DISSERTATION

Titel der Dissertation

„Distance Of Probability Measures
And Respective Continuity Properties
Of Acceptability Functionals"

Verfasser

Dipl.-Ing. Mag. Alois Pichler

Angestrebter akdaemischer Grad

Doctor of Philosophy (PhD)

Wien, im November 2010

Studienkennzahl laut Studienblatt:	A 094 136
Dissertationsgebiet laut Studienblatt:	Statistik und Operations Research
Betreuer:	Univ.-Prof. Mag. Dr. Georg Ch. Pflug

富美枝のために

Contents

[1]

[1] Thesis-2010-09

Introduction

Stochastic programming – or stochastic optimization, as it is often referred to – is a framework for modelling optimization problems that involve uncertainty in some way. From historic perspective, stochastic programming already appeared at the very beginning of (linear) programming and it may be considered as one of the main driving forces in the research and investigations on optimization in general, even for particular problems in linear programming. In recent years then stochastic optimization was employed for even more problems, among them for example financial and economic applications.

Particularly in economic environments the stochastic nature is often intrinsic, as for example economic cycles or stocks cannot be foreseen, they have to be modelled in some way. – On the other side there is a strong wish and evident desire to better understand,

(i) to *model*,

(ii) to *measure* and

(iii) to *manage*

economic changes and related risks. This holds true for private persons with a basic interest in economics, thus not only for fund managers in the privileged and responsible position to manage a certain portfolio.

This work is dedicated to stochastic optimization and investigates problems in stochastic programming, which particularly arise in economic environments:

At the beginning a general framework is provided, which allows measuring uncertainty in an appropriate and convenient way. Then risk functionals are introduced, which allow to quantify risk, which go along with economic decisions. Risk functionals have gained a lot of interest in recent years, as they have very strong regularizing properties, which are of crucial interest in a stochastic environment, even leading to robust optimization. It is the further purpose to measure these risks, which are immanent to problems, where uncertainty plays a prominent role.

Some following sections then will investigate continuity properties and clarify the question, how the probability distribution impacts the solution of the stochastic problem in consideration. We will give conditions allowing controlling the impact of the pertaining distributions on the decisions.

In the situation where stochastic information is not available at all or missing to some extend, a kind of rule of thumb, which is intuitively clear, is proved, stating that it is better to equally distribute ones funds than to expose oneself on single stocks.

Some further investigations on approximating probability distributions complete the work: The purpose of this last section then is to reduce the general, multidimensional probability distributions to finite distributions, and to make them available for computational investigations.

The three-step process mentioned above – *to model, to measure and to manage* – was introduced by Prof. Pflug. It somehow reminds to Georg W. F. Hegel's dialectic (the triad) of *thesis, antithesis and synthesis*, the basis of getting to knowledge.

In this sense I feel deeply indebted to Prof. Pflug. I want to thank him for fundamental discussions, for incitement and co-operation during the period where the foundations of the first step in this tree-step-process – the thesis – were laid.

A. Pichler

Vienna, May 2010

Part I.

Continuity Properties Of Acceptability Functionals

1. Stochastic Funding

In this section we shall address and elaborate the concept of distances, in particular distances of probability measures. Actually there exists a broad variety of distance concepts, Rachev lists 76 metrics for probability spaces in [44]. We have found that the concept of Wasserstein distances is adapted to the problems we have in mind, and this is the reason why we will elaborate on this particular distance.

1.1. Metric Spaces

We consider a Polish space, i.e. a complete and separable space with metrizable topology; the pair $(\Omega,\ d)$ denotes such a space Ω, equipped with its respective distance function d.

1.1.1. Product Space

The product of finitely many Polish spaces $\left(\Omega^T,\ d\right) := \times_{t\in T}\ (\Omega_t,\ d_t)$ is a Polish space as well, and several distance functions metrisize the same topology, for example

- ▷ $d(x,y) := d_p(x,y) := \left(\sum_t w_t \cdot d_t(x_t,y_t)^p\right)^{1/p}$ $(p \geq 1)$ or
- ▷ $d(x,y) := d_\infty(x,y) := \max_t\ w_t \cdot d_t(x_t,y_t)$;

the weights are strictly positive, $w_t > 0$.

1.1.2. Wasserstein Distance

Given a Polish space we shall consider probability measures on its Borel sets. The collection of all probability measures, which satisfy for some – and thus any – $\omega_0 \in \Omega$ the moment-like condition

$$\int_\Omega d(\omega_0,\omega)^r\, \mathbb{P}[\mathrm{d}\omega] < \infty \tag{1.1}$$

is denoted by $\mathcal{P}_r(\Omega; d)$.

On this space of probability measures define the function

$$\mathsf{d}_r(\mathbb{P}_1,\mathbb{P}_2; d) := \left(\inf\left\{\int_{\Omega\times\Omega} d(\omega_1,\omega_2)^r\, \pi[\mathrm{d}\omega_1,\mathrm{d}\omega_2]\right\}\right)^{\frac{1}{r}}, \tag{1.2}$$

where the infimum is taken over all (bivariate) probability measures π on $\Omega \times \Omega$ which have respective marginals, that is

$$\pi [A \times \Omega] = \mathbb{P}_1 [A] \text{ and } \pi [\Omega \times B] = \mathbb{P}_2 [B]$$

for all measurable sets $A \subseteq \Omega$ and $B \subseteq \Omega$. We shall call such a measure π a *transport plan.*

d_r is called r^{th}-*Wasserstein distance.* It is well-defined, as for example the product measure[1]

$$\pi := \mathbb{P}_1 \otimes \mathbb{P}_2$$

has the required marginals and whence

$$\mathsf{d}_r (\mathbb{P}_1, \mathbb{P}_2; d)^r \leq \int_\Omega \int_\Omega d (\omega_1, \omega_2)^r \, \mathbb{P}_1 [\mathrm{d}\omega_1] \, \mathbb{P}_2 [\mathrm{d}\omega_2] \, .$$

A very comprehensive and beautiful discussion and treatment of this function d_r can be found in Villanis beautiful books ([56] and [57]), but we want to mention the books by Rachev and Rüschendorf as well, [45][2] and [20].

We shall use the properties that the infimum in (1.2) is actually attained, and $\mathsf{d}_r (.,\ .; d)$ turns out to be a metric on the space $\mathcal{P}_r (\Omega; d)$, so particularly satisfies the triangle inequality

$$\mathsf{d}_r (\mathbb{P}, \mathbb{Q}; d) \leq \mathsf{d}_r \left(\mathbb{P}, \tilde{\mathbb{Q}}; d\right) + \mathsf{d}_r \left(\tilde{\mathbb{Q}}, \mathbb{Q}; d\right) .$$

Remark 1.1. We are using the symbol d for the distance in the original space Ω, and $\mathsf{d}_r (.; d)$ to account for the distance on probabilities in $\mathcal{P}_r (\Omega; d)$. However, if no confusion may occur in the given context, we will omit the additional argument in the sequel and simply write $\mathsf{d}_r (\mathbb{P}, \mathbb{Q}) = \mathsf{d}_r (\mathbb{P}, \mathbb{Q}; d)$ for the distance on $\mathcal{P}_r$ specified by d.

In honor of G. Monge[3] (cf. [34]) and Leonid Kantorovich[4] (cf. [26]) the distance d_r is sometimes called *Monge-Kantorovich distance* of order r, and d_2 is called quadratic Wasserstein distance as well. Moreover, the distance d_1 is also called Kantorovich-Rubinstein distance and sometimes denoted $\mathsf{d}_{KA} := \mathsf{d}_1$.

Lemma 1.2 (Monotonicity and convexity)**.**

(i) Suppose that $r_1 \leq r_2$, then $\mathsf{d}_{r_1} (\mathbb{P}, \mathbb{Q}) \leq \mathsf{d}_{r_2} (\mathbb{P}, \mathbb{Q})$.

(ii) The Wasserstein distance is r-convex[5] in any of its components, that is to say for $0 \leq \lambda \leq 1$ we have that

$$\mathsf{d}_r (\mathbb{P}, (1 - \lambda) \mathbb{Q}_0 + \lambda \mathbb{Q}_1)^r \leq (1 - \lambda) \, \mathsf{d}_r (\mathbb{P}, \mathbb{Q}_0)^r + \lambda \, \mathsf{d}_r (\mathbb{P}, \mathbb{Q}_1)^r ,$$

[1] $(\mathbb{P}_1 \otimes \mathbb{P}_2) [A \times B] := \mathbb{P}_1 [A] \cdot \mathbb{P}_2 [B]$ defines a $\sigma-$additive measure due to the Hahn-Kolmogorov theorem.

[2] For a different concept of distance in an actuarial context cf. [41].

[3] Gaspard Monge (1746 - 1818) investigated how to efficiently construct dugouts.

[4] L. Kantorovich was awarded the price in Economic Sciences in Memory of Alfred Nobel in 1975.

[5] For the notion of r-concavity (r-convexity) see [53].

and

$$\begin{aligned}&\mathsf{d}_r\left(\mathbb{P},\left(1-\lambda\right)\mathbb{Q}_0+\lambda\mathbb{Q}_1\right)\\&\quad\leq(1-\lambda)^{\frac{1}{r}}\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_0\right)+\lambda^{\frac{1}{r}}\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_1\right)\\&\quad\leq\max\left\{\lambda,1-\lambda\right\}^{\frac{1}{r}-1}\left((1-\lambda)\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_0\right)+\lambda\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_1\right)\right).\end{aligned}$$

Remark 1.3. It should be noted that convexity in the traditional sense is actually achieved for the Kantorovich distance ($r=1$),

$$\mathsf{d}_{KA}\left(\mathbb{P},\,(1-\lambda)\,\mathbb{Q}_0+\lambda\mathbb{Q}_1\right)\leq(1-\lambda)\,\mathsf{d}_{KA}\left(\mathbb{P},\mathbb{Q}_0\right)+\lambda\,\mathsf{d}_{KA}\left(\mathbb{P},\mathbb{Q}_1\right);$$

for the general Wasserstein distance ($r>1$), however, a correction factor

$$\max\left\{\lambda,1-\lambda\right\}^{\frac{1}{r}-1}>1$$

has to be accepted.

Proof. Observe that $\frac{1}{\frac{r_2}{r_2-r_1}}+\frac{1}{\frac{r_2}{r_1}}=1$. By use of Hölder's inequality (cf. Proposition 10.1 in the Appendix)

$$\begin{aligned}\int d^{r_1}\mathrm{d}\pi&=\int 1\cdot d^{r_1}\mathrm{d}\pi\\&\leq\left(\int 1^{\frac{r_2}{r_2-r_1}}\mathrm{d}\pi\right)^{\frac{r_2-r_1}{r_2}}\cdot\left(\int d^{r_1\frac{r_2}{r_1}}\mathrm{d}\pi\right)^{\frac{r_1}{r_2}}\\&=\left(\int d^{r_2}\mathrm{d}\pi\right)^{\frac{r_1}{r_2}}.\end{aligned}$$

Thus, $\left(\int d^{r_1}\mathrm{d}\pi\right)^{\frac{1}{r_1}}\leq\left(\int d^{r_2}\mathrm{d}\pi\right)^{\frac{1}{r_2}}$ for every measure π, which proves the first assertion.

As for the second let π_0 and π_1 be measures chosen with adequate marginals in such way that

$$\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_0\right)^r=\int d\left(x,y\right)^r\pi_0\left[\mathrm{d}x,\mathrm{d}y\right]\ \text{and}\ \mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_1\right)^r=\int d\left(x,y\right)^r\pi_1\left[\mathrm{d}x,\mathrm{d}y\right].$$

The probability measure $\pi_\lambda:=(1-\lambda)\,\pi_0+\lambda\pi_1$ then has the marginals $\mathbb{P}$ and $\mathbb{Q}_\lambda:=(1-\lambda)\,\mathbb{Q}_0+\lambda\mathbb{Q}_1$, and

$$\begin{aligned}&\mathsf{d}_r(\mathbb{P},\,(1-\lambda)\mathbb{Q}_0+\lambda\mathbb{Q}_1)^r\\&\quad\leq\int d\left(x,y\right)^r\pi_\lambda\left[\mathrm{d}x,\mathrm{d}y\right]\\&\quad=(1-\lambda)\int d\left(x,y\right)^r\pi_0\left[\mathrm{d}x,\mathrm{d}y\right]+\lambda\int d\left(x,y\right)^r\pi_1\left[\mathrm{d}x,\mathrm{d}y\right]\\&\quad=(1-\lambda)\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_0\right)^r+\lambda\,\mathsf{d}_r\left(\mathbb{P},\mathbb{Q}_1\right)^r.\end{aligned}$$

The other statements follow as $x\mapsto x^{\frac{1}{r}}$ is concave and by employing Hölder's inequality. □

Remark 1.4. To note an important consequence: All evaluations are continuous with respect to d_r, provided they are continuous with respect to $\mathsf{d}_1 = \mathsf{d}_{KA}$, the Kantorovich-distance. As an example consider the next lemma:

Lemma 1.5. *Consider a linear space equipped with a norm (that is $d(x,y) := \|x-y\|$, for example $(\Omega, d) = (\mathbb{R}^m, \|.\|)$, then*

$$\|\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}\| \le \mathsf{d}_r(\mathbb{P}_1, \mathbb{P}_2)$$

where $\mu_{\mathbb{P}} := \mathbb{E}_{\mathbb{P}}[\mathsf{Id}]$ is the barycentre with respect to the measure $\mathbb{P}$ – provided it exists[6]*.*

Proof. The proof for the Kantorovich distance ($r = 1$) is an application of Jensen's inequality as the norm is a convex function:

$$\begin{aligned} \|\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}\| &= \left\| \int x\, \mathbb{P}_1[\mathrm{d}x] - \int y\, \mathbb{P}_2[\mathrm{d}y] \right\| \\ &= \left\| \int (x-y)\ \pi[\mathrm{d}x, \mathrm{d}y] \right\| \\ &\le \int \|x-y\|\, \pi[\mathrm{d}x, \mathrm{d}y]. \end{aligned}$$

Taking the infimum over all measures π with appropriate marginals $\mathbb{P}_1$ and $\mathbb{P}_2$ gives the assertion, as $\|\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}\| \le \mathsf{d}_1(\mathbb{P}_1, \mathbb{P}_2) \le \mathsf{d}_r(\mathbb{P}_1, \mathbb{P}_2)$. □

Remark 1.6. This formula gives rise to the interpretation, that every particle has to be transported – on average – at least the distance of the barycentres $\mu_{\mathbb{P}_1} - \mu_{\mathbb{P}_2}$.

1.2. Comparison Of Wasserstein With Weak* Topology

A Polish space may be considered as a natural subspace of probability measures on the same space. The embedding is elaborated in the next statement and involves Dirac's point measure $\delta_x[A] := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$.

Theorem 1.7 (Embedding)**.** *The mapping*

$$\begin{aligned} i\colon\ (\Omega, d) &\to (\mathcal{P}_r(\Omega; d), \mathsf{d}_r) \\ \omega &\mapsto \delta_\omega \end{aligned}$$

is an isometric embedding for all $1 \le r < \infty$.

[6] $\mathsf{Id}(\omega) := \omega$ is the identity

Proof. There is just one single measure with the marginals δ_{ω_1} and δ_{ω_2}, which is the transport plan $\pi = \delta_{\omega_1} \otimes \delta_{\omega_2}$, that is $\pi[A \times B] = \delta_{\omega_1}[A] \cdot \delta_{\omega_2}[B]$. Thus

$$\begin{aligned} \mathsf{d}_r(\delta_{\omega_1}, \delta_{\omega_2})^r &= \int d(\omega, \omega')^r \, \delta_{\omega_1}[\mathrm{d}\omega] \otimes \delta_{\omega_2}[\mathrm{d}\omega'] \\ &= d(\omega_1, \omega_2)^r . \end{aligned}$$

□

Remark 1.8. It should be noted that Ω may be considered as a *strict, closed subset* of $\mathcal{P}_r$, $\Omega \subsetneqq \mathcal{P}_r$, as one verifies straight forward that

$$\mathsf{d}_r(\delta_x, \mathbb{P})^r = \int d(x, \omega)^r \, \mathbb{P}[\mathrm{d}\omega] > 0,$$

whenever $\mathbb{P} \neq \delta_x$: As above, there is just one unique transport plan, which is $\pi = \delta_x \otimes \mathbb{P}$.

For additional insight it may be mentioned that Ω is even uniformly bounded away from simple convex combinations in the following sense:

$$\inf\left\{\mathsf{d}_r\left(\frac{1}{2}(\delta_x + \delta_y), \delta_\omega\right) :\ \omega \in \Omega\right\} \geq \frac{1}{2} d(x, y) :$$

This follows from the fact that there is again just a single transport plan to transport $\frac{1}{2}(\delta_x + \delta_y)$ to δ_ω, which is the measure $\pi := \frac{1}{2}\delta_x \otimes \delta_\omega + \frac{1}{2}\delta_x \otimes \delta_\omega$. Whence

$$\begin{aligned} \mathsf{d}_r\left(\frac{1}{2}(\delta_x + \delta_y), \delta_\omega\right) &\geq \mathsf{d}_{KA}\left(\frac{1}{2}(\delta_x + \delta_y), \delta_\omega\right) \\ &= \int d(\omega_1, \omega_2)\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_1] \\ &= \frac{1}{2} d(x, \omega) + \frac{1}{2} d(\omega, y) \\ &\geq \frac{1}{2} d(x, y) \end{aligned}$$

due to Lemma 1.2 and the triangle inequality.

The Wasserstein distance d_r induces a topology on $\mathcal{P}_r(\Omega; d)$ which we want to denote as τ_{d_r} here.

We may consider $\mathcal{P}_r(\Omega; d)$ additionally as a subset of the dual of $C_b(\Omega)$, the set of all bounded and continuous functions on Ω. We thus have the weak* topology available, which we denote as $\sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right)$.

In comparing these two topologies there is a slight barrier: Note, that

$$\begin{aligned} \mathcal{P}_r(\Omega; d) = &\left\{\mathbb{P} \in C_b(\Omega)^* :\ \int 1 \mathrm{d}\mathbb{P} = 1 \text{ and } \int \varphi \mathrm{d}\mathbb{P} \geq 0 \text{ for all } \varphi \geq 0\right\} \\ &\cap \left\{\mathbb{P} \in C_b(\Omega)^* :\ \int d(\omega, \omega_0)^r \, \mathbb{P}[\mathrm{d}\omega] < \infty\right\}. \end{aligned}$$

The defining moment condition (1.1) imposed (the latter condition) indicates that $\mathcal{P}_r(\Omega; d)$ is possibly *not closed* in the topology $\sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right)$, unless $\omega \mapsto d(\omega, \omega_0)^r$ is a bounded function itself.

To illustrate the key differences between the weak* topology

$$\sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right)$$

and the topology induced by the Wasserstein distance on $\mathcal{P}_r(\Omega)$ consider the sequence of measures (with some mass disappearing to infinity)

$$\mathbb{P}_n := \left(1 - \frac{1}{n^r}\right)\delta_0 + \frac{1}{n^r}\delta_n$$

on the real line $\mathbb{R}$:

(i) By definition, $(\mathbb{P}_n)_n$ converges in weak* topology $\sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right)$ to $\mathbb{P}$ if, and only if, (iff) for all $\varphi \in C_b(\Omega)$ (φ bounded and continuous) $\int \varphi \mathrm{d}\mathbb{P}_n \to \int \varphi \mathrm{d}\mathbb{P}$. The sequence $(\mathbb{P}_n)_n$ in consideration thus converges to δ_0 in the weak* sense, as

$$\begin{aligned}\int \varphi \mathrm{d}\mathbb{P}_n &= \left(1 - \frac{1}{n^r}\right)\varphi(0) + \frac{1}{n^r}\varphi(n) \\ &\to \varphi(0) = \int \varphi \mathrm{d}\delta_0,\end{aligned}$$

that is to say:

$$\mathbb{P}_n \to \delta_0 \qquad \sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right).$$

(ii) Employing the usual absolute value $|.|$ as a distance on the real line we find that

$$\begin{aligned}\mathsf{d}_r(\mathbb{P}_n, \delta_0; |.|)^r &= \int |\omega|^r \, \mathbb{P}_n[\mathrm{d}\omega] \\ &= \left(1 - \frac{1}{n^r}\right)\cdot 0 + \frac{1}{n^r}\cdot n^r \\ &= 1.\end{aligned}$$

Whence,

$$\mathbb{P}_n \not\to \delta_0 \qquad \tau_{\mathsf{d}_r}.$$

(iii) On $\mathbb{R}$, however, there is the *equivalent* metric $|x|' := \min\{|x|, 1\}$. Using this metric to compute the Wasserstein distance one obtains

$$\begin{aligned}\mathsf{d}_r\left(\mathbb{P}_n, \delta_0; |.|'\right)^r &= \int |\omega|'^r \, \mathbb{P}_n[\omega] \\ &= \left(1 - \frac{1}{n^r}\right)\cdot 0 + \frac{1}{n^r}1 \\ &\to 0,\end{aligned}$$

and thus

$$\mathbb{P}_n \to \delta_0 \qquad \tau_{\mathsf{d}_r(.,\|\|')}.$$

Thus, starting even with an equivalent metric on Ω, the topologies on $\mathcal{P}_r(\Omega; d)$ are *not* necessarily *equivalent*.

(iv) One may understand measures as linear functionals of the form

$$\begin{aligned} \mathbb{P} : C_b(\Omega) &\to \mathbb{R} \\ \varphi &\mapsto \int \varphi \mathrm{d}\mathbb{P} \end{aligned}$$

on the space of bounded, continuous function $C_b(\Omega)$ as well, and equip the latter space as usual with the Banach-space norm

$$\|\varphi\|_\infty := \sup_{\omega\in\Omega} |\varphi(\omega)|$$

for $\varphi \in C_b(\Omega)$. In this situation

$$\|\mathbb{P}\| = \sup_{\|\varphi\|_\infty \le 1} \left| \int \varphi \mathrm{d}\mathbb{P} \right|,$$

which means for the measures in consideration

$$\begin{aligned} \|\mathbb{P}_n - \mathbb{P}\| &= \sup_{\|\varphi\|_\infty \le 1} \left| \varphi(0)\left(1 - \frac{1}{n^r}\right) + \varphi(n)\frac{1}{n^r} - \varphi(0) \right| \\ &= \frac{1}{n^r} \sup_{\|\varphi\|_\infty \le 1} |\varphi(n) - \varphi(0)| \\ &\to 0. \end{aligned}$$

That is to say

$$\mathbb{P}_n \to \delta_0 \qquad \tau_{\|.\|}$$

in the strong (norm) topology.

Summarizing, we have weak* convergence and even strong convergence, but the Wasserstein distance in general is something different.

The next theorem will elaborate the generalities of these peculiar patterns with the result, that the weak* topology $\sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right)$ is – in general and as its name indicates – strictly *weaker* than the topology induced by the Wasserstein distance: $\tau^* \subsetneq \tau_{\mathsf{d}_r}$; whenever d is bounded, then the topologies coincide,

$$\tau_{\mathsf{d}_r} = \sigma\left(\mathcal{P}_r(\Omega; d), C^b(\Omega)\right).$$

A crucial tool to unify the topologies on subsets will be the uniform tightness condition (1.3) below.

Theorem 1.9 (Wasserstein metrizes the weak* topology)**.** *Let $(\mathbb{P}_n)_n$ be a sequence of measures in $\mathcal{P}_r(\Omega)$, and let $\mathbb{P} \in \mathcal{P}_r(\Omega)$. Then the following are equivalent:*

(i) $\mathsf{d}_r(\mathbb{P}_n, \mathbb{P}) \xrightarrow[n\to\infty]{} 0$,

(ii) $\mathbb{P}_n \xrightarrow[n\to\infty]{} \mathbb{P}$ *in weak* sense, and* $\mathbb{P}_n$ *satisfies the following* uniform tightness *condition: For some (and thus any)* $\omega_0 \in \Omega$,

$$\limsup_{n\to\infty} \int_{d(\omega_0,\omega)\geq R} d(\omega_0,\omega)^r\, \mathbb{P}_n[\mathrm{d}\omega] \xrightarrow[R\to\infty]{} 0. \tag{1.3}$$

Proof. As for the proof we refer to Theorem 7.12 in Villani [56]. □

Notice, that one may – as we did in the exemplary introduction above – always replace d by the bounded distance, say, $d'(x,y) := \frac{d(x,y)}{1+d(x,y)}$ or $d'(x,y) := \min\{1,\, d(x,y)\}$. So in this situation the uniform tightness condition is trivial, and Wasserstein thus metrizes weak* convergence on the whole of $\mathcal{P}_r(\Omega)$. In this situation $\mathcal{P}_r(\Omega)$ is closed in the topology $\sigma\left(\mathcal{P}_r(\Omega; d),\, C^b(\Omega)\right)$, the topologies coincide.

Essential for our intentions on quantization is to have simple approximations available. The next theorem clarifies the role of quantizers to a sufficient extent.

Initial proofs of the statement involve the weaker Prohorov distance and deep results of Kolmogorov in [1]; the following proof by elementary means is partially adapted from [7].

Lemma 1.10 (Auxiliary lemma)**.** *Let* $w \geq 0$*. Then*

$$(x^p + w\,y^p)^{\frac{1}{p}} \leq \begin{cases} (x^r + w\,y^r)^{\frac{1}{r}} & r \leq p \\ (1+w)^{\frac{1}{p}-\frac{1}{r}}\,(x^r + w\,y^r)^{\frac{1}{r}} & r \geq p \end{cases} \tag{1.4}$$

and

$$(x+y)^r \leq (1+w)^{r-1}\left(x^r + w^{1-r} y^r\right). \tag{1.5}$$

Proof. To accept (1.4) notice, by differentiating,

$$\begin{aligned} 0 &= \frac{\partial}{\partial y}\frac{(x^p + wy^p)^{1/p}}{(x^r + wy^r)^{1/r}} \\ &= \frac{(x^p + wy^p)^{\frac{1}{p}-1}}{(x^r + wy^r)^{\frac{1}{r}+1}}\frac{w}{y}\left(x^r y^p - x^p y^r\right) \end{aligned}$$

implies that there is an extremum at $x = y$. Comparing with the value on the boundary ($x, y \to 0$, $x, y \to \infty$) gives the first statement.

As for (1.5) substitute $p \leftarrow 1$, $y \leftarrow y/w$ in (1.4). □

Theorem 1.11. *If* (Ω, d) *is separable, then* $(\mathcal{P}_r(\Omega), \mathsf{d}_r)$ *is separable and all finite, discrete measures* $\sum_{\omega\in\mathcal{Q}} \mathbb{P}_\omega \delta_\omega$ *($\mathcal{Q}$ finite) are dense.*

Proof. Let $\mathbb{P} \in \mathcal{P}_r(\Omega)$ be any measure and choose $\epsilon > 0$. Due to separability there is a dense sequence ω_n, and $\Omega = \bigcup_{n=1}^{\infty} B_\epsilon(\omega_n)$. Now successively define the partition

$$\Omega_n := B_\epsilon(\omega_n) \setminus \bigcup_{j=1}^{n-1} \Omega_j$$

covering Ω again, but these sets are pairwise disjoint. Set $\mathbb{P}_n := \mathbb{P}[\Omega_n]$ and, due to the σ-additivity of the measure, $\sum_{n=1}^{\infty} \mathbb{P}_n = 1$.

One verifies that the measure $\mathbb{P}_\epsilon := \sum_{n=1}^{\infty} \mathbb{P}_n \delta_{\omega_n} \in \mathcal{P}_r(\Omega)$, because

$$\begin{aligned}\left(\int d(\omega_0, \omega)^r \, \mathbb{P}_\epsilon[\mathrm{d}\omega]\right)^{\frac{1}{r}} &= \mathsf{d}_r(\delta_{\omega_0}, \mathbb{P}_\epsilon) \\ &\leq \mathsf{d}_r(\delta_{\omega_0}, \mathbb{P}) + \mathsf{d}_r(\mathbb{P}, \mathbb{P}_\epsilon) \\ &= \int d(\omega_0, \omega) \, \mathbb{P}[\mathrm{d}\omega] + \mathsf{d}_r(\mathbb{P}, \mathbb{P}_\epsilon) \\ &\leq K + \left(\int d(\omega, \omega')^r \, \pi_\epsilon[\mathrm{d}\omega, \mathrm{d}\omega]\right)^{\frac{1}{r}} \\ &\leq K + \epsilon\end{aligned}$$

when transporting Ω_n to $\{\omega_n\}$, which is established by the transport plan $\pi_\epsilon[A \times B] := \sum_{n=1} \mathbb{P}[A \cap \Omega_n] \cdot \delta_{\omega_n}[B]$. Whence,

$$\int d(\omega, \omega_0)^r \, \mathbb{P}_\epsilon[\mathrm{d}\omega] = \sum_{n=1}^{\infty} \mathbb{P}_n d(\omega_n, \omega_0)^r < \infty$$

is finite and so $\sum_{n=1}^{\infty} \mathbb{P}_n \delta_{\omega_n} \in \mathcal{P}_r(\Omega)$.

We thus may choose N_ϵ big enough such that $\sum_{n=N_\epsilon+1}^{\infty} \mathbb{P}_n d(\omega_n, \omega_0)^r < \epsilon^r$. The discrete *and finite* measure

$$\sum_{n=1}^{N_\epsilon} \mathbb{P}_n \delta_{\omega_n} + \left(\sum_{n=N_\epsilon+1} \mathbb{P}_n\right) \cdot \delta_{\omega_0}$$

approximates $\mathbb{P}$ sufficiently good: choose the transport plan

$$\begin{aligned}\pi[A \times B] = &\sum_{n=1}^{N_\epsilon} \mathbb{P}[A \cap \Omega_n] \cdot \delta_{\omega_n}[B] \\ &+ \mathbb{P}\left[A \setminus \bigcup_{n=1}^{N_\epsilon} \Omega_n\right] \cdot \delta_{\omega_0}[B],\end{aligned}$$

then

$$\int d(\omega,\omega')^r \pi[\mathrm{d}\omega,\mathrm{d}\omega'] =$$

$$= \sum_{n=1}^{N_\epsilon} \int_{\Omega_n} d(\omega,\omega_n)^r \mathbb{P}[\mathrm{d}\omega] + \sum_{n=N_\epsilon+1} \int_{\Omega_n} d(\omega,\omega_0)^r \mathbb{P}[\mathrm{d}\omega]$$

$$\le \sum_{n=1}^{N_\epsilon} \int_{\Omega_n} d(\omega,\omega_n)^r \mathbb{P}[\mathrm{d}\omega] +$$

$$+ \sum_{n=N_\epsilon+1} \int_{\Omega_n} (d(\omega,\omega_n) + d(\omega_n,\omega_0))^r \mathbb{P}[\mathrm{d}\omega]$$

and further, using (1.5) from the auxiliary lemma with $w = 1$,

$$\int d(\omega,\omega')^r \pi[\mathrm{d}\omega,\mathrm{d}\omega'] =$$

$$\le 2^{r-1} \sum_{n=1} \int_{\Omega_n} d(\omega,\omega_n)^r \mathbb{P}[\mathrm{d}\omega]$$

$$+ 2^{r-1} \sum_{n=N_\epsilon+1} \int_{\Omega_n} d(\omega_n,\omega_0)^r \mathbb{P}[\mathrm{d}\omega]$$

$$\le 2^{r-1} \int_{\Omega} \epsilon^r \mathbb{P}[\mathrm{d}\omega] + 2^{r-1} \sum_{n=N_\epsilon+1} \mathbb{P}_n d(\omega_n,\omega_0)^r$$

$$\le 2^{r-1}\epsilon^r + 2^{r-1}\epsilon^r.$$

This establishes that all finite measures are dense.

