following equivalent representations of the above defined functionals hold:*

$$M_{gp}(x) = \sup_{B\in\mathcal{B}} \big(\langle g,p\rangle_B - xp(B)\big) \tag{43}$$

$$A_{gp}(t) = \text{conv-hypo}\left(\sup_{B:p(B)=t} \langle g,p\rangle\right),\ t\in[0,1] \tag{44}$$

$$S_{qp}(x) = \sup_{B\in\mathcal{B}} \big(q(B) - xp(B)\big) \tag{45}$$

$$T_{qp}(t) = \text{conv-hypo}\left(\sup_{B:p(B)\le t} q(B)\right),\ t\in[0,1] \tag{46}$$

where conv-hypo *is the convex hull of the hypo-graph.*

Proof:
Since $L_g(x)$ is a Borel set, (43) follows from (32) and (33) of Lemma 6.
To show (45), we notice that continuous functions in $\mathcal{C}_{[0,1]}$ can be approximated from below by characteristic functions of Borel sets. Hence,

$$\sup_{g\in\mathcal{C}_{[0,1]}} \big(\langle g,q\rangle - \langle xg,p\rangle\big) = \sup_{B\in\mathcal{B}} \big(\langle \mathbb{1}_B,q\rangle_* - x\langle \mathbb{1}_B,p\rangle_*\big) = \sup_{B\in\mathcal{B}} \big(q(B) - xp(B)\big)$$

This shows (45).
To show (44) we use definition (24) and (43) and the well known fact that the change of the order of sup and inf gives the convex (or concave) hull of the expression. The "concave hull" is the convex hull of the hypo-graph of the function. We have

$$\begin{aligned}
A_{gp}(t) &= \inf_{x\in\mathbb{R}}(xt + M_{gp}(x)) = \inf_{x\in\mathbb{R}}(xt + \sup_{B\in\mathcal{B}}\big(\langle g,p\rangle - xp(B)\big)) = \\
&= \inf_{x\in\mathbb{R}}\sup_{B\in\mathcal{B}}(xt + \langle g,p\rangle - xp(B)) = \\
&= \text{conv-hypo}\left(\sup_{B\in\mathcal{B}}\inf_{x\in\mathbb{R}}(xt + \langle g,p\rangle - xp(B))\right) = \\
&= \text{conv-hypo}\left(\sup_{B\in\mathcal{B}}\big(\langle g,p\rangle + \inf_{x\in\mathbb{R}} x(t-p(B))\big)\right) = \\
&= \text{conv-hypo}\left(\sup_{B\in\mathcal{B}}\big(\langle g,p\rangle - \chi_{\{t=p(B)\}}\big)\right) = \text{conv-hypo}\left(\sup_{B:p(B)=t}\langle g,p\rangle_B\right)
\end{aligned}$$

(46) can be shown in a similar way using definition (25) and (45).

$$\begin{aligned}
T_{qp}(t) &= \inf_{x\in\mathbb{R}}(xt + S_{gp}(x)) = \inf_{x\in\mathbb{R}}(xt + \sup_{B\in\mathcal{B}}\big(q(B) - xp(B)\big)) = \\
&= \text{conv-hypo}\left(\sup_{B\in\mathcal{B}}\big(\langle g,p\rangle + \inf_{x\in\mathbb{R}_+} x(t-p(B))\big)\right) = \\
&= \text{conv-hypo}\left(\sup_{B\in\mathcal{B}}\big(\langle g,p\rangle - \chi_{p(B)\le t}\big)\right) = \text{conv-hypo}\left(\sup_{B:p(B)\le t} q(B)\right)
\end{aligned}$$

Here χ_A is the characteristic function in convex analysis, equal 0 on A and $+\infty$ otherwise. $\square$
Remark: From the last two steps of the proof we can conclude

$$\begin{aligned} A_{gp}(t) &= \sup_{B:p(B)=t} \langle g,p\rangle_B, \ t \in E_p \\ T_{qp}(t) &= \sup_{B:p(B)\leq t} q(B), \ t \in E_p \end{aligned}$$

3.5 Majorization inequalities

For a practical test, whether a pair of objects is majorized by another pair, some simple inequalities, similar to the classical inequalities (17), we prove the following theorems. Note, these theorems are tests for a majorization, if the existence of a Markov operator – Conjectures 2 and 3 from subsection 3.6 – is proved to be true.

3.5.1 Inequalities for a pair of a continuous function and a probability measure

A Markov operator is band conserving. Therefore, we have $[g_{\min}, g_{\max}] \subset [g'_{\min}, g'_{\max}]$. For the following we choose two real numbers c_+ and c_- satisfying

$$-\infty < c_- \leq g'_{\min} \leq g_{\min} \leq g_{\max} \leq g'_{\max} \leq c_+ \leq +\infty \ .$$

Taking c_- and c_+ as integral bounds, all integrals are taken over the range of $g = \mathbf{M}g'$ as well as g'.

Theorem 4 *Let g and g' be two functions from $\mathcal{C}(\mathcal{Z})$ and p and p' be two probability measures from $\mathcal{S}^*$ satisfying $\langle g',p'\rangle = \langle g,p\rangle$. Then, the following statements are equivalent:*

(i) $M_{gp}(x) \leq M_{g'p'}(x)$ *for all* $x \in [c_-, c_+]$

(ii) $A_{gp}(t) \leq A_{g'p'}(t)$ *for all* $t \in [0,1]$

(iii) $\langle F(g),p\rangle \leq \langle F(g'),p'\rangle$ *(Karamata's inequality)*

(iv) For any differentiable non-decreasing $h(x)$ we have

$$\int_{c_-}^{c_+} h(x)F_{gp}(x)dx \leq \int_{c_-}^{c_+} h(x)F_{g'p'}(x)dx \quad \textit{(Hardy-Littlewood inequality)}$$

(v) For any differentiable non-decreasing $r(t)$ we have

$$\int_0^1 r(t)G_{gp}(t)dt \leq \int_0^1 r(t)G_{g'p'}(t)dt$$

Proof: $F(x) = (s-x)_+$ is a special convex function. This shows $(iii) \implies (i)$.
$(i) \implies (iii)$ is a consequence from the fact that every convex function can be represented as a weak limit of a combination of piecewise affine functions, i.e., functions of type $(s-x)_+$. An alternative possibility is to use identity (42). Subtracting them for (g,p) and (g',p') we obtain

$$\langle F(g),p\rangle - \langle F(g'),p'\rangle = \int_{c_-}^{c_+} (M_{gp}(x) - M_{g'p'}(x))F''(x)dx$$

and the claim follows, since $F'' \geq 0$.
We show $(i) \implies (ii)$. We have for every $x \in [c_-, c_+]$, $M_{gp}(x) \leq M_{g'p'}(x)$. Hence for all x and $t \in [0,1]$

$$xt + M_{gp}(x) \leq xt + M_{g'p'}(x)$$

and therefore from Definition (24) $A_{gp}(t) \leq A_{g'p'}(t)$.
$(ii) \implies (i)$ can be shown in the same manner, using Corollary (1).
To show $(iv) \iff (i)$ we write

$$\begin{aligned}\int_{c_-}^{c_+} h(x)F_{g'p'}(x)dx &= \int_{c_-}^{c_+} h'(x)M_{g'p'}(x)dx - h(c)\langle g', p'\rangle \\ \int_{c_-}^{c_+} h(x)F_{gp}(x)dx &= \int_{c_-}^{c_+} h'(x)M_{gp}(x)dx - h(c)\langle g, p\rangle\end{aligned}$$

Subtraction gives

$$\begin{aligned}\int_{c_-}^{c_+} h(x)\big(F_{g'p'}(x) - F_{gp}(x)\big)dx &= \int_{c_-}^{c_+} h'(x)\big(M_{g'p'}(x) - M_{gp}(x)\big)dx - \\ &- h(c)\big(\langle g', p'\rangle - \langle g, p\rangle\big)\end{aligned}$$

and, since $\langle g', p'\rangle = \langle g, p\rangle$,

$$\int_{c_-}^{c_+} h(x)\big(F_{g'p'}(x) - F_{gp}(x)\big)dx = \int_{c_-}^{c_+} h'(x)\big(M_{g'p'}(x) - M_{gp}(x)\big)dx$$

Now, $(i) \implies (iv)$ follows because of $h'(x) \geq 0$. $(iv) \implies (i)$ follows setting for $h'(x)$ a sequence, tending to the point measure δ_x.
$(i) \implies (v)$ follows in a similar way as $(i) \implies (iv)$. □

3.5.2 Inequalities for a pair of measures

Analogously,we can derive a couple of majorization inequalities for the functionals, depending of a pair of measures. Similarly, we choose two real numbers c_+ and c_- satisfying

$$0 \leq c_- \leq X_{qp} \leq X_{q'p'} \leq 1 \leq Y_{q'p'} \leq Y_{qp} \leq c_+ \leq +\infty\ .$$

Theorem 5 *Let q and q' and p and p' be two pairs of probability measures from $\mathcal{S}^*$. Then, the following statements are equivalent:*

(i) $S_{q'p'}(x) \leq S_{qp}(x)$ *for all* $x \geq 0$

(ii) $T_{q'p'}(t) \leq T_{qp}(t)$ *for all* $t \in [0,1]$

(iii) $H^*[q', p'] \leq H^*[q, p]$

(iv) For any differentiable non-decreasing $h(x)$ we have

$$\int_{c_-}^{c_+} h(x)U_{q'p'}(x)dx \leq \int_{c_-}^{c_+} h(x)U_{qp}(x)dx$$

(v) For any differentiable non-decreasing $r(t)$ we have

$$\int_0^1 r(t)V_{q'p'}(t)dt \leq \int_0^1 r(t)V_{qp}(t)dt$$

Proof: The proof is a verbatim repetition of the corresponding statements of Theorem 4 □

Corollary 4 *Given a Markov operator $\mathbf{M}$. If a measure q has a density with respect to p, then $\mathbf{M}^*q$ has a density with respect to $\mathbf{M}^*p$.*

Proof: If $q \ll p$, then $Z_{qp} = 0$. From Theorem 5 we have $S_{q'p'}(x) \leq S_{qp}(x)$ for $q' = \mathbf{M}^*q$ and $p' = \mathbf{M}^*p$. Since $S_{q'p'}(x) \geq 0$ we have $Z_{q'p'} = 0$. Hence $q' \ll p'$. □

3.6 A couple of conjectures

For a closed theory, it would be necessary to prove the following statements. We formulate them here as conjectures.