Now choose rational numbers $\tilde{\mathbb{P}}_k \in \mathbb{Q} \cap [0,1]$ with $\left|\tilde{\mathbb{P}}_k - \mathbb{P}_k\right| \le \frac{\epsilon}{2^k \max_{i\neq j}^{N_\epsilon} d(\omega_i,\omega_j)^p}$ for $k = 0, 1 \ldots N_\epsilon$ having unit sum (0 is included here with $\mathbb{P}_0 := \sum_{k=N_\epsilon+1} \mathbb{P}_k$). Then

$$\mathsf{d}_r\left(\sum_{k=0}^{N_\epsilon} \mathbb{P}_k \delta_{\omega_k}, \sum_{k=0}^{N_\epsilon} \tilde{\mathbb{P}}_k \delta_{\omega_k}\right) < 2\epsilon$$

This establishes separability, because all computations here are restricted to the sequence ω_n, being dense in the entire Ω. □

Theorem 1.12. *Let (Ω, d) be a Polish space, then $(\mathcal{P}_r(\Omega; d), \mathsf{d}_r)$ is a Polish space again.*

Proof. The space is metrizable and we have established separability. The only ingredient missing which makes $(\mathcal{P}_r(\Omega; d), \mathsf{d}_r)$ Polish is completeness, which is established in [7]. We shall mention only that – interestingly – compactness and Prohorov's theorem are required to derive the result. □

2. Kantorovich Duality

Discrete measures are *dense* with respect to the Wasserstein distance; we thus shall investigate the distance of two discrete measures in some more detail, and initially in some more generality:

Consider the discrete measures $\mathbb{P} := \sum_s \mathbb{P}_s \delta_{\omega_s}$ and $\mathbb{Q} := \sum_t \mathbb{Q}_t \delta_{\omega'_t}$ and – within this setting – the problem

$$\begin{array}{rl} \text{minimize} & \\ \text{(in } \pi) & \langle \pi, c\rangle := \sum_{s,t} \pi_{s,t} c_{s,t} \\ \text{subject to} & \sum_t \pi_{s,t} = \mathbb{P}_s, \\ & \sum_s \pi_{s,t} = \mathbb{Q}_t, \\ & \pi_{s,t} \geq 0, \end{array} \tag{2.1}$$

where $c_{s,t} := c(\omega_s, \omega'_t)$ is a matrix derived from the a general cost-function c, representing the costs related to the transportation of masses from ω_s to ω'_t and $\langle \pi, c\rangle = \sum_{s,t} \pi_{s,t} c_{s,t}$ is the objective (we would like to point out the reference [6] for a similar discretization and approach).

Notice that this is just (1.2) in the discrete setting for the particular cost-function

$$c_{s,t} := d(\omega_s, \omega'_t)^r .$$

Observe further that

$$\sum_{s,t} \pi_{s,t} = \sum_s \mathbb{P}_s = \sum_t \mathbb{Q}_t = 1,$$

π thus is a probability measure, representing a transport plan from $\mathbb{P}$ to $\mathbb{Q}$.

(2.1) is actually a *linear program*, one may thus compute the objective involving the dual program as well – cf. subsection 10.3.3 in the Appendix. The respective dual of problem (2.1) is

$$\begin{array}{rl} \text{maximize} & \\ \text{(in } \lambda, \mu) & \sum_s \mathbb{P}_s \lambda_s + \sum_t \mathbb{Q}_t \mu_t \\ \text{subject to} & \lambda_s + \mu_t \leq c_{s,t}, \end{array} \tag{2.2}$$

having the same objective value as (2.1) (cf. 10.3.3 in the Appendix).

It does not come as a big surprise that this result holds true in much more generality:

Theorem 2.1 (Kantorovich duality)**.** *Let X and Y be Polish spaces, $\mathbb{P}$ ($\mathbb{Q}$, resp.) a probability measure on X (Y, resp.), and $c\colon X\times Y\to\mathbb{R}_{\geq 0}$ be non-negative, lower semi-continuous*[1] *cost function. Then*

$$\inf_{\pi}\int_{X\times Y} c\left(x,y\right)\pi\left[\mathrm{d}x,\mathrm{d}y\right]=\sup_{\lambda,\mu}\int_X \lambda\left(x\right)\mathbb{P}\left[\mathrm{d}x\right]+\int_Y \mu\left(y\right)\mathbb{Q}\left[\mathrm{d}y\right], \tag{2.3}$$

where

(i) $\pi\left[A\times Y\right]=\mathbb{P}\left[A\right]$ *and* $\pi\left[X\times B\right]=\mathbb{Q}\left[B\right]$ *has adjusted marginals, and*

(ii) $\lambda\in\mathbb{L}^1\left(X,\mathbb{P}\right)$ *and* $\mu\in\mathbb{L}^1\left(Y,\mathbb{Q}\right)$ *are such that*

$$\lambda\left(x\right)+\mu\left(y\right)\leq c\left(x,y\right)$$

almost everywhere.

Proof. Again, the proof is contained in [56]. □

2.1. Maximal Kantorovich Potential

The dual variables in numerical computations appear very unhandy, as they are, dependent on the actual solver, often unpredictable high or low. This subsection is to better understand this pattern and to overcome this difficulty in a natural way.

Define the *c-concave* functions

$$\begin{aligned}\lambda_t^c &:= \min_s c_{s,t}-\lambda_s,\\ \mu_s^c &:= \min_t c_{s,t}-\mu_t\end{aligned} \tag{2.4}$$

and notice, that – given an optimal solution (λ,μ) of (2.2) – the pairs (λ,λ^c), (μ^c,μ) and particularly (λ^{cc},λ^c) maximize (2.2) as well.

The dual (2.2) thus rewrites in the simple form

$$\text{maximize (in }\lambda)\quad \sum_s \mathbb{P}_s\lambda_s^{cc}+\sum_t \mathbb{Q}_t\lambda_t^c\ .$$

This observation analogously holds true in a continuous environment as well, the $c-$*concave* functions then are

$$\begin{aligned}\lambda^c\left(t\right) &:= \min_s c\left(s,t\right)-\lambda\left(s\right),\\ \mu^c\left(t\right) &:= \min_t c\left(s,t\right)-\mu\left(t\right).\end{aligned}$$

This gives rise to the following

[1] that is the sets $\{c>t\}$ are open for all t, or equivalently, iff $c\left(y_0\right)\leq\liminf_{y\to y_0}c\left(y\right)$ whenever $y\to y_0$.

Definition 2.2 (maximal Kantorovich potential)**.** A function $\lambda \in \mathbb{L}^1(\Omega, \mathbb{P})$ is a *maximal Kantorovich potential* if (λ, λ^c) is a maximizing pair for (2.3) ((2.2), respectively).

It turns out that the maximal Kantorovich potential is not unique. However, we will deduce some natural bounds for the maximal Kantorovich potential in the sequel.

Indeed, for some fixed constant C, $\lambda_s + C$ and $\mu_t - C$ are solutions of the dual problem (2.2) as well with the same objective value, because $\mathbb{P}$ and $\mathbb{Q}$ are probabilities: We may thus normalize the dual variables and assume for a moment – without loss of generality – that $\max_t \mu_t = 0$.

With this fixing we obtain that

$$\lambda_s = \min_t c_{s,t} - \mu_t \geq \min_t c_{s,t},$$

and further that

$$\begin{aligned} 0 &= \max_t \mu_t \\ &= \max_t \min_{s'} c_{s',t} - \lambda_{s'} \\ &\leq \min_{s'} \max_t c_{s',t} - \lambda_{s'}, \\ &\leq \max_t c_{s,t} - \lambda_s, \end{aligned}$$

which is combined

$$\min_t c_{s,t} \leq \lambda_s \leq \max_t c_{s,t}.$$

These bounds are easier to handle in a linear program than the initial fixing, we thus replace the fixing $\max_t \mu_t = 0$ by $\min_t c_{s,t} \leq \lambda_s \leq \max_t c_{s,t}$ in the sequel to control the dual variable and helping to avoid any losses in significance for the dual variable in the numerical application. This is a relaxation, however, it will turn out to be almost equivalent.

Those bounds may be incorporated in the initial problem by adding two new variables (vectors) $\rho^{(1)}$ and $\rho^{(2)}$: The modified problem is

$$\begin{array}{rl} \text{minimize} & \\ (\text{in } \pi, \rho^{(1)} \text{ and } \rho^{(2)}) & \sum_{s,t} \pi_{s,t} c_{s,t} + \sum_s \rho_s^{(1)} \max_t c_{s,t} - \sum_s \rho_s^{(2)} \min_t c_{s,t} \\ \text{subject to} & \sum_t \pi_{s,t} + \rho_s^{(1)} - \rho_s^{(2)} = \mathbb{P}_s, \\ & \sum_s \pi_{s,t} = \mathbb{Q}_t, \\ & \pi_{s,t} \geq 0,\ \rho_s^{(1)} \geq 0,\ \rho_s^{(2)} \geq 0. \end{array}$$

This augmented problem now has the desired dual

$$\begin{array}{rl} \text{maximize} & \\ (\text{in } \lambda, \mu) & \sum_s \mathbb{P}_s \lambda_s + \sum_t \mathbb{Q}_t \mu_t \\ \text{subject to} & \lambda_s + \mu_t \leq c_{s,t}, \\ & \min_t c_{s,t} \leq \lambda_s \leq \max_t c_{s,t} \end{array} \tag{2.5}$$

and it follows further that

$$\begin{aligned}\mu_t &= \min_s c_{s,t} - \lambda_s \\ &\leq \min_s c_{s,t} - \min_{t'} c_{s,t'} \\ &\leq 0,\end{aligned}$$

when employing this particular index s which minimizes $\min_t c_{s,t}$; moreover,

$$\begin{aligned}\mu_t &= \min_s c_{s,t} - \lambda_s \\ &\geq \min_s c_{s,t} - \max_{t'} c_{s,t'} \\ &\geq \min_{s,t} c_{s,t} - \max_{s,t} c_{s,t}.\end{aligned}$$

Summarizing, we may – without loss of generality – require the dual variables to satisfy the relations

$$\min_{s,t} c_{s,t} \leq \lambda_s \leq \max_{s,t} c_{s,t}$$

and

$$\min_{s,t} c_{s,t} - \max_{s,t} c_{s,t} \leq \mu_t \leq 0,$$

recovering almost the initial fixing.

The costs for numerical accuracy of the dual variable within this approach, however, are two additional vectors in the optimization problem, which is not increasing the problem's dimensions, and not significantly increasing the problem's size.

As some numerical solvers always require lower and upper bounds for the variables, the modified problem (2.5) is just appropriate for those solvers.

2.2. The Kantorovich-Rubinstein Theorem

Consider the particular situation of a cost function c being induced by a distance d,

$$c(\omega, \omega') := d(\omega, \omega')$$

and assume – without loss of generality – that the supporting points of $\mathbb{P}$ and $\mathbb{Q}$ coincide (otherwise one may simply add missing points with mass zero). Due to the reverse triangle inequality $(d(\omega_s, \omega) - d(\omega_s, \omega') \leq d(\omega, \omega'))$ the map

$$\omega \mapsto d(\omega_s, \omega) - \lambda_s$$

is continuous with Lipschitz constant 1. The function

$$\lambda^d(\omega') := \inf_s d(\omega_s, \omega') - \lambda_s$$

thus is a Lipschitz-1-function as well, and particularly thus

$$\lambda^d(\omega') - \lambda^d(\omega) \leq d(\omega', \omega).$$

Whence,

$$-\lambda^d(\omega) \le \inf_{\omega'} d(\omega', \omega) - \lambda^d(\omega') \le -\lambda^d(\omega)$$

(choose $\omega' = \omega$ in the infimum), which in turn means that

$$\lambda^{dd}(\omega) = -\lambda^d(\omega) \tag{2.6}$$

for the function $\lambda^{dd}(\omega) := \inf_{\omega'} d(\omega', \omega) - \lambda^d(\omega')$.

Notice now, that $\lambda^d(\omega_t) = \lambda_t^c$ and $\lambda^{dd}(\omega_s) = \lambda_s^{cc}$ in the setting of (2.4), and λ_s^{cc} satisfies

$$\sup_{\omega_s \neq \omega_{s'}} \frac{|\lambda_s^{cc} - \lambda_{s'}^{cc}|}{d(\omega_s, \omega_{s'})} \le 1.$$

In this situation (c a distance) and using (2.6) the dual problem thus rewrites

$$\begin{array}{cl} \text{maximize} \\ \text{(in } \lambda) & \sum_s \mathbb{P}_s \lambda_s - \sum_s \mathbb{Q}_s \lambda_s \\ \text{subject to} & \lambda_s - \lambda_{s'} \le d(\omega_s, \omega_{s'}), \end{array}$$

or

$$\begin{array}{cl} \text{maximize} \\ \text{(in } \lambda) & \mathbb{E}_{\mathbb{P}}[\lambda] - \mathbb{E}_{\mathbb{Q}}[\lambda] \\ \text{subject to} & L(\lambda) \le 1, \end{array}$$

where $L(\lambda) := \sup_{s \neq s'} \frac{|\lambda_s - \lambda_{s'}|}{d(\omega_s, \omega_{s'})}$ is the Lipschitz constant.

Again it does not come as a big surprise that this statement holds in more generality as well, as the discrete measures are dense with respect to the Wasserstein distance. This is the content of the Kantorovich-Rubinstein theorem:

Theorem 2.3 (Kantorovich-Rubinstein Theorem)**.** *Let $X = Y$ be a Polish space and $c = d$ a lsc. metric. Then*

$$\inf_\pi \int_{X \times X} d(x, y)\, \pi[\mathrm{d}x, \mathrm{d}y] = \sup_\lambda \int_X \lambda(x)\, \mathbb{P}[\mathrm{d}x] - \int_X \lambda(x)\, \mathbb{Q}[\mathrm{d}x],$$

where

(i) π has the adjusted marginals $\mathbb{P}$ and $\mathbb{Q}$, and

(ii) $\lambda \in L^1(\mathrm{d}\,|\mathbb{P} - \mathbb{Q}|)$ with Lipschitz constant bounded by 1,

$$L(\lambda) := \sup \frac{\lambda(x) - \lambda(y)}{d(x, y)} \le 1.$$

Remark 2.4. To put it in different words: The pair $\left(\lambda^{dd}, -\lambda^d\right)$ is a *maximal Kantorovich potential* for the cost function induced by a distance.

2.3. Subdifferential, And Derivative Of The Norm

In order to investigate the problems in mind and to properly formalize the desired results we need to extend the notion of derivatives, which is usually accomplished by subgradients in the context of convex (or concave) functions – and this is the content of this section.

Definition 2.5. Given a normed vector space X with dual X^*, the sub-differential of a $\mathbb{R}$-valued function $f\colon X \to \mathbb{R}$ is the collection of all sub-gradients, that is to say

$$\partial f(x) = \{u^* \in X^* \colon f(z) - f(x) \geq u^*(z-x) \text{ for all } z \in X\}.$$

The most prominent convex function in the context of normed spaces is of course the norm itself, and we will investigate its derivative in particular: According to the famous theorem of Hahn-Banach there is, for every $x \in X$ in a normed space (Banach space) X a continuous, linear functional $\mathbf{HB}_x \in X^*$ in its dual such that

$$|\mathbf{HB}_x(h)| \leq \|h\| \text{ and } \mathbf{HB}_x(x) = \|x\| \tag{2.7}$$

for all $h \in X$ (that is to say the norm in the dual is one, $\|\mathbf{HB}_x\| = 1$, where the norm is the Lipschitz constant

$$\|\lambda\| := \sup_{h \neq 0} \frac{|\lambda(h)|}{\|h\|} = L(\lambda)$$

for the linear functional λ). This functional $\mathbf{HB}_x$ may be found on a constructive basis for finite dimensional Banach spaces by induction on the dimension of the space; in an infinite dimensional space the existence is based on the axiom of choice (original work of the theorem guaranteeing existence may be found in [5] and [22]).

The norm is a convex function and has a subdifferential, and $\mathbf{HB}_x$ is a potential derivative: The next statement clarifies this topic in a bit more detail:

Theorem 2.6. *Let $\mathbf{HB}_x$ be the Hahn-Banach functional in x, then the following holds true:*

(i) $\mathbf{HB}_x$ *is a subgradient of the norm at x, $\mathbf{HB}_x \in \partial \|.\|(x)$*[2]*.*

(ii) Suppose that $x \mapsto \mathbf{HB}_x$ is weak continuous (i.e. in continuous in the topology $\sigma(X^*, X)$) in a neighbourhood of x, then $\mathbf{HB}_x$ is the derivative of the norm in x, that is*

$$\lim_{t \to 0,\, t \in \mathbb{R}} \frac{\|x + th\| - \|x\|}{t} = \mathbf{HB}_x(h)$$

for any $h \in X$ – the sub-differential is a singleton;

(iii) For any $h \in X^*$

$$\mathbf{HB}_{\mathbf{HB}_x}(h) = h\left(\frac{x}{\|x\|}\right). \tag{2.8}$$

[2]In the sequel we will sometimes abuse notation and write $\mathbf{HB}_x$ for a functional, but for the entire subgradient $\partial \|.\|(x)$ as well.

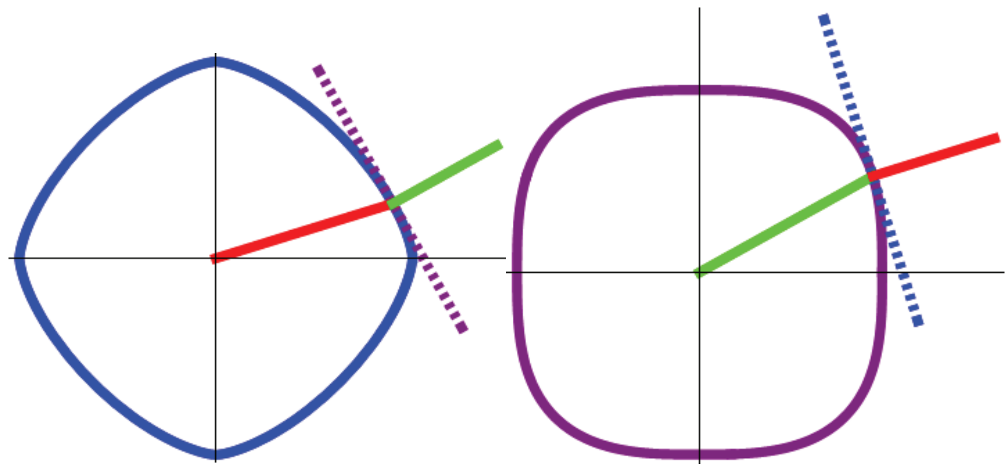

Figure 2.1.: Ball of the norm with radius $\|x\|$ and the sub-gradient $\mathbf{HB}_x$ (left); ball of the dual and $\mathbf{HB}_x$ with sub-gradient $\frac{x}{\|x\|}$ (right). The respective vectors are parallel.

Proof. Observe that

$$\begin{aligned} \|z\| &\geq \mathbf{HB}_x(z) \\ &= \mathbf{HB}_x(x) + \mathbf{HB}_x(z-x) \\ &= \|x\| + \mathbf{HB}_x(z-x) \end{aligned}$$

for any $z \in X$, which is the defining equation for an element (sub-gradient) to qualify for the sub-differential $\partial \|.\|(x)$ of the norm.

Let $t \in \mathbb{R}$, then

$$\begin{aligned} \mathbf{HB}_x(h) &= \frac{\mathbf{HB}_x(x+th) - \mathbf{HB}_x(x)}{t} \\ &\leq \frac{\|x+th\| - \|x\|}{t} \\ &\leq \frac{\mathbf{HB}_{x+th}(x+th) - \mathbf{HB}_{x+th}(x)}{t} \\ &= \mathbf{HB}_{x+th}(h) \\ &\xrightarrow[t\to 0]{} \mathbf{HB}_x(h), \end{aligned}$$

establishing the second assertion.

The latter statement is obvious as

$$\|x\| = \sup_{\|f^*\|\leq 1} f^*(x)$$

and equality is obtained for the particular choice $f^* = \mathbf{HB}_x$. □

Remark 2.7. In the space $\mathbb{R}^m$ we will adopt the usual notation and write $f^*(x) = f^\top x = \langle f, x\rangle$ similarly, depending on the given context. Given x, the Hahn-Banach statement gives an element ω_x such that $\langle \omega_x, h\rangle \le \|\omega_x\| \cdot \|h\|$ and $\langle \omega_x, x\rangle = \|\omega_x\| \cdot \|x\|$.

Example 2.8 (Derivative of the l^p-norm)**.** Consider the norm

$$\|x\|_p := \left(\sum w_t \cdot |x_t|^p\right)^{\frac{1}{p}}$$

on $X = \mathbb{R}^m$ for some positive weights $w_t > 0$. One verifies that

$$\mathbf{HB}_x = \left(w_t \cdot \left|\frac{x_t}{\|x\|_p}\right|^{p-1} \cdot \operatorname{sign} x_t\right)_t,$$

that is $\mathbf{HB}_x(h) = \sum_t w_t \left|\frac{x_t}{\|x\|_p}\right|^{p-1} \cdot \operatorname{sign}(x_t) \cdot h_t$.

For the all-one-vector $\mathbb{1}$ one particularly finds that

$$\mathbf{HB}_{\mathbb{1}} = \frac{w^\top}{\|\mathbb{1}\|_p^{p-1}} = \frac{w^\top}{\left(\sum_t w_t\right)^{1-\frac{1}{p}}} \in \partial \|.\| (\mathbb{1}),$$

which is the unique subgradient if $1 \le p < \infty$. This rewrites as

$$\frac{\mathbf{HB}_{\mathbb{1}}}{\|\mathbb{1}\|_p} = \frac{w^\top}{\sum_t w_t} = \frac{w^\top}{w^\top \mathbb{1}},$$

and this functional corresponds to weighting its arguments with weights w.

For the usual weights $w_t = 1$ the observation further simplifies to

$$\frac{\mathbf{HB}_{\mathbb{1}}}{\|\mathbb{1}\|_p} = \frac{\mathbb{1}^\top}{m}, \tag{2.9}$$

where m is the dimension of the space – this will turn out to be essential in a subsequent result.

Another noteworthy special case is $p = 2$: In this situation

$$\mathbf{HB}_x = \left(\frac{w_t x_t}{\|x\|_2}\right)_t,$$

which further reduces to

$$\mathbf{HB}_x = \frac{x^*}{\|x\|_2}$$

for the usual weights $w_t = 1$, which is the same as $\mathbf{HB}_x(h) = \frac{x^\top h}{\|x\|_2} = \sum \frac{x_t h_t}{\|x\|_2}$.

Remark 2.9. (The sub-differential is possibly empty in the predual). Recall that the dual of c_0 (the set of sequences $a\colon \mathbb{N} \to \mathbb{R}$ which tend to zero, $a_n \to 0$) is $c_0^* := l^1$ (the set of sequences summable in absolute values), whose dual in turn is $c_0^{**} = l^\infty$ (the set of bounded sequences).

Choose $x := \sum_{n\in\mathbb{N}} 2^{-n} e_n \in l^1$ and observe that $\mathbf{HB}_x := (\operatorname{sign} x_n)_{n\in\mathbb{N}}$ satisfies (2.7), that is $\mathbf{HB}_x(x) = \sum_n |x_n| = \|x\|_{l^1}$, $\mathbf{HB}_x(y) \le \sum_n |y_n| = \|y\|_{l^1}$ for any other vector $y \in l^1$ and $\mathbf{HB}_x \in l^\infty \backslash c_0$.

It is easy to observe that $\mathbf{HB}_x \in l^\infty \backslash c_0$, whenever the support $\{n\colon x_n \neq 0\}$ is not finite; moreover, no such vector exists such that $\mathbf{HB}_x \notin l^\infty \backslash c_0$, provided infinite support of x.

In this situation thus

(i) $\partial \|.\|(x) \neq \emptyset$, because ($\mathbf{HB}_x \in \partial \|.\|(x) \neq \emptyset$), but

(ii) $\partial \|.\|(x) \cap c_0 = \emptyset$ in the predual c_0 of l^1.

Similar examples can be given for the non-reflexive space $\mathbb{L}^1$, which is isometrically embedded in the strictly larger space of finite (but not σ-finite) measures $ba = (\mathbb{L}^\infty)^*$, although a bit more involving.

3. Different Representations Of The Acceptability Functional

In this section we shall elaborate the concept of acceptability functionals, which is a concept still in actual research. Acceptability functionals will allow quantifying risk and to base an investor's decision, whether or not a risk is acceptable.

3.1. Definitions

We recall the definition of the distribution function for consistency reasons:

Definition 3.1. (Cumulative distribution function)

(i) The function $G_Y(x) := \mathbb{P}[Y \leq x]$ is the *cumulative distribution function (cdf.)* of a random variable Y, and

(ii) $G_Y^{-1}(\alpha) := \inf\{x\colon G_Y(x) \geq \alpha\}$ is the *quantile function*, the generalized inverse of Y's cdf..

The function G_Y is continuous from the right, and G_Y^{-1} is continuous from the left. Some authors (cf. [53]) use the term *left α-quantile* for $G_Y^{-1}(\alpha)$.

Often we shall suppress the index Y accounting for the random variable if the random variable in consideration is obvious from the context.

The next, useful statement follows Pflug and Römisch [39], but rather dates back to [17].

Proposition 3.2. *The following assertions hold true:*

(i) For all $0 < \alpha < 1$ and $x \in \mathbb{R}$

$$G\left(G^{-1}(\alpha)\right) \geq \alpha \text{ and } G^{-1}(G(x)) \leq x.$$

(ii) Let U be uniform on the same probability space (that is $\mathbb{P}[U \in [a,b]] = b - a$ for all $0 \leq a \leq b \leq 1$) and independent from Y. Then the augmented random variable

$$F(Y,U) := (1-U)\cdot G_Y(Y-) + U \cdot G_Y(Y) \tag{3.1}$$

is uniformly distributed as well and

$$Y = G_Y^{-1}(F(Y,U)) \tag{3.2}$$

a.s.

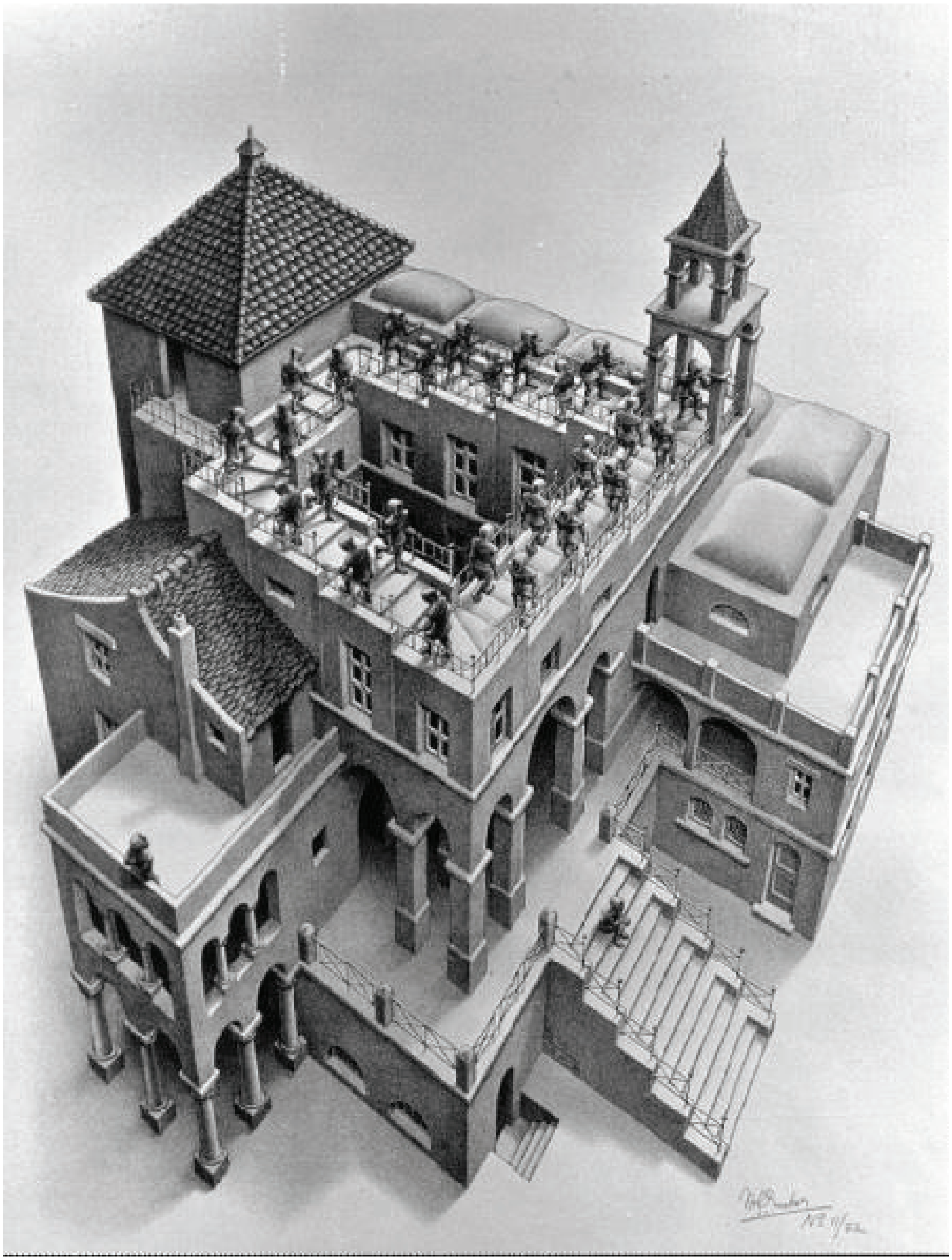

Figure 3.1.: M. C. Escher (1898 - 1972): Ascending and descending. Not monotone, non acceptable staircase.

Here,

$$G(x-) := \lim_{x' \to x,\, x' \leq x} G(x')$$

is the left-sided limit of G at x. In general

$$G(x-) \leq G(x),$$

with equality (exactly) in all points of continuity of G.

Remark 3.3. In the latter proposition it is sufficient to require U to be uniform on the (countable) set

$$\bigcup_{\{t \in \mathbb{R}\colon G(t-) < G(t)\}} G^{-1}(\{t\}),$$

which can possibly be achieved easier than independence.

The next definition of an acceptability functional basically follows Pflug and Römisch [39], in the notion of convex risk measures it can be found in Föllmer [18] as well. This notion generalizes the concept of coherent risk measures, which was introduced in Artzner et al., [3, 2].

Definition 3.4 (Acceptability Functional)**.** An *acceptability functional* is a $\mathbb{R} \cup \{-\infty\}$-valued mapping defined on a linear space $\mathcal{Y}$ of random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$ if the following defining properties hold true:

(i) Translation equivariance: $\mathcal{A}(Y + c) = \mathcal{A}(Y) + c$ for all random variables $Y \in \mathcal{Y}$ and constant random variables c $(c(\omega) \equiv c)$;

(ii) Concavity: $\mathcal{A}((1-\lambda) Y_0 + \lambda Y_1) \geq (1-\lambda) \mathcal{A}(Y_0) + \lambda \mathcal{A}(Y_1)$ for all random variables Y_0, Y_1 and $0 \leq \lambda \leq 1$;

(iii) Monotonicity: $\mathcal{A}(Y_1) \leq \mathcal{A}(Y_2)$ whenever $Y_1 \leq Y_2$ almost surely.

Remark 3.5. In the literature $\mathcal{A}$ is sometimes referred to as *monetary utility function*, and the monotonicity property is frequently referred to in stating $\mathcal{A}$ is *non-decreasing*.

Remark 3.6. It is sufficient to require *translation equivariance* for just one single $Y_0 \in \mathcal{Y}$, as the property *translation equivariance* then can be proved to hold for any $Y \in \mathcal{Y}$. This follows from a statement, which Prof. T. Rockafellar mentioned in a recent private communication he had with Prof. G. Pflug – cf. Proposition 10.3 in the Appendix.

Definition 3.7. A functional $\mathcal{A}$ on random variables is said to be *version independent* if it is defined for both, Z and $\tilde{Z}$, and additionally $\mathcal{A}(Z) = \mathcal{A}\left(\tilde{Z}\right)$, whenever Z and $\tilde{Z}$ have the same cumulative distribution function.

Remark 3.8. A term, which is frequently used likewise to express version independence, is *law invariance*.

An outstanding example of a version independent acceptability functional is the *average value at risk* at level α denoted $\mathbb{A}\mathrm{V@R}_\alpha$[1].

Definition 3.9 (Average value at risk)**.** Let Y denote a random variable.

(i) The average value at risk at level $\alpha > 0$ is defined via the random variable's distribution function G_Y (quantile function G_Y^{-1}) as

$$\mathbb{A}\mathrm{V@R}_\alpha(Y) := \frac{1}{\alpha}\int_0^\alpha G_Y^{-1}(u)\,\mathrm{d}u.$$

(ii) The average value at risk at level $\alpha = 0$ is

$$\mathbb{A}\mathrm{V@R}_0(Y) := \operatorname{ess\,inf}(Y).$$

This functional is obviously version independent. It is moreover just a special case of the *distortion acceptability functional* $\mathcal{A}_H$:

Definition 3.10 (Distortion Acceptability Functional)**.** The distortion acceptability functional is the Stieltjes integral of the form

$$\mathcal{A}_H(Y) = \int_0^1 G_Y^{-1}(u)\,\mathrm{d}H(u),$$

for some H which we assume to be bounded, right continuous and increasing on $[0,1]$.

Usually we shall assume in addition that H obeys the representation

$$H(u) = \int_0^u h(u')\,\mathrm{d}u'$$

for some non-increasing $h(.)$ (H thus is concave).

Remark 3.11. As indicated, the $\mathbb{A}\mathrm{V@R}_\alpha$ is a distortion acceptability functional as well, it is obtained by the particular choice $H_\alpha(u) := \min\left\{\frac{u}{\alpha}, 1\right\}$ and the respective density $h_\alpha(u') := \frac{1}{\alpha}\mathbb{1}_{[0,\alpha]}(u')$ (cf. Figure 3.2).

---

[1] The

- ▷ *Average Value at Risk*

is sometimes also called

- ▷ *conditional value at risk (for the additional representation* $\mathbb{A}V@R_\alpha(Y) = \mathbb{E}[Y : Y \leq V@R_\alpha(Y)]$*),*
- ▷ *expected shortfall,*
- ▷ *tail value-at-risk* or newly
- ▷ *super-quantile* (of course *sub-quantile* could be justified as well by simply changing the sign).
- ▷ Actuaries tend to use the term *Conditional Tail Expectation* (CTE).

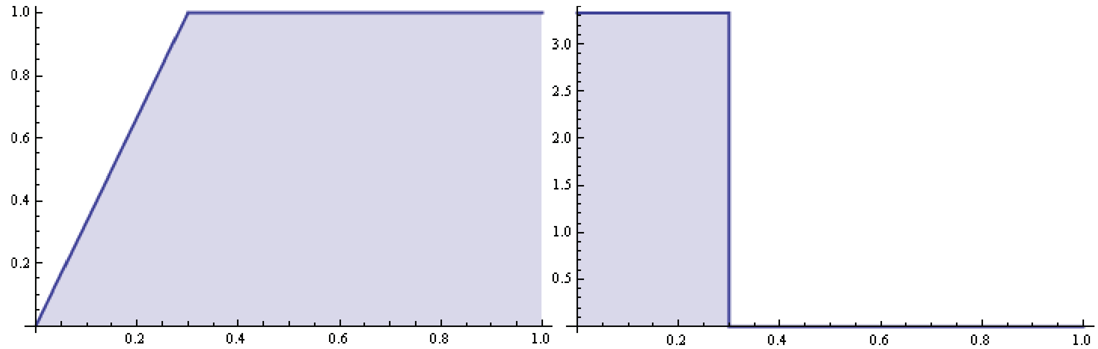

Figure 3.2.: Distortion function H and density h for the average value at risk at level $\alpha = 30\%$.

Definition 3.12 (Coupling and Fréchet bounds)**.** Let Y_1 and Y_2 denote random variables.

(i) Y_1 and Y_2 are coupled in a comonotone[2] way (with respect to $\mathbb{P}$) if

$$\mathbb{P}[Y_1 \leq y_1, Y_2 \leq y_2] = \min\{\mathbb{P}[Y_1 \leq y_1], \mathbb{P}[Y_1 \leq y_2]\};$$

(ii) Y_1 and Y_2 are coupled in an antimonotone[3] way (with respect to $\mathbb{P}$) if

$$\mathbb{P}[Y_1 \leq y_1, Y_2 \leq y_2] = \max\{0, \mathbb{P}[Y_1 \leq y_1] + \mathbb{P}[Y_1 \leq y_2] - 1\}.$$

The next lemma will be essential in further investigations, a similar statement is included in [25]; for copulas in general we refer to [35].[4]

Lemma 3.13 (Extension of Hoeffding's Lemma)**.** *Let X be a random variable and π have identical marginals, $\pi_1 = \pi_2 = \mathbb{P}$.*

(i) Assume that Y^c is coupled in a comonotone *way with X (with respect to $\mathbb{P}$). Then for any random variable Y which has the same distribution as Y^c*

$$\int X(x) \cdot Y(x)\, \mathbb{P}[\mathrm{d}x] \leq \int X(x) \cdot Y^c(x)\, \mathbb{P}[\mathrm{d}x] \tag{3.3}$$

and

$$\int X(x) \cdot Y(y)\, \pi[\mathrm{d}x, \mathrm{d}y] \leq \int X(x) \cdot Y^c(x)\, \mathbb{P}[\mathrm{d}x];$$

(ii) If Y^a is coupled in an antimonotone *way with X (with respect to $\mathbb{P}$), then for any random variable Y which has the same distribution as Y^a*

$$\int X(x) \cdot Y^a(x)\, \mathbb{P}[\mathrm{d}x] \leq \int X(x) \cdot Y(x)\, \mathbb{P}[\mathrm{d}x] \tag{3.4}$$

[2] Sometimes also called maximum copula.

[3] Sometimes also called minimum copula or anti-comonotone instead.

[4] For an actuarial application to combined annuities cf. [42].

and

$$\int X(x) \cdot Y^a(x)\, \mathbb{P}[\mathrm{d}x] \leq \int X(x) \cdot Y(y)\, \pi[\mathrm{d}x, \mathrm{d}y].$$

Moreover, for any Y there is a random variable Y^c (Y^a, resp.) on the same probability space, coupled in a comonotone (antimonotone, resp.) way with X, which has the same distribution as Y.

Proof. (For completeness we include the proof here – cf. [23])

Let Y be any other random variable and choose an *independent* copy X' (Y', resp.) of X (Y, resp.). Recall that

$$CoV(X, Y) = \frac{1}{2}\mathbb{E}_{\mathbb{P}}\left[(X - X')(Y - Y')\right]$$

and observe that

$$\int_{-\infty}^{\infty} \mathbb{1}_{\{X' \leq u\}}(\omega) - \mathbb{1}_{\{X \leq u\}}(\omega)\, \mathrm{d}u = X'(\omega) - X(\omega).$$

Thus

$$\begin{aligned} CoV(X, Y) =& \frac{1}{2}\int \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\mathbb{1}_{\{X \leq u\}} - \mathbb{1}_{\{X' \leq u\}}\right)\left(\mathbb{1}_{\{Y \leq v\}} - \mathbb{1}_{\{Y' \leq v\}}\right) \mathrm{d}u\mathrm{d}v\mathrm{d}\mathbb{P} \\ =& \frac{1}{2}\int \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_{\{X \leq u\}}\mathbb{1}_{\{Y \leq v\}} - \mathbb{1}_{\{X' \leq u\}}\mathbb{1}_{\{Y \leq v\}} \\ & - \mathbb{1}_{\{X \leq u\}}\mathbb{1}_{\{Y' \leq v\}} + \mathbb{1}_{\{X' \leq u\}}\mathbb{1}_{\{Y' \leq v\}} \mathrm{d}u\mathrm{d}v\mathrm{d}\mathbb{P} \\ =& \frac{1}{2}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2G(u, v) - 2G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v \\ =& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) - G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v, \end{aligned}$$

where $G(u, v) = \mathbb{P}[X \leq u, Y \leq v]$ is the bivariate distribution function of the pair (X, Y)[5]. Choosing the Fréchet bounds, that is

$$\max\{0,\, G(u) + G(v) - 1\} \leq G(u, v) \leq \min\{G(u), G(v)\},$$

establishes that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} & \max\{0,\, G(u) + G(v) - 1\} - G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v \\ & \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) - G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v \\ & \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min\{G(u),\, G(v)\} - G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v, \end{aligned}$$

[5]It is worth mentioning that neither $\int G(u, v)\, \mathrm{d}u\mathrm{d}v$ nor $\int G(u)\, G(v)\, \mathrm{d}u\mathrm{d}v$ converge on its own.

and thus

$$\mathbb{E}_{\mathbb{P}}[X \cdot Y] \leq \mathbb{E}_{\mathbb{P}}[X \cdot Y^c] \tag{3.5}$$

for the comonotone coupling, and

$$\mathbb{E}_{\mathbb{P}}[X \cdot Y^a] \leq \mathbb{E}_{\mathbb{P}}[X \cdot Y]$$

for the antimonotone coupling.

Consider now the extended space $\Omega \times \Omega$ and extend the random variables by $\tilde{X}(x, y) := X(x)$ and $\tilde{Y^c}(x, y) := Y^c(x)$. The cumulative distribution function stays unchanged, as

$$\pi\left[\tilde{X} \leq u\right] = \pi[\{X \leq u\} \times \Omega] = \mathbb{P}[X \leq u]$$

and

$$\pi\left[\tilde{Y^c} \leq u\right] = \pi[\{Y^c \leq u\} \times \Omega] = \mathbb{P}[Y^c \leq u].$$

Moreover $\left(\tilde{X}, \tilde{Y^c}\right)$ are coupled in a comonotone way with respect to π as well, because

$$\begin{aligned}
\pi\left[\tilde{X} \leq u, \tilde{Y^c} \leq v\right] &= \pi[\{X \leq u, Y^c \leq v\} \times \Omega] \\
&= \mathbb{P}[X \leq u, Y^c \leq v] \\
&= \min\{\mathbb{P}[X \leq u], \mathbb{P}[Y^c \leq v]\} \\
&= \min\left\{\pi\left[\tilde{X} \leq u\right], \pi\left[\tilde{Y^c} \leq v\right]\right\}
\end{aligned}$$

by definition of $\tilde{X}$ and $\tilde{Y^c}$.

Now let Y be any random variable on Ω with the same cdf. as Y^c and define the extension via $\tilde{Y}(x, y) := Y(y)$. Apply (3.5) with π replaced by $\mathbb{P}$ to obtain that

$$\begin{aligned}
\int\int X(x) Y(y) \pi[\mathrm{d}x, \mathrm{d}y] &= \mathbb{E}_\pi\left[\tilde{X} \cdot \tilde{Y}\right] \\
&\leq \mathbb{E}_\pi\left[\tilde{X} \cdot \tilde{Y^c}\right] \\
&= \int\int X(x) Y^c(x) \pi[\mathrm{d}x, \mathrm{d}y] \\
&= \int\int X(x) Y^c(x) \mathbb{P}[\mathrm{d}x],
\end{aligned}$$

which is the desired result.

The result for the antimonotone coupling is analogous.

As regards the existence of a coupling which is extremal for the inequalities in consideration just observe that

$$Y^c := G_Y^{-1}(F(X, U))$$

and

$$Y^a := G_Y^{-1}(1 - F(X, U))$$

do the required job (cf. (3.1)). □

Corollary 3.14 (Separation Of Variables)**.** *Let A be a version independent functional on random variables (all defined on the same space Ω) such that whenever $A(Y) > -\infty$ and $G_{\tilde{Y}} = G_Y$, then $A(Y) = A\left(\tilde{Y}\right)$. Then*

$$\sup_{A(Y)>-\infty} \int X(x) \cdot Y(y)\, \pi[\mathrm{d}x, \mathrm{d}y] + A(Y) \leq \sup_{A(Y)>-\infty} \int X(x) \cdot Y(x)\, \mathbb{P}[\mathrm{d}x] + A(Y)$$

and

$$\inf_{A(Y)>-\infty} \int X(x) \cdot Y(x)\, \mathbb{P}[\mathrm{d}x] + A(Y) \leq \inf_{A(Y)>-\infty} \int X(x) \cdot Y(y)\, \pi[\mathrm{d}x, \mathrm{d}y] + A(Y),$$

where π is any bi-variate measure with identical marginals, $\pi_1 = \pi_2 = \mathbb{P}$.

Proof. Let Y be any random variable with $A(Y) > -\infty$. Consider the random variable Y^c, which is coupled in a comonotone way with X. From Lemma 3.13 and version independence we deduce that

$$\sup_{\tilde{Y}\sim Y} \int X(x) \cdot \tilde{Y}(x)\, \mathbb{P}[\mathrm{d}x] + A\left(\tilde{Y}\right) = \int X(x) \cdot Y^c(x)\, \mathbb{P}[\mathrm{d}x] + A(Y^c),$$

where the supremum is over all random variables $\tilde{Y}$ which have the same distribution as Y – that is to say the supremum is attained by Y^c.

From the extension of Lemma 3.13 we deduce further that

$$\sup_{\tilde{Y}\sim Y} \int X(x) \cdot \tilde{Y}(y)\, \pi[\mathrm{d}x, \mathrm{d}y] + A\left(\tilde{Y}\right) \leq \int X(x) \cdot Y^c(x)\, \mathbb{P}[\mathrm{d}x] + A(Y^c).$$

Taking the supremum of all respective random variables Ys finally establishes the result.

The other inequality is analogous by choosing the corresponding antimonotone coupling. □

3.2. Rearrangement And Ordering

We will see in the sequel that the $\mathbb{AV@R}$-functional is very central in the theory of acceptability functionals and it appears from different angles. We will establish here a first, very useful and powerful representation for the $\mathbb{AV@R}$.

Definition 3.15. Let q be a bounded and measurable function on $[0,1]$, $q\colon [0,1] \to \mathbb{R}$. The *h-distorted quantile* of q is the function

$$\begin{aligned} q_h\colon\ & [0,1] \to \mathbb{R} \\ & x \mapsto \min \operatorname*{argmax}_{t\in\mathbb{R}}\ x \cdot t - \int_0^1 (t - q(u))^+ h(u)\,\mathrm{d}u, \end{aligned} \tag{3.6}$$

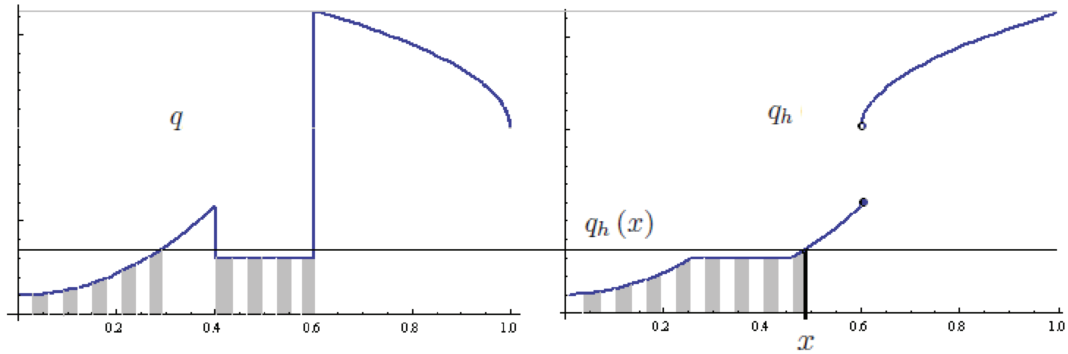

Figure 3.3.: Exemplary function q and its ordered rearrangement q_h.

where t – for consistency reasons – is restricted to $\operatorname{ess\,inf}_h q \le t \le \operatorname{ess\,sup}_h q$ (if those bounds exist)[6], [7].

Remark 3.16. There might be more t-s qualifying for the arg max in (3.6), however, the minimum ensures that q_h will be continuous from the left.

Remark 3.17. The term *distorted quantile* (reordering) q_h of the initial function q is motivated by equation (3.8) and the statements *(i)* and *(v)* in the next proposition. In Figure 3.3 we have depicted an exemplary function q and the related *rearrangement* q_h for the function $h(t) = 1$ to illustrate the relation.

Proposition 3.18. *Suppose h is positive and measurable. The h-distorted quantile q_h then obeys these following properties:*

(i) *q_h is non-decreasing and continuous from the left (lower semi-continuous);*

(ii) *$\operatorname{ess\,inf}_h q \le q_h \le \operatorname{ess\,sup}_h q$;*

(iii) *The supremum in (3.6) is attained at $t = q_h(x)$ satisfying*

$$x = \int_{\{q \le q_h(x)\}} h(u)\,\mathrm{d}u; \tag{3.7}$$

(iv) *The relations*

$$x \cdot q_h(x) = \int_0^x q_h(u)\,\mathrm{d}u + \int_0^1 (q_h(x) - q(u))^+ h(u)\,\mathrm{d}u$$

and the rearrangement property

$$\int_0^x q_h(u)\,\mathrm{d}u = \int_{\{q \le q_h(x)\}} q(u)\,h(u)\,\mathrm{d}u \tag{3.8}$$

hold true for all $x \in [0,1]$.

[6] The positive part is defined as $x^+ := \max\{x, 0\}$.

[7] Essential infimum (supremum, resp.) with respect to the measure $h\mathrm{d}\lambda$ having h as density with respect to the Lebesgue measure λ.

(v) If q is non-decreasing and $h(x) = 1$, then $q = q_h$ except on (countable many) points where q is not continuous.

Remark. $q\colon [0,1] \mapsto \mathbb{R}$ may be considered itself to be a random variable on $[0,1]$ or the quantile function of a random variable: Assume that h is a density[8] for the Lebesgue measure on $[0,1]$. From (3.7) the particular interpretation derives that $\mathbb{P}_h[q \leq t] := \int_{\{q\leq t\}} h(u)\,\mathrm{d}u = q_h^{-1}(t)$ is the cumulative distribution function of q, but measured with density h instead.

Proof. Observe first that $q_h(0) = \operatorname{ess\,inf}_h q$ and $q_h(1) = \operatorname{ess\,sup}_h q$ (if they exist).

Then, for x fixed, the function

$$F(t;x) := t \cdot x - \int_0^1 (t - q(u))^+ h(u)\,\mathrm{d}u \tag{3.9}$$

is continuous in t, so the supremum is attained at, say, $t^*(x) = q_h(x)$.

This function (3.9) is moreover concave in t and thus there is a subdifferential. For this argument we necessarily have that $x = \int_{\{q\leq t^*\}} h(u)\,\mathrm{d}u$ (derivative of (3.9) with respect to t, cf. [46]) – otherwise, there would be a strictly negative or positive slope in (3.6), and $t^*(x)$ would not be optimal. This shows that $x \mapsto t^*(x)$ is a non-decreasing function as well. Whence, q_h is non-decreasing and we have that

$$x = \int_{\{q\leq q_h(x)\}} h(u)\,\mathrm{d}u. \tag{3.10}$$

Replacing x by the (right continuous) inverse $x \leftarrow q_h^{-1}(t)$ this rewrites

$$q_h^{-1}(t) = \int_{\{q\leq t\}} h(u)\,\mathrm{d}u. \tag{3.11}$$

To evaluate the function (3.9) at its optimal point $q_h(x)$ we get

$$\begin{aligned}
x \cdot q_h(x) - \int_0^1 (q_h(x) - q(u))^+ h(u)\,\mathrm{d}u =& \\
=& x \cdot q_h(x) - \int_{\{q\leq q_h(x)\}} (q_h(x) - q(u))\, h(u)\,\mathrm{d}u \\
=& x \cdot q_h(x) - q_h(x) \cdot \int_{\{q\leq q_h(x)\}} h(u)\,\mathrm{d}u \\
&+ \int_{\{q\leq q_h(x)\}} q(u)\, h(u)\,\mathrm{d}u \\
=& x \cdot q_h(x) - q_h(x) \cdot x + \int_{\{q\leq q_h(x)\}} q(u)\, h(u)\,\mathrm{d}u \\
=& \int_{\{q\leq q_h(x)\}} q(u)\, h(u)\,\mathrm{d}u
\end{aligned}$$

[8]To guarantee the measure to be a probability measure it might be necessary to rescale by $\int h \mathrm{d}\lambda$

by virtue of (3.10).

An equivalent expression for what just was derived is[9]

$$\begin{aligned}\int_0^x q_h(u)\,\mathrm{d}u &= x\cdot q_h(x) - \int_0^{q_h(x)} q_h^{-1}(t)\,\mathrm{d}t\\ &= x\cdot q_h(x) - \int_0^{q_h(x)}\int_{\{q\le t\}} h(u)\,\mathrm{d}u\mathrm{d}t\end{aligned}$$

according to (3.11).[10] By Fubini's theorem thus

$$\int_0^x q_h(u)\,\mathrm{d}u = x\cdot q_h(x) - \int_0^1 (q_h(x)-q(u))^+ h(u)\,\mathrm{d}u,$$

which proves the statement. □

The next statement builds on a dual representation which was developed by Dana in [13], the dual has been established by Pflug in [37].

Corollary 3.19. *Let $\mathcal{A}_H(Y) = \int G^{-1}\mathrm{d}H$ be a distortion acceptability functional. Then*

$$\begin{aligned}\mathcal{A}_H(Y) &= \max_{t\in\mathbb{R}} t - \mathbb{E}\left[(t-Y)^+ h(G(Y)-U_Y)\right], \qquad (3.12)\\ &= \max_{t\in\mathbb{R},\,U \text{ uniform}} t - \mathbb{E}\left[(t-Y)^+ h(U)\right]\end{aligned}$$

where $U_Y = F(Y,U)$ is the augmented random variable[11] *as defined in* (3.2).

Proof. We know from Theorem 3.18 that

$$\max_t t\cdot x - \int_0^1 (t-q(u))^+ h(u)\,\mathrm{d}u = \int_{\{q\le q_h(x)\}} q(u)\,h(u)\,\mathrm{d}u,$$

[9] Young's inequality $\int_0^x f(u)\,\mathrm{d}u + \int_0^y f^{-1}(t)\,\mathrm{d}t \ge x\cdot y$ is an equality for $y=f(x)$ and increasing f.

[10] Assuming differentiability we could proceed straight forward and differentiate to get

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}\, x\cdot q_h(x) - \int_0^1 (q_h(x)-q(u))^+ h(u)\,\mathrm{d}u =\\ &= q_h(x) + x\cdot q_h'(x) - \int_{\{q\le q_h(x)\}} h(u)\,\mathrm{d}u\cdot q_h'(x)\\ &= q_h(x) + x\cdot q_h'(x) - x\cdot q_h'(x)\\ &= q_h(x).\end{aligned}$$

Integrating again establishes the remaining relations.

[11] Notice that $U_Y = 0$ for a probability space without atoms. For this reasons we will sometimes in the sequel consider non-atomic probability spaces solely, without stating this explicitly.

or

$$\max_t t - \frac{1}{x}\int_0^1 (t - q(u))^+ h(u)\,\mathrm{d}u =$$

$$= \frac{1}{x}\int_{\{q \le q_h(x)\}} q(u)\,h(u)\,\mathrm{d}u$$

$$= \frac{\int_{\{q \le q_h(x)\}} q(u)\,h(u)\,\mathrm{d}u}{\int_{\{q \le q_h(x)\}} h(u)\,\mathrm{d}u}.$$

Now choose $q := G_Y^{-1}$ and $x := 1$ to get

$$\max_t t - \int_0^1 \left(t - G_Y^{-1}(u)\right)^+ h(u)\,\mathrm{d}u = \int G_Y^{-1}(u)\,h(u)\,\mathrm{d}u,$$

which may be restated as

$$\int u h^G(u)\,\mathrm{d}G_Y(u) = \max_t t - \int_0^1 (t-u)^+ h^G(u)\,\mathrm{d}G_Y(u)$$

using the function $h^G(u) = \begin{cases} h(G(u)) & \text{if } G(u-) = G(u) \\ \frac{H(G(u)) - H(G(u-))}{G(u) - G(u-)} & \text{if } G(u-) < G(u). \end{cases}$

Whence,

$$\begin{aligned} \mathcal{A}_H(Y) &= \mathbb{E}\left[Y\, h^G(G(Y) - U_Y)\right] \\ &= \max_t t - \mathbb{E}\left[(t-Y)^+ h^G(G(Y) - U_Y)\right]. \end{aligned}$$

Recall, that $G(Y) - U_Y$ is uniformly distributed, and $h^G(G(Y) - U_Y)$ and Y are coupled in an antimonotone way, as h is decreasing. In view of (3.4) this establishes the result. □

As a corollary we have the following representation of the expected shortfall:

Corollary 3.20 (Concave representation of 𝔸V@R)**.** *The average value at risk obeys the representation*

$$\mathbb{A}V@R_\alpha(Y) = \max_t t - \frac{1}{\alpha}\mathbb{E}\left[(t-Y)^+\right], \tag{3.13}$$

and

$$V@R_\alpha(Y) \in \operatorname*{argmax}_t t - \frac{1}{\alpha}\mathbb{E}\left[(t-Y)^+\right].$$

Proof. Choose the function $h_\alpha(x) := \begin{cases} \frac{1}{\alpha} & \text{if } x \le \alpha \\ 0 & \text{if } x > \alpha \end{cases}$ and observe that the infimum in (3.6) is attained at $q_{h_\alpha}(1) = \operatorname{ess\,sup}_{h_\alpha} G_Y^{-1}$ for the function $q = G_Y^{-1}$, that is

$$\begin{aligned} \mathrm{V@R}_\alpha(Y) &= \operatorname*{ess\,sup}_{h_\alpha} G_Y^{-1} \\ &= q_{h_\alpha}(1). \end{aligned}$$

But in the situation $\{Y \leq q_{h_\alpha}(1)\}$ we have that $h^G(Y) = \frac{1}{\alpha}$, and thus (3.12) simplifies to

$$\mathbb{AV@R}_\alpha(Y) = \max_t t - \frac{1}{\alpha}\mathbb{E}\left[(t-Y)^+\right],$$

which completes the proof. □

For additional insight we offer this alternative proof:

Proof.

$$\begin{aligned}
\mathbb{AV@R}_\alpha(Y) &= \frac{1}{\alpha}\int_0^\alpha G^{-1}(u)\,\mathrm{d}u \\
&= \frac{1}{\alpha}\left(\int_0^\alpha G^{-1}(u)\,\mathrm{d}u + \int_\alpha^1 G^{-1}(\alpha)\,\mathrm{d}u - (1-\alpha)\,G^{-1}(\alpha)\right) \\
&= G^{-1}(\alpha) + \frac{1}{\alpha}\left(\int_0^1 \min\left\{G^{-1}(u),\, G^{-1}(\alpha)\right\}\mathrm{d}u - G^{-1}(\alpha)\right) \\
&= G^{-1}(\alpha) + \frac{1}{\alpha}\int_0^1 \min\left\{G^{-1}(u) - G^{-1}(\alpha),\, 0\right\}\mathrm{d}u \\
&= \frac{1}{\alpha}\left(G^{-1}(\alpha)\cdot\alpha - \int_0^1 \max\left\{G^{-1}(\alpha) - G^{-1}(u),\, 0\right\}\mathrm{d}u\right).
\end{aligned}$$

The function $\alpha \mapsto G^{-1}(\alpha)$ is non-decreasing, in view of (3.6) we thus find that

$$G^{-1}(\alpha) \in \operatorname*{argmax}_t\, t\cdot\alpha - \int_0^1 \left(t - G^{-1}(u)\right)^+ \mathrm{d}u.$$

Whence,

$$\begin{aligned}
\mathbb{AV@R}_\alpha(Y) &= \frac{1}{\alpha}\max_t\, t\cdot\alpha - \int_0^1 \left(t - G^{-1}(u)\right)^+ \mathrm{d}u \\
&= \max_t\, t - \frac{1}{\alpha}\mathbb{E}\left[(t-Y)^+\right].
\end{aligned}$$

□

3.3. A Rockafellar-and-Uryasev Type Representation

The representation (3.13) for the $\mathbb{AV@R}$ and its importance for linear programming was realized by Rockafellar and Uryasev in the original paper [48], as well in Pflug's [36]. We have elaborated the following generalization for distortions.