Conjecture 1 *For a given Markov operator $\mathbf{M}$ and a probability measure p, there is a Markov operator $\mathbf{X}$, satisfying $\mathbf{Q}_{p'}\mathbf{X} = \mathbf{M}^*\mathbf{Q}_p$ with $p' = \mathbf{M}^*p$.*

Conjecture 2 *Given two pairs of a continuous function and a probability measure (g, p) and (g', p') with $\langle g, p\rangle = \langle g', p'\rangle$. For all $x \in \mathbb{R}$ the inequalities $M_{gp}(x) \leq M_{g'p'}(x)$ hold. Then, there exists a Markov operator $\mathbf{M}$ with $g = \mathbf{M}g'$ and $p' = \mathbf{M}^*p$.*

Conjecture 3 *Given two pairs of measures (q, p) and (q', p') with $q \neq p$. For all $x \in \mathbb{R}$ the inequalities $S_{q'p'}(x) \leq S_{qp}(x)$ hold. Then, there exists a Markov operator $\mathbf{M}$ with $q' = \mathbf{M}^*q$ and $p' = \mathbf{M}^*p$.*

Conjecture 4 *Given two concave functions $T_1(t), T_2(t)$ on $t \in [0, 1]$ with $T_1(1) = T_2(1)$ and $T_1(t) \leq T_2(t)$ for all $t \in [0, 1]$. Then, there exist four probability measures p, q, p', q' with $T_1 = T_{q'p'}$ and $T_2 = T_{qp}$.*

Remark: Birkhoff's Theorem states that there exists a doubly stochastic matrix, being the convex combination of the extremal elements of doubly stochastic matrices – the permutations. In general, such a theorem can not be expected. A Markov operator, even if it exists, is not a convex combination of the extremal elements of Markov operators, since this set is to small. In the worst case, for example, if $\mathcal{Z}$ is a connected subset of $\mathbb{R}^n$, it consists only of one element, namely the identity.

3.7 Discussion of the classical theory as a special case

The definition of majorization in linear algebra starts with the partial ordering of two given vectors x and y in $\mathbb{R}_n$, i.e. with the definition of two permutation matrices $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$. From a physical point of view, a permutation matrix defines a renumbering of the states of the underlying state space. Since in general, the permutation matrices $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ are different, this implies that the state space is renumbered in a different way what is of cause unphysical unless, the states of the space are indistinguishable. In such a space only the uniform measure

– we denote it $\mathbb{1}^* = \frac{1}{n}(1, 1, ..., 1)$ – has a physical sense. This is the field of application of the classical majorization theory.
In finite dimension, the space $\mathbb{R}_n$ can be identified with its dual $\mathbb{R}_n^*$ unless a norm, different to the euclidian, is defined. Thus, we cannot give to the vectors x and y a physical meaning as extensive or intensive value. Moreover, dealing with doubly stochastic matrices, it is not clear, whether $\mathbf{D}$ or $\mathbf{D}^*$ has to be understood as a Markov matrix. If we consider $\mathbf{D} = \mathbf{M}$ as Markov matrix, $\mathbf{D}^*\mathbb{1}^* = \mathbb{1}^*$ means that the uniform measure is the invariant measure of the adjoint. With $y = \mathbf{D}x$, $y \prec_{\mathbb{1}} x$ means $(x, \mathbb{1}^*) \prec (y, \mathbb{1}^*)$.
If we consider $\mathbf{D}^* = \mathbf{M}$ as Markov matrix, then $\mathbf{D}^{**} = \mathbf{D} = \mathbf{M}^*$, since the finite dimensional euclidian space is reflexive $\mathbb{R}_n^{**} = \mathbb{R}_n$. Now, $\mathbf{D}\mathbb{1}^* = \mathbb{1}^*$ is the equation for the invariant measure and $y \prec_{\mathbb{1}} x$ means $(y, \mathbb{1}^*) \prec^* (x, \mathbb{1}^*)$.
For a given Markov operator, it is possible to single out a special measure, namely an invariant measures. Without a pre-specified Markov operator we have to identify a reference measure for the topological space. Actually, this assumes that we have in $\mathcal{Z}$ a further structure, for example, $\mathcal{Z}$ is a topological group. Then, Haar's measure can be single out.
As every Markov operator, a doubly stochastic matrix satisfy the maximum principle (is band conserving). Thus, the components of $y = \mathbf{D}x$ are less spread out then the components of x. The same is true for $\mathbf{D}^*$. But the adjoint of a Markov operator does not satisfy the maximum principle. The result of such an operator is "nearer" to an invariant measure of this operator. For a doubly stochastic matrix the invariant measure of the adjoint is constant. Thus, being "nearer" to the constant is actually the same as to be less spread out.
An equivalent of Birkhoff's Theorem, stating that any doubly stochastic matrix is a convex combination of permutation matrices, is trivial: A Markov matrix is the convex combination of deterministic Markov matrices – matrices with exactly one 1 in each row.
The substantial part of Birkhoff's Theorem is the description of the extremal elements (permutaions) of the convex subset of doubly stochastic matrices. For a given Markov operator with invariant measure μ it can be found a doubly stochastic matrix $\mathbf{D}$ with the connection

$$\mathbf{Q}_\mu(\mathbf{M} - \mathbf{I}) = \mathbf{D} - \mathbf{I}$$

(here $\mathbf{I}$ is the identity). Thus, $\mathbf{M} - \mathbf{I}$ can be written as a convex combination of permutations, multiplied by $\mathbf{Q}_\mu^{-1}$. This seems to be not very interesting.
The infinite dimensional analogon of a stochastic matrix is a Markov operator acting in the Banach space of continuous functions on a topological space. In a general Banach space we have no operation "transponation". Defining the adjoint of an operator as its "transposed", we have to declare what means "conserving of constant functions". Some authors define doubly stochastic operators as Markov operators, which adjoint has the uniform distribution as the invariant measure. But this is a special case as well, because it is not clear, what is the uniform distribution in general, for example, if $\mathcal{Z}$ is disconnected. Again, this requires an additional structure in $\mathcal{Z}$.
Thus, the notion of doubly stochastic matrices cannot be generalized in a natural manner.
Karamata's inequality does not require ordering and can be generalized. Clearly, inequality (6) is this generalization. However, inequality $M_{gp}(x) \leq M_{g'p'}(x)$ (Theorem 4) is not a generalization of inequality (17). For the special case of a doubly stochastic matrix, inequality $M_{gp}(x) \leq M_{g'p'}(x)$ means: $y_1 + ... + y_{i(c)} \leq x_1 + ... + x_{j(c)}$ for any real c, where $i(c)$ and $j(c)$ are the largest indices with $y_i > c$ and $x_j > c$.

In integration theory, majorization is defined via level sets and rearrangements of functions in a function space on a measure, fixed from the beginning. Like in the finite dimensional case the rearrangement of a function is, actually, the rearrangement of the underlying state space and therefore un-physically.
Moreover, majorization in integration theory is based on the pointwise comparison of the distribution function. This requires the restriction to non-atomic or pure atomic and uniform measures.
We prefer a construction based on the main functionals. That allows to consider general probability measures. In Appendix A examples with different supporting points are given.
In integration theory, the object $\int_0^t g^*(t')dt'$ is defined for the absolute value $|g|$. We defined $A_{gp}(\cdot)$ for g. Thus, if $g_{\min}$ is negative, $A_{gp}(\cdot)$ can be a non-decreasing function.

In the literature there are already proposed generalizations of majorization. We refer here to some of them.
A kind of majorization between two vectors with respect to a fixed measure was used by several authors. Actually, [4] seems to be the first one.
Grinberg [5] proposed a majorization of pairs of vectors ("weighted arrays") similar to our Definition 1 for the finite dimensional case. In contrast to [5] where the convex functions $|g-x|$ are used for the approximation of convex functions, we used the convex functions $(g-x)_+$. This seems more natural, because a Markov operator is band conserving.
[5] as well as the majority of the authors propose a definition of majorization in the opposite direction, where "$\prec$" means "$\leq$" in the corresponding inequalities. The direction we used was already proposed by physicists (see [17] for a collection of references). In our definition $(g', p') \prec (g, p)$ holds, if a functional of (g', p') is larger then the same functional of (g, p). This is not typical. Nevertheless, we prefer our definition for a couple of reasons: At first, it coincides with the direction of time. Thus, this is the right direction from a physical point of view.
It seems, the confusion comes from the fact that the direction of the action of a Markov operator is ignored. It acts backwards in time and the result is less "spread out" before the action.
Second, given two convex functions and their conjugates, we have equality in Young's inequality on the monotone graph:

$$\begin{aligned} F_1(x) + F_1^*(y) &= xy \\ F_2(x) + F_2^*(y) &= xy \end{aligned}$$

Obviously we have

$$F_1(x) \leq F_2(x) \iff F_2^*(y) \leq F_1^*(y)$$

Thus, the direction of the inequality changes with the change of the type of the compared objects. Nevertheless, the situation "1" is related to "2" is the same. There is no reason, to denote the first case "1" $\prec$ "2", but the second case "2" $\prec$ "1".
And last, comparing pairs, the inequalities $H[g', p'] \geq H[g, p]$ and $H^*[q', p'] \leq H^*[q, p]$ show that a entropy functional decreases for states but increases for observables and states. Thus, the influence of an observable on the entropy is larger than that of a state.

Appendix A
The discrete and non-uniform distributed case

The case, when $\mathcal{Z} = \{z_1, ..., z_n\}$ is a finite set with discrete topology, is a typical case, when the usual methods for problems with non-atomic measures fail. For a better understanding of the abstract theorems in the previous sections, we summarize here the main definitions and results for the discrete case. As it is well known, in this case all functions on $\mathcal{Z}$ are continuous, $\mathcal{C}(\mathcal{Z}) = \mathbb{R}_n$ and the set of probability measures is the simplex $\mathcal{S}_n^* \subset \mathbb{R}_n^*$. We write $\mathbb{R}_n$ and $\mathbb{R}_n^*$ to distinguish the different norms in the n-dim space.
We fix a function $g \in \mathcal{C}(\mathcal{Z})$ and a measure $p \in \mathcal{S}_n^*$. Since the topology is discrete, we can renumber the points $z_1, ..., z_n$ by means of a permutation $\mathbf{\Pi}_g$ in such a way that

$$-\infty = g_{n-1} < g_n \leq g_{n-1} \leq ... \leq g_2 \leq g_1 < g_0 = \infty$$

This implies a numbering of the components of the measure $p = (p_1, ..., p_n)$. This numbering has nothing to do with an ordering of the values p_i, which are different, in general. This is a reason, why it is not possible to start with two ordered vectors. A simultaneously ordering of two vectors requires equal values p_i (equal distribution) and means – from a physical point of view – indistinguishably states z_i.
To compare two function-measures-pairs we have to end up with functions that forget the ordering during its derivation.
The discrete case is interesting for numerical applications, too. A typical situation is the approximation of a continuous function $g \in \mathcal{C}(\mathcal{Z})$ by an elementary function, i.e., an element of $\mathcal{C}^{**}$ taking only a finite number of values.
Let (B_i) be a disjunct covering of $\mathcal{Z}$ (i.e., $B_i \in \mathcal{B}(\mathcal{Z})$, $B_i \cap B_j = \emptyset$, $\bigcup_{i=1}^n B_i = \mathcal{Z}$) and

$$g = \sum_{i=1}^{n} g_i \mathbb{1}_{B_i}$$

Considering such a function, we cannot distinguish different states in B_i. Thus, all states in B_i from g's-point of view look like a single state z_i. Given, in addition, a measure $p \in \mathcal{S}^*(\mathcal{Z})$, we can define $p_i = p(B_i)$ and are in the situation of a finite discrete space. Considering two such functions, we have, in general, two different finite states with different dimensions n_1 and n_2.