Theorem 3.21. *Let $\mathcal{A}_H\{G\} = \int_0^1 G^{-1}(p)\,\mathrm{d}H(p)$ be as above with h non-decreasing and the corresponding measure non-atomic. Then the representation*

$$\begin{aligned}\mathcal{A}_H(Y) &= \sup_{y\in\mathbb{R}} y - \mathbb{E}\left[(\mu_H(y) - \mu_H(Y))^+\right]\\ &= \sup_{y\in\mathbb{R}} y - \mathbb{E}\left[\int_{\min(y,Y)}^{y} \frac{H(G(u))}{G(u)}\mathrm{d}u\right]\end{aligned}$$

holds true, where $\mu_H(y) = \int_p^y \frac{H(G(u))}{G(u)}\mathrm{d}u$.

The supremum is attained at $\arg\min\{y:\ H(G(y)) = 1\}$.

Proof. For future reference recall the general identities to compute the expectation

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_0^1 f\left(G^{-1}(p)\right)\mathrm{d}p\\ &= \int f(x)\,\mathrm{d}G(x)\\ &= \int f(x)\,G'(x)\,\mathrm{d}x\\ &= f(a) + \int_a^\infty f'(x)\,(1-G(x))\,\mathrm{d}x \text{ if } \mathbb{P}[f(X)\geq f(a)] = 1\\ &= f(b) - \int_{-\infty}^b f'(x)\,G(x)\,\mathrm{d}x \text{ if } \mathbb{P}[f(X)\leq f(b)] = 1,\end{aligned} \tag{3.14}$$

where the latter two identities are applications of the product rule.

Define

$$\begin{aligned}f^*(p) &:= \int_0^{H^{-1}(p)} G^{-1}(x)\,h(x)\,\mathrm{d}x \qquad (3.15)\\ &= \int_{-\infty}^{G^{-1}\left(H^{-1}(p)\right)} u\cdot h(G(u))\;G'(u)\,\mathrm{d}u\\ &= p\cdot G^{-1}\left(H^{-1}(p)\right) - \int_{-\infty}^{G^{-1}\left(H^{-1}(p)\right)} H(G(u))\,\mathrm{d}u\end{aligned}$$

and $\mu(y) := \int_p^y \frac{H(G(u))}{G(u)}\mathrm{d}u$, thus $\mu'(y) = \frac{H(G(y))}{G(y)}$. As

$$\mathbb{P}[\min\{\mu(Y),\mu(p)\}\leq\mu(p)] = 1,$$

we may apply (3.14) to get

$$\begin{aligned}\mathbb{E}[\min\{\mu(Y),\mu(p)\}] &= \mu(p) - \int_{-\infty}^p \frac{H(G(y))}{G(y)}G(y)\,\mathrm{d}y\\ &= \mu(p) - \int_{-\infty}^p H(G(y))\,\mathrm{d}y.\end{aligned}$$

Whence,

$$\begin{aligned}
f^*(p) &= p\cdot G^{-1}\left(H^{-1}(p)\right)+\mathbb{E}\left[\min\left\{\mu(Y),\mu\left(G^{-1}\left(H^{-1}(p)\right)\right)\right\}\right]\\
&\quad -\mu\left(G^{-1}\left(H^{-1}(p)\right)\right)\\
&= p\cdot G^{-1}\left(H^{-1}(p)\right)+\mathbb{E}\left[\min\left\{\mu(Y)-\mu\left(G^{-1}\left(H^{-1}(p)\right)\right),0\right\}\right]\\
&= p\cdot G^{-1}\left(H^{-1}(p)\right)-\mathbb{E}\left[\max\left\{\mu\left(G^{-1}\left(H^{-1}(p)\right)\right)-\mu(Y),0\right\}\right].
\end{aligned}$$

Notice that f^* is convex (because both, G^{-1} and h are non-decreasing in (3.15)) and $f^{*\prime}(p)=G^{-1}\left(H^{-1}(p)\right)$; the function $f(y):=\sup_p y\cdot p-f^*(p)$ thus attains its maximum at

$$p^*=H(G(y)). \tag{3.16}$$

This means

$$\begin{aligned}
f(y) &= y\cdot H(G(y))-H(G(y))\cdot G^{-1}\left(H^{-1}(H(G(y)))\right)\\
&\quad +\mathbb{E}\left[\max\left\{\mu\left(G^{-1}\left(H^{-1}(H(G(y)))\right)\right)-\mu(Y),0\right\}\right]\\
&= \mathbb{E}\left[\max\left\{\mu(y)-\mu(Y),0\right\}\right]\\
&= \mu(y)+\mathbb{E}\left[\max\left\{-\mu(Y),-\mu(y)\right\}\right]\\
&= \mu(y)-\mathbb{E}\left[\min\left\{\mu(Y),\mu(y)\right\}\right].
\end{aligned}$$

From the elementary Fenchel-Moreau-Rockafellar theorem (cf. [47]) we get in addition that

$$f^*(p)=\sup_y y\cdot p-f(y).$$

The distortion acceptability function may be stated as

$$\begin{aligned}
\mathcal{A}_H(Y) &= \int_0^1 G_Y^{-1}(p)\,\mathrm{d}H(p)\\
&= \int_0^1 G_Y^{-1}(p)\,h(p)\,\mathrm{d}p\\
&= f^*(1),
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathcal{A}_H(G) &= \sup_y y-f(y)\\
&= \sup_y y-\mu(y)+\mathbb{E}\left[\min\left\{\mu(Y),\mu(y)\right\}\right]\\
&= \sup_y y+\mathbb{E}\left[\min\left\{\mu(Y)-\mu(y),0\right\}\right]\\
&= \sup_y y-\mathbb{E}\left[\max\left\{\mu(y)-\mu(Y),0\right\}\right]\\
&= \sup_y y-\mathbb{E}\left[(\mu(y)-\mu(Y))^+\right].
\end{aligned}$$

The supremum may be restricted to $H(G(y)) < 1$ due to continuity of both, H and G, and (3.16). □

The representation

$$\mathbb{AV@R}_\alpha\{G\} = \sup_y y - \frac{1}{\alpha}\mathbb{E}\left[(y-Y)^+\right] \tag{3.17}$$

though (cf. Rockafellar and Uryasev, [48]) is an immediate consequence of Theorem 3.21, as

$$\begin{aligned}\int_{\min(y,Y)}^{y} \frac{H(G(u))}{G(u)}\mathrm{d}u &= \frac{1}{\alpha}(y-\min\{y,Y\}) \\ &= \frac{1}{\alpha}(y+\max\{-y,-Y\}) \\ &= \frac{1}{\alpha}\max\{0,y-Y\} \\ &= \frac{1}{\alpha}(y-Y)^+\end{aligned}$$

whenever $y \leq G^{-1}(\alpha)$.

The representation (3.17) exposes the $\mathbb{AV@R}$ functional again, because this representation does *not* depend explicitly on G_Y, whereas in contrast to other functions H, μ_H *does* depend on G_Y.

Corollary 3.22. *Suppose $\mathcal{A}_H$ has the representation*

$$\mathcal{A}_H(Y) = \int_0^1 G_Y^{-1}(p)\,\mathrm{d}H(p)$$

for some non-decreasing H with $H(0)=0$ and $H(1)=1$, then

$$\mathbb{AV@R}_\alpha(Y) \leq \mathcal{A}_H(Y)$$

whenever $\alpha \leq \frac{1}{L(H)}$.

Proof. Note first, that $H \leq \bar{H} := -(-H)^{**}$ (for the bi-conjugate see section 10.3.2 on the Fenchel-transform) with equality in the points $\{0,1\}$ (equality holds always, if H is concave). Apply the previous theorem to $\bar{H}$ and observe that $\frac{\bar{H}(G(x))}{G(x)} =$

$\frac{\bar{H}(G(x))-\bar{H}(0)}{G(x)-0} \leq L\left(\bar{H}\right)$. Thus

$$\begin{aligned}
\sup_{y\in\mathbb{R}} y - \mathbb{E}\left[\int_{\min(y,Y)}^{y} \frac{H\left(G\left(u\right)\right)}{G\left(u\right)} \mathrm{d}u\right] \\
&\geq \sup_{y\in\mathbb{R}} y - \mathbb{E}\left[\int_{\min(y,Y)}^{y} L\left(\bar{H}\right) \mathrm{d}u\right] \\
&= \sup_{y\in\mathbb{R}} y - L\left(\bar{H}\right)\cdot \mathbb{E}\left[y - \min\left(y,Y\right)\right] \\
&= \sup_{y\in\mathbb{R}} y - \frac{1}{1/L\left(\bar{H}\right)}\cdot \mathbb{E}\left[\max\left(0, y-Y\right)\right] \\
&= \mathbb{AV@R}_{\frac{1}{L\left(\bar{H}\right)}}\left\{G\right\} \\
&\geq \mathbb{AV@R}_{\alpha}\left\{G\right\}
\end{aligned}$$

whenever $\alpha \leq \frac{1}{L\left(\bar{H}\right)}$; this, however, holds particularly true for $\alpha \leq \frac{1}{L(H)}$, because $L\left(H\right) \geq L\left(\bar{H}\right)$ and $\mathbb{AV@R}_{\alpha}$ is increasing in α. □

3.4. Dual Representation

In the discussion so far we had a strong focus on distortion acceptability functionals.

We shall investigate dual representations for general, concave acceptability functional in the sequel. This is a more general concept, which applies for distortions in particular, as we will elaborate.

To continue the investigations let us introduce some terms:

Definition 3.23 (Dual pairing and conceptually related terms)**.** Let $\mathcal{Y}$ and $\mathcal{Z}$ be locally convex vector spaces.

(i) A *dual pair* is a triple $(\mathcal{Y}, \mathcal{Z}, \langle .|.\rangle)$ consisting of two vector spaces $\mathcal{Y}$ and $\mathcal{Z}$ and a bi-linear form
$$\langle .|.\rangle : \mathcal{Y}\times\mathcal{Z}\to\mathbb{R};$$

(ii) A function $h:\mathcal{Y}\to\mathbb{R}\cup\{-\infty,\infty\}$ is called *proper* if
- ▹ its domain $\mathrm{dom}\left(h\right) := \{h<+\infty\}$ is not empty and
- ▹ $\{h=-\infty\}$ is empty;

(iii) A $\mathbb{R}\cup\{+\infty\}$-valued ($\mathbb{R}\cup\{-\infty\}$-valued, respectively) function h is said to be *lower semi-continuous* (lsc.) (*upper semi-continuous*, usc., resp.) provided that all sets
$$\{h>t\} \qquad (\{h<t\}\,,\ \text{resp.})$$
are open for any $t\in\mathbb{R}$[12];

[12]Equivalently, iff $h\left(y_0\right) \leq \liminf_{y\to y_0} h\left(y\right)$ ($\limsup_{y\to y_0} h\left(y\right) \leq h\left(y_0\right)$, resp.) whenever $y\to y_0$.

(iv) The largest lsc. function which is less than f is denoted lsc$f$ (the smallest usc. function which is bigger than f is usc$f$, resp.);

(v) The *conjugate* h^* of h is the (convex and lsc.) function

$$h^*(z) := \sup_{y\in\mathcal{Y}} \langle y|z\rangle - h(y)$$

with range $\mathbb{R}\cup\{+\infty\}$;

(vi) The *bi-conjugate* h^{**} of h is the (convex and lsc.) function

$$h^{**}(y) := \sup_{z\in\mathcal{Z}} \langle y|z\rangle - h^*(z),$$

again with range $\mathbb{R}\cup\{+\infty\}$.

Theorem 3.24 (Fenchel-Moreau-Rockafellar)**.** *Let f be proper and convex, then*

$$f^{**} = \operatorname{lsc} f.$$

(See, e.g., Rockafellar [47, Theorem 5] or Aubin and Ekeland [4, Theorem 4.4.2] for a proof.)

Example 3.25. A Banach space $\mathcal{Y} := X$, together with its dual $\mathcal{Z} := X^*$ and the natural bi-linear form

$$\langle y|z^*\rangle := z^*(y)$$

is a dual pair.

Given this setting one computes the conjugate of the norm, which is the lsc.-function

$$\begin{aligned} \|.\|^*(z^*) &= \sup_{x\in X} \langle x|z^*\rangle - \|x\| \\ &= \begin{cases} 0 & \text{if } \|z^*\|\le 1 \\ +\infty & \text{if } \|z^*\| > 1. \end{cases} \end{aligned}$$

As the norm is continuous and thus lsc. the Fenchel-Moreau-Rockafellar Theorem 3.24 now states that

$$\|y\| = \sup_{\|z^*\|\le 1} z^*(y) = s_{\{z^*:\, \|z^*\|\le 1\}}(y),$$

where $s_B(y) := \sup_{z^*\in B} z^*(y)$ is often called *support function of the set B*.

Example 3.26. For a sample space Ω with a probability measure $\mathbb{P}$ the spaces $\mathbb{L}^p(\Omega,\mathbb{P})$ and $\mathbb{L}^{p'}(\Omega,\mathbb{P})$ ($\frac{1}{p}+\frac{1}{p'}=1$ and $1\le p<\infty$) again form a natural *dual pair*, when employing the natural bi-linear form

$$\langle Z|Y\rangle := \mathbb{E}[Y\cdot Z].$$

Recall that an acceptability functional is concave, due to the Fenchel-Moreau Theorem 3.24 it may be represented via its dual. In a general situation we will find that the bi-dual dominates the original functional $\mathcal{A}$, and they coincide for upper lower semi-continuous (usc.) functionals. The general form is

$$\mathcal{A}(Y) = \inf\left\{\mathbb{E}[Y \cdot Z] - A(Z) : A(Z) > -\infty\right\}.$$

Definition 3.27 (Dual Representation)**.** We shall use the following terms:

(i) A representation of form

$$\mathcal{A}(Y) = \inf\left\{\mathbb{E}[Y \cdot Z] - A(Z) : Z \in \mathcal{Z}\right\}$$

is called a *dual representation* of $\mathcal{A}$.

(ii) For a functional A on random variables the set

$$\mathcal{Z}_A := \{Z \in \mathcal{Z} : A(Z) > -\infty\}$$

is called *supergradient set* of A, or *risk envelope*.

(iii) A is said to be *version independent*, if for any Z and $\tilde{Z}$ having the same distribution then $A(Z) = A\left(\tilde{Z}\right)$ and either both or none is in $\mathcal{Z}_{\mathcal{A}}$.

Remark. As explicitly outlined in [24] A is version independent iff $\mathcal{A}$ is version independent.

3.5. Derivative Of The Acceptability Functional

The dual representations imposed for the acceptability functional allow to characterize the derivative.

Theorem 3.28. *Let $\mathcal{A}$ be a lower semi-continuous (lsc.) and concave acceptability functional on a normed, linear space with range $\mathbb{R}$. Then $\partial\mathcal{A}(Y)$ is not empty*[13].

Proof. Consider the hypo-graph of $\mathcal{A}$,

$$hyp(\mathcal{A}) = \{(Y', \alpha') : \alpha' < \mathcal{A}(Y')\},$$

which obeys these following properties:

(i) $hyp(\mathcal{A})$ is convex, because $\mathcal{A}$ is concave;

(ii) $hyp(\mathcal{A})$ is open, because

$$hyp(\mathcal{A}) = \bigcup_{t\in\mathbb{R}} \{\mathcal{A} > t\} \times (-\infty, t)$$

and $\mathcal{A}$ is lsc.

[13] Note that $\mathcal{A}$ is finitely valued, its range is $\bar{\mathbb{R}}\setminus\{-\infty\}$.

For fixed Y there is – due to the separating hyperplane theorem (a geometric version of the Hahn-Banach theorem) – a closed hyperplane separating the point $(Y, \mathcal{A}(Y))$ from the open set $hyp(\mathcal{A})$, that is there is a pair (u^*, v^*) with the property

$$u^*(Y - Y') + v^* \cdot (\mathcal{A}(Y) - \alpha') > 0$$

for all $(Y', \alpha') \in hyp(\mathcal{A})$: v^* is a number, and u^* is continuous, as the hyperplane is closed.

Notice that $v^* > 0$, because

$$u^*(Y - Y) + v^* \cdot \underbrace{(\mathcal{A}(Y) - \alpha')}_{>0} > 0$$

for the particular choice $(Y, \alpha') \in hyp(\mathcal{A})$. Whence,

$$\tilde{u}^*(Y' - Y) + \mathcal{A}(Y) - \alpha' > 0$$

for $\tilde{u}^* := -\frac{1}{v^*}u^*$ and $\alpha' < \mathcal{A}(Y')$.

Let α' tend to $\mathcal{A}(Y')$, $\alpha' \to \mathcal{A}(Y')$ to establish that

$$\tilde{u}^*(Y' - Y) + \mathcal{A}(Y) - \mathcal{A}(Y') \geq 0:$$

this rewrites as

$$\mathcal{A}(Y') \leq \mathcal{A}(Y) + \tilde{u}^*(Y' - Y),$$

which is the defining equation for $\tilde{u}^*$ to qualify for the subdifferential $\partial\mathcal{A}(Y)$. The subdifferential therefore is non-empty. □

Ruszczyński and Shapiro have derived the following more general result for the special case of Banach-lattices in [52, Proposition 3.1]; see as well [53, page 264 and Theorem 7.79] for a more detailed elaboration of the proof. Their result is based on the classical *Baire Category Theorem* and the *Klee-Nachbin-Namioka* Theorem. The latter states that positive linear functionals are continuous (cf. Borwein's[14] paper *Automatic continuity and openness of convex relations in* [8] for an additional reference).

For an extension to multivariate risk mappings it is rewarding, indeed, to consult Ruszczyński and Shapiro in [51, 50].

Theorem 3.29. *Suppose that $\mathcal{A} : \mathcal{Y} \to \bar{\mathbb{R}}$ is a proper acceptability functional on the Banach-lattice $\mathcal{Y}$. Then $\mathcal{A}$ is continuous and subdifferentiable on the interior of its domain.*

[14] Jonathan M. Borwein, together with his brother Peter Borwein, is rather well-known for some compelling algorithms to compute π and their contribution to the Riemann zeta function $\zeta(s) = \sum_{n=1} \frac{1}{n^s}$. They were rewarded the Chauvent and Hasse prizes during their collaboration at Dalhousie University in Halifax (Canada).

It is made evident now that the subgradient of a concave lsc. functional $\mathcal{A}$ is not empty. But what about the sub-differential of A? The set $\partial A(z)$ is obviously non-empty neither, whenever we may identify X with X^{**}, that is for reflexive spaces.

This may not hold true, however, for other spaces: To this end recall Remark 2.9: given $0 \neq z^* \in l^1$ with non-finite support, $\mathbf{HB}_{z^*} \notin c_0$, and we further will not find any vector $\tilde{x} \in c_0$ such that

$$\begin{aligned}\|z^*\| &= \sup_{x \in c_0} z^*(x) - \|.\|^*(x) \\ &= \sup_{\|x\| \le 1} z^*(x) \\ &= z^*(\tilde{x}),\end{aligned}$$

the supremum is *not* attained and – in view of the next theorem – the subdifferential in the pre-dual is empty.

It is the content of the next theorem to further characterize the sub-differential for reflexive spaces ($1 < r < \infty$), although the statement is valid for $r = 1$ as well (but not for $r = \infty$ in general):

Theorem 3.30. *Let $\mathcal{A}$ be usc. with values in $\mathbb{R}$ and dual to A on $\mathbb{L}^r$ for some $1 \le r < \infty$ and $\mathcal{Z}_{\mathcal{A}} \subseteq \mathbb{L}^{r'}$ ($\frac{1}{r} + \frac{1}{r'} = 1$), then*

$$\partial \mathcal{A}(Y) = \operatorname{argmin}\left\{\mathbb{E}[Y \cdot Z] - A(Z) : Z \in \mathcal{Z}_{\mathcal{A}}\right\}.$$

The linear mapping $\mathcal{A}'(Y)(h) := \mathbb{E}[h \cdot Z_Y]$ is a subdifferential of $\mathcal{A}$ if $Z_Y \in \partial \mathcal{A}(Y)$, and $\mathcal{A}'(Y) \in \partial \mathcal{A}(Y)$ is a subgradient of the acceptability function $\mathcal{A}$.

Proof. Consider $Z_Y \in \operatorname{argmin}\left\{\mathbb{E}[Y \cdot Z] - A(Z) : Z \in \mathcal{Z}_{\mathcal{A}}\right\}$. By linearity of $\mathcal{A}'(Y)$

$$\begin{aligned}\mathcal{A}(Y) + \mathcal{A}'(Y)\left(\tilde{Y} - Y\right) &= \mathbb{E}[Y \cdot Z_Y] - A(Z_Y) + \mathbb{E}\left[\left(\tilde{Y} - Y\right) \cdot Z_Y\right] \\ &= \mathbb{E}\left[\tilde{Y} \cdot Z_Y\right] - A(Z_Y).\end{aligned}$$

As $\mathcal{A}\left(\tilde{Y}\right)$ is the infimum over all such Z thus

$$\mathcal{A}(Y) + \mathcal{A}'(Y)\left(\tilde{Y} - Y\right) \ge \mathcal{A}\left(\tilde{Y}\right),$$

that is

$$\mathcal{A}'(Y) \in \partial \mathcal{A}(Y)$$

for the concave function $\mathcal{A}$.

Conversely, let $\tilde{Z}_Y \in \partial \mathcal{A}(Y)$, that is $\tilde{Z}_Y \in (\mathbb{L}^r)^* = \mathbb{L}^{r'}$ for $1 \le r < \infty$, and hence there is a function $Z_Y \in \mathbb{L}^{r'}$ such that

$$\mathcal{A}(Y) + \mathbb{E}\left[Z_Y\left(\tilde{Y} - Y\right)\right] \ge \mathcal{A}\left(\tilde{Y}\right)$$

for all $\tilde{Y} \in \mathbb{L}^r$. Equivalently,

$$\mathbb{E}\left[Z_Y Y\right] - \mathcal{A}\left(Y\right) \leq \inf_{\tilde{Y}} \mathbb{E}\left[Z_Y \tilde{Y}\right] - \mathcal{A}\left(\tilde{Y}\right) = A\left(Z_Y\right),$$

and thus

$$\mathcal{A}\left(Y\right) \geq \mathbb{E}\left[Z_Y \cdot Y\right] - A\left(Z_Y\right),$$

which in turn means that $Z_Y \in \operatorname{argmin}\left\{\mathbb{E}\left[Y \cdot Z\right] - A\left(Z\right) :\ Z \in \mathcal{Z}_{\mathcal{A}}\right\}$. □

A proof by different means is contained in Ruszczyński and Shapiro [52, page 437].

Theorem 3.31. *Given a usc. functional $\mathcal{A}$ with dual A on a reflexive space, then*

$$Y \in \partial A\left(\partial \mathcal{A}\left(Y\right)\right) \text{ and } Z \in \partial \mathcal{A}\left(\partial A\left(Z\right)\right).$$

Proof. This is immediate, as $\mathbb{E}\left[Y \cdot Z\right] = \mathcal{A}\left(Y\right) + A\left(Z\right)$. □

3.6. Kusuoka's Representation

The $\mathbb{AV@R}$-functionals are extreme points among all distortion acceptability functionals, this is another remarkable fact exposing $\mathbb{AV@R}$. The corresponding Choquet representation is

$$\begin{aligned} \mathcal{A}_H\left(Y\right) &= \int_0^1 G_Y^{-1}\left(u\right) \mathrm{d}H\left(u\right) \\ &= \int_0^1 \mathbb{AV@R}_\alpha\left(Y\right) \mathrm{d}M\left(\alpha\right), \end{aligned}$$

where M is monotonically increasing, satisfying $M\left(0\right) = 0$ and $M\left(1\right) = H\left(1\right)$.

Kusuoka (cf. [30]) was able to reveal a general representation in the form

$$\mathcal{A}\left(Y\right) = \inf\left\{\int_0^1 \mathbb{AV@R}_\alpha\left(Y\right) \mathrm{d}m_G\left(\alpha\right) :\ G \in \mathcal{G}\right\},$$

where $\{m_G : G \in \mathcal{G}\}$ is a family of measures (not necessarily probability measures). His results were brought later in a more general context in [24] and the supplementary note [55].

We shall state the theorem as it is contained in [24] without proof. However, Kusuoka already obtained the essential ingredient, which is the transition from the dual representation to Kusuoka's presentation – and vice versa. We will demonstrate how this transition works in general, the rest then are non-trivial technicalities, to which we take our hat off.

The Kusuoka representation links the dual representation to a representation of a convex-combination involving the $\mathbb{AV@R}_\alpha$ at all levels α by using the fact that $\alpha \mapsto \mathbb{AV@R}_\alpha\left(Y\right)$ is an increasing function, as is made obvious by (3.17).

Theorem 3.32 (Kusuoka's Representation)**.** *Suppose that* $(\Omega, \Sigma, \mathbb{P})$ *is a standard probability space, and let* $\mathcal{A} : \mathbb{L}^\infty(\Omega, \Sigma, \mathbb{P}) \to \mathbb{R}$ *be a version independent acceptability functional. The following then are equivalent:*

(i) There is a version independent, usc. and concave functional $A : \mathbb{L}^1(\Omega, \Sigma, \mathbb{P}) \to \mathbb{R}$ *such that* $A(Y) > -\infty$ *whenever* $Y \geq 0$ *and* $\mathbb{E}[Y] = 1$[15]*, and*

$$\mathcal{A}(Y) = \inf_{Z \in \mathbb{L}^1} \{\mathbb{E}[Y \cdot Z] - A(Z)\} \text{ for } Y \in \mathbb{L}^\infty;$$

(ii) There is a convex function $v : \mathcal{P}([0,1]) \to [0, \infty]$ *such that*

$$\mathcal{A}(Y) = \inf_{m \in \mathcal{P}([0,1])} \left\{ \int_0^1 \mathbb{A}V@R_\alpha(Y) \, \mathrm{d}m(\alpha) + v(m) \right\} \text{ for } Y \in \mathbb{L}^\infty.$$

Sketch of the Proof. To establish the relation between the first and second assertion consider the set $D_\searrow$ of non-increasing, right-continuous, $\mathbb{R}$-valued functions on $]0,1]$ such that $f(1) = 0$ and $\|f\|_1 = \int_0^1 f(x)\,\mathrm{d}x = 1$. Define the map

$$\begin{aligned} T : D_\searrow(]0,1]) &\to \mathcal{M}(]0,1]) \\ f &\mapsto m_f \end{aligned}$$

where the measure m_f is defined by $\mathrm{d}m_f(x) = -x\mathrm{d}f(x)$. Using integration by parts one verifies that

$$\begin{aligned} \|m_f\|_1 &= m_f(]0,1]) \\ &= \int \mathrm{d}m_f \\ &= -\int x\mathrm{d}f(x) \\ &= x \cdot f(x)\big|_0^1 + \int_0^1 f(x)\,\mathrm{d}x \\ &= 1; \end{aligned}$$

notice in addition that $T(h_\alpha) = \delta_\alpha$ on $]0,1]$, where $h_\alpha(x) = \frac{1}{\alpha}\mathbb{1}_{]0,\alpha[}(x)$, moving the $\mathbb{A}$V@R a bit closer.

The relation then is established by

$$A(Y) := \begin{cases} -v\left(T\left(-G_{-Y}^{-1}\right)\right) & \text{if } Y \geq 0 \text{ and } \mathbb{E}[Y] = 1, \\ -\infty & \text{else.} \end{cases}$$

Conversely let $v(m) := \mathcal{V}(T^{-1}(m))$, where $\mathcal{V}\left(-G_{-Y}^{-1}\right) := -A(Y)$.

This is a bijection, allowing to pass from A to v and backwards – but this is not more than a motivation for the proof of Kusuoka's representation. The proof further incorporates antimonotone couplings, as elaborated in Lemma 3.13, in an essential way. □

[15] i.e. A is well-defined on densities.

4. Continuity

This section is dedicated to the investigation of continuity properties of the acceptability function.

In an initial stage usual continuity properties of the acceptability functional are investigated.

Further, as the acceptability functional is defined on a probability space, the question how the underlying probability measure influences the returning result of the acceptability functional is of interest as well; the other chapter will investigate this question then in detail.

4.1. Continuity Of The Acceptability Functional

Monotonicity and translation equivariance in the definition of the acceptability functional are already strong regularizing properties, as the following immediate lemma reveals.

Lemma 4.1. *Suppose that $\mathcal{Y} = \mathbb{L}^p(\Omega, \Sigma, \mathbb{P})$ and there is one almost surely bounded random variable $\tilde{Y} \in \mathbb{L}^\infty(\Omega, \Sigma, \mathbb{P})$ such that $\mathcal{A}\left(\tilde{Y}\right) > -\infty$. Then $\mathcal{A}(Y) > -\infty$ for any $Y \in \mathbb{L}^\infty(\Omega, \Sigma, \mathbb{P})$ and $\mathcal{A}$ is Lipschitz-1 on this subspace, that is*

$$|\mathcal{A}(Y_1) - \mathcal{A}(Y_2)| \le \|Y_1 - Y_2\|_\infty .$$

Proof. Observe that $\tilde{Y} \le \tilde{c}$ for some $\tilde{c} < \infty$. Whence $-\infty < \mathcal{A}\left(\tilde{Y}\right) \le \mathcal{A}(\tilde{c})$ by monotonicity, thus further $-\infty < \mathcal{A}(c)$ for any real number c by translation equivariance.

Next choose $c \le Y$ a.s., therefore $-\infty < \mathcal{A}(c) \le \mathcal{A}(Y)$, so $\mathcal{A}$ is $\mathbb{R}$-valued for any $Y \in \mathbb{L}^\infty$.