A.1 The functionals of a function and a measures

We consider the case $\mathcal{Z} = \{z_1, ..., z_n\}$ a function $g = (g_1, ..., g_n) \in \mathcal{C}(\mathcal{Z})$ and a measure $p = (p_1, ..., p_n)$. Since the topology is discrete, we can renumber the points $z_1, ..., z_n$ is such a way that $g_n \leq g_{n-1} \leq ... \leq g_2 \leq g_1$.
Now, the functional $M_{gp}(x) = \langle (g-x)_+, p \rangle$ is equivalent to

$$M_{gp}(x) = \sum_{j:g_j > x} (g_j - x)p_j = \sum_{j=1}^{i-1} (g_j - x)p_j = \sum_{j=1}^{i-1} g_j p_j - x \sum_{j=1}^{i-1} p_j = m_{i-1} - x t_{i-1}$$

where we defined $i = i(x)$ such that $g_i \leq x < g_{i-1}$ and

$$\begin{aligned} m_0 &= 0,\ m_i = \sum\nolimits_{j=1}^{i} g_j p_j \\ t_0 &= 0,\ t_i = \sum\nolimits_{j=1}^{i} p_j \end{aligned}$$

We have $m_n = \langle g, p\rangle$ and $t_n = 1$.
Equivalently, we have

$$M_{gp}(x) = \max_{0\le i\le n}(m_i - xt_i)$$

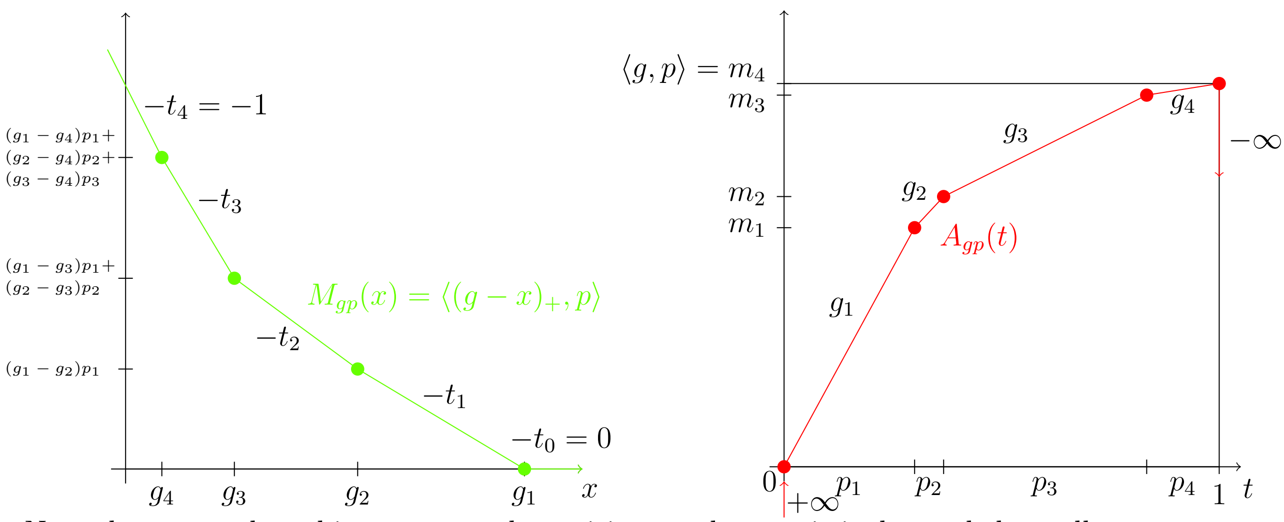
This is a convex piecewise affine function defined by the two monotone sequences $(g_n, g_{n-1}, ..., g_2, g_1)$ (abscissa) and $(1 = -t_n, -t_{n-1}, ..., -t_2-, t_1-, t_0 = 0)$ (slope). It goes throw the points $(g_i, m_i - g_i t_i)$.
For the convex conjugate of such a function we have (see subsection A.3)

$$M^*_{gp}(t) \;=\; \sup_x \big(xt - M_{gp}(x)\big) = \max_{0\le i\le n}(g_i t_i - m_i + g_i t)$$

Therefor we have

$$A_{gp}(t) = -M^*_{gp}(-t) = \min_{0\le i\le n}(m_i + g_i t - g_i t_i)$$

This is a concave piecewise affine function defined by the two monotone sequences $(0 = t_0, t_1, t_2, ..., t_{n-1}, t_n = 1)$ (abscissa)and $(g_1, g_2, ..., g_{n-1}, g_n)$ (slope). It goes throw the points (t_i, m_i).
The following pictures show an example with $n = 4$. We took a function g with $g_i \ge 0$. Therefore m_i and $A_{gp}(t)$ are monotone increasing. In general, $A_{gp}(t)$ is concave but not monotone. For example, see the picture on page 21.



Note that g_i can be arbitrary, not only positive numbers as it is demanded usually.
Not depending on g and p we have the fixed values $M_{gp}(+\infty) = 0$ and $A_{gp}(0) = 0$. Moreover, we have $A_{gp}(1) = \langle g, p\rangle$.
For $x \in [g_i, g_{i-1}]$, $t \in [t_{i-1}, t_i]$ we have

$$A_{gp}(t) - g_i t = \sum_{j=1}^{i-1}(g_j - g_i)p_j = M_{gp}(x) + (x - g_i)t_{i-1}$$

and therefore

$$A_{gp}(t) \;=\; M_{gp}(x) + xt - (t - t_{i-1})(x - g_i) \le M_{gp}(x) + xt$$

Thus, $(t - t_{i-1})(x - g_i)$ is the defect of Young's inequality.

At the intervall endpoints we have

$$
\begin{aligned}
A_{gp}(t_{i-1}) &= M_{gp}(g_i) + g_i t_{i-1} \\
A_{gp}(t_i) &= M_{gp}(g_i) + g_i t_i
\end{aligned}
$$

hence

$$
\begin{aligned}
A_{gp}(t_i) - A_{gp}(t_{i-1}) &= g_i p_i \\
M_{gp}(g_{i+1}) - M_{gp}(g_i) &= (g_i - g_{i+1}) t_i
\end{aligned}
$$

A.1.1 The subdifferentials

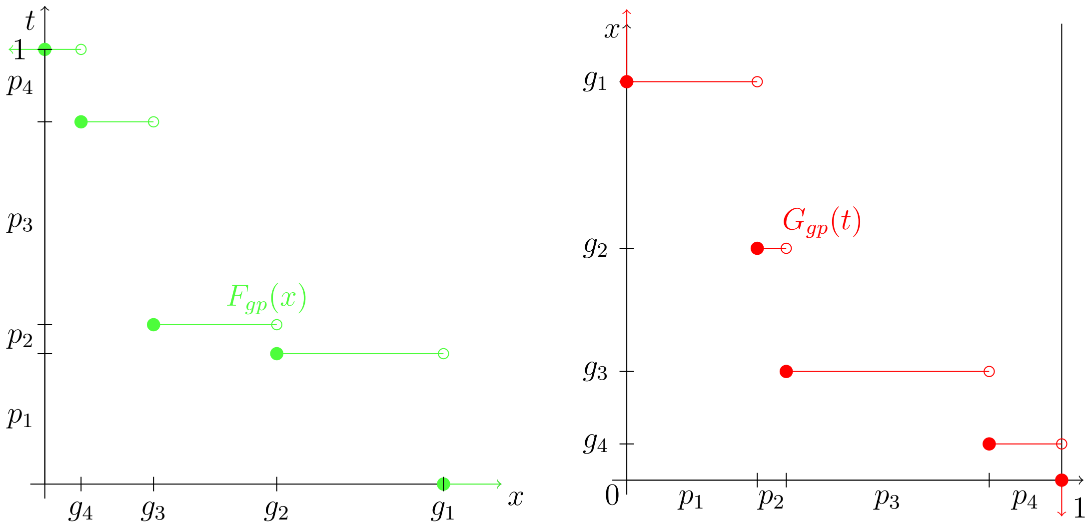
The subdifferentials of $-\partial M_{gp}(x)$ and $\partial A_{gp}(t)$ connect the abscissa and slope values ($0 = t_0, t_1, t_2, ..., t_{n-1}, t_n = 1$) and $(g_1, g_2, ..., g_{n-1}, g_n)$.
The functions

$$
\begin{aligned}
F_{gp}(x) &= -\sup \partial M_{gp}(x) \\
G_{gp}(t) &= \inf \partial A_{gp}(t)
\end{aligned}
$$

are connected with $M_{gp}(x)$ and $A_{gp}(t)$ via

$$
M_{gp}(x) = \int_x^\infty F_{gp}(x')dx', \; A_{gp}(t) = \int_0^t G_{gp}(t')dt'
$$

The aim of the definition of $F_{gp}(x)$ and $G_{gp}(t)$ is to handle the subdifferentials as functions. This are precisely the functions, arising by the definition of $F_{gp}(x)$ and $G_{gp}(t)$ as distribution function and its inverse.



The subdifferentials constitue a so called maximal monotone graph Φ_{gp}:

$$
(x,t) \in \Phi_{gp} \iff \begin{array}{l} x = g_k, \; t \in [t_{k-1}, t_k] \\ x \in [g_k, g_{k-1}], \; t = t_{k-1} \end{array}
$$

In particulare, the points (g_{k-1}, t_{k-1}), (g_k, t_{k-1}) and (g_k, t_k) and the line segments between these points belong to Φ_{gp}.

A.1.2 Level sets, distribution function and its inverse

As usual, the functions $F_{gp}(x)$ and $G_{gp}(t)$ are defined, starting with the definition oof level sets

$$L_g(x) = g^{-1}\big((x,\infty)\big) = \big\{z \in \mathcal{Z} | g(z) > x\big\}$$

We have $n+1$ level sets:

$$\begin{aligned}
L_g(x) &= L_0 = \emptyset, && x \geq g_1 \\
&= L_1 = B_1, && g_1 > x \geq g_2 \\
&= L_2 = B_1 \cup B_2, && g_2 > x \geq g_3 \\
&\vdots \\
&= L_{n-1} = B_1 \cup \cdots \cup B_{n-1}, && g_{n-1} > x \geq g_n \\
&= L_n = \mathcal{Z}, && g_n > x
\end{aligned}$$

Using the level sets we can define the so-called distribution function $F_{gp}(x) = p(L_g(x))$

$$\begin{aligned}
F_{gp}(x) &= p(L_0) = t_0 = 0, && x \geq g_1 \\
&= p(L_1) = t_1 = p_1, && g_1 > x \geq g_2 \\
&= p(L_2) = t_2 = p_1 + p_2, && g_2 > x \geq g_3 \\
&\vdots \\
&= p(L_{n-1}) = t_{n-1} = p_1 + \ldots + p_{n-1}, && g_{n-1} > x \geq g_n \\
&= p(L_n) = t_n = 1, && g_n > x
\end{aligned}$$

$$F_{gp}(x) = p(L_g(x)) = \sum_{i=0}^{n} t_i \mathbb{1}_{[g_{i+1}, g_i)}(x)$$

In the same way we can define the inverse of the distribution function $G_{gp}(t)$.