To observe continuity choose c_1 and c_2 such that $c_1 \le Y_1 - Y_2 \le c_2$ a.s.. From monotonicity and translation equivariance then follows that

$$\begin{aligned} \mathcal{A}(Y_1) &= \mathcal{A}(Y_2 + Y_1 - Y_2) \\ &\le \mathcal{A}(Y_2 + c_2) \\ &= \mathcal{A}(Y_2) + c_2, \end{aligned}$$

and

$$\begin{aligned} \mathcal{A}(Y_2) &= \mathcal{A}(Y_1 + Y_2 - Y_1) \\ &\le \mathcal{A}(Y_1 - c_1) \\ &= \mathcal{A}(Y_1) - c_1, \end{aligned}$$

combined thus

$$c_1 \le \mathcal{A}(Y_1) - \mathcal{A}(Y_2) \le c_2.$$

Changing the role of Y_1 and Y_2 reveals the assertion. □

Remark 4.2. Notice that the latter lemma does *not* imply that $\mathcal{A}$ is continuous with respect to the norm $\|.\|_p$ for $p < \infty$: It is correct that $\|.\|_{p_1} \le \|.\|_{p_2}$ whenever $p_1 \le p_2$, but the identity is not a continuous embedding.

Not as immediate as the latter Lemma, but an immediate corollary to Theorem 3.29 is the following

Corollary 4.3. *Let $\mathcal{A} : \mathbb{L}^p \to \mathbb{R}$ ($1 \le p < \infty$) be a real-valued acceptability functional. Then $\mathcal{A}$ is continuous and subdifferentiable on the entire $\mathbb{L}^p$.*

This justifies the next stability statement:

Theorem 4.4. *Let $\mathcal{A}$ be an acceptability functional on $\mathbb{L}^p(\Omega, \Sigma, \mathbb{P})$, $1 \le p < \infty$. Then*

$$\left|\mathcal{A}(Y) - \mathcal{A}\left(\tilde{Y}\right)\right| \le \left\|Y - \tilde{Y}\right\|_p \cdot \max\left\{\inf_{Z\in\partial\mathcal{A}(Y)} \|Z\|_{p'}, \inf_{Z\in\partial\mathcal{A}(\tilde{Y})} \|Z\|_{p'}\right\};$$

in particular

$$\left|\mathcal{A}(Y) - \mathcal{A}\left(\tilde{Y}\right)\right| \le \left\|Y - \tilde{Y}\right\|_p \cdot \sup_{A(Z)>-\infty} \|Z\|_{p'},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$.

Proof. Let $Z \in \partial\mathcal{A}(Y)$ be chosen, that is

$$\mathcal{A}(Y) \le \mathcal{A}\left(\tilde{Y}\right) + \mathbb{E}\left[Z \cdot \left(Y - \tilde{Y}\right)\right].$$

By Hölder's inequality thus

$$\begin{aligned}\mathcal{A}(Y) - \mathcal{A}\left(\tilde{Y}\right) &\le \mathbb{E}\left[Z \cdot \left(Y - \tilde{Y}\right)\right] \\ &\le \|Z\|_p \cdot \left\|Y - \tilde{Y}\right\|_{p'}.\end{aligned}$$

Interchanging the role of Y and $\tilde{Y}$ gives the assertion. □

4.2. Continuity With Respect To Changing The Measure

Theorem 4.5 (The average value at risk is continuous with respect to the Wasserstein-metric - I.)**.** *Let Y be a $\mathbb{R}$-valued random variable which is Hölder continuous* *with constant $L_\beta(Y)$ on some Polish space. Then $\mathbb{A}V@R$ is Hölder-continuous with constant $\frac{L_\beta(Y)}{\alpha^{\frac{\beta}{r}}}$, that is*

$$\left|\mathbb{A}V@R_{\alpha,\mathbb{P}}(Y)-\mathbb{A}V@R_{\alpha,\mathbb{Q}}(Y)\right|\leq\frac{L_\beta(Y)}{\alpha^{\frac{\beta}{r}}}\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta$$

for any $1\leq r<\infty$.

Proof. Recall that

$$\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{P}}(Y)=\max\left\{a-\frac{1}{\alpha}\mathbb{E}_\mathbb{P}\left[(a-Y)^+\right]:\ a\in\mathbb{R}\right\}\tag{4.1}$$

and the maximum is attained at $a^*\in\mathrm{V@R}_\alpha(Y)$ and so satisfies $\mathbb{P}[Y\leq a^*]\leq\alpha$[1]. Then choose a measure π minimizing $\mathsf{d}_r(\mathbb{P},\mathbb{Q})$ and a^* maximizing (4.1). With this choice

$$\begin{aligned}
&\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{P}}(Y)-\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{Q}}(Y)\\
&\qquad=a^*-\frac{1}{\alpha}\mathbb{E}_\mathbb{P}\left[(a^*-Y)^+\right]-\max_a a-\frac{1}{\alpha}\mathbb{E}_\mathbb{Q}\left[(a-Y)^+\right]\\
&\qquad\leq a^*-\frac{1}{\alpha}\mathbb{E}_\mathbb{P}\left[(a^*-Y)^+\right]-a^*+\frac{1}{\alpha}\mathbb{E}_\mathbb{Q}\left[(a^*-Y)^+\right]\\
&\qquad=\frac{1}{\alpha}\mathbb{E}_\mathbb{Q}\left[(a^*-Y)^+\right]-\frac{1}{\alpha}\mathbb{E}_\mathbb{P}\left[(a^*-Y)^+\right]\\
&\qquad=\frac{1}{\alpha}\int\left((a^*-Y(\omega_1))^+-(a^*-Y(\omega_2))^+\right)\cdot\\
&\qquad\qquad\cdot 1_{\{(\omega_1,\omega_2):\,Y(\omega_1)\leq a^*,\,Y(\omega_2)\leq a^*\}}\pi\left[\mathrm{d}\omega_1,\mathrm{d}\omega_2\right].
\end{aligned}$$

Observe now that $x\mapsto(a^*-x)^+$ is Lipschitz-continuous with Lipschitz-constant 1, so $(a^*-x)^+-(a^*-y)^+\leq|x-y|$. Thus,

$$\begin{aligned}
&\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{P}}(Y)-\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{Q}}(Y)\\
&\qquad\leq\frac{1}{\alpha}\int\left|Y(\omega_1)-Y(\omega_2)\right|\cdot\\
&\qquad\qquad\cdot 1_{\{(\omega_1,\omega_2):\,Y(\omega_1)\leq a^*,\,Y(\omega_2)\leq a^*\}}\pi\left[\mathrm{d}\omega_1,\mathrm{d}\omega_2\right]\\
&\qquad\leq\frac{1}{\alpha}\int L_\beta(Y)\cdot d(\omega_1,\omega_2)^\beta\cdot 1_{\{(\omega_1,\omega_2):\,Y(\omega_1)\leq a^*,\,Y(\omega_2)\leq a^*\}}\pi\left[\mathrm{d}\omega_1,\mathrm{d}\omega_2\right].
\end{aligned}$$

[1]For atomic probability spaces the augmented random variable $F(Y,U)$ has to be involved.

Now one may apply Hölder's inequality with $\frac{1}{\frac{r}{\beta}}+\frac{1}{\frac{r}{r-\beta}}=1$ to get

$$\begin{aligned}
&\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{P}}(Y)-\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{Q}}(Y)\\
&\quad\leq\frac{L_\beta(Y)}{\alpha}\cdot\left(\int d(\omega_1,\omega_2)^r\,\pi[\mathrm{d}\omega_1,\mathrm{d}\omega_2]\right)^{\frac{\beta}{r}}\cdot\\
&\qquad\cdot\left(\int \mathbb{1}_{\{(\omega_1,\omega_2):\,Y(\omega_1)\leq a^*,\,Y(\omega_2)\leq a^*\}}^{\frac{r}{r-\beta}}\pi[\mathrm{d}\omega_1,\mathrm{d}\omega_2]\right)^{\frac{r-\beta}{r}}\\
&\quad\leq\frac{L_\beta(Y)}{\alpha}\cdot\left(\int d(\omega_1,\omega_2)^r\,\pi[\mathrm{d}\omega_1,\mathrm{d}\omega_2]\right)^{\frac{\beta}{r}}\cdot\\
&\qquad\cdot\left(\int \mathbb{1}_{\{(\omega_1,\omega_2):\,Y(\omega_1)\leq a^*\}}\pi[\mathrm{d}\omega_1,\mathrm{d}\omega_2]\right)^{\frac{r-\beta}{r}}\\
&\quad=\frac{L_\beta(Y)}{\alpha}\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta\cdot\mathbb{P}[Y\leq a^*]^{\frac{r-\beta}{r}}\\
&\quad\leq\frac{L_\beta(Y)}{\alpha}\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta\cdot\alpha^{1-\frac{\beta}{r}}\\
&\quad=\frac{L_\beta(Y)}{\alpha^{\frac{\beta}{r}}}\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta.
\end{aligned}$$

Reversing the role of $\mathbb{P}$ and $\mathbb{Q}$ finally establishes the result. □

Corollary 4.6 (Continuity of the acceptability functional with respect to the Wasserstein distance – II)**.** *Suppose $\mathcal{A}$ is an acceptability functional with Kusuoka-representation*

$$\mathcal{A}_{\mathbb{P}}(Y)=\inf_{G\in\mathcal{G}}\int_0^1\mathbb{A}V@R_{\alpha,\mathbb{P}}(Y)\,\mathrm{d}m^G(\alpha),$$

where $\mathcal{G}$ is a set of positive measures.

Then $\mathcal{A}$ is continuous with respect to the Wasserstein distance, provided that

$$K:=\sup_{G\in\mathcal{G}}\int_0^1\frac{1}{\alpha^{\frac{\beta}{r}}}\mathrm{d}m^G(\alpha)<\infty$$

is finite:

$$|\mathcal{A}_{\mathbb{P}}(Y)-\mathcal{A}_{\mathbb{Q}}(Y)|\leq\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta\cdot L_\beta(Y)\cdot\sup_{G\in\mathcal{G}}\int_0^1\frac{1}{\alpha^{\frac{\beta}{r}}}\mathrm{d}m^G(\alpha).\tag{4.2}$$

Proof. The proof is straight forward: choose $G_\varepsilon\in\mathcal{G}$ such that

$$\int_0^1\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{Q}}(Y)\,\mathrm{d}m^{G_\varepsilon}(\alpha)<\mathcal{A}_{\mathbb{Q}}(Y)+\varepsilon.$$

Then

$$\begin{aligned}\mathcal{A}_{\mathbb{P}}(Y)-\mathcal{A}_{\mathbb{Q}}(Y)=&\\ &\leq\int_0^1 \mathbb{AV@R}_{\alpha,\mathbb{P}}(Y)\,\mathrm{d}m^{G_\varepsilon}(\alpha)-\int_0^1 \mathbb{AV@R}_{\alpha,\mathbb{Q}}(Y)\,\mathrm{d}m^{G_\varepsilon}(\alpha)+\varepsilon\\ &\leq\int_0^1 \mathbb{AV@R}_{\alpha,\mathbb{P}}(Y)-\mathbb{AV@R}_{\alpha,\mathbb{Q}}(Y)\,\mathrm{d}m^{G_\varepsilon}(\alpha)+\varepsilon\\ &\leq L_\beta(Y)\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta\cdot\int\frac{1}{\alpha^{\frac{\beta}{r}}}\mathrm{d}m^{G_\varepsilon}(\alpha)+\varepsilon\\ &\leq K\cdot L_\beta(Y)\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta+\varepsilon.\end{aligned}$$

Let $\varepsilon\to 0$ and interchange the role of $\mathbb{P}$ and $\mathbb{Q}$ to observe the desired assertion. □

Remark 4.7. This next example demonstrates that it is necessary to somehow bound α away from 0, because

$$\mathbb{P}\mapsto\mathbb{AV@R}_{0,\mathbb{P}}$$

is *not* continuous in general: Define the measure $\mathbb{P}_n:=\left(1-\frac{1}{n^{r+1}}\right)\delta_0+\frac{1}{n^{r+1}}\delta_{-n}$ and observe that

$$\begin{aligned}\mathsf{d}_r(\delta_0,\ \mathbb{P}_n) &= \left(\frac{1}{n^{r+1}}n^r\right)^{\frac{1}{r}}\\ &= n^{-\frac{1}{r}}\to 0\end{aligned}$$

but, however, for the simple random variable $Y(\omega)=\omega$,

$$\mathbb{AV@R}_{0;\delta_0}(Y)-\mathbb{AV@R}_{0;\mathbb{P}_n}(Y)=0+n=n.$$

$\mathbb{P}\to\mathbb{AV@R}_{0,\mathbb{P}}$ thus is far from being continuous with respect to d_r for $1\leq r<\infty$.

The next theorem elaborates the geometry of a general acceptability functional in sufficient detail.

Theorem 4.8 (The acceptability functional is continuous with respect to the Wasserstein-metric – III)**.** *Let $\mathcal{A}$ be version independent, and let Y be a $\mathbb{R}$-valued random variable which is Hölder continuous, that is $|Y(\omega)-Y(\omega')|\leq L_\beta(Y)\cdot d(\omega,\omega')^\beta$ for some $\beta\leq 1$. Then $\mathcal{A}$ is Hölder-continuous as well with respect to changing the measure, precisely*

$$\mathcal{A}_{\mathbb{Q}}(Y)-\mathcal{A}_{\mathbb{P}}(Y)\leq L_\beta(Y)\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})^\beta\cdot\inf_{Z\in\partial\mathcal{A}_{\mathbb{P}}(Y)}\|Z\|_{r'_\beta} \tag{4.3}$$

whenever $1\leq r<\infty$ and $r'_\beta\geq\frac{r}{r-\beta}$.[2]

In particular

$$|\mathcal{A}_{\mathbb{P}}(Y)-\mathcal{A}_{\mathbb{Q}}(Y)|\leq L_\beta(Y)\cdot\mathsf{d}_r(\mathbb{P},\mathbb{Q})\cdot\sup_{A(Z)>-\infty}\|Z\|_{r'_\beta}.$$

[2] r'_β satisfies $\frac{\beta}{r}+\frac{1}{r'_\beta}\leq 1$ and is just the usual Hölder-conjugate exponent for $\beta=1$.

Remark 4.9. It should be noticed that the norm $\|Z\|_q$ is version independent, as

$$\begin{aligned} \|Z\|_q^q &= \mathbb{E}\left[|Z|^q\right] \\ &= q \cdot \int_0^\infty t^{q-1} \mathbb{P}\left[|Z| \geq t\right] \mathrm{d}t \\ &= q \cdot \int_0^\infty t^{q-1} \left(1 - \mathbb{P}\left[Z \leq t\right] + \mathbb{P}\left[Z \leq -t\right]\right) \mathrm{d}t \end{aligned}$$

by integration by parts. Due to the assumption that $\mathcal{A}$ is version independent it follows that

$$\sup_{A(Z) > -\infty} \|Z\|_{p_\beta}$$

is symmetric in $\mathbb{P}$ and $\mathbb{Q}$, and for convenience we thus may suppress the index in $A_{\mathbb{Q}}(Z) > -\infty$.

Before we prove the latter theorem we want to mention the following essential corollary:

Corollary 4.10 (The acceptability functional is continuous with respect to the Wasserstein-metric – IV)**.** *Let $\mathcal{A}$ be version independent on a Polish space, Y a $\mathbb{R}$-valued random variable which is Hölder continuous, that is $|Y(\omega) - Y(\omega')| \leq L_\beta(Y) \cdot d(\omega, \omega')^\beta$ for some $\beta \leq 1$. Then the map*

$$\mathbb{P} \mapsto \mathcal{A}_{\mathbb{P}}(Y)$$

is τ_{d_r}-continuous provided that $\sup_{A(Z) > -\infty} \|Z\|_{r'_\beta} < \infty$.

Proof of Theorem 4.8. Recall the dual representation

$$\mathcal{A}_{\mathbb{P}}(Y) = \inf\left\{\mathbb{E}_{\mathbb{P}}\left[Y \cdot Z\right] - A_{\mathbb{P}}(Z) : \ Z \text{ in } \mathcal{Z}_{\mathcal{A}}\right\}$$

and recall that there is an optimal $Z_Y \in \partial \mathcal{A}_{\mathbb{P}}(Y)$ such that

$$\mathbb{E}_{\mathbb{P}}\left[Y \cdot Z_Y\right] - A_{\mathbb{P}}(Z_Y) \geq \mathcal{A}_{\mathbb{P}}(Y).$$

For any other dual variable $Z \in \mathcal{Z}_{\mathcal{A}_{\mathbb{Q}}}$ thus

$$\begin{aligned} \mathcal{A}_{\mathbb{Q}}(Y) - &\mathcal{A}_{\mathbb{P}}(Y) \\ &\leq \mathbb{E}_{\mathbb{Q}}\left[Y \cdot Z\right] - A_{\mathbb{Q}}(Z) - \mathbb{E}_{\mathbb{P}}\left[Y \cdot Z_Y\right] + A_{\mathbb{P}}(Z_Y) \\ &= \int Y(\omega_2) \cdot Z(\omega_2)\, \mathbb{Q}\left[\mathrm{d}\omega_2\right] - \int Y(\omega_1)\, Z_Y(\omega_1)\, \mathbb{P}\left[\mathrm{d}\omega_1\right] - A_{\mathbb{Q}}(Z) + A_{\mathbb{P}}(Z_Y) \\ &= \inf_\pi \int Y(\omega_2) \cdot Z(\omega_2) - Y(\omega_1)\, Z_Y(\omega_1)\, \pi\left[\mathrm{d}\omega_1, \mathrm{d}\omega_2\right] - A_{\pi_2}(Z) + A_{\pi_1}(Z_Y), \end{aligned}$$

where the infimum is over all measure π with marginals $\pi_1 = \mathbb{P}$ and $\pi_2 = \mathbb{Q}$.

Taking the infimum over all random variables $Z \in \mathcal{Z}_{\mathcal{A}_{\mathbb{Q}}}$ one obtains

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) - &\mathcal{A}_{\mathbb{P}}(Y) \\ &\leq \inf_{Z \in \mathcal{Z}_{\mathcal{A}}} \inf_{\pi} \int Y(\omega_2) \cdot Z(\omega_2) - Y(\omega_1) Z_Y(\omega_1)\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2] \\ &\qquad A_{\pi_2}(Z) + A_{\pi_1}(Z_Y).\end{aligned}$$

The Corollary 3.14 to Hoeffding's Lemma (cf. Lemma 3.13) allows to separate the second variable $Z(\omega_2)$. We may rephrase the latter inequality accordingly as

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) - &\mathcal{A}_{\mathbb{P}}(Y) \\ &\leq \inf_{\tilde{\pi}} \inf_{Z \in \mathcal{Z}_{\mathcal{A}}} \int Y(\omega_2) \cdot Z(\omega_3) - Y(\omega_1) Z_Y(\omega_1)\, \tilde{\pi}[\mathrm{d}\omega_1, \mathrm{d}\omega_2, \mathrm{d}\omega_3] + \\ &\quad - A_{\tilde{\pi}_2}(Z) + A_{\tilde{\pi}_1}(Z_Y),\end{aligned}$$

where $\tilde{\pi}$ has the additional, third marginal $\tilde{\pi}_3 = \mathbb{Q}$, together with the other marginals $\tilde{\pi}_1 = \mathbb{P}$ and $\tilde{\pi}_2 = \mathbb{Q}$.

Let $G_Y(z) := \mathbb{P}[Z_Y \leq z]$ and $G(z) := \mathbb{Q}[Z \leq z]$ be the respective cumulative distribution functions of Z_Y and Z. Let U be independent of Z and define the random variable

$$F(Z, U) := (1 - U) \cdot G(Z-) + U \cdot G(Z),$$

which is uniformly distributed, and moreover $G^{-1}(F(Z, U)) = Z$ a.s. by Proposition 3.2.

Further define the $\mathbb{R}$-valued random variable

$$Z' := G_Y^{-1}(F(Z, U)),$$

as above $G_Y^{-1}(p) := \inf\{u \colon G_Y(u) \geq p\}$ ($G^{-1}(p) := \inf\{u \colon G(u) \geq p\}$, resp.) and recall that $G_Y\left(G_Y^{-1}(p)\right) \geq p$ and $G_Y^{-1}(G_Y(z)) \leq z$. Thus

$$\begin{aligned}\mathbb{Q}[Z' \leq z] &= \mathbb{Q}\left[G_Y^{-1}(F(Z, U)) \leq z\right] \\ &= \mathbb{Q}[F(Z, U) \leq G_Y(z)] \\ &= G_Y(z) \\ &= \mathbb{P}[Z_Y \leq z].\end{aligned}$$

It follows that Z_Y and Z' have the same distribution given their respective measures, that is

$$\mathbb{Q}[Z' \leq z] = \mathbb{P}[Z_Y \leq z].$$

From this observation and due to the assumption of version-independence one deduces first that

$$A_{\tilde{\pi}_2}(Z') = A_{\tilde{\pi}_1}(Z_Y),$$

and further, by restricting to random variables of the particular form Z' and by choosing the appropriate coupled measure (we link the third marginal ($\mathrm{d}y$) with the first marginal ($\mathrm{d}z$)), that

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) - &\mathcal{A}_{\mathbb{P}}(Y)\\ &\le \inf_{\tilde{\pi}} \inf_{Z} \int Y(\omega_2)\cdot Z'(\omega_3) - Y(\omega_1)\, Z_Y(\omega_1)\, \tilde{\pi}[\mathrm{d}\omega_1, \mathrm{d}\omega_2, \mathrm{d}\omega_3]\\ &\le \inf_{\tilde{\pi}} \int Y(\omega_2)\cdot Z_Y(\omega_1) - Y(\omega_1)\, Z_Y(\omega_1)\, \tilde{\pi}[\mathrm{d}\omega_1, \mathrm{d}\omega_2, \mathrm{d}\omega_3]\\ &= \inf_{\pi} \int Y(\omega_2)\cdot Z_Y(\omega_1) - Y(\omega_1)\, Z_Y(\omega_1)\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2].\end{aligned}$$

Now note that $\frac{1}{\frac{r}{\beta}} + \frac{1}{\frac{r}{r-\beta}} = 1$, we thus may apply Hölder's inequality to get

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) &- \mathcal{A}_{\mathbb{P}}(Y)\\ &\le \inf_{\pi} \int (Y(\omega_2) - Y(\omega_1))\cdot Z_Y(\omega_1)\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2]\\ &\le \inf_{\pi} \left(\int |Y(\omega_2) - Y(\omega_1)|^{\frac{r}{\beta}}\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2]\right)^{\frac{\beta}{r}} \left(\int |Z_Y(\omega_1)|^{\frac{r}{r-\beta}}\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2]\right)^{\frac{r-\beta}{r}},\end{aligned}$$

the infimum being computed over all measures with marginals $\pi_1 = \mathbb{P}$ and $\pi_2 = \mathbb{Q}$.

Utilizing Hölder's continuity and taking the infimum over all measures with the respective marginals finally establishes that

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) &- \mathcal{A}_{\mathbb{P}}(Y)\\ &\le \inf_{\pi} L_\beta(Y)\cdot \left(\int d(\omega_2, \omega_1)^r\, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2]\right)^{\frac{\beta}{r}} \left(\int |Z_Y|^{\frac{r}{r-\beta}}\, \mathrm{d}\mathbb{P}\right)^{\frac{r-\beta}{r}}\\ &\le L_\beta(Y)\cdot \mathsf{d}_r(\mathbb{P}, \mathbb{Q})^\beta \cdot \|Z_Y\|_{r'_\beta}\end{aligned}$$

because $\|Z\|_{r'_\beta} \ge \|Z\|_{\frac{r-\beta}{r}}$ as $r'_\beta \ge \frac{r}{r-\beta}$.

As we may accept any random variables $Z_Y \in \partial\mathcal{A}(Y)$ one thus obtains

$$\begin{aligned}\mathcal{A}_{\mathbb{Q}}(Y) &- \mathcal{A}_{\mathbb{P}}(Y)\\ &\le L_\beta(Y)\cdot \mathsf{d}_r(\mathbb{P}, \mathbb{Q})^\beta \cdot \inf_{Z\in\partial\mathcal{A}(Y)} \|Z\|_{r'_\beta}.\end{aligned}$$

Interchanging the role of $\mathbb{P}$ and $\mathbb{Q}$ finally establish the other result, that is

$$|\mathcal{A}_{\mathbb{Q}}(Y) - \mathcal{A}_{\mathbb{P}}(Y)| \le L_\beta(Y)\cdot \mathsf{d}_r(\mathbb{P}, \mathbb{Q})^\beta \cdot \sup_{A(Z)>-\infty} \|Z\|_{r'_\beta}$$

holds in particular. □

Using the latter result the continuity of the $\mathbb{A}$V@R functional follows by a straightforward computation:

Corollary 4.11 (The average value at risk is continuous with respect to the Wasserstein-metric)**.** *Let Y be a $\mathbb{R}$-valued random variable which is Hölder continuous for some $\beta \le 1$. Then $\mathbb{A}V@R_\alpha$ is Hölder-continuous with constant $\frac{L_\beta(Y)}{\alpha^{\frac{\beta}{r}}}$ with respect to changing the measure, that is*

$$\left|\mathbb{A}V@R_{\alpha,\mathbb{P}}\left(Y\right) - \mathbb{A}V@R_{\alpha,\mathbb{Q}}\left(Y\right)\right| \le \frac{L_\beta\left(Y\right)}{\alpha^{\frac{\beta}{r}}} \cdot \mathsf{d}_r\left(\mathbb{P},\mathbb{Q}\right)$$

for any $1 \le r < \infty$.

Proof. The proof consists of evaluating $\|h\|_{r'_\beta}$ for the function h_α, because

$$\begin{aligned}
\|Z\|_{r'_\beta} &= \|h_\alpha\left(U\right)\|_{\frac{r}{r-\beta}} \\
&= \left(\int_0^\alpha \frac{1}{\alpha^{\frac{r}{r-\beta}}} \mathrm{d}x\right)^{\frac{r-\beta}{r}} \\
&= \left(\alpha^{1-\frac{r}{r-\beta}}\right)^{\frac{r-\beta}{r}} \\
&= \alpha^{\frac{r-\beta}{r}-1} \\
&= \alpha^{-\frac{\beta}{r}}.
\end{aligned}$$

□

Remark 4.12*.* Notice that we get in particular

$$\left|\mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{P}}\left(Y\right) - \mathbb{A}\mathrm{V@R}_{\alpha,\mathbb{Q}}\left(Y\right)\right| \le \frac{L\left(Y\right)}{\alpha} \cdot \mathsf{d}_{KA}\left(\mathbb{P},\mathbb{Q}\right)$$

for the $\mathbb{A}$V@R and the Kantorovich metric ($r = 1$; cf. Pflug and Wozabal, [40]).

5. Modulus Of Continuity

In order to better understand the behaviour of the acceptability functional we shall further determine the driving constants in (4.3), which are

$$\inf_{Z\in\partial\mathcal{A}(Y)} \|Z\|_{r'_\beta}\,, \tag{5.1}$$

$$\sup_{A(Z)>-\infty} \|Z\|_{r'_\beta}\,. \tag{5.2}$$

and (4.2). For some exemplary, but typical and relevant acceptability functionals we will compute these numbers in the sequel. Moreover, we will discuss cases in which the bounds found are sharp.

5.1. Distortion Risk Functional

Recall that $\mathcal{A}_H$ is called a *distortion acceptability functional*, provided it can be stated as a Stieltjes integral of the form

$$\mathcal{A}_H\left\{G\right\} = \int_0^1 G^{-1}\left(u\right)\mathrm{d}H\left(u\right)$$

for some H which we assume to be bounded, right continuous and increasing on $[0,1]$.

Usually we shall assume that H has the representation $H\left(u\right) = \int_0^u h\left(u'\right)\mathrm{d}u'$ for some non-increasing $h\left(u\right)$ (H thus is convex), and the $\mathbb{AV@R}_\alpha$ is obtained by the particular choice $H_\alpha\left(g\right) := \min\left\{\frac{g}{\alpha}, 1\right\}$ and its respective density $h_\alpha\left(u\right) := \frac{1}{\alpha}1_{[0,\alpha]}\left(u\right)$.

The dual representation of the distortion acceptability functional[1] is

$$\mathcal{A}_H\left(Y\right) = \inf\left\{\mathbb{E}\left[Y\cdot h\left(U\right)\right] : U \text{ uniform in } [0,1]\right\}.$$

It is straight forward that for $Z = h\left(U\right)$,

$$\begin{aligned}\|Z\|_{r'} &= \|h\left(U\right)\|_{r'}\\ &= \left(\int_0^1 h\left(u\right)^{r'}\mathrm{d}u\right)^{\frac{1}{r'}}\\ &= \|h\|_{r'}\end{aligned}$$

[1] $U:\Omega\to\mathbb{R}$ is uniform iff $\mathbb{P}\left[U\in[a,b]\right] = b-a$.

is a constant for any uniform U, and thus $\sup_{A(Z)>-\infty}\|Z\|_{r'}=\|h\|_{r'}$.

As the $\mathbb{A}$V@R is a distortion acceptability functional as well for the particular function $h(u)=\frac{1}{\alpha}\mathbb{1}_{[0,\alpha]}(u)$, thus

$$\begin{aligned}\|h\|_{r'} &= \left(\int_0^\alpha\left(\frac{1}{\alpha}\right)^{r'}\mathrm{d}u\right)^{\frac{1}{r'}}\\ &= \left(\alpha^{1-r'}\right)^{\frac{1}{r'}}\\ &= \frac{1}{\alpha^{\frac{1}{r}}},\end{aligned}$$

as was already elaborated.

5.2. The $\mathbb{A}$V@R At Level 0

The average value at risk at level 0 is by definition

$$\mathbb{A}\mathrm{V@R}_0(Y):=\operatorname{ess\,inf}(Y).$$

The dual of this functional is well know, and given as

$$\begin{aligned}\mathbb{A}\mathrm{V@R}_0(Y) &= \inf\{\mathbb{E}[Y\cdot Z_Y]:\mathbb{E}[Z_Y]=1,\ Z_Y\geq 0\text{ a.s.}\}\\ &= s_{\{Z_Y:\,\mathbb{E}[Z_Y]=1,\,Z_Y\geq 0\text{ a.s.}\}}(Y)\end{aligned}$$

where the infimum is *not* attained in general[2] (see [39] and [24] for further treatment of this functional):

This functional is of particular interest for $Y\in\mathbb{L}^\infty(\Omega,\mathbb{P})\subseteq\mathbb{L}^p(\Omega,\mathbb{P})$, because otherwise $\mathbb{A}\mathrm{V@R}_0$ might not even be defined. Z_Y thus is in the space $Z_Y\in\mathbb{L}^\infty(\Omega,\mathbb{P})^*=ba(\Omega,\Sigma,\mathbb{P})$, the space of bounded and finitely-additive[3] measures on Ω, which are absolutely continuous with respect to $\mathbb{P}$. But our duality theory requires $Z_Y\in\mathbb{L}^1(\Omega,\mathbb{P})$ anyhow.

Put

$$Z_Y(\omega):=\frac{\mathbb{1}_{\{Y=\operatorname{ess\,inf}Y\}}(\omega)}{\mathbb{P}[\{Y=\operatorname{ess\,inf}Y\}]}$$

[2] The dual representation of $\mathbb{A}\mathrm{V@R}_\alpha$ is $\mathbb{A}\mathrm{V@R}_\alpha(Y)=\inf\left\{\mathbb{E}[Y\cdot Z_Y]:\mathbb{E}[Z_Y]=1,\,0\leq Z_Y\leq\frac{1}{\alpha}\text{ a.s.}\right\}=s_{C_\alpha}(Y)$, where $C_\alpha=\left\{Z_Y\in\mathbb{L}^\infty:\mathbb{E}[Z_Y]=1,\,0\leq Z_Y\leq\frac{1}{\alpha}\text{ a.s.}\right\}$ is weak* compact by Alaoglu's Theorem. This is in significant contrast to $\mathbb{A}\mathrm{V@R}_0$, as $\{Z_Y\in\mathbb{L}^\infty:\mathbb{E}[Z_Y]=1,\,Z_Y\geq 0\text{ a.s.}\}$ is not weak* compact, particularly not bounded.

[3] Recall that *finitely-additive* is weaker than *countably-additive*: The Theorem of Radon-Nikodym identifies countably-additive, $\mathbb{P}$-absolutely continuous measures with $\mathbb{L}^1(\Omega,\mathbb{P})$, which is the predual space in the given environment.

(if this makes sense at all) and observe that

$$\begin{aligned}\mathbb{AV@R}_0\left(\tilde{Y}\right) &= \operatorname{ess\,inf}\left(\tilde{Y}\right)\\ &\leq \mathbb{E}\left[\tilde{Y}\cdot Z_Y\right]\\ &= \mathbb{E}\left[\left(\tilde{Y}-Y\right)\cdot Z_Y\right]+\operatorname{ess\,inf}\left(Y\right)\\ &= \mathbb{E}\left[\left(\tilde{Y}-Y\right)\cdot Z_Y\right]+\mathbb{AV@R}_0\left(Y\right),\end{aligned}$$

the defining equation for the subgradient and in this situation $Z_Y \in \partial\mathbb{AV@R}_0\left(Y\right)$.

Recall that Theorem 4.8 requires the quantity $\left\|Z_Y\right\|_{r'}$, which evaluates to

$$\begin{aligned}\left\|Z_Y\right\|_{r'} &= \left(\frac{\mathbb{P}\left[\left\{Y=\operatorname{ess\,inf}\left(Y\right)\right\}\right]}{\mathbb{P}\left[\left\{Y=\operatorname{ess\,inf}\left(Y\right)\right\}\right]^{r'}}\right)^{\frac{1}{r'}}\\ &= \frac{1}{\mathbb{P}\left[\left\{Y=\operatorname{ess\,inf}\left(Y\right)\right\}\right]^{\frac{1}{r}}}.\end{aligned}$$

This bound is of course useless except for the particular situation

$$\mathbb{P}\left[\left\{Y=\operatorname{ess\,inf}\left(Y\right)\right\}\right]>0.$$

More, however, cannot be expected, because $\mathbb{P}\mapsto\mathbb{AV@R}_{0;\mathbb{P}}$ is *not* continuous, as we have already observed in Remark 4.7.

5.3. Deviation Risk Functional

Intimately connected with acceptability functionals $\mathcal{A}$ are deviation risk functionals, they appear in some situations a bit more natural to handle. Often they are called *pure risk functional*, or simply *deviation functional* as well.

As regards the notation we shall follow basically Pflug and Römisch [38] here.

Definition 5.1 (Deviation risk functional)**.** A real valued mapping $\mathcal{D}$ on random variables $Y\in\mathcal{Y}$ is a *deviation risk functional* provided that $\mathbb{E}-\mathcal{D}$ is an acceptability functional. Equivalently,

- ▷ (Translation invariance) $\mathcal{D}\left(Y+c\right)=\mathcal{D}\left(Y\right)$ for all random variables Y and real numbers $c\in\mathbb{R}$,
- ▷ (Convexity) $\mathcal{D}\left(\left(1-\lambda\right)Y_0+\lambda Y_1\right)\leq\left(1-\lambda\right)\mathcal{D}\left(Y_0\right)+\lambda\mathcal{D}\left(Y_1\right)$ for $0\leq\lambda\leq1$ and Y_0, $Y_1\in\mathcal{Y}$, and
- ▷ (Monotonicity) $\mathbb{E}\left[Y_0\right]-\mathcal{D}\left(Y_0\right)\leq\mathbb{E}\left[Y_1\right]-\mathcal{D}\left(Y_1\right)$ provided that $Y_0\leq Y_1$ almost everywhere.

Remark 5.2. One verifies that $\rho\cdot\mathcal{D}$ is a deviation risk functional as well, provided that $0\leq\rho\leq1$.

function $h(u)$	Fenchel conjugate $h^*(v) = \sup_{u\in\mathbb{R}} u\cdot v - h(u)$
$h(u) = b\cdot u^+ - a\cdot(-u)^+$	$h^*(v) = \begin{cases} 0 & \text{if } a\le v\le b \\ +\infty & \text{else} \end{cases}$
$h(u) = \lvert u\rvert$	$h^*(v) = \begin{cases} 0 & \text{if } \lvert v\rvert\le 1 \\ +\infty & \text{else} \end{cases}$
$h(u) = \frac{1}{p}\lvert u\rvert^p,\ (p>1)$	$h^*(v) = \frac{1}{q}\lvert v\rvert^q$ for $\frac{1}{p}+\frac{1}{q}=1$
$h(u) = \frac{1}{p}\left((-u)^+\right)^p,\ (p>1)$	$h^*(v) = \frac{1}{q}\left((-u)^+\right)^q$
$h(u) = \alpha+\beta u+\gamma\cdot f(\delta u+\epsilon)$	$h^*(v) = -\alpha-\epsilon\frac{v-\beta}{\delta}+\gamma\cdot f^*\left(\frac{v-\beta}{\delta\gamma}\right)$ for $\gamma>0$
$h(u) = u+\frac{1}{\gamma}\left(e^{-\gamma u}-1\right)$	$h^*(v) = \begin{cases} \frac{v}{\gamma}+\frac{1-v}{\gamma}\ln(1-v) & \text{if } v\le 1 \\ +\infty & \text{else} \end{cases}$

Table 5.1.: List of exemplary convex functions and their Fenchel conjugates.

Remark 5.3. The deviation risk functional obeys a dual representation, which comes along with the dual representation of acceptability functionals:

Given the representation

$$\mathcal{D}(Y) = \sup_{D(Z)<\infty}\left\{\mathbb{E}[Y\cdot Z] - D(Z)\right\},$$

where $D(Z) := \mathcal{D}^*(Z) = \sup\left\{\mathbb{E}[Y\cdot Z] - \mathcal{D}(Y) : Y\in\mathcal{Y}\right\}$ is the respective Fenchel-Moreau dual, and $\mathcal{A} := \mathbb{E} - \rho\cdot\mathcal{D}$[4], then

$$\mathcal{A}(Y) = \inf_{D\left(\frac{1-Z}{\rho}\right)<\infty}\left\{\mathbb{E}[Y\cdot Z] + \rho D\left(\frac{1-Z}{\rho}\right)\right\},$$

that is to say the relation $A(Z) = -\rho D\left(\frac{1-Z}{\rho}\right)$ for the dual variables holds true (cf. Pflug and Römisch, [39, Section 2.5]).

In this setting the interesting quantity, the modulus of continuity thus rewrites

$$\sup_{A(Z)>-\infty}\|Z\|_{r'} = \sup_{D(Z)<\infty}\|1-\rho Z\|_{r'}. \tag{5.3}$$

In the following discussion we will use the dual representations elaborated by Pflug in [38] and use these representations then to determine if changing the measure is continuous with respect to the Wasserstein distance for a few selected acceptability functionals.

The Deviation Functional $\mathcal{D}(Y) = \mathbb{E}\left[h\left(Y-\mathbb{E}[Y]\right)\right]$

Let h denote a non-negative convex function satisfying $h(0)=0$. The deviation risk functional

$$\mathcal{D}(Y) := \mathbb{E}\left[h\left(Y-\mathbb{E}[Y]\right)\right]$$

[4]Compositing the objective functional in this described way is sometimes called *scalarization*.

was investigated by Markowitz first for the particular situation $h(u) = |u|^p$ and $p = 2$ in [31].

Its dual representation is given by

$$\begin{aligned}\mathcal{D}(Y) &= \mathbb{E}\left[h\left(Y - \mathbb{E}[Y]\right)\right]\\ &= \sup_{\mathbb{E}[Z]=0} \mathbb{E}[Y \cdot Z] - D_{h^*}(Z),\end{aligned}$$

the convex dual of $\mathcal{D}$ thus takes the particular form

$$D(Z) = \begin{cases}\inf_{a\in\mathbb{R}} \mathbb{E}\left[h^*(Z-a)\right] & \text{if } \mathbb{E}[Z]=0\\ \infty & \text{else,}\end{cases}$$

where h^* is the Fenchel-dual of h (cf. [39]), $h^*(v) = \sup\{u \cdot v - h(u) : u \in \mathbb{R}\}$, and $D_{h^*}(Z) = \inf_{a\in\mathbb{R}} \mathbb{E}\left[h^*(Z-a)\right]$.

To investigate the continuity of the related acceptability functional consider first the random variables

$$Z_n := \frac{n}{\mathbb{P}[A]}\mathbb{1}_A - \frac{n}{\mathbb{P}[A^{\complement}]}\mathbb{1}_{A^{\complement}},$$

which satisfy $\mathbb{E}[Z_n] = 0$ and thus are feasible for any measurable set A satisfying $0 < \mathbb{P}[A] < 1$. Notice that

$$\begin{aligned}\left\|1 - \rho Z_n\right\|_{r'} &\geq \left\|1 - \rho Z_n\right\|_1\\ &\geq 2n\rho - 1.\end{aligned}$$

In order to keep

$$\sup_{D(Z)<\infty} \left\|1 - \rho Z\right\|_{r'}$$

bounded (cf. (5.3)) these random variables Z_n have to be excluded, which in turn means that necessarily $D(Z_n) = \infty$ has to hold for large n, that is $D_{h^*}(Z_n) = \infty$. As

$$\begin{aligned}D(Z_n) &= D_{h^*}(Z_n)\\ &= \inf_{a\in\mathbb{R}} \mathbb{E}\left[h^*(Z_n - a)\right]\\ &= \inf_{a\in\mathbb{R}} h^*\left(\frac{n}{\mathbb{P}[A]} - a\right)\mathbb{P}[A] + h^*\left(-\frac{n}{\mathbb{P}[A^{\complement}]} - a\right)\mathbb{P}\left[A^{\complement}\right]\end{aligned}$$

we necessarily conclude that $\{h^* < \infty\}$ has to stay bounded, i.e. compact.

We may conclude that if the deviation risk functional $\mathbb{E}\left[h\left(Y - \mathbb{E}[Y]\right)\right]$ is continuous with respect to changing the measure for the Wasserstein distance, then necessarily $\{h^* < \infty\}$ is bounded.

This latter requirement particularly excludes the variance ($h(u) = u^2$) and all functionals of the form $\mathcal{D}(Y) = \mathbb{E}\left[|Y - \mathbb{E}[Y]|^p\right]$ for any $p > 1$.

Remark 5.4. For completeness we shall give an additional word to the variance in the sequel.

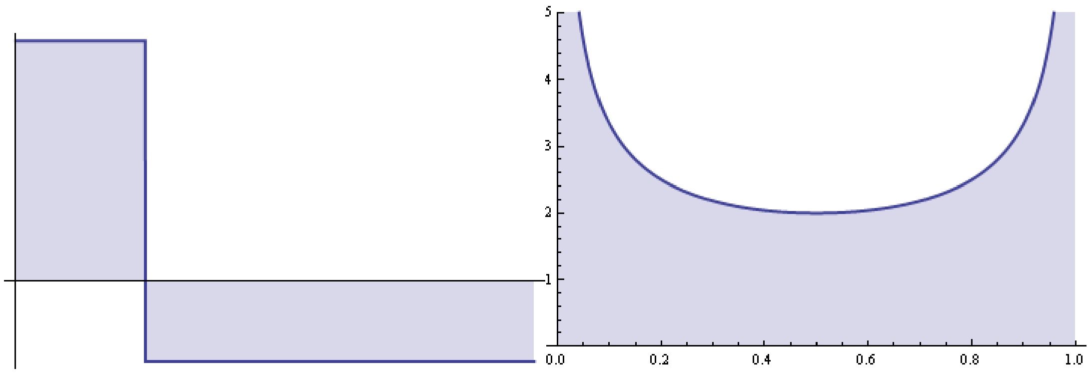

Figure 5.1.: The random variables Z_n (left) and $\|Z_1\|_2$ as a function of $\mathbb{P}[A]$ (right).

$\mathcal{D}(Y) := \mathbb{M}ad(Y)$ Is Continuous

The mean average deviation is defined as $\mathbb{M}ad[Y] := \mathbb{E}[|Y - \mathbb{E}[Y]|]$, which is the latter situation for $h(u) := |u|$. Note, that

$$h^*(u) = \begin{cases} 0 & \text{if } -1 \le u \le 1 \\ \infty & \text{else} \end{cases}$$

satisfies $\{h^* < \infty\} = [-1, 1]$ and thus is bounded, thus hope raises that $\mathbb{E} - \rho \cdot \mathbb{M}ad$ is continuous.

In fact, choose $p = 1$, with dual parameter $p' = \infty$, and thus $-2 \le Z \le 2$, as $\mathbb{E}[Z] = 0$ and $D_{h^*}(Z) < \infty$. Whence

$$\sup_{D(Z)<\infty} \|1 - \rho Z\|_\infty \le 1 + 2\rho.$$

To summarize:

$$\begin{aligned} |\mathcal{A}_{\mathbb{P}}(Y) - \mathcal{A}_{\mathbb{Q}}(Y)| &\le L(Y) \cdot (1 + 2\rho) \cdot d_1(\mathbb{P}, \mathbb{Q}) \\ &\le L(Y) \cdot (1 + 2\rho) \cdot d_r(\mathbb{P}, \mathbb{Q}), \end{aligned}$$

where

$$\mathcal{A}(Y) = \mathbb{E}[Y] - \rho \cdot \mathbb{M}ad[Y].$$

We conclude that $\mathbb{E} - \rho \cdot \mathbb{M}ad$ is continuous with respect to changing the measure when employing the Wasserstein distance.

$\mathcal{D}(Y) := \mathbb{E}|Y - \mathbb{E}Y|^p$ Is Not Continuous

Define the measures $\mathbb{P}_n := \frac{1}{2}\delta_{-n} + \frac{1}{2}\delta_n$ and $\mathbb{Q} := \delta_0$ on the simple space $(\mathbb{R}, |.|)$ and observe that

$$\mathsf{d}_r(\mathbb{P}_n, \mathbb{Q}) = \left(\frac{1}{2}n^r + \frac{1}{2}n^r\right)^{\frac{1}{r}} = n.$$

Consider the random variable $Y = \mathsf{Id}$. Then $\mathbb{E}_{\mathbb{P}_n}[Y] = 0$. For $h(u) := |u|^p$, however,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[Y] - \rho \cdot \mathbb{E}_{\mathbb{Q}}[|Y - \mathbb{E}_{\mathbb{Q}}[Y]|^p] - \mathbb{E}_{\mathbb{P}_n}[Y] + \rho \cdot \mathbb{E}_{\mathbb{P}_n}[|Y - \mathbb{E}_{\mathbb{P}_n}[Y]|^p] \\ &= 0 - 0 - 0 + \rho \cdot n^p \\ &= \rho \cdot n^p \\ &\geq \rho \cdot n^{p-1} \cdot \mathsf{d}_r(\mathbb{P}_n, \mathbb{Q}),\end{aligned}$$

and thus the functional

$$\mathcal{A}(Y) := \mathbb{E}[Y] - \rho \cdot \mathbb{E}[|Y - \mathbb{E}[Y]|^p]$$

in general is *not* Lipschitz-continuous for $p > 1$ with respect to changing the measure.

$\mathcal{D}(Y) := \mathbb{V}ar(Y)$ Is Not Continuous

This is just a special case of the previous situation for $p = 2 > 1$: The functional

$$\mathcal{A}(Y) := \mathbb{E}[Y] - \rho \cdot \mathbb{V}ar[Y]$$

is *not* continuous with respect to changing the measure.

The Central Deviation $\mathcal{D}(Y) := \|Y - \mathbb{E}[Y]\|_p$ Is Continuous

The functional $\mathcal{D}(Y) := \|Y - \mathbb{E}[Y]\|_p$ is convex as well, the dual is

$$\begin{aligned}\mathcal{D}(Y) &= \sup\{\mathbb{E}[Y \cdot Z] : \mathbb{E}[Z] = 0,\ D_{p'}(Z) \leq 1\} \\ &= \sup\{\mathbb{E}[Y \cdot Z] : D(Z) < \infty\},\end{aligned}$$

where

$$D_{p'}(Z) = \inf_{a \in \mathbb{R}} \|Z - a\|_{r'} \tag{5.4}$$

and thus

$$D(Z) = \begin{cases} 0 & \text{if } \mathbb{E}[Z] = 0 \text{ and } \inf_{a\in\mathbb{R}} \|Z - a\|_{p'} \leq 1 \\ \infty & \text{else.}\end{cases}$$

Now notice that $\mathbb{E}[Z] = 0$ and $\|Z - a^*\|_{p'} = \inf_{a\in\mathbb{R}} \|Z - a\|_{p'} \leq 1$ imply that

$$\begin{aligned} 0 &= \mathbb{E}[Z] \\ &\leq |\mathbb{E}[Z - a^*]| = |\mathbb{E}[Z] - a^*| \\ &\leq \mathbb{E}[|Z - a^*|] \\ &= \|Z - a^*\|_1 \\ &\leq \|Z - a^*\|_{p'} \\ &\leq 1,\end{aligned}$$

and thus $|a^*| \leq 1 + |\mathbb{E}[Z]| = 1$ for the optimal a^* in (5.4). Moreover,

$$\begin{aligned} \|Z\|_{p'} &\leq |a^*| + \|Z - a^*\|_{p'} \\ &\leq 1 + 1 \\ &= 2, \end{aligned}$$

and it follows that

$$\sup_{D(Z)<\infty} \|1 - \rho Z\|_{p'} \leq 1 + 2\rho.$$

To summarize,

$$\begin{aligned} |\mathcal{A}_{\mathbb{P}}(Y) - \mathcal{A}_{\mathbb{Q}}(Y)| &\leq L(Y) \cdot (1 + 2\rho) \cdot d_1(\mathbb{P}, \mathbb{Q}) \\ &\leq L(Y) \cdot (1 + 2\rho) \cdot d_r(\mathbb{P}, \mathbb{Q}), \end{aligned}$$

for the acceptability functional $\mathcal{A}(Y) = \mathbb{E}[Y] - \rho \cdot \|Y - \mathbb{E}[Y]\|_p$.

Improved Bound For The Standard Deviation $\mathcal{D}(Y) := \|Y - \mathbb{E}[Y]\|_2$

The latter bound, however, can be improved in the situation $p = p' = 2$, as

$$\inf_{a \in \mathbb{R}} \|Z - a\|_2 = \|Z\|_2$$

(that is $a^* = \mathbb{E}[Z] = 0$) and whence

$$\begin{aligned} \|1 - \rho Z\|_2^2 &= \mathbb{E}\left[(1 - \rho Z)^2\right] \\ &= 1 - 2\rho \mathbb{E}[Z] + \rho^2 \mathbb{E}\left[Z^2\right] \\ &\leq 1 + \rho^2, \end{aligned}$$

and therefore

$$\sup_{D(Z)<\infty} \|1 - \rho Z\|_2 \leq \sqrt{1 + \rho^2}.$$

Given Y, the random variable minimizing

$$\begin{aligned} \mathcal{A}(Y) &:= \mathbb{E}[Y] - \rho \sqrt{\mathbb{V}ar[Y]} \\ &= \inf \left\{ \mathbb{E}[Y \cdot Z] : \mathbb{E}[Z] = 1, \mathbb{E}\left[Z^2\right] \leq 1 + \rho^2 \right\} \end{aligned}$$

may be given explicitly as $Z_Y = 1 - \frac{\rho}{\sqrt{\mathbb{V}ar[Y]}} (Y - \mathbb{E}[Y])$; moreover, $\|Z_Y\|_2 = \sqrt{1 + \rho^2}$.

The Lower Central Semideviation $\mathcal{D}(Y) := \left\|(\mathbb{E}[Y]-Y)^+\right\|_p$

Consider the functional $\mathcal{D}(Y) := \left\|(\mathbb{E}[Y]-Y)^+\right\|_p$. In this situation we have that

$$\left\|(\mathbb{E}[Y]-Y)^+\right\|_p = \sup\left\{\mathbb{E}[Y\cdot Z]:\ Z=\mathbb{E}[V]-V,\ V\geq 0 \text{ and } \|V\|_{p'}\leq 1\right\}$$

and

$$D(Z) = \begin{cases} 0 & \text{if } Z=\mathbb{E}[V]-V,\ V\geq 0 \text{ and } \|V\|_{p'}\leq 1\\ \infty & \text{else.}\end{cases}$$

It is comparably easy thus to establish that

$$\begin{aligned}\sup_{D(Z)<\infty}\|1-\rho Z\|_{r'} &\leq 1+\rho\cdot\|V-\mathbb{E}[V]\|_{r'}\\ &\leq 1+\rho\cdot\left(\|V\|_{r'}+\mathbb{E}[V]\right)\\ &\leq 1+2\rho\cdot\|V\|_{p'}\end{aligned}$$

whenever $r\geq p$ and again we have continuity for the acceptability functional of the particular form

$$\mathcal{A}(Y)=\mathbb{E}[Y]-\rho\cdot\left\|(Y-\mathbb{E}[Y])^-\right\|_p.$$

Minimal Loss Functional

Consider the deviation risk functional (with dual representation) of the form (with dual representation)

$$\begin{aligned}\mathcal{D}(Y):&=\min_{a\in\mathbb{R}}\mathbb{E}[h(Y-a)]\\ &=\sup\left\{\mathbb{E}[Y\cdot Z]-\mathbb{E}[h^*(Z)]:\ \mathbb{E}[Z]=0\right\},\end{aligned}$$

that is

$$D(Z)=\begin{cases}\mathbb{E}[h^*(Z)] & \text{if } \mathbb{E}[Z]=0\\ \infty & \text{else.}\end{cases}$$

Again and as above consider the random variables $Z_n := \frac{n}{\mathbb{P}[A]}\mathbb{1}_A - \frac{n}{\mathbb{P}[A^\complement]}\mathbb{1}_{A^\complement}$ which has $\mathbb{E}[Z_n]=0$, but $\|1-\rho Z\|_{p'}>2n\rho-1$. In order to keep

$$\sup_{D(Z)<\infty}\|1-\rho Z\|_{p'_\beta}$$

bounded we necessarily have to exclude those variables Z_n, that is to say $\{h^*<\infty\}$ has to stay bounded from at least one side. This again excludes all function of the form $h(u)=|u|^p$.

Entropic Acceptability Functional $\mathcal{A}(Y) = -\frac{1}{\gamma}\ln\mathbb{E}\left[e^{-\gamma Y}\right]$

The entropic acceptability functional is a special case of the previous section for the choice $h(u) = u + \frac{1}{\gamma}\left(e^{-\gamma u} - 1\right)$.

In this situation

$$\begin{aligned}\mathcal{D}(Y) &= \min_{a\in\mathbb{R}} h(Y-a)\\ &= \mathbb{E}[Y] + \frac{1}{\gamma}\ln\mathbb{E}\left[e^{-\gamma Y}\right],\end{aligned}$$

the minimum is attained for $a^* = -\frac{1}{\gamma}\ln\mathbb{E}\left[e^{-\gamma Y}\right]$. The dual representation is

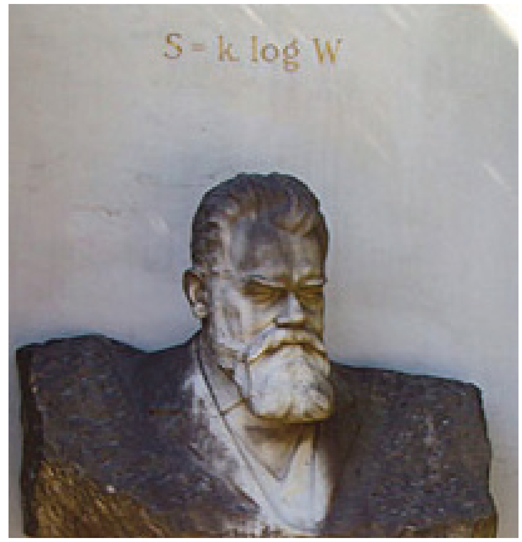

Figure 5.2.: Ludwig Boltzmann's epitaph.

$$\begin{aligned}\mathcal{D}(Y) &= \sup_{\mathbb{E}[Z]=0} \mathbb{E}[Y\cdot Z] - h^*(Z)\\ &= \sup \mathbb{E}[Y\cdot Z] - D(Z)\end{aligned}$$

for $D(Z) = \begin{cases}\mathbb{E}[h^*(Z)] & \text{if } \mathbb{E}[Z]=0\\ \infty & \text{else.}\end{cases}$

Now we learn from table 5.1 that $h^*(v) = \infty$ for $v > 1$, and thus

$$\sup_{D(Z)<\infty} \|1-\rho Z\|_{r'} = \sup_{\mathbb{E}[Z]=0,\, Z\le 1} \|1-\rho Z\|_{r'}.$$

Choose any measurable set A and consider the random variable $Z_A = \mathbb{1}_A - \frac{\mathbb{P}[A]}{1-\mathbb{P}[A]}\mathbb{1}_{A^\complement}$ satisfying $Z_A \le 1$ and $\mathbb{E}[Z_A] = 0$.

Then

$$\|1-\rho\cdot Z_A\|_{r'} = \left((1-\rho)\,\mathbb{P}[A] + \left(1+\rho\frac{\mathbb{P}[A]}{1-\mathbb{P}[A]}\right)^{r'}(1-\mathbb{P}[A])\right)^{\frac{1}{r'}};$$

this quantity evaluates as

$$\|1-\rho\cdot Z_A\|_{r'} = \begin{cases}1 & \text{if } r'=1\\ \text{unbounded} & \text{else,}\end{cases}$$

an indicator for the acceptability functional not to be Lipschitz with respect to changing the measure.

Indeed, let $\mathbb{P}_n := \left(1-\frac{1}{n^r}\right)\delta_0 + \frac{1}{n^r}\delta_{-n}$. Then

$$\begin{aligned}\mathsf{d}_r(\mathbb{P}_n,\delta_0) &= \left(\int |\omega-0|^r\,\mathbb{P}_n[\mathrm{d}\omega]\right)^{\frac{1}{r}}\\ &= \left(0+n^r\frac{1}{n^r}\right)^{\frac{1}{r}} = 1.\end{aligned}$$

However,

$$\mathcal{A}_{\delta_0}(\mathsf{Id}) = -\frac{1}{\gamma}\ln\mathbb{E}_{\delta_0}\left[e^{-\gamma\mathsf{Id}}\right] = 0,$$

but

$$\begin{aligned}\mathcal{A}_{\mathbb{P}_n}(\mathsf{Id}) &= -\frac{1}{\gamma}\ln\mathbb{E}_{\mathbb{P}_n}\left[e^{-\gamma\mathsf{Id}}\right]\\ &= -\frac{1}{\gamma}\ln\left(\left(1-\frac{1}{n^r}\right)e^0+\frac{1}{n^r}e^{\gamma n}\right)\\ &= -n+\frac{r}{\gamma}\ln n + o(1),\end{aligned}$$

so we do not have continuity for any $r \geq 1$.

Summary

This following table summarizes the particular results established above:

Acceptability Functional	Modulus of Continuity
$AV@R_\alpha(Y)$	$\alpha^{-\frac{1}{r}}$
$\mathcal{A}_h(Y)$	$\|h\|_r$
$\mathcal{A}(Y) = -\frac{1}{\gamma}\ln\mathbb{E}\left[e^{-\gamma Y}\right]$	not continuous
$\mathcal{A} = \mathbb{E} - \rho\cdot\mathcal{D}$ for Deviation $\mathcal{D}$	
$\mathcal{D}(Y) = \mathbb{M}ad(Y)$	$1+2\rho$
$\mathcal{D}(Y) = \mathbb{E}[h(Y-\mathbb{E}[Y])]$	necessary condition: $\{h^* < \infty\}$ bounded
$\mathcal{D}(Y) = \mathbb{E}[(Y-\mathbb{E}[Y])^p]$	not continuous for $p > 1$
$\mathcal{D}(Y) = \mathbb{V}ar(Y)$	not continuous
$\mathcal{D}(Y) = \|Y-\mathbb{E}[Y]\|_p$	$1+2\rho$
$\mathcal{D}(Y) = \|Y-\mathbb{E}[Y]\|_2$	$\sqrt{1+\rho^2}$
$\mathcal{D}(Y) = \left\|(\mathbb{E}[Y]-Y)^+\right\|_p$	$1+2\rho$ if $p \leq r$
$\mathcal{D}(Y) = \min_{a\in\mathbb{R}}\mathbb{E}[h(Y-a)]$	necessary condition: $\{h^* < \infty\}$ bounded from left or right

6. Ambiguity

In the investigations above we have discussed the problem

$$\begin{array}{rl} \text{minimize} & \\ \text{(in } \mathbb{Q}) & \mathcal{A}_{\mathbb{Q}}(Y) \\ \text{subject to} & \mathsf{d}_r(\mathbb{Q},\mathbb{P}) \leq K. \end{array} \tag{6.1}$$

and found a lower bound, which was of the form

$$\mathcal{A}_{\mathbb{P}}(Y) - L(Y)\cdot K\cdot \inf_{Z\in\partial\mathcal{A}_{\mathbb{P}}(Y)} \|Z\|_{r'}.$$

We shall drive the investigations further now and give some very general situations, for which the bounds achieved are sharp. Further, we shall characterize the measure, such that the problem in consideration attains its minimal value. It will turn out that these measures have an interesting description as a transport plan.