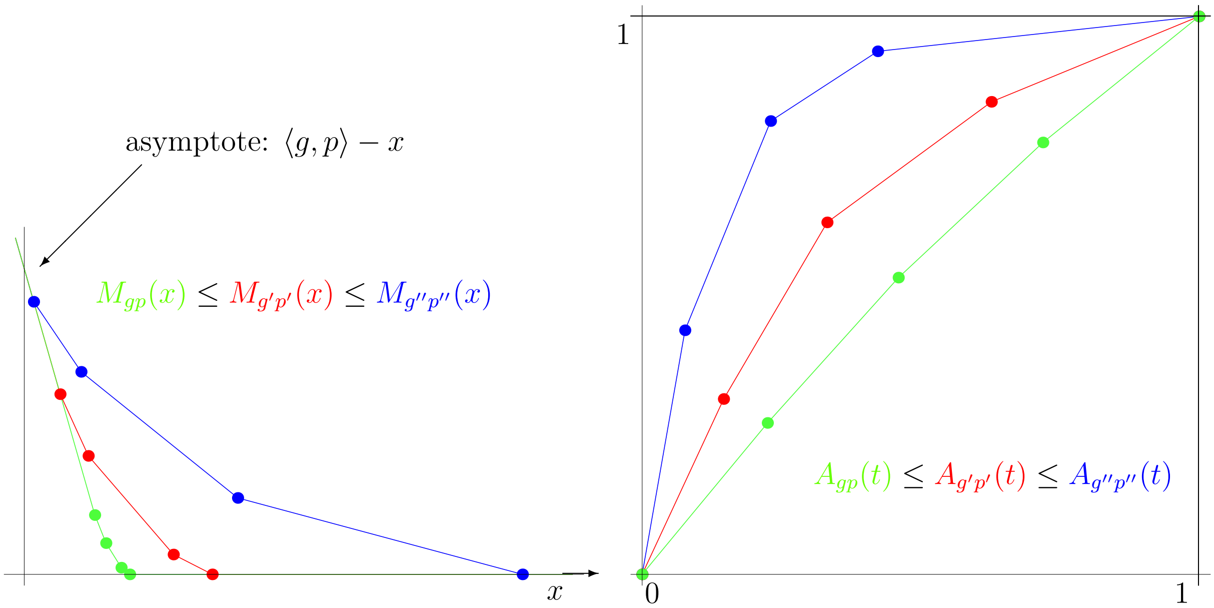
$$\begin{aligned}
C_{gp}(t) &= \{x \in \mathbb{R} | F_{gp}(x) \leq t\} = \\
&= C_{gp}(t_k) = [g_{k+1}, g_k), \quad t_k \leq t < t_k
\end{aligned}$$

A.1.3 Majorization

We consider a function $g \in \mathbb{R}_n$ and a measure $p \in \mathcal{S}_n^*$ and a Markov matrix $\mathbf{M}$ and set $p' = \mathbf{M}^* p$, $p'' = \mathbf{M}^* p'$, $g = \mathbf{M} g'$, $g' = \mathbf{M} g''$. Then, we have the following majorization: $(p'', g'') \prec (p', g') \prec (p, g)$. This is connected with the inequalities

$$M_{gp}(x) \leq M_{g'p'}(x) \leq M_{g''p''}(x), \;\; A_{gp}(t) \leq A_{g'p'}(t) \leq A_{g''p''}(t) \; .$$

This is illustrated in the following pictures
Note that the nodes, lie not on lines, because the measures are not uniform ones.



A.2 The functionals of two measures

We consider the case $\mathcal{Z} = \{z_1, ..., z_n\}$, and two measures $p = (p_1, ..., p_n)$ and $q = (q_1, ..., q_n)$. Since the topology is discrete, we can renumber the points $z_1, ..., z_n$ is such a way that

$$\frac{q_n}{p_n} \leq \frac{q_{n-1}}{p_{n-1}} \leq ... \leq \frac{q_2}{p_2} \leq \frac{q_1}{p_1}$$

If some $p_i = 0$, we have $\frac{q_i}{p_i} = +\infty$. This implies $\frac{q_j}{p_j} = +\infty$ for all $j < i$. The points with vanishing measure are the first ones.
Now, the functional $S_{qp}(x) = \sup_B \big(q(B) - xp(B)\big)$ is equivalent to

$$S_{qp}(x) = \max_{0 \leq i \leq n} (s_i - xt_i)$$

where we defined $i = i(x)$ such that $\frac{q_i}{p_i} \leq x < \frac{q_{i-1}}{p_{i-1}}$ and

$$\begin{aligned} s_0 &= 0, \; s_i = \sum\nolimits_{j=1}^{i} q_j \\ t_0 &= 0, \; t_i = \sum\nolimits_{j=1}^{i} p_j \end{aligned}$$

We have $s_n = t_n = 1$.

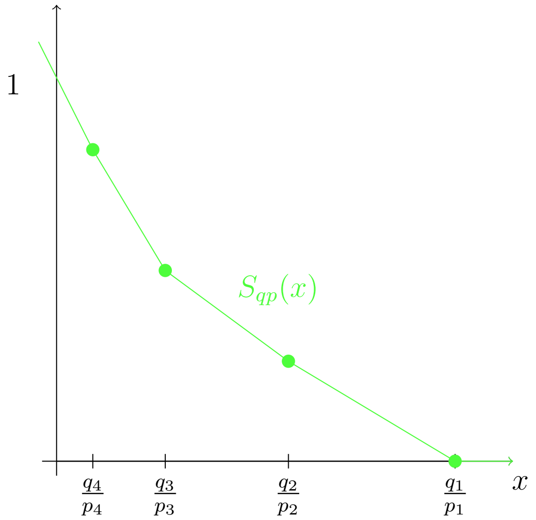
$S_{qp}(x)$ is a convex piecewize affin function going throw the points $\left(\frac{q_i}{p_i}, s_i - t_i\frac{q_i}{p_i}\right)$.
For the convex conjugate of such a function we have

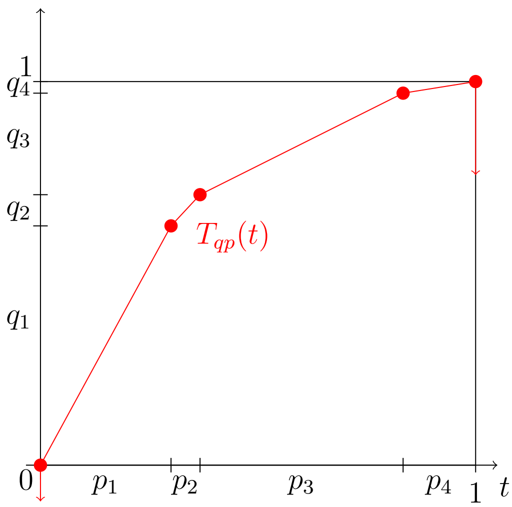
$$S^*_{qp}(t) \;=\; \sup_x \big(xt - S_{qp}(x)\big) = \max_{0\le i\le n}\left(\frac{q_i}{p_i}t_i - s_i + \frac{q_i}{p_i}t\right)$$

Therefor we have

$$T_{qp}(t) = -S^*_{qp}(-t) = \min_{0\le i\le n}\left(s_i + \frac{q_i}{p_i}t - \frac{q_i}{p_i}t_i\right)$$

This is a concave and monotone (in contrast to $A_{gp}(t)$)) piecewize affin function going throw the points (t_i, s_i).
The following pictures show an example with $n = 4$.





Note that g_i can be arbitrary, not only positive numbers as it is demanded usually.
Not depending on g and p we have the fixed values
$M_{qp}(+\infty) = 0$, $A_{qp}(0) = 0$ and $A_{qp}(1) = 1$.
For $x \in [g_i, g_{i-1}]$, $t \in [t_{i-1}, t_i]$ we have

$$A_{qp}(t) - g_i t = \sum_{j=1}^{i-1}(g_j - g_i)p_j = M_{qp}(x) + (x - g_i)t_{i-1}$$

and therefore

$$A_{qp}(t) \;=\; M_{qp}(x) + xt - (t - t_{i-1})(x - g_i) \le M_{qp}(x) + xt$$

Thus, $(t - t_{i-1})(x - g_i)$ is the defect of Young's inequality.
At the intervall endpoints we have

$$\begin{aligned} A_{qp}(t_{i-1}) &= M_{qp}(g_i) + g_i t_{i-1} \\ A_{qp}(t_i) &= M_{qp}(g_i) + g_i t_i \end{aligned}$$

hence

$$\begin{aligned} A_{qp}(t_i) - A_{qp}(t_{i-1}) &= g_i p_i \\ M_{qp}(g_{i+1}) - M_{qp}(g_i) &= (g_i - g_{i+1})t_i \end{aligned}$$

A.2.1 The subdifferentials

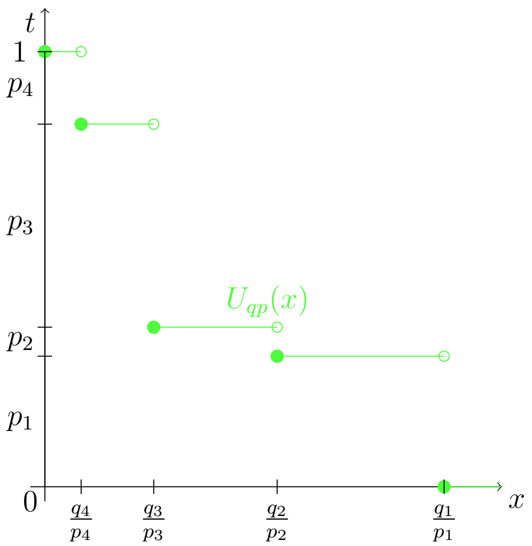
The subdifferentials of $-\partial S_{qp}(x)$ and $\partial T_{qp}(t)$ connect the abscissa and slope values $(0 = t_0, t_1, t_2, ..., t_{n-1}, t_n = 1)$ and $\left(\frac{q_1}{p_1}, \frac{q_2}{p_2}, ..., \frac{q_{n-1}}{p_{n-1}}, \frac{q_n}{p_n}\right)$.
The functions