Moreover, situations will occur where the bounds are not attained. For some of them we will prove, that no such bound exists in general, the problem thus is not continuous.

Before turning to the general situation, however, we shall start with the simpler problem

$$\begin{array}{rl} \text{minimize} & \\ \text{(in } \mathbb{Q}) & \mathbb{E}_{\mathbb{Q}}[Y] \\ \text{subject to} & \mathsf{d}_r(\mathbb{P},\mathbb{Q}) \leq K \\ & \mathbb{Q}\in\mathcal{P}_r(\Omega) \end{array} \tag{6.2}$$

to develop the strategy and the notion.

As above, let $(\Omega,\Sigma,\mathbb{P})$ denote a probability triple. We shall assume in addition that Ω is a linear space, for example $(\mathbb{R}^m,\|.\|)$, equipped with an appropriate norm function $\|.\|$.

On this space there is the usual notion of a dual Ω^*, collecting all linear, continuous functionals on Ω. Notice, that any linear functional $Y\colon\Omega\to\mathbb{R}$ is a random variable itself, and $Y\in\Omega^*$.

Given a linear random variable Y on $\mathbb{R}^m$, $Y(\omega)$ then represents an inner product like evaluation for an atom $\omega\in\Omega$: Y may be represented as $Y(\omega)=\sum_{s=1}^m Y_s\omega_s = Y^\top\omega$ for the vector $Y=(Y_1,Y_2,\dots Y_m)\in\mathbb{R}^m$.

It is obvious that for any such Y there is a vector

$$\omega_Y \in \underset{\|\omega\|=1}{\operatorname{argmax}} \frac{Y(\omega)}{\|\omega\|} \subseteq \mathbb{R}^m;$$

in full generality, to break a fly on the wheel, this is the Hahn-Banach Theorem for the reflexive $\mathbb{R}^m$, whose unit ball is compact.

Theorem 6.1. *If Ω is linear, equipped with a norm and Y a linear functional, then, for all $1 \le r < \infty$, the bound*

$$\mathbb{E}_{\mathbb{P}}[Y] - K \cdot L(Y)$$

is sharp, the minimizing measure is the push-forward (image measure of $\mathbb{P}$)[1]

$$\mathbb{Q}^* := T_*(\mathbb{P}) = \mathbb{P} \circ T^{-1}$$

for the (affine linear) transport map $T(\omega) := \omega - K \cdot \omega_Y$.

Proof. By Kantorovich's famous duality theorem

$$\begin{aligned} \left|\mathbb{E}_{\mathbb{P}}[Y] - \mathbb{E}_{\mathbb{Q}}[Y]\right| &\le L(Y) \cdot \mathsf{d}_{KA}(\mathbb{P}, \mathbb{Q}) \\ &= L(Y) \cdot \mathsf{d}_1(\mathbb{P}, \mathbb{Q}) \qquad (6.3) \\ &\le L(Y) \cdot \mathsf{d}_r(\mathbb{P}, \mathbb{Q}) \qquad (6.4) \end{aligned}$$

for $r \ge 1$, establishing that $\mathbb{E}_{\mathbb{Q}}[Y] \ge \mathbb{E}_{\mathbb{P}}[Y] - L(Y) \cdot \mathsf{d}_r(\mathbb{P}, \mathbb{Q}) \ge \mathbb{E}_{\mathbb{P}}[Y] - L(Y) \cdot K$.

To observe that this bound is sharp indeed let Ω and Y be linear,

$$\mathbb{Q}^* := T_*(\mathbb{P}) = \mathbb{P} \circ T^{-1}$$

and define the transport plan

$$\pi := (\mathsf{Id} \times T)_*(\mathbb{P}),$$

that is $\pi[A \times B] = \mathbb{P}[A \cap T^{-1}(B)]$, where

$$(\mathsf{Id} \times T)(\omega) := (\omega, T(\omega)).$$

The Wasserstein distance of $\mathbb{P}$ and $\mathbb{Q}^*$ is bounded by K, because

$$\begin{aligned} \mathsf{d}_r(\mathbb{P}, \mathbb{Q}^*)^r &\le \int d(\omega_1, \omega_2)^r \, \pi[\mathrm{d}\omega_1, \mathrm{d}\omega_2] \\ &= \int d(\omega, T(\omega))^r \, \mathbb{P}[\mathrm{d}\omega] \\ &= \int \|K \cdot \omega_Y\|^r \, \mathbb{P}[\mathrm{d}\omega] \\ &= K^r \|\omega_Y\|^r \\ &= K^r. \end{aligned}$$

[1] Villani rather uses the notation $T\#\mathbb{P} := T_*(\mathbb{P})$ for the push-forward measure. We, however, have the impression that $T_*(\mathbb{P})$ is more convenient, the notation $\mathbb{P}^T$ is in frequent use as well.

Given this measure $\mathbb{Q}^*$ the objective of the primal function is

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}^*}[Y] &= \int Y(\omega)\,\mathbb{P}\circ T^{-1}[\mathrm{d}\omega] \\ &= \int Y(T(\omega))\cdot\mathbb{P}[\mathrm{d}\omega] \\ &= \int Y(\omega - K\cdot\omega_Y)\,\mathbb{P}[\mathrm{d}\omega] \\ &= \mathbb{E}_{\mathbb{P}}[Y] - K\cdot Y(\omega_Y) \\ &= \mathbb{E}_{\mathbb{P}}[Y] - K\cdot L(Y),\end{aligned}$$

which is the minimum value we can achieve in view of (6.4). □

We shall now turn to the general situation.

Theorem 6.2 (Optimal transport plan)**.** *Let Y be linear on a linear space equipped with a norm.*

(i) Consider the problem

$$\begin{array}{rl}\textit{minimize} \\ (\textit{in } \mathbb{Q}) & \mathcal{A}_{\mathbb{Q}}(Y) \\ \textit{subject to} & \mathsf{d}_r(\mathbb{Q},\mathbb{P}) \leq K.\end{array}$$

Then the minimal value is $\mathcal{A}_{\mathbb{P}}(Y) - K\cdot L(Y)\cdot\min_{Z\in\partial\mathcal{A}(Y)}\|Z\|_{r'}$.

(ii) For $1<r<\infty$ the minimizing measure is given by the push-forward

$$\mathbb{Q}^* := T_*(\mathbb{P}) = \mathbb{P}\circ T^{-1},$$

where T is the transport map

$$\begin{aligned}T(\omega): &= \omega - K\cdot\left|\frac{Z_Y(\omega)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}}\cdot\operatorname{sign} Z_Y(\omega)\cdot\omega_Y \\ &= \omega - \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\cdot|Z_Y(\omega)|^{\frac{r}{r-1}-2}\cdot Z_Y(\omega)\cdot\omega_Y\end{aligned} \tag{6.5}$$

(the random variable $Z_Y\in\operatorname{argmin}\left\{\|Z\|_{\frac{r}{r-1}} : Z\in\partial\mathcal{A}_{\mathbb{P}}(Y)\right\}$ is the smallest (in norm), and optimal dual variable (Lagrange multiplier)).

(iii) For $r=1$ the optimal transport map is

$$T(\omega) := \omega - K\cdot\frac{\mathbb{1}_{\{|Z_Y|=\|Z_Y\|\}}(\omega)}{\mathbb{P}[|Z_Y|=\|Z_Y\|]}\cdot\operatorname{sign} Z_Y(\omega)\cdot\omega_Y.$$

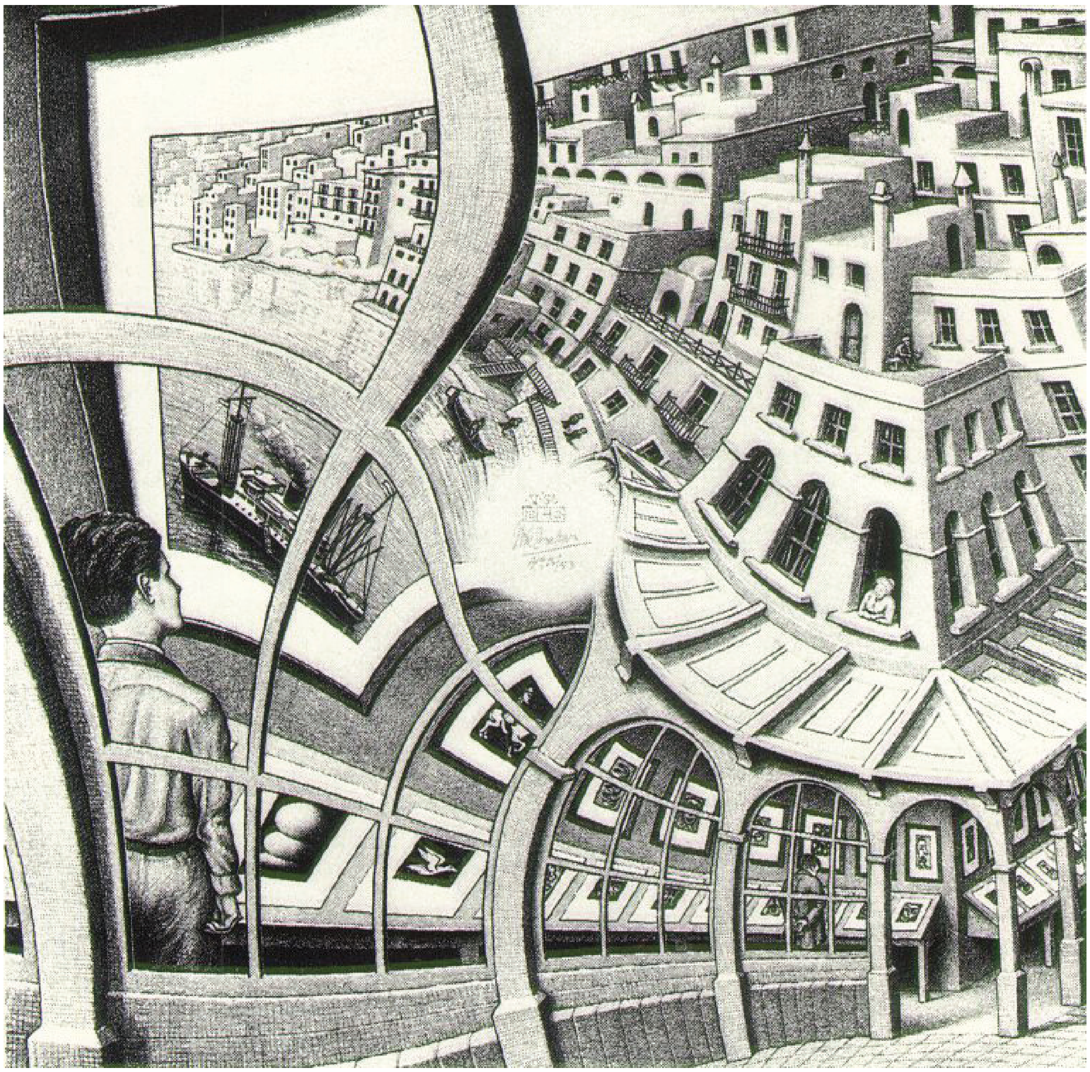

Figure 6.1.: Distortions in M. C. Escher's (1898 - 1972) drawing "Gallery".

Proof. Again consider the transport plan

$$\pi := (\mathsf{Id} \times T)_* (\mathbb{P})$$

with the marginals $\mathbb{P}$ and $\mathbb{Q}^*$; observe that

$$\begin{aligned}
\mathsf{d}_r (\mathbb{P}, \mathbb{Q}^*)^r &\leq \int \|\omega - \omega'\|^r \, \pi [\mathrm{d}\omega, \mathrm{d}\omega'] \\
&= \int \|\omega - T(\omega)\|^r \, \mathbb{P} [\mathrm{d}\omega] \\
&= \int \left\| K \cdot \left| \frac{Z_Y(\omega)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \omega_Y \right\|^r \mathbb{P} [\mathrm{d}\omega] \\
&= \frac{K^r}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}}} \cdot \int |Z_Y|^{\frac{r}{r-1}} \, \mathrm{d}\mathbb{P} \\
&= K^r,
\end{aligned}$$

that is to say $\mathbb{Q}^*$ has an accepted distance from $\mathbb{P}$.

We shall observe now that the transport map T is injective:

Choose ω_1 and ω_2 and note that

$$\begin{aligned}
&T(\omega_1) - T(\omega_2) = \\
&= \omega_1 - \omega_2 + \\
&- K \cdot \omega_Y \left(\left| \frac{Z_Y(\omega_1)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(\omega_1) - \left| \frac{Z_Y(\omega_2)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(\omega_2) \right).
\end{aligned}$$

One may assume – without loss of generality – that $Z_Y(\omega_1) \leq Z_Y(\omega_2)$ (otherwise reverse them) and distinguish the following two situations:

(i) If $Z_Y(\omega_1) = Z_Y(\omega_2)$, then $T(\omega_1) - T(\omega_2) = \omega_1 - \omega_2$ and T thus is injective on this subset.

(ii) If $Z_Y(\omega_1) < Z_Y(\omega_2)$, then $Y(\omega_1) \geq Y(\omega_2)$ a.s., because Y and Z_Y are coupled in an antimonotone way. In this situation

$$\begin{aligned}
&Y(T(\omega_1) - T(\omega_2)) = \\
&\quad = Y(\omega_1 - \omega_2) + \\
&\quad\quad - K \|Y\| \left(\left| \frac{Z_Y(\omega_1)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(\omega_1) - \left| \frac{Z_Y(\omega_2)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(\omega_2) \right) \\
&\quad > Y(\omega_1 - \omega_2) \\
&\quad \geq 0,
\end{aligned}$$

because the map $x \mapsto \operatorname{sign}(x) \cdot |x|^{\frac{1}{r-1}}$ is increasing.

Whence, $T(\omega_1) \neq T(\omega_2)$ unless $\omega_1 = \omega_2$.

Define the random variable

$$Z_Y^T := \mathbb{E}\left[Z_Y | \, T \right]$$

by conditional expectation (conditioning as factorization, a by-product of the Radon-Nikodym theorem), which is a nice way to circumvent discussions about subsets, which are not contained in T's image set. Due to its definition Z_Y^T obeys the defining property

$$\begin{aligned} \int_{T^{-1}(B)} Z_Y \mathrm{d}\mathbb{P} &= \int_{T^{-1}(B)} \mathbb{E}\left[Z_Y | \, T \right] \circ T \mathrm{d}\mathbb{P} \\ &= \int_B \mathbb{E}\left[Z_Y | \, T \right] \mathrm{d}T_* (\mathbb{P}) \\ &= \int_B Z_Y^T \mathrm{d}\mathbb{Q}^* \end{aligned} \tag{6.6}$$

for all measurable sets B (cf. [58]). Notice, that

$$\begin{aligned} \int_{T^{-1}(B)} Z_Y \mathrm{d}\mathbb{P} &= \int_B Z_Y^T \mathrm{d}\mathbb{Q}^* \\ &= \int_B Z_Y^T \mathrm{d}T_* (\mathbb{P}) \\ &= \int_{T^{-1}(B)} Z_Y^T \circ T \mathrm{d}\mathbb{P} \end{aligned}$$

by the change of variable formula again and for all measurable sets B, thus

$$Z_Y = Z_Y^T \circ T$$

$\mathbb{P}$-a.e., and

$$Z_Y^T = Z_Y \circ T^{-1}$$

$\mathbb{Q}^*$-a.e as T is injective.

One deduces from (6.6) further that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left[Z_Y \cdot \mathbb{1}_{\{Z_Y \leq t\}} \right] &= \int_{T^{-1}T\{Z_Y \leq t\}} Z_Y \mathrm{d}\mathbb{P} \\ &= \int_{T\{Z_Y \leq t\}} Z_Y^T \mathrm{d}\mathbb{Q}^* \\ &= \mathbb{E}_{\mathbb{Q}^*}\left[Z_Y^T \cdot \mathbb{1}_{\left\{Z_Y^T \leq t\right\}} \right], \end{aligned}$$

and whence

$$t - \frac{1}{\alpha} \mathbb{E}_{\mathbb{P}}\left[(t - Z_Y)^+ \right] = t - \frac{1}{\alpha} \mathbb{E}_{\mathbb{Q}^*}\left[\left(t - Z_Y^T \right)^+ \right],$$

which is a well-known identity – cf. Corollary 3.20. Taking the maximum with respect to t, it will be attained for the same t at the left and at the right:

$$\begin{aligned} G_{Z_Y}^{-1}(\alpha) &= \operatorname*{argmax}_t t - \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}\left[(t - Z_Y)^+\right] \\ &= \operatorname*{argmax}_t t - \frac{1}{\alpha}\mathbb{E}_{\mathbb{Q}^*}\left[\left(t - Z_Y^T\right)^+\right] \\ &= G_{Z_Y^T}^{-1}(\alpha), \end{aligned}$$

(a.e.) and so it follows that Z_Y and Z_Y^T have the same cumulative distribution function under their respective measures, $\mathbb{P}[Z_Y \le z] = \mathbb{Q}^*\left[Z_Y^T \le z\right]$, so finally $A_{\mathbb{Q}^*}\left(Z_Y^T\right) = A_{\mathbb{P}}(Z_Y)$ (there are obviously easier ways to introduce Z_Y^T and to show that the distributions coincide, but we consider this as elegant).

As Z_Y is optimal, $\mathcal{A}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}[Y \cdot Z_Y] - A_{\mathbb{P}}(Z_Y)$ and thus further

$$\begin{aligned} &\mathcal{A}_{\mathbb{Q}^*}(Y) - \mathcal{A}_{\mathbb{P}}(Y) = \\ &\qquad \le \mathbb{E}_{\mathbb{Q}^*}\left[Y \cdot Z_Y^T\right] - A_{\mathbb{Q}^*}\left(Z_Y^T\right) - \mathbb{E}_{\mathbb{P}}[Y \cdot Z_Y] + A_{\mathbb{P}}(Z_Y) \\ &\qquad = \mathbb{E}_{\mathbb{Q}^*}\left[Y \cdot Z_Y^T\right] - \mathbb{E}_{\mathbb{P}}[Y \cdot Z_Y] \\ &\qquad = \mathbb{E}_{\mathbb{P}}[(Y \circ T) \cdot Z_Y] - \mathbb{E}_{\mathbb{P}}[Y \cdot Z_Y] \\ &\qquad = \mathbb{E}_{\mathbb{P}}[(Y(T - \mathsf{Id})) \cdot Z_Y], \end{aligned}$$

by linearity of Y. Using $\omega_Y(Y) = \|Y\| = L(Y)$ one finds further that

$$\begin{aligned} &\mathcal{A}_{\mathbb{Q}^*}(Y) - \mathcal{A}_{\mathbb{P}}(Y) = \\ &\qquad \le \mathbb{E}_{\mathbb{P}}\left[Y\left(-K \cdot \left|\frac{Z_Y(\omega)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \operatorname{sign} Z_Y(\omega) \cdot \omega_Y\right) \cdot Z_Y\right] \\ &\qquad = -\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\|Y\| \cdot \mathbb{E}_{\mathbb{P}}\left[|Z_Y(\omega)|^{\frac{1}{r-1}} \cdot |Z_Y|\right] \\ &\qquad = -\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}}\|Y\| \cdot \mathbb{E}_{\mathbb{P}}\left[|Z_Y|^{\frac{r}{r-1}}\right] \\ &\qquad = -\frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \cdot L(Y) \cdot \|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}} \\ &\qquad \le -K \cdot L(Y) \cdot \min_{Z \in \partial\mathcal{A}(Y)} \|Z_Y\|_{\frac{r}{r-1}}, \end{aligned}$$

whence

$$\mathcal{A}_{\mathbb{Q}^*}(Y) - \mathcal{A}_{\mathbb{P}}(Y) \le -K \cdot L(Y) \cdot \min_{Z \in \partial\mathcal{A}(Y)} \|Z_Y\|_{\frac{r}{r-1}}.$$

In view of (4.3) this is smallest difference achievable.

For the Kantorovich distance ($r = 1$) the proof needs a slight modification, it may read as follows:

$$\begin{aligned}
\mathsf{d}_{KA}(\mathbb{P}, \mathbb{Q}^*) &\leq \int \|\omega - T(\omega)\| \, \mathbb{P}[\mathrm{d}\omega] \\
&= \int \left\| K \cdot \mathbb{1}_{\{|Z_Y| = \|Z_Y\|\}}(\omega) \frac{\operatorname{sign} Z_Y(\omega)}{\mathbb{P}[|Z_Y| = \|Z_Y\|]} \cdot \omega_Y \right\| \mathbb{P}[\mathrm{d}\omega] \\
&= K \cdot \int \frac{\mathbb{1}_{\{|Z_Y| = \|Z_Y\|\}}(\omega)}{\mathbb{P}[|Z_Y| = \|Z_Y\|]} \mathbb{P}[\mathrm{d}\omega] \\
&= K.
\end{aligned}$$

On the other side,

$$\begin{aligned}
\mathcal{A}_{\mathbb{Q}^*}(Y) - \mathcal{A}_{\mathbb{P}}(Y) &= \\
&= \mathbb{E}_{\mathbb{P}} \left[Y \left(-K \cdot \mathbb{1}_{\{|Z_Y| = \|Z_Y\|\}}(\omega) \frac{\operatorname{sign} Z_Y(\omega)}{\mathbb{P}[|Z_Y| = \|Z_Y\|]} \cdot \omega_Y \right) \cdot Z_Y \right] \\
&= -K \cdot \|Y\| \cdot \int \frac{\mathbb{1}_{\{|Z_Y| = \|Z_Y\|\}} |Z_Y|}{\mathbb{P}[|Z_Y| = \|Z_Y\|]} \mathrm{d}\mathbb{P} \\
&= -K \cdot L(Y) \cdot \|Z_Y\|_\infty \\
&\leq -K \cdot L(Y) \cdot \min_{Z_Y \in \partial \mathcal{A}(Y)} \|Z_Y\|_\infty ,
\end{aligned}$$

which establishes the result in this particular case. □

6.1. Interpretation And Further Discussion

6.1.1. Similarity And Relation With Optimal Transport Maps In $\mathbb{R}^m$.

The optimal transport map T in (6.7) – at least to some extent – strikingly reminds to the following result in the context of optimal transportation. We cite the theorem from [1] in a laxly way to make the key ingredients apparent.

Theorem 6.3 (Optimal transport maps in $\mathbb{R}^m$). *Let $\mathbb{P}$ and $\mathbb{Q}$ denote probability measures on $\mathbb{R}^m$ and assume the distance function of the particular form*

$$c(x, y) = \varphi(x - y)$$

for some strictly convex function $\varphi : \mathbb{R}^m \to [0, \infty)$. Then the optimal measure for the Kantorovich problem (2.3) takes the particular form $\pi = (\mathsf{Id} \times T)_*(\mathbb{P})$, where

$$T(x) = x - (\nabla\varphi)^{-1} (\nabla\lambda(x)) \tag{6.7}$$

and λ is the respective maximal Kantorovich potential (cf. section 2.1).

Before we proceed to sketching the proof let us point out the similarities with (6.5): Choose

$$\varphi(x) := \frac{\|Z_Y\|_{\frac{1}{r-1}}}{r \cdot K^{r-1}} \cdot \|x\|_r^r ,$$

thus

$$\begin{aligned}
\nabla\varphi(x) &= \frac{\|Z_Y\|_{\frac{1}{r-1}}}{K^{r-1}} \cdot \|x\|_r^{r-1} \left(\left| \frac{x_t}{\|x\|_r} \right|^{r-1} \cdot \frac{x_t}{|x_t|} \right)_t \\
&= \frac{\|Z_Y\|_{\frac{1}{r-1}}}{K^{r-1}} \cdot \left(|x_t|^{r-2} \cdot x_t \right)_t
\end{aligned}$$

by Example 2.8. Whence,

$$\begin{aligned}
(\nabla\varphi)^{-1}(y) &= \frac{K}{\|Z_Y\|_{\frac{1}{r-1}}^{\frac{1}{r-1}}} \cdot |y|^{\frac{1}{r-1}-1} \cdot y \\
&= \frac{K}{\|Z_Y\|_{\frac{1}{r-1}}^{\frac{1}{r-1}}} \cdot |y|^{\frac{r}{r-1}-2} \cdot y,
\end{aligned}$$

and the optimal transport map according (6.7) thus is given as

$$T(x) = x - \frac{K}{\|Z_Y\|_{\frac{1}{r-1}}^{\frac{1}{r-1}}} \cdot |\nabla\lambda(x)|^{\frac{r}{r-1}-2} \cdot \nabla\lambda(x) .$$

Our result in (6.5) is completely similar, but the dual variable Z_Y taking the role of $\nabla\lambda(x)$ instead, the gradient of the maximal Kantorovich potential. Notice, that λ itself is the dual variable of the problem (2.3).

Problem 6.4. As a consequence we would appreciate to have the maximal Kantorovich potential in the particular form

$$\begin{aligned}
\lambda(x) &= \int_{x_0}^{x} \mathbf{HB}_Y(\dot{\mathrm{r}}(t))\, Z_Y(\mathrm{r}(t))\, \mathrm{d}t \\
&= \int_{x_0}^{x} Z_Y(\mathrm{r})\, \mathbf{HB}_Y(\mathrm{dr}) ,
\end{aligned}$$

where $t \mapsto \mathrm{r}(t)$ is a proper parametrisation of any path connecting x and some fixed initial point x_0. It is open – to our knowledge – to characterize the Kantorovich potential accordingly.

Sketch of the proof of Theorem 6.3. Recall that

$$\begin{aligned}
\lambda(x) + \lambda^c(y) &\leq c(x,y) \\
&= h(x-y)
\end{aligned}$$

and $x' \mapsto h(x'-y) - \lambda(x')$ attains its minimum (which is equal to $\lambda^c(y)$) at x. By differentiating both sides we obtain that

$$\nabla\lambda(x) = h'(x-y)\cdot \mathsf{Id},$$

which immediately reveals that

$$y = x - (\nabla h)^{-1}(\nabla\lambda(x)),$$

which in turn is (6.7).

It remains to be shown that the push-forward equation $\mathbb{Q} = T_*(\mathbb{P})$ holds true. This is non-trivial, however, some mystery is revealed by writing $\pi[\mathrm{d}x, \mathrm{d}y] = \mathbb{P}[\mathrm{d}x]\cdot \delta_{\{y=T(x)\}}[\mathrm{d}y]$. □

A complete proof may be found, as mentioned, in [56] and [1].

Remark 6.5 ($\mathbb{Q}^*$'s density). Let $\Omega = \mathbb{R}^m$, $\mathbb{P}$ ($\mathbb{Q}^*$, resp.) have Lebesgue density $f_{\mathbb{P}}$ ($f_{\mathbb{Q}^*}$, resp.) and let g be any random variable. Then

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^*}[g] &= \int g\mathrm{d}\mathbb{Q}^* \\
&= \int g\mathrm{d}\mathbb{P}\circ \mathrm{T}^{-1} \\
&= \int g(T)\,\mathrm{d}\mathbb{P} \\
&= \int g(T(x))\,f_{\mathbb{P}}(x)\,\mathrm{d}x \\
&= \int g(x)\,\frac{f_{\mathbb{P}}(T^{-1}x)}{|\det T'(T^{-1}(x))|}\mathrm{d}x,
\end{aligned}$$

($\det T'$ is the volume of the Jacobian) and on the other side

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^*}[g] &= \int g\mathrm{d}\mathbb{Q}^* \\
&= \int g(x)\,f_{\mathbb{Q}^*}(x)\,\mathrm{d}x
\end{aligned}$$

for all measurable functions g; we conclude that the respective densities coincide, that is

$$f_{\mathbb{Q}^*}(x) = \frac{f_{\mathbb{P}}(T^{-1}(x))}{|\det T'(T^{-1}(x))|}. \tag{6.8}$$

This is the bases to display distributions via their density in the sequel.

This observation exposes that the construction is in line with transforming a $\mathbb{R}^m$- valued random variable X, say, to $T\circ X$.

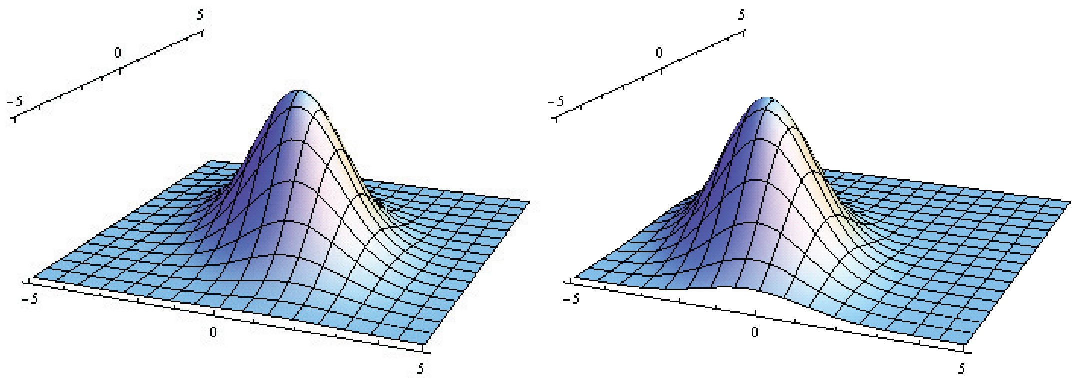

Figure 6.2.: For the expectation, the worst measure is a simple translate. Here, $Y(\omega_1, \omega_2) = \omega_1 + \omega_2$ and whence the direction of the translation is $-\omega_Y = -\frac{1}{\sqrt{2}}(1,1)$ for the Euclidean distance.

6.1.2. Expectation

Recall that the expectation is a concave functional as well, $\mathcal{A} := \mathbb{E}$, so it is included as well in Theorem 6.2. This was already established earlier, but the latter theorem expresses the optimal transport map in terms of Z_Y. The subgradient of the expectation is the simple function $Z_Y(\omega) := 1$, because obviously

$$\mathcal{A}\left(\tilde{Y}\right) \leq \mathcal{A}(Y) + \mathbb{E}\left[Z_Y\left(\tilde{Y} - Y\right)\right].$$

The transport map reduces to the simple translation $T(\omega) := \omega - K \cdot \omega_Y$ in this situation for all $1 \leq r < \infty$, which is exemplary depicted in Figure 6.2.

6.1.3. Distortions

For distortions we have elaborated so far that Z_Y is coupled in an antimonotone way with Y and moreover $Z_Y = h(U)$: We thus can give the dual variable as $Z_Y = h(G_Y(Y))$[2], and the transport map rewrites

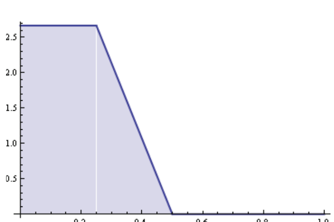

Figure 6.3.: Distortion function h.

[2]Recall that h is *de*creasing.

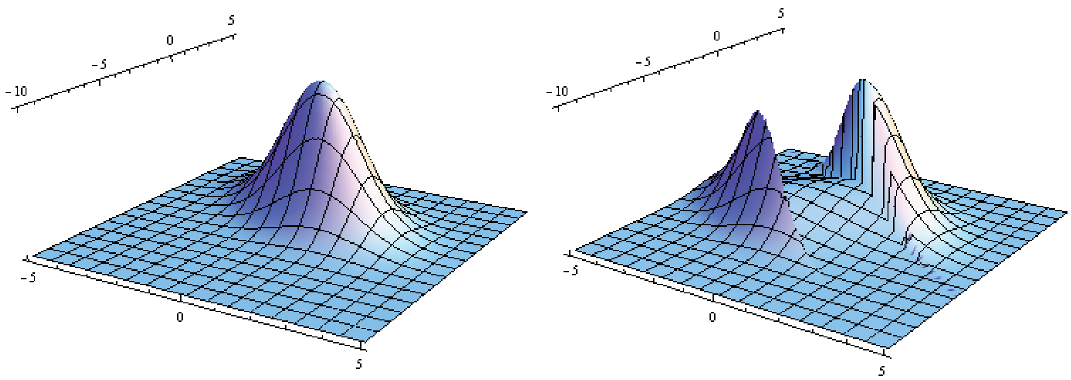

Figure 6.4.: Initial and distorted probability measure for h as in Fig. 6.3: 50 % stay at the same place, 25% of the mass is simply being shifted in direction $-\omega_Y$, and the remaining 25% are brutally distorted in between (Y as in the previous example).

$$T(\omega) = \omega - K \cdot \left| \frac{h(1 - G_Y(Y(\omega)))}{\|h\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \omega_Y$$

for non-negative distortion functions h. Together with (6.8) in Remark 6.5 this enables us to illustrate the geometry by plotting some densities, which we want to do here in providing some examples.[3]

6.1.4. The $\mathbb{A}$V@R

As for $\mathbb{A}\mathrm{V@R}_\alpha$ the optimal dual variable basically is $Z_Y = \mathbb{1}_{\left\{Y \le G_Y^{-1} \alpha\right\}}$. The transport map, again for all $1 \le r < \infty$, is

$$T(\omega) = \omega - \frac{K}{\alpha} \cdot \omega_Y \cdot \mathbb{1}_{\left\{Y < G_Y^{-1}(\alpha)\right\}}(\omega),$$

[3] Let $\mathbb{P}$ denote a multivariate normal probability measure with mean μ and covariance matrix Σ ($\mathbb{P} \sim \mathcal{N}(\mu, \Sigma)$) and Y a linear functional of the form $Y(\omega) = Y^\top \omega = \sum_i Y_i \omega_i$, then $\mathbb{P}^Y \sim \mathcal{N}\left(Y^\top \mu, Y^\top \Sigma Y\right)$, that is $\mathbb{P}^Y \sim \mathcal{N}\left(\sum_i Y_i \mu_i, \sum_{i,j} Y_i \Sigma_{i,j} Y_j\right)$; whence, $G_Y(y) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{y} e^{-\frac{1}{2}\frac{\left(x - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x$ and $G_Y(Y(\omega)) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{Y^\top \omega} e^{-\frac{1}{2}\frac{\left(x - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x$ is a $\mathbb{R}$-valued random variable.

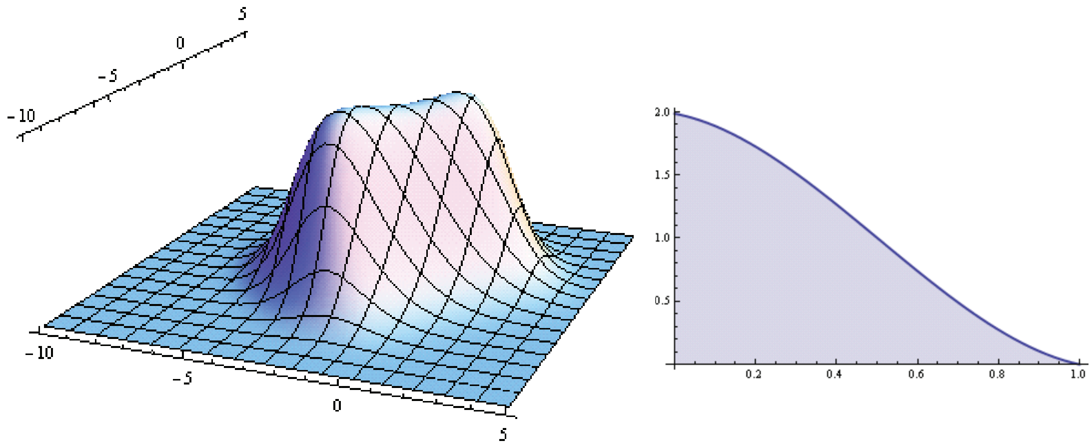

Figure 6.5.: Resulting probability distribution for the distortion function indicated in the second plot.

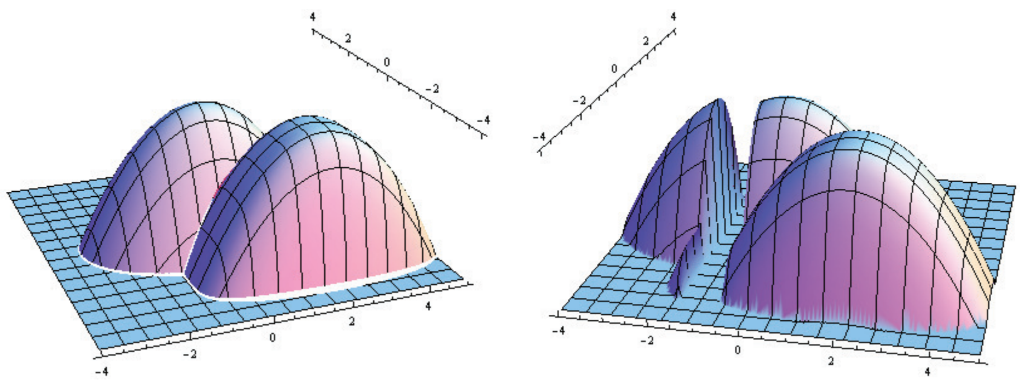

Figure 6.6.: Initial and split probability measure (with two modes), as it is worst with respect to the $\mathbb{AV}@\mathrm{R}$. Displayed from different perspectives.

which again includes the expectation for $\alpha = 1$. And the particular case $r = 1$ is included here naturally.

This transport map simply splits the sample space according the α-quantile: Those samples, which *do not* contribute to the computation of $\mathbb{AV}@\mathrm{R}_\alpha$ (which have quantile $\left(\left\{Y > G_Y^{-1}(\alpha)\right\}\right)$, are left unchanged on their place, while all other samples, which *do* contribute to the $\mathbb{AV}@\mathrm{R}_\alpha$ $\left(\left\{Y(\omega) < G_Y^{-1}(\alpha)\right\}\right)$, are being simply worsened by shifting them the distance $\frac{K}{\alpha}$; moreover, all of them are being shifted

- ▷ in *parallel*
- ▷ in the *same direction* $-\omega_Y$ and
- ▷ the *same distance* $\frac{K}{\alpha}$.

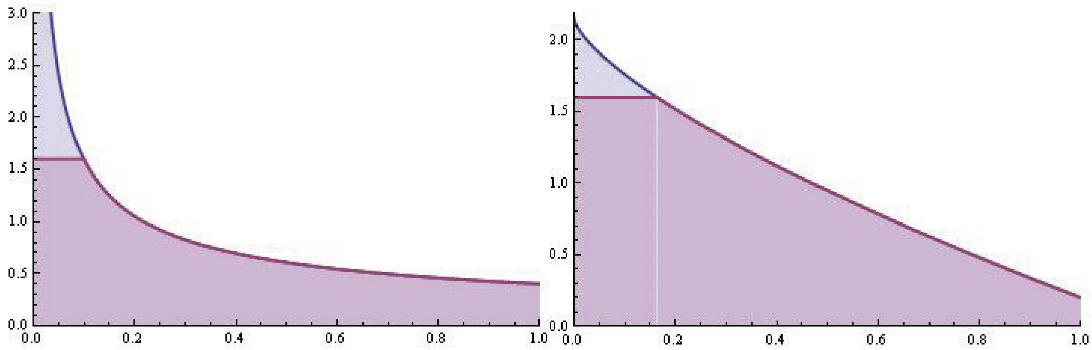

Figure 6.7.: Exemplary shape for a bounded, and an unbounded distortion h. The area under both charts is one.

6.2. The Distorted Functional $\mathcal{A}_H$

The optimal transport map is given by Theorem 6.2, provided that $r > 1$: that is to say to give the worst measure is not a difficulty at all, provided we compare the respective functionals in the Wasserstein metric.

As we have seen already the situation changes for the Kantorovich metric, as we have to require that

$$\mathbb{P}\left[|Z_Y| = \|Z_Y\|_\infty\right] > 0$$

in this situation. We shall continue the discussion at this point and elaborate the continuity properties for the Kantorovich distance further.

Recall that $Z_Y = h(U)$ for some uniform distribution U for the distorted acceptability functional. The latter condition $\mathbb{P}\left[\{|Z_Y| = \|Z_Y\|_\infty\}\right] > 0$ thus holds iff

$$\lambda\left\{|h| = \|h\|_\infty\right\} > 0,$$

where λ is the Lebesgue measure on [0, 1]: this, as we have seen above, is particularly true for $\mathbb{A}$V@R's distortion function $h_a = \frac{1}{\alpha}\mathbb{1}_{[0,\alpha]}$, and the optimal measure can be given in this situation as indicated.

We shall now further discuss the properties in the situation where h is

(i) unbounded, or

(ii) h is bounded, but not flat at its top.

It will turn out that the first problem is pretty easy, whereas the second involves tough mathematical results.

6.2.1. Unbounded Distortions

Theorem 6.6. *Suppose that h is not bounded and Y is linear on a linear space. Then the problem*

$$\begin{array}{rl} \textit{minimize} & \\ \textit{(in } \mathbb{Q}) & \mathcal{A}_{H;\mathbb{Q}}(Y) \\ \textit{subject to} & \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \le K \end{array}$$

is not bounded neither, i. e. the solution is $-\infty$*.*

Proof. Consider the measures

$$\mathbb{Q}_n := T_n \# \mathbb{P},$$

where the transport plans

$$T_n(\omega) := \omega - K \cdot \frac{\mathbb{1}_{\{|Z_Y| \ge n\}}(\omega)}{\mathbb{P}[|Z_Y| \ge n]} \cdot \operatorname{sign} Z_Y(\omega) \cdot \omega_Y$$

are given by cutting the (possibly sub-optimal) dual variable Z_Y.

As above notice that $\mathsf{d}_{KA}(\mathbb{Q}_n,\mathbb{P}) = K$, but

$$\begin{aligned} \mathcal{A}_{H,\mathbb{Q}_n}(Y) - \mathcal{A}_{H,\mathbb{P}}(Y) &\le -K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\{|Z_Y| \ge n\}}(\omega)}{\mathbb{P}[|Z_Y| \ge n]} \mathbb{P}[\mathrm{d}\omega] \\ &\le -K \cdot L(Y) \cdot n, \end{aligned}$$

and the problem thus does not allow a bounded (real) solution. □

6.2.2. Bounded Distortions

Theorem 6.7. *Let Y be a (continuous) linear functional on* $(\mathbb{R}^m, \|.\|)$*. Moreover assume that h is bounded, but* $\lambda\{|h| = \|h\|_\infty\} = 0$ *(cf. Figure 6.7). Then the problem*

$$\begin{array}{rl} \textit{minimize} & \\ \textit{(in } \mathbb{Q}) & \mathcal{A}_{H;\mathbb{Q}}(Y) \\ \textit{subject to} & \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \le K \end{array} \tag{6.9}$$

is bounded, but there does not *exist a measure* $\mathbb{Q}$ *with* $\mathsf{d}_{KA}(\mathbb{P},\mathbb{Q}) \le K$ *attaining the minimum in* (6.9)*, that is to say the respective* argmin*-set is empty.*

Remark 6.8. Notice, that the latter statement holds true on finite dimensional spaces, so there is no chance on infinite dimensional spaces neither to find a minimizing measure.

Proof. Define the set

$$C := \operatorname{argmin}\{\mathcal{A}_{H;\mathbb{Q}}(Y) : \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \le K\},$$

which is the argmin-set, consisting of all measures minimizing the problem (cf. (6.1))

$$\begin{array}{rl} \text{minimize} \\ \text{(in } \mathbb{Q}) & \mathcal{A}_{H;\mathbb{Q}}(Y) \\ \text{subject to} & \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \leq K \end{array}$$

in consideration.

In order to prove the statement by contradiction suppose that C were not empty. As we already know the minimum value of the problem precisely we may write

$$C = \{\mathbb{Q}\colon \mathcal{A}_{H;\mathbb{Q}}(Y) - \mathcal{A}_{H;\mathbb{P}}(Y) = -K \cdot L(Y) \cdot \|h\|_\infty, \, \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \leq K\}.$$

Further one may write

$$C = \bigcap_{n>1} C_n,$$

where the sets C_n originate from the relaxed problem

$$C_n = \left\{\mathbb{Q}\colon \mathcal{A}_{H;\mathbb{Q}}(Y) - \mathcal{A}_{H;\mathbb{P}}(Y) \leq -K \cdot L(Y) \cdot \left(\|h\|_\infty - \frac{1}{n}\right), \, \mathsf{d}_{KA}(\mathbb{Q},\mathbb{P}) \leq K\right\};$$

those sets C_n are certainly non-empty.

Consider the measures

$$\mathbb{Q}_n := (T_n)_*(\mathbb{P}),$$

defined via the transport maps

$$T_n(\omega) := \omega - K \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(\omega)}{\mathbb{P}\left[|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right]} \cdot \operatorname{sign} Z_Y(\omega) \cdot \omega_Y$$

by appropriately cutting the dual variable Z_Y at its top.

By the same reasoning as above they satisfy $\mathsf{d}_{KA}(\mathbb{Q}_n,\mathbb{P}) = K$ by construction, and

$$\begin{aligned} \mathcal{A}_{H,\mathbb{Q}_n}(Y) - \mathcal{A}_{H,\mathbb{P}}(Y) &\leq -K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(\omega)}{\mathbb{P}\left[|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right]} \mathbb{P}[\mathrm{d}\omega] \\ &\leq -K \cdot L(Y) \cdot \left(\|Z_Y\|_\infty - \frac{1}{n}\right), \end{aligned}$$

and thus $\mathbb{Q}_n \in C_n$.

As $(\mathbb{R}^m, \|.\|)$ is locally compact the space of continuous functions vanishing at infinity $C_0(\mathbb{R}^m, \|.\|)$ is a Banach space, which Riesz' theorem identifies with the space of regular Borel measures.

The probability measures $\mathbb{Q}_n$ may be considered themselves as elements of this dual via the setting

$$\begin{aligned} \mathbb{Q}_n\colon C(\mathbb{R}^m) &\to \mathbb{R} \\ \varphi &\mapsto \int \varphi \mathrm{d}\mathbb{Q}_n, \end{aligned}$$

but moreover

$$|\mathbb{Q}_n(\varphi)| \leq \int \|\varphi\|_\infty \,\mathrm{d}\mathbb{Q}_n = \|\varphi\|_\infty$$

for any function $\varphi \in C_0(\mathbb{R}^m, \|.\|)$, and thus $\|\mathbb{Q}_n\| \leq 1$: That is to say all those measures $\mathbb{Q}_n$ are within the unit ball $B_1(0)$ of the dual of $C_0(\mathbb{R}^m, \|.\|)$.

Alaoglu's theorem states that the closed unit ball $B_1(0)$ in the dual is weakly* compact, thus there is an accumulation point $\mathbb{Q} \in B_1(0)$ such that

$$\mathbb{Q}_{n_k} \to \mathbb{Q}$$

in the weak* topology for some sub-sequence $(n_k)_k$. Again by Riesz' theorem $\mathbb{Q}$ has a representation as a measure, although not necessarily as a probability measure.

We shall prove next that C is convex. This holds true, because

(i) the distance d_{KA} is convex for the situation $r = 1$ (cf. Lemma 1.2 and its subsequent remark), and

(ii) $\mathbb{Q} \mapsto \mathcal{A}_{H,\mathbb{Q}}(Y)$ is convex: to accept this consider $\mathbb{Q}_0 \in C$, $\mathbb{Q}_1 \in C$, define $\mathbb{Q}_\lambda := (1-\lambda)\mathbb{Q}_0 + \lambda\mathbb{Q}_1$ and observe that the distribution functions

$$\begin{aligned} G_{Y,\lambda}(z) : &= \mathbb{Q}_\lambda[Y \leq z] \\ &= (1-\lambda)\mathbb{Q}_0[Y \leq z] + \lambda\mathbb{Q}_1[Y \leq z] \\ &= (1-\lambda) G_{Y,0}(z) + \lambda G_{Y,1}(z) \end{aligned}$$

are convex-combinations. Whence

$$\begin{aligned} \mathcal{A}_{H,\mathbb{Q}_\lambda}(Y) &= \int G_{Y,\lambda}(z)\,\mathrm{d}H(z) \\ &= \int (1-\lambda) G_{Y,0}(z) + \lambda G_{Y,1}(z)\,\mathrm{d}H(z) \\ &= (1-\lambda)\int G_{Y,0}(z)\,\mathrm{d}H(z) + \lambda\int G_{Y,1}(z)\,\mathrm{d}H(z) \\ &= (1-\lambda)\mathcal{A}_{H,\mathbb{Q}_0}(Y) + \lambda\mathcal{A}_{H,\mathbb{Q}_0}(Y) \end{aligned}$$

is a convex combination as well.

So C is convex. By Mazur's theorem the norm-closure and its weak* closure coincide for convex sets,

$$\mathbb{Q} \in \overline{C}^{\text{weak}*} = \overline{C}^{\|\cdot\|},$$

we thus deduce in particular that

$$\|\mathbb{Q}\| = 1,$$

and the limiting measure $\mathbb{Q}$ thus is a probability measure.

Now define the increasing sets $\Omega_n := \left\{ |Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n} \right\}$. Observe that

$$\begin{aligned} \mathbb{Q}_n \left[\bigcup_j \Omega_j \right] &\ge \mathbb{Q}_n \left[\Omega_n \right], \\ &\ge \mathbb{P} \left[\left\{ |Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n} \right\} \right] \\ &= \lambda \left\{ |h| \le \|h\|_\infty - \frac{1}{n} \right\} \\ &\to 1 \end{aligned}$$

due to our assumptions, and particularly because $Z_Y = h(U)$. Whence $\mathbb{Q}_n \left[\bigcup_j \Omega_j \right] = 1$, and consequently $\mathbb{Q} \left[\bigcup_j \Omega_j \right] = 1$, because

$$\mathbb{Q}_{n_k} \to \mathbb{Q}$$

in the weak* topology. By construction (recall the definition of the transport map T_n), $\mathbb{Q}_n$ and $\mathbb{P}$ coincide on any Ω_n, so $\mathbb{Q}$ and $\mathbb{P}$ coincide on every set $A \subseteq \bigcup_k \Omega_k$. This, however, means $\mathbb{Q} = \mathbb{P}$, because

$$\mathbb{Q} \left[\bigcup_j \Omega_j \right] = \mathbb{P} \left[\bigcup_j \Omega_j \right] = 1.$$

This is a contradiction, because the measure $\mathbb{P}$ certainly is not optimal for the problem (6.9).

Whence, C is the empty set,

$$C = \emptyset,$$

and there is no optimal measure $\mathbb{Q}$ for the problem (6.9). □

Remark 6.9. As in the examples above we have again found measures such that

$$\mathbb{Q}_{n_k} \to \mathbb{P}$$

in the weak* topology and in norm, but

$$\mathbb{Q}_{n_k} \not\to \mathbb{P}$$

in the Wasserstein d_r distance. In particular

$$\mathsf{d}_r \left(\mathbb{Q}_n, \mathbb{P} \right) = 1,$$

and

$$\mathcal{A}_{H,\mathbb{Q}_n}(Y) - \mathcal{A}_{H,\mathbb{P}}(Y) \le -K \cdot L(Y) \cdot \left(\|h\|_\infty - \frac{1}{n} \right) < 0$$

and $\mathbb{Q} \mapsto \mathcal{A}_{H;\mathbb{Q}}(Y)$ is d_r-continuous.

6.3. Implications For Investment Strategies

We have elaborated in previous sections that the quantity

$$\inf_{Z\in\mathcal{A}_H(Y)} \|Z\|_{r'}$$

does not depend on Y in a lot of situations, for example for all distortion acceptability functionals, as

$$\inf_{Z\in\mathcal{A}(Y)} \|Z\|_{r'} = \|h\|_{r'}$$

(including the $\mathbb{AV@R}$).

This fact is quite remarkable for portfolio optimization as we want to elaborate now.

To this end consider again a linear space $\Omega = \mathbb{R}^m$, where any $\omega \in \Omega$ may be interpreted as a return-vector

$$\omega = (\omega_s)_{s=1}^m = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{pmatrix}$$

of m stocks. Any subset of potential returns is assigned a probability by the measure $\mathbb{P}$ on Ω.

A random variable $Y \in \Omega^*$ (Y linear) is just an investment strategy, as

$$Y(\omega) = \sum_{s=1}^m Y_s \cdot \omega_s$$

represents the return of the portfolio, any of the stocks s being weighted with the respective exposure Y_s according the portfolio decomposition. A usual budget constraint then is given as

$$Y(\mathbb{1}) = 1,$$

which means that the total investment in the set of m stocks represents 100 % of the budget available.

In case short-selling has to be excluded as well, then the additional conditions read

$$Y(e_s) \geq 0 \qquad (s \in \{1, 2, \dots m\}),$$

where e_s is the s^{th}-unit vector in $\mathbb{R}^m$, representing stock s.

Notice, that within this setting a typical portfolio optimization problem reads

$$\begin{array}{rl} \text{maximize} \\ \text{(in } Y\text{)} & \mathbb{E}[Y] \\ \text{subject to} & Y \in \Omega^* \text{ linear}, \\ & \mathcal{A}(Y) \geq q, \\ & Y(\mathbb{1}) = 100\%, \\ & (Y(e_s) \geq 0) \end{array}$$

where no more risk than q is accepted, and risk is being measured by $\mathcal{A}$.

In a typical situation the measure $\mathbb{P}$ can be observed from the past, and it is available as an empirical measure (for a convex approximation approach of this or similar problems (chance constraints) we would like to refer the reader to Campi [12]: work, which is built – at least partially – on [11]). To base an investment decision only on this measure would be somewhat myopic, because close deviations or similar observations could have appeared as likely as the observations themselves (cf. [27]). For this it is evident that measures, which are close, should be taken into account as well.

The next theorem elaborates, that accepting more measures will finally lead to a simple investment decision, which consists of equally weighting all exposures in the portfolio. This is contained in the next corollary for the choice

$$e = \mathbb{1},$$

as in this situation

$$\frac{\mathbf{HB}_{\mathbb{1}}}{\|\mathbb{1}\|} = \frac{\mathbb{1}^\top}{m}$$

due to (2.9) in the introductory example. But the assignment $\frac{\mathbb{1}^\top}{m}$ just means equal weights for any of the stocks.

In preparation it should be mentioned that $\frac{\mathbf{HB}_{\mathbb{1}}}{\|\mathbb{1}\|}$ is the solution of the following optimization problem:

Lemma 6.10. *For $e \neq 0$ the problem*

$$\begin{array}{rl} \text{minimize} & \\ \text{(in } Y) & \|Y\| \\ \text{subject to} & Y \in \Omega^* \text{ linear}, \\ & Y(e) = 1 \end{array}$$

has objective value $\frac{1}{\|e\|}$, which is attained at $\tilde{Y} = \frac{\mathbf{HB}_e}{\|e\|}$.

Proof. Notice, that

$$1 = Y(e) \leq \|Y\| \, \|e\|,$$

and whence $\|Y\| \geq \frac{1}{\|e\|}$.

However, $\left\|\tilde{Y}\right\| = \left\|\frac{\mathbf{HB}_e}{\|e\|}\right\| = \frac{1}{\|e\|}$, and so the assertion is immediate. □

Corollary 6.11 (It is asymptotically optimal to equally invest in all available assets)**.** *Consider $\mathbb{R}^m$, equipped with the norm $\|x\| := \left(\sum_{s=1}^m |x_s|^p\right)^{\frac{1}{p}}$, let $x \mapsto \mathbf{HB}_x$ be continuous in a neighbourhood of $\mathbb{1}$ and suppose that*

$$\inf_{Z \in \mathcal{A}(Y)} \|Z\|_{r'}$$

is independent of Y. Then

$$\lim_{K\to\infty} \operatorname*{argmax}_{Y(\mathbb{1})=1} \min_{\mathsf{d}_r(\mathbb{P},\mathbb{Q})\le K} \mathcal{A}_{\mathbb{Q}}(Y) = \frac{\mathbb{1}^\top}{m},$$

where the maximum is over all linear functionals Y satisfying $Y(\mathbb{1}) = 1$.

We shall prove the statement in slightly some more generality:

Corollary 6.12. *Let $x \mapsto \mathbf{HB}_x$ be continuous in a neighbourhood of $0 \neq e \in \Omega$ and suppose that*

$$\inf_{Z\in\mathcal{A}(Y)} \|Z\|_{r'}$$

is independent of Y. Then

$$\lim_{K\to\infty} \operatorname*{argmax}_{Y(e)=1} \min_{\mathsf{d}_r(\mathbb{P},\mathbb{Q})\le K} \mathcal{A}_{\mathbb{Q}}(Y) = \frac{\mathbf{HB}_e}{\|e\|},$$

where the maximum is over all linear functionals Y satisfying $Y(e) = 1$.

Proof. Recall first that the inner problem

$$\min_{\mathsf{d}_r(\mathbb{P},\mathbb{Q})\le K} \mathcal{A}_{\mathbb{Q}}(Y)$$

has the (finite) solution

$$\mathcal{A}_{\mathbb{P}}(Y) - K \cdot L(Y) \cdot \inf_{Z\in\partial\mathcal{A}(Y)} \|Z\|_{r'} = \mathcal{A}_{\mathbb{P}}(Y) - K' \cdot L(Y)$$

for some constant $K' := K \cdot \inf_{Z\in\partial\mathcal{A}(Y)} \|Z\|_{r'}$. The outer problem

$$\max_{Y(e)=1} \min_{\mathsf{d}_r(\mathbb{P},\mathbb{Q})\le K} \mathcal{A}_{\mathbb{Q}}(Y) = \max_{Y(e)=1} \mathcal{A}_{\mathbb{P}}(Y) - K' \cdot L(Y)$$

has Lagrangian

$$L(Y,\lambda) = \mathcal{A}_{\mathbb{P}}(Y) - K' \cdot L(Y) - \lambda(1 - Y(e));$$

In the saddle point $\left(\tilde{Y}, \tilde{\lambda}\right)$ necessarily the derivative of the Lagrangian with respect to any direction h has to vanish, that is the equations

$$\begin{aligned} 0 &= L'\left(\tilde{Y}, \tilde{\lambda}\right)(h) \\ &= \mathcal{A}'_{\mathbb{P}}\left(\tilde{Y}\right)(h) - K' \cdot \mathbf{HB}_{\tilde{Y}}(h) + \tilde{\lambda} \cdot h(e) \\ &= \mathbb{E}\left[h \cdot Z_{\tilde{Y}}\right] - K' \cdot \mathbf{HB}_{\tilde{Y}}(h) + \tilde{\lambda} \cdot h(e), \\ 1 &= \tilde{Y}(e) \end{aligned}$$

have to hold *simultaneously* for all directions h, where $\tilde{Y}$ denotes any optimal solution and $Z_{\tilde{Y}}$ the random variable minimizing the dual problem

$$\mathcal{A}_{\mathbb{P}}\left(\tilde{Y}\right) = \inf_{A(Z)>-\infty} \mathbb{E}\left[\tilde{Y} \cdot Z\right] - A(Z)$$

for this optimal choice $\tilde{Y}$.

Then define the $Z_{\tilde{Y}}$-barycenter of $\mathbb{P}$, $\mu_{Z_{\tilde{Y}}} := \mathbb{E}_{\mathbb{P}}\left[\mathsf{Id} \cdot Z_{\tilde{Y}}\right]$, and observe that

(i) $\mu_{Z_{\tilde{Y}}}$ exists, as

$$\begin{aligned}\left\|\mu_{Z_{\tilde{Y}}}\right\| &= \left\|\mathbb{E}_{\mathbb{P}}\left[\mathsf{Id}\cdot Z_{\tilde{Y}}\right]\right\| \\ &\leq \left(\int \|x\|^r \mathbb{P}\left[\mathrm{d}x\right]\right)^{\frac{1}{r}} \left(\int \left|Z_{\tilde{Y}}\right|^{\frac{r}{r-1}}\right)^{\frac{r-1}{r}} \\ &= \left(\int d\left(0,x\right)^r \mathbb{P}\left[\mathrm{d}x\right]\right)^{\frac{1}{r}} \cdot \left\|Z_{\tilde{Y}}\right\|_{\frac{r}{r-1}} \\ &< \infty,\end{aligned}$$

independently of Y.

(ii) Moreover notice that

$$\mathbb{E}_{\mathbb{P}}\left[h\cdot Z_{\tilde{Y}}\right] = h\left(\mu_{Z_{\tilde{Y}}}\right),$$

as h is linear.

Then choose $\lambda_K > 0$ so that $K = \|\mu_{Z_Y} + \lambda_K \cdot e\|$ (which is certainly possible if K is big enough, say $K > \|\mu_{Z_Y}\|$) and define

$$Y_K^* := \frac{\mathbf{HB}_{\mu_{Z_Y}+\lambda_K\cdot e}}{\mathbf{HB}_{\mu_{Z_Y}+\lambda_K\cdot e}\left(e\right)}.$$

The pair (Y_K^*, λ_K) solves the Lagrangian as well, it has the desired properties, because

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}}\left[h\cdot Z_Y\right] - K\cdot\mathbf{HB}_{Y_K^*}\left(h\right) + \lambda_K\cdot h\left(e\right) = \\ &\qquad = h\left(\mu_{Z_Y}\right) - K\cdot\frac{h\left(\mu_{Z_Y}+\lambda_K\cdot