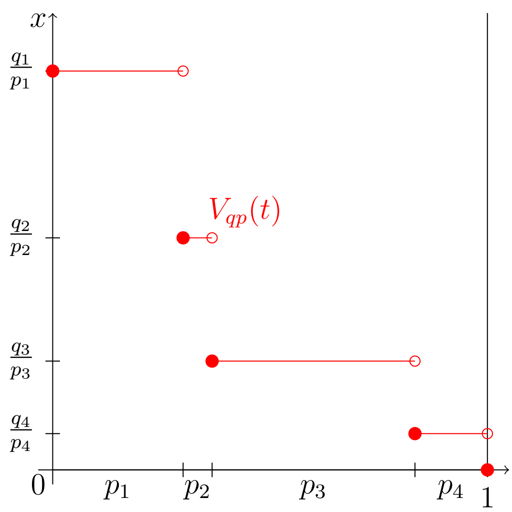
$$\begin{aligned} U_{qp}(x) &= -\sup \partial S_{qp}(x) \\ V_{qp}(t) &= \inf \partial T_{qp}(t) \end{aligned}$$

are connected with $S_{qp}(x)$ and $T_{qp}(t)$ via

$$S_{qp}(x) = 1 - \int_0^x U_{qp}(x')dx', \ T_{qp}(t) = 1 - \int_t^1 V_{qp}(t')dt'$$

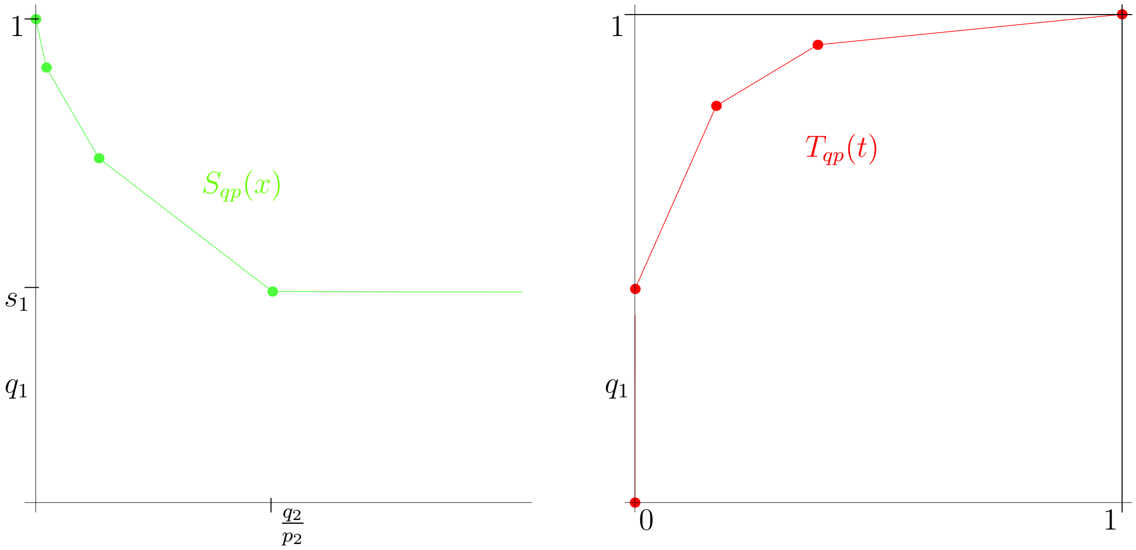
The aim of the definition of $U_{qp}(x)$ and $V_{qp}(t)$ is to handle the subdifferentials as functions.





A.2.2 Degenrate measures

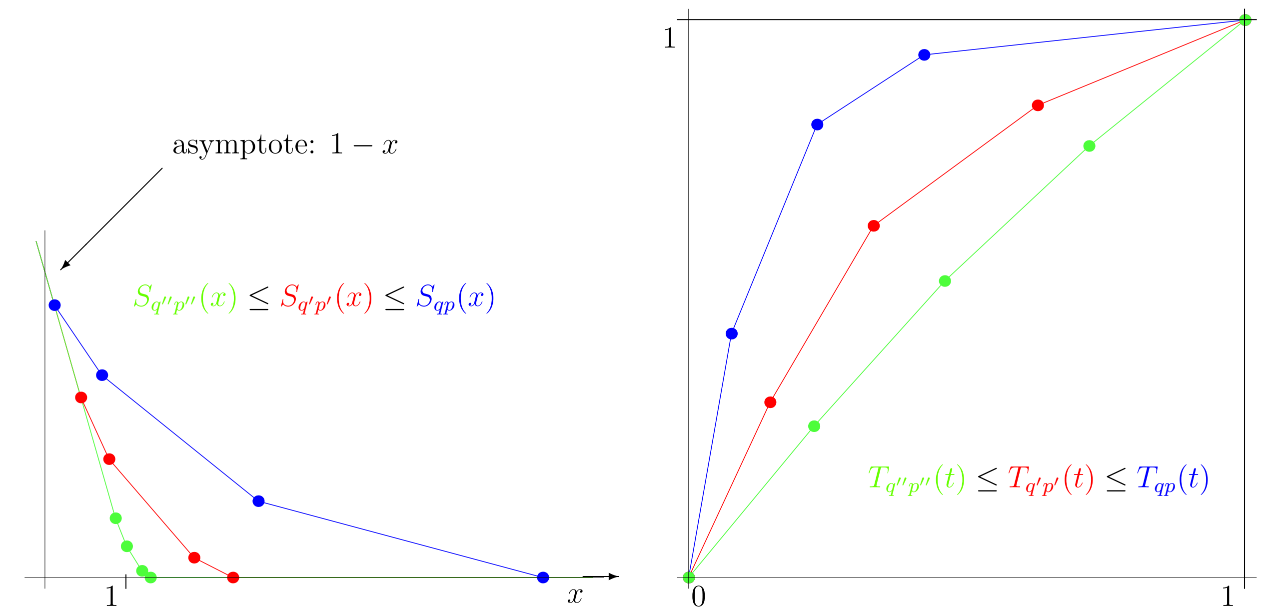
In the case, when measure p is degenerate, i.e., there is a Borel set B with $p(B) = 0$ but $q(B) > 0$ there is no radon-Nikodym derivative q/p. Nevertheless, the functions $S_{qp}(x)$ and $T_{qp}(t)$ are well defined, but have some special properties. We illustrate here the case, when $p_1 = 0$, $q_1 > 0$ and $p_i > 0$ for $i = 2, ..., n$. Clearly, $\frac{q_1}{p_1} = \infty$ is the largest number. As a result $S_{qp}(\infty) = q_1 > 0$ and $T_{qp}(0) = q_1 > 0$.



The case, when $p_1 = 0$ and $q_1 = 0$ means that state z_1 plays no role and can be omit. The state space becomes $(n-1)$- dimensional.

A.2.3 Majorization

We consider two given measures $p, q \in \mathcal{S}^*$ and a Markov operator $\mathbf{M}$ and set $p' = \mathbf{M}^* p$, $p'' = \mathbf{M}^* p'$, $q' = \mathbf{M}^* q$, $q'' = \mathbf{M}^* q'$. The majorization inequalities between the corresponding functions $S(x)$ and $T(t)$ are shown in the following pictures.
Note that the nodes, lie not on lines, because the measures are not uniform ones.



A.3 An excurs to piecewize affine functions

In this Appendix, we showed that in the finite case majorization is an partial order between piecewize affine functions, coinciding at two given endpoints. The reversal is true as well:
If two piecewize affine functions coincide at the endpoints and satisfy an inequality, then they are connected via a Markov operator in a special way. We show this in the Appendix B. Here we collect some facts on piecewize affine functions.
Convex piecewize affine functions are an invariant set for the Legendre-Fenchel tranform, which makes them especially important for approximations of general convex functions.
We consider convex functions $t = F(x)$, its subdifferential $y = \varphi(x)$, its inverse $x = \varphi^{-1}(y)$ und its conjugate convex $s = F^*(y)$.
A convex piecewize affine functions can be given via

- via two monotone sequences $(x_i)_{i=1}^n$ (abscissa) and $(y_i)_{i=0}^n$ (slopes);
- as tha maximum over a set of lines $F(x) = \max_i(a_i + y_i x)$;
- as a polygon $(x_i, t_i)_{i=0}^{n+1}$

For our application we have to consider the connection between the first two possibilities.

A.3.1 The function given as monotone sequences

The easiest way to define convex piecewize affine functions can be given via monotone sequences of abscissas x_i and slopes y_i

$$\begin{array}{ccccccccc} y_0 & \leq & y_1 & \leq & y_2 & \leq & \cdots & \leq & y_n \\ & & x_1 & \leq & x_2 & \leq & \cdots & \leq & x_n \end{array} \tag{47}$$

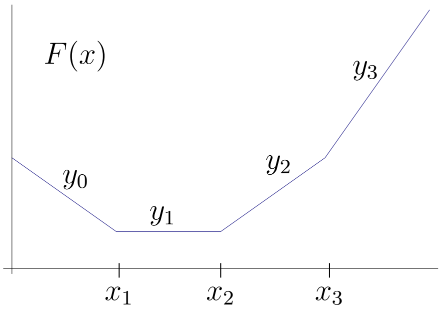
with the general property

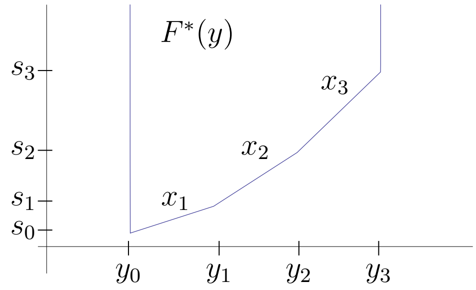
$$(x_{j-1} - x_j)(y_{i-1} - y_i) \geq 0, \;\; i = 1, ..., n$$

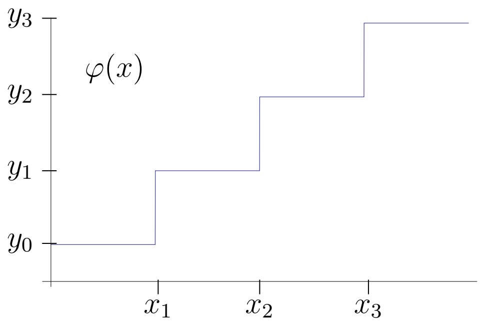
For completeness, we set $x_0 = -\infty$ and $x_{n+1} = \infty$.
The conjugate convex of such a function can be easely calculated. We denote a convexe piecewize affine function F defined by sequences of abscissas x_i and slopes y_i by $F[x, y]$. Since the subdifferentials of a convexe function and its conjugate constitute the same sets, we have $F^* = F^*[y, x]$ for a given $F[x, y]$. Similar notation we use for concave functions defined by sequences with different monotonicity.
In this notation, the defined functionals reads

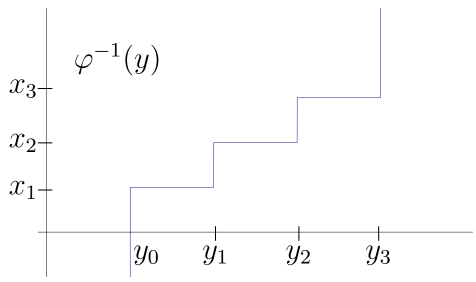
$$\begin{aligned} M_{gp} &= M_{gp}[g, -t] \;, \\ A_{gp} &= M_{gp}[t, g] \;, \\ S_{qp} &= S_{qp}[q/p, -t] \;, \\ M_{gp} &= M_{gp}[t, q/p] \;. \end{aligned}$$

The following pictures are an illustation of a typical example of $F(x)$, $\varphi(x)$, $\varphi^{-1}(y)$ and $F^*(y)$.









A.3.2 The function given as the maximum of lines

A function $F(x)$ given as

$$F(x) = \max_{0 \le i \le n} (a_i + y_i x) \ ,$$

i.e. as a maximum over $n+1$ lines $a_i + y_i x$ with given sequences $(a_i)_{i=0}^n$ and $(y_i)_{i=0}^n$, is allways a convex piecewize affine function. It may happens that some of the lines play no role for the maximum. Unlike the previous case, a condition to avoid such lines is more complicated. We assume that no lines can be omited. Then,

$$F(x) = a_i + y_i x, \ \ x_i \le x \le x_{i+1}$$

with some $x_0, ..., x_{n+1}$ that must be calculated. We have $x_0 = -\infty$, $x_{n+1} = +\infty$. The other x_i have to satisfy

$$a_{i-1} + y_{i-1} x_i = a_i + y_i x_i \ .$$

Hence, we obtain the abszissas

$$x_i = \frac{a_{i-1} - a_i}{y_i - y_{i-1}} \ .$$

Obviously, the slopes are y_i. Corresponding to (47) the sequences x_i and y_i have to increases monotonly, i.e.

$$\begin{aligned} x_i \le x_{i+1} &\iff \frac{a_{i-1}-a_i}{y_i - y_{i-1}} \le \frac{a_i - a_{i+1}}{y_{i+1}-y_i} \\ &\iff 0 \le a_{i-1}(y_i - y_{i+1}) + a_i(y_{i+1} - y_{i-1}) + a_{i+1}(y_{i-1} - y_i) \\ &\iff 0 \le y_{i-1}(a_{i+1} - a_i) + y_i(a_{i-1} - a_{i+1}) + y_{i+1}(a_i - a_{i-1}) \\ &\iff a_{i-1}(y_{i+1} - y_i) + a_{i+1}(y_i - y_{i-1}) \le a_i(y_{i+1} - y_{i-1}) \\ &\iff a_i \ge a_{i-1}\frac{y_{i+1}-y_i}{y_{i+1}-y_{i-1}} + a_{i+1}\frac{y_i - y_{i-1}}{y_{i+1}-y_{i-1}} \end{aligned}$$

This is a concavity condition for a_i with respect to y_i.
For the values in the points of intersection we have

$$t_i = F(x_i) = a_{i-1} + y_{i-1}x_i = a_i + y_i x_i = \frac{a_{i-1}y_i - a_i y_{i-1}}{y_{i+1} - y_{i-1}}$$

The subdifferential is the polygon

$$\begin{aligned} \Phi = (-\infty, y_0) &\to (x_1, y_0) \to (x_1, y_1) \to (x_2, y_1) \to (x_2, y_2) \to \cdots \\ &\to (x_{i-1}, y_{i-1}) \to (x_i, y_{i-1}) \to (x_i, y_i) \to \cdots \to (x_n, y_n) \to (+\infty, y_n) \end{aligned}$$

The convex conjugate can be written as a maximum over lines as well

$$\begin{aligned} F^*(y) &= \sup_x \big(xy - F(x)\big) = \max_{0\le i\le n}\left(\sup_{x_i \le x \le x_{i+1}} (xy - a_i - y_i x)\right) = \\ &= \max_{0\le i\le n}\left(-a_i + \sup_{x_i\le x\le x_{i+1}} x(y - y_i)\right) = \\ &= \max_{0\le i\le n}\left(-a_i + \begin{cases} x_i(y-y_i), & y \le y_i \\ x_{i+1}(y - y_i), & y_i \le y \end{cases}\right) = \\ &= \max_{0\le i\le n}\left(\begin{cases} -a_i + x_i(y-y_i), & y \le y_i \\ -a_i + x_{i+1}(y - y_i), & y_i \le y \end{cases}\right) = \\ &= \max_{1\le i\le n}\big(-a_i - x_i y_i + x_i y\big) = \\ &= \max_{1\le i\le n}\big(b_i + x_i y\big) \end{aligned}$$

with $b_i = -a_i - x_i y_i$. We obtain

$$F^*(y) = b_i + x_i y = -a_i - x_i y_i, \;\; y_{i-1} \le y \le y_i$$

and $F^*(y_i) = s_i = -a_i$. Analogously we obtain $t_i = -b_i$.

A.3.3 The function given as a polygon

Given the points

$$(t_0, s_0), (t_1, s_1), (t_2, s_2), ..., (t_n, s_n) \;.$$

Since in our application we used cancave functions, we look for the concave hull $T(t)$ of this points. We start with the ansatz

$$T(t) = \min_{0\leq k\leq n-1} (y_k + h_{k+1} t)$$

with

$$\begin{aligned} h_{k+1} &= \frac{s_{k+1} - s_k}{t_{k+1} - t_k} \\ y_k &= s_k - h_{k+1} t_k = \frac{s_k t_{k+1} - s_{k+1} t_k}{t_{k+1} - t_k} \end{aligned}$$

We get

$$\begin{aligned} T(t) &= \min_{0\leq k\leq n-1} \left(\frac{s_k t_{k+1} - s_{k+1} t_k + t s_{k+1} - t s_k}{t_{k+1} - t_k} \right) = \\ &= \min_{0\leq k\leq n-1} \left(\frac{s_k (t_{k+1} - t) + s_{k+1} (t - t_k)}{t_{k+1} - t_k} \right) \end{aligned}$$

This is just the convex combination of the values. $y_k + h_{k+1} t$ are line segments. We have $T(t_k) = s_k$ as expected. This is a piecewize affine function given via $t_0, t_1, ..., t_n$ (abscissa) and $h_1, h_2, ..., h_n$ (slopes).
As in the previous case, it may happens that some of the points (t_i, s_i) are not necessary. To avoid this situation, the points have to satisfy the convexity condition

$$s_k t_{k-1} - s_{k+1} t_{k-1} - s_{k-1} t_k + s_{k+1} t_k + s_{k-1} t_{k+1} - s_k t_{k+1} \geq 0 \ .$$

A.3.4 Summary

We summarize the transitions from one representation into the other in a table:

	$F(x)$	$F^*(y)$
mon. sequences	(x_i, y_i)	(y_i, x_i)
lines	$\max_{0\leq i\leq n} (-s_i + y_i x)$	$\max_{1\leq i\leq n} (-t_i + x_i y)$
polygon points	(x_i, t_i)	(y_i, s_i)
	$x_i = \frac{s_i - s_{i-1}}{y_i - y_{i-1}}$	$y_i = \frac{t_i - t_{i-1}}{x_i - x_{i-1}}$
	$t_i = x_i y_i - s_i$	$s_i = x_i y_i - t_i$
	$= \frac{s_i y_{i-1} - s_{i-1} y_i}{y_i - y_{i-1}}$	$= \frac{t_i x_{i-1} - t_{i-1} x_i}{x_i - x_{i-1}}$
	$\max_{0\leq i\leq n} \left(t_{i-1} \frac{x_i - x}{x_i - x_{i-1}} + t_i \frac{x - x_{i-1}}{x_i - x_{i-1}} \right)$	$\max_{1\leq i\leq n} \left(s_{i-1} \frac{y_i - y}{y_i - y_{i-1}} + s_i \frac{y - y_{i-1}}{y_i - y_{i-1}} \right)$

Appendix B
The construction of a general Markov matrix

We denote in the following by $y \prec_{\mathbb{1}} x$ the classical majorization of two given vectors in $\mathbb{R}^n$. We have $y \prec_{\mathbb{1}} x$ if and only if there is a doubly stochastic matrix $\mathbf{D}^*$ with $y = \mathbf{D}^* x$. The Robin Hood method is a tool to construct $\mathbf{D}^*$ for two given vectors x and y with $y \prec_{\mathbb{1}} x$. This method works because a doubly stochastic matrix can be presented as a product of matrices of a special form. Those special matrices can be found in a constructive and intuitive way, this is the Robin Hood method.
The matrix

$$\mathbf{T} = \alpha \mathbf{I} + (1-\alpha)\mathbf{C}(j,i)$$

where $\alpha \in [0,1]$ and $\mathbf{C}(j,i)$ is a transposition, i.e., a matrix interchanging the i-th and j-th states, is a special doubly stochastic matrix, the convex combination of the identity and transposition.
It can be shown that any doubly stochastic matrix $\mathbf{D}^*$ can be presented as a product

$$\mathbf{D}^* = \mathbf{T}_k \cdots \mathbf{T}_1$$

The Robin Hood method finds successivelly the matrices $\mathbf{T}_i$. From this representation follows Birkhoff's famos theorem, stating that any doubly stochastic matrix is a convex combination of permutations matrices, or equivalently, the permutations matrices are the extremal points of the convex set of doubly stochastic matrices. We have

$$\begin{aligned} \mathbf{D} &= \mathbf{T}_k \cdots \mathbf{T}_1 = \big(\alpha_k \mathbf{I} + (1-\alpha_k)\mathbf{C}(i_k,j_k)\big) \cdots \big(\alpha_1 \mathbf{I} + (1-\alpha_1)\mathbf{C}(i_1,j_1)\big) = \\ &= \sum q_i \prod_l \mathbf{C}(i_l,j_l) = \sum q_i \mathbf{G}_i \end{aligned}$$

where we used that a permutations matrix can be written as a product of transpositions.
All this statements fail in the case of a general Markov matrix. Even the simultaneous ordering of two vectors – neccesary for the definition of the majorization what is the starting point of the Robin Hood method – is not possible if the underlying state space has no uniform measure. In the following, we propose a method inspired by the intuitive physical definition of probabilities as convex combinations of the results of a finite number of experiments. This allowds us to use Robin Hood method for general Markov matrices after the transformation in a state space with natural uniform measure.

B.1 The classical Robin Hood method

Given two sequences x and y with $y \prec_{\mathbb{1}} x$, we construct a doubly stochastic matrix $\mathbf{D}^*$ with $y = \mathbf{D}^* x$ by the so-called Robin Hood method.
The Robin Hood method starts with two ordered vectors $x^{\downarrow}$ and $y^{\downarrow}$. To achieve this, at first one has to rearange x and y in decreasing order, i.e. one has to permute the vectors, each of them by its own permutation. Let $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ be two permutation matrices such that $\mathbf{\Pi}_x x = x^{\downarrow}$ and $\mathbf{\Pi}_y y = y^{\downarrow}$. Actually, a permutation matrix defines a renumbering of the states of the underlying state space. Since in general, the permutation matrices $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ are diffent, this implies that the state space is renumbered in a diffent way what is of cause unphysical unless, the states of

the space are indistinguishable. In such a space only the uniform measure $\mathbb{1}^*$ has a physical sense. This is the field of application of the classical majorization theory.
Given two monotone decreasing sequences x and y with $y \prec_{\mathbb{1}} x$, we construct a doubly stochastic matrix $\mathbf{D}^*$ with $y = \mathbf{D}^* x$ constructing a sequence of matrices

$$\mathbf{T} = \lambda \mathbf{I} + (1-\lambda)\mathbf{C}_{ji}$$

equalizing the sequences x and y starting from their middle. After constructing a matrix $\mathbf{T}$, a new vector $x' = \mathbf{T}x$ is defined and the procedure is repeated with the vectors x' and y. The algoritm terminates after a finite number k of steps when $x' = \mathbf{T}x = y$. We obtain $\mathbf{D} = \mathbf{T}_k \cdots \mathbf{T}_1$
We describe one iteration step: Starting with two vectors x and y

$$\begin{array}{ccccccccccccc} x_1 & \geq & x_2 & \geq & \dots & \geq & x_j & \geq & \dots & \geq & x_i & \geq & \dots & \geq & x_n \\ \vee & & \vee & & & & \vee & & & & \wedge & & & & \wedge \\ y_1 & \geq & y_2 & \geq & \dots & \geq & y_j & \geq & \dots & \geq & y_i & \geq & \dots & \geq & y_n \end{array}$$

we define the indizees i and j as the nearest idizees, for which y and x are not balance out yet. Thus,

$$j = \max_k (x_k > y_k), \quad i = \min_k (x_k < y_k)$$

Since $\sum_i x_i = \sum_i y_i$, we have $j < n$ and $i > 1$. We set

$$\lambda = \max \left\{ \frac{y_j - x_i}{x_j - x_i}, \frac{x_j - y_i}{x_j - x_i} \right\} \tag{48}$$

(What value is taken is equivalent to the question wether $y_i + y_j$ is larger then $x_i + x_j$.)
It follows $\lambda x_j + (1-\lambda)x_i \geq \lambda x_i + (1-\lambda)x_j$ and $\lambda \geq \frac{1}{2}$, because of

$$\frac{y_j - x_i}{x_j - x_i} + \frac{x_j - y_i}{x_j - x_i} = 1 + \frac{y_j - y_i}{x_j - x_i} \geq 1$$

We define $\mathbf{T} = \lambda \mathbf{I} + (1-\lambda)\mathbf{C}_{ji}$. Then

$$x'\mathbf{T}x = (x_1, \dots, x_{j-1}, y_j, ..., x_j + x_i - y_j, x_{i+1}, \dots, x_n)$$

if in (48) $\lambda = \frac{y_j - x_i}{x_j - x_i}$ or

$$\mathbf{T}x = (x_1, \dots, x_{j-1}, x_j + x_i - y_i, ..., y_i, x_{i+1}, \dots, x_n)$$

otherwize. In any case, the number of nonequal values in the vectors x and y is reduced at least by 1. In the case, when $x_j + x_i = y_j + y_i$ this number reduces by 2.
Finally, the desired doubly sochastic matrix is $\mathbf{D} = \mathbf{\Pi}_y^{-1} \mathbf{D}_0 \mathbf{\Pi}_x$.

B.2 Rational probabilities

From a physical point of view the probability of finding i_k times a state z_k during n equivalent experiments is $p_k = \frac{i_k}{n}$. It is natural to define the statistical state of the physical system as

$$z = \frac{i_1}{n}\ 'z_1' + ... + \frac{i_m}{n}\ 'z_m' = p_1\ 'z_1' + ... + p_m\ 'z_m'$$

with $i_1 + ... + i_m = n$ or $p_i \geq 0$ and $p_1 + ... + p_m = 1$. Thus, in a natural way arise rational probabilities as collections of experiment results with equal weights. Passing to the limit $n \to \infty$ we obtain general probability measures in Banach spaces.
Conversely, starting with m rational numbers $p_1, ..., p_m$ with $p_i \geq 0$ and $p_1 + ... + p_m = 1$ we can construct a homogeneuos state space with n states, by taking n as the least common multiple of the denominators of the p_i. Doing so, we obtain m integers $i_k = p_k n$ with $i_1 + ... + i_m = n$. From a physical point of view this means that we consider instead of one state z_k, i_k copies of this state. Instead of m-dimensional spaces $\mathcal{C} = \mathbb{R}_m$ and $\mathcal{C}^* = \mathbb{R}^*_m$ we obtain n-dimensional space. $\mathcal{C} = \mathbb{R}_n$ and $\mathcal{C}^* = \mathbb{R}^*_n$. Since the result of every experiment is counted with the same weight, the uniform measure $\mathbb{1}^*_n \in \mathcal{S}^*_n$ plays a special role.

B.3 Special Markov matrices

B.3.1 Special Markov matrices for rational measures

Let $n = i_1 + ... + i_m$, $p = \frac{1}{n}(i_1, ..., i_m)$ a measure in $\mathcal{S}^*_m$ and $\mathbb{1}^* = \frac{1}{n}(1, ..., 1)$ the uniform measure in $\mathcal{S}^*_n$. We define two matrices $\mathbf{G}_p$ and $\mathbf{H}_p$, depending on a given measure p,

$$\mathbf{G} = \left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array}\right), \; \mathbf{H} = \left(\begin{array}{ccc|ccc|c|ccc} \frac{1}{i_1} & \cdots & \frac{1}{i_1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{i_2} & \cdots & \frac{1}{i_2} & \cdots & 0 & \cdots & 0 \\ & \vdots & & & \vdots & & \vdots & & \vdots & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \frac{1}{i_m} & \cdots & \frac{1}{i_m} \end{array}\right)$$

$\mathbf{G}_p$ consists of $i_1, ..., i_m$ equal rows, $\mathbf{H}_p$ of equal columns. Obviously, $\mathbf{G}$ and $\mathbf{H}$ are Markov matrices $\mathbf{G}_p \in \mathcal{M}(\mathbb{R}_m, \mathbb{R}_n)$, $\mathbf{H}_p \in \mathcal{M}(\mathbb{R}_n, \mathbb{R}_m)$ with the properties

$$\mathbf{G}_p \mathbb{1}_m = \mathbb{1}_n, \; \mathbf{G}^*_p \mathbb{1}^*_n = p, \; \mathbf{H}_p \mathbb{1}_n = \mathbb{1}_m, \; \mathbf{H}^*_p p = \mathbb{1}^*_n, \; \mathbf{H}_p \mathbf{G}_p = \mathbf{I}_m, \; \mathbf{G}^*_p \mathbf{H}^*_p = \mathbf{I}_m \tag{49}$$

Here, $*$ denotes transponation.
If we have $i_j = 0$ for some j, we omit the states z_j, getting a problem with few dimensions. This case means that the states z_j did not occur in the experiment.

B.3.2 A doubly stochastic matrix from two measures

Given two measures p and q and a Markov matrix $\mathbf{M}$ with $q = \mathbf{M}^* p$, we define a matrix $\mathbf{D} : \mathbb{R}_n \to \mathbb{R}_n$ by

$$\mathbf{D} = \mathbf{G}_p \mathbf{M} \mathbf{H}_q \; . \tag{50}$$

Here n is the least common multiple of all denominators of p and q. We denote by $\mathcal{D}$ the set of doubly stochastic matrices in $\mathbb{R}_n$. We have $\mathbf{D} \in \mathcal{D}$. This follows form

$$\begin{aligned} \mathbf{D}\mathbb{1}_n &= \mathbf{G}_p \mathbf{M} \mathbf{H}_q \mathbb{1}_n = \mathbf{G}_p \mathbf{M} \mathbb{1}_m = \mathbf{G}_p \mathbb{1}_m = \mathbb{1}_n \\ \mathbf{D}^* \mathbb{1}^*_n &= \mathbf{H}^*_q \mathbf{M}^* \mathbf{G}^*_p \mathbb{1}^*_n = \mathbf{H}^*_q \mathbf{M}^* p = \mathbf{H}^*_q q = \mathbb{1}^*_n \end{aligned}$$

(positivity is obvious), where we used the properties (49).

Considering matrices $\mathbf{D} : \mathbb{R}_n \to \mathbb{R}_n$, we restrict ourself to a special case to simplifies the notations. In general, the Markov matrix $\mathbf{M}$ can act between different spaces $\mathbb{R}_{m_1}$ and $\mathbb{R}_{m_2}$ and the doubly stochastic matrix $\mathbf{D}$ can act between different spaces $\mathbb{R}_{n_1}$ and $\mathbb{R}_{n_2}$, too. This allows us to don't care about the least common multiple of all denominators of p and q. n_1 is the lcm of q and n_2 is the lcm of p. Moreover, possible states we have to omit, are also covered by this construction. The connection of the spaces in the general case is illustrated in the diagram.

$$\begin{array}{ccc} \mathbb{R}_{m_1} & \xrightarrow{\mathbf{M}} & \mathbb{R}_{m_2} \\ \mathbf{H}_q \uparrow & & \downarrow \mathbf{G}_p \\ \mathbb{R}_{n_1} & \xrightarrow[\mathbf{D}]{} & \mathbb{R}_{n_2} \end{array}$$

B.4 Construction of a Markov matrix from two pairs of a continuous function and a probability measure

Given two pairs composed of a continuous function and a probability measure (g, q) and (f, p) in $\mathbb{R}_m$ with $(g, q) \prec (f, p)$. We look for a Markov matrix $\mathbf{M}$ with

$$\begin{aligned} f &= \mathbf{M}g \\ q &= \mathbf{M}^* p \end{aligned}$$

We construct such a $\mathbf{M}$ by transforming the promblem in a homogeneus space $\mathbb{R}_n$. At first we look for a heuristic $\mathbf{D} \in \mathcal{D}$ related to this problem and using the construction from the previous subsection.
From $f = \mathbf{M}g$ and (49) it follows that for any measure α we have $f = \mathbf{M}g = \mathbf{M}\mathbf{H}_\alpha\mathbf{G}_\alpha g$. Hence, for any measure β we have

$$\mathbf{G}_\beta f = \mathbf{G}_\beta \mathbf{M}\mathbf{H}_\alpha\mathbf{G}_\alpha g$$

Wie define

$$\begin{aligned} \mathbf{D} &= \mathbf{G}_\beta\mathbf{M}\mathbf{H}_\alpha \\ y &= \mathbf{G}_\beta f \\ x &= \mathbf{G}_\alpha g \end{aligned}$$

We have $\mathbf{D} \in \mathcal{D}$. To satisfy $q = \mathbf{M}^* p$ we set, according to (50), $\alpha = q$ and $\beta = p$.
Let $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ be two permutation matrices such that $\mathbf{\Pi}_x x = x^\downarrow$ and $\mathbf{\Pi}_y y = y^\downarrow$.
This suggests the following algorithm:

1) From p and q we construct four matrices $\mathbf{G}_p^*$ and $\mathbf{H}_p^*$ and $\mathbf{G}_q^*$ and $\mathbf{H}_q^*$ such that $\mathbf{G}_p^*\mathbb{1}^* = p$, $\mathbf{H}_q^* q = \mathbb{1}^*$, $\mathbf{G}_q^*\mathbb{1}^* = q$ and $\mathbf{H}_p^* p = \mathbb{1}^*$.

2) We set $y = \mathbf{G}_p f$ and $x = \mathbf{G}_q g$.

3) We define permutations $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_y$ such that $\mathbf{\Pi}_x x = x^\downarrow$ and $\mathbf{\Pi}_y y = y^\downarrow$ are ordered in decreasing order.

4) We construct $\mathbf{D}_0 \in \mathcal{D}$ using the Robin-Hood method and satisfying $y^\downarrow = \mathbf{D}_0 x^\downarrow$.

5) We set $\mathbf{M} = \mathbf{H}_p\mathbf{D}\mathbf{G}_q$ with $\mathbf{D} = \mathbf{\Pi}_y^{-1}\mathbf{D}_0\mathbf{\Pi}_x$.

For the proof we have to show $\mathbf{M} \in \mathcal{M}$, $q = \mathbf{M}^* p$ and $f = \mathbf{M}g$.

1) As a product of three Markov matrices, $\mathbf{M}$ is a Markov matrix, too.

2) Since $\mathbf{M}^* = \mathbf{G}_q^*\mathbf{D}^*\mathbf{H}_p^*$ we conclude

$$\mathbf{M}^*p = \mathbf{G}_q^*\mathbf{D}^*\mathbf{H}_p^*p = \mathbf{G}_q^*\mathbf{D}^*\mathbb{1}^* = \mathbf{G}_q^*\mathbb{1}^* = q\ .$$

3) From the definitions of x and y it follows

$$\mathbf{M}g = \mathbf{H}_p\mathbf{D}\mathbf{G}_q g = \mathbf{H}_p\mathbf{D}x = \mathbf{H}_p y = \mathbf{H}_p\mathbf{G}_p f = f$$

The new definition of majorization is connected with the old one via

$$(g,q) \prec (f,p) \iff \mathbf{G}_p f \prec_{\mathbb{1}} \mathbf{G}_q g$$

This will be proved in the next subsection.
Note that rationallity is only a property of the measures. The functions f and g can be arbitrary, even irrational.

B.5 Majorization and the existence of a Markov matrix

Theorem 6 in this subsection gives the equivalence of inequalities between pairs of monotone sequences, majorization and the existence of Markov matrices, connecting them.
Given two pairs of monotone sequences

$$\begin{aligned} 0 &= s_0 < s_1 < ... < s_{m_1} = 1 \\ & \quad g_{m_1} < ... < g_1 \end{aligned}$$

and

$$\begin{aligned} 0 &= t_0 < t_1 < ... < t_{m_2} = 1 \\ & \quad f_{m_2} < ... < f_1 \end{aligned}$$

We define two sequences (measures)

$$\begin{aligned} q_i &= s_i - s_{i-1},\ i = 1, ..., m_1 \\ p_i &= t_i - t_{i-1},\ i = 1, ..., m_2 \end{aligned}$$

and $y = \mathbf{G}_p f \in \mathbb{R}_{n_2}$, $x = \mathbf{G}_q g \in \mathbb{R}_{n_1}$. This is the general situation, illustated in the diagram on page 50. We have the following

Theorem 6 *Let F be an arbitrary convex function. The following statements are equivalent:*

(i) $\sum_{i=1}^{m_2} p_i F(f_i) \leq \sum_{i=1}^{m_1} q_i F(g_i)$

(ii) $\frac{1}{n_2} \sum_{j=1}^{n_2} F(y_j) \leq \frac{1}{n_1} \sum_{j=1}^{n_1} F(x_j)$

(iii) $(g, q) \prec (f, p)$

(iv) $y \prec_{\mathbb{1}} x$

(v) $M_{gq}(x) \geq M_{fp}(x)$ *for* $x \in \mathbb{R}$

(vi) $A_{gq}(t) \geq A_{fp}(t)$ *for* $t \in [0, 1]$

(vii) $\exists\, \mathbf{M} \in \mathcal{M}(\mathbb{R}_{m_2}, \mathbb{R}_{m_1})$ *with* $f = \mathbf{M}g$, $p = \mathbf{M}^* q$.

Proof:
$(i) \iff (iii)$ is the new definition of majorization.
$(i) \iff (v) \iff (vi)$ follows from Theorem 4
$(ii) \iff (iv)$ follows from the classical majorization theory in linear algebra.
We prove $(ii) \iff (i)$: We have $p_i = \alpha_i / n_2$, $q_i = \beta_i / n_1$. $\mathbf{G}_p f$ is a vector, containing α_i times f_i. Thus we have

$$\frac{1}{n_2} \sum_{j=1}^{n_2} F(y_j) = \sum_{i=1}^{m_2} \frac{\alpha_i}{n_2} F(f_i)$$

Analogously follows a similar equality for the right had sides of the inequalities. This shows $(ii) \iff (i)$.
We show $(ii) \implies (vii)$. From (ii) follows (iv) and therefore the existence of a doubly stochastic matrix $\mathbf{D}$ with $y = \mathbf{D}x$. Setting $\mathbf{M} = \mathbf{H}_p \mathbf{D} \mathbf{G}_q$, we showed in the previous subsection that $\mathbf{M}$ is a Markov matrix and $f = \mathbf{M}g$, $p = \mathbf{M}^* q$ hold.
$(vii) \implies (iii)$ is the definition of majorization □

B.6 Construction of a Markov matrix from two pairs of measures

Given two pair of measures (t, q) and (s, p) in $\mathbb{R}_m$ with $(t, q) \prec^* (s, p)$. We look for a Markov matrix $\mathbf{M}$ with

$$\begin{aligned} q &= \mathbf{M}^* p \\ t &= \mathbf{M}^* s \end{aligned}$$

For simplicity, we assume $q_i > 0$, $p_i > 0$ for $i = 1, ..., m$. The general case can be handled in the same way but the vectors x and y can not be defined explicitely but as soltions of some linear systems, in general underdetermined ones.
We reduce this case to the just investigated case, seeking for a Markov matrix $\mathbf{X}$ satisfying

$$\begin{aligned} p &= \mathbf{X}^* q \\ f &= \mathbf{X} g \end{aligned}$$

with $\mathbf{Q}_q f = t$ and $\mathbf{Q}_p g = s$. Here $\mathbf{Q}_q$ is the multiplication operator.

Clearly, the matrix $\mathbf{X}$ is the physical inverse one to $\mathbf{M}$. Both matrices are connected via

$$\begin{aligned} \mathbf{Q}_q\mathbf{X} &= \mathbf{M}^*\mathbf{Q}_p \\ \mathbf{Q}_p\mathbf{M} &= \mathbf{X}^*\mathbf{Q}_q \end{aligned}$$

Now, we have the following algorithm:

1) From p, q, r, s we construct two vectors $f = \mathbf{Q}_q^{-1}t$ and $g = \mathbf{Q}_p^{-1}s$.

2) We construct a Markov matrix $\mathbf{X}$ from $p = \mathbf{X}^*q$ and $f = \mathbf{X}g$ from two pair of a continuous function and a probability measure. From the general theory we know that $(g, q) \prec (f, p)$.

3) We set $\mathbf{M} = \mathbf{Q}_p^{-1}\mathbf{X}^*\mathbf{Q}_q$.

We have $\mathbf{M}^* = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}\mathbf{G}_p\mathbf{Q}_p^{-1}$.
For the proof we have to show $\mathbf{M} \in \mathcal{M}$, $q = \mathbf{M}^*p$ and $t = \mathbf{M}^*s$.

1) Proof of $q = \mathbf{M}^*p$:

$$\mathbf{M}^*p = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}\mathbf{G}_p\mathbf{Q}_p^{-1}p = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}\mathbf{G}_p\mathbb{1} = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}\mathbb{1} = \mathbf{Q}_q\mathbf{H}_q\mathbb{1} = \mathbf{Q}_q\mathbb{1} = q$$

2) Proof of $t = \mathbf{M}^*s$:

$$\mathbf{M}^*s = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}\mathbf{G}_p\mathbf{Q}_p^{-1}s = \mathbf{Q}_q\mathbf{H}_q\mathbf{D}x = \mathbf{Q}_q\mathbf{H}_q y = \mathbf{Q}_q\mathbf{H}_q\mathbf{G}_q\mathbf{Q}_q^{-1}t = \mathbf{Q}_q\mathbf{Q}_q^{-1}t = t$$

3) We have $\mathbf{M} = \mathbf{Q}_p^{-1}\mathbf{G}_p^*\mathbf{D}^*\mathbf{H}_q^*\mathbf{Q}_q$. Consequentely,

$$\mathbf{M}\mathbb{1} = \mathbf{Q}_p^{-1}\mathbf{G}_p^*\mathbf{D}^*\mathbf{H}_q^*\mathbf{Q}_q\mathbb{1} = \mathbf{Q}_p^{-1}\mathbf{G}_p^*\mathbf{D}^*\mathbf{H}_q^*q = \mathbf{Q}_p^{-1}\mathbf{G}_p^*\mathbf{D}^*\mathbb{1}^* = \mathbf{Q}_p^{-1}\mathbf{G}_p^*\mathbb{1}^* = \mathbf{Q}_p^{-1}p = \mathbb{1}$$

The new definition of majorization is connected with the old one via

$$(q, t) \prec^* (p, s) \iff \mathbf{G}_q\mathbf{Q}_q^{-1}t \prec_{\mathbb{1}} \mathbf{G}_p\mathbf{Q}_p^{-1}s$$

References

[1] W. Arendt, A. Grabosch, G. Greiner, U. Groh, H.P. Lotz, U. Moustakas, R. Nagel, F. Neubrander, U. Schlotterbeck, *One-parameter Semigroups of Positive Operators*, Springer 1986.

[2] Colin Bennett, Robert C. Sharpley, *Interpolation of Operators*, Academic Press, Orlando, 1988.

[3] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, SIAM, 1987.

[4] L. Fuchs, A new proof of an inequality of Hardy-Littlewood-Polya, Mat. Tidsskr. B. 1947. p. 53-54

[5] D. Grinberg et al.,
Weighted majorization and a result stronger than Fuchs – MathLinks topic #104714

[6] G. H. Hardy, J. E. Littlewood, G. Polya, *Inequalities*, Cambridge University Press, 1952.

[7] S. Kaplan, *The Bidual of* $C(X)$ *I*, North Holland, 1985.

[8] S. Kaplan, *Lebesgue Theory in the Bidual of* $C(X)$, American Mathematical Society, 1996.

[9] Elliott H. Lieb, Michael Loss, *Analysis*, Graduate Studies in Mathematics, vol. 14

[10] Albert W. Marshall, Ingram Olkin, Barry C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer, 2nd edition, 2009.

[11] A. Papoulis *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, 1984.

[12] Z. Semadeni, *Banach Spaces of Continuous Functions*, Warszawa 1971.

[13] J. Michael Steele, *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*, Cambridge University Press, 2004.

[14] H. Stephan, *Lyapunov functions for positive linear evolution problems*, ZAMM Z. Angew. Math. Mech., 85 (2005) pp. 766-777

[15] H. Stephan, A mathematical framework for general classical systems and time irreversibility as its consequence. *WIAS Berlin 2011*; Preprint n°1629.

[16] H. Stephan, Microscopic evolution equations in statistical physics. *WIAS Berlin 2014*; Preprint

[17] C. Zylka, Über eine neue Halbordnung, Leibniz-Online, 3(2007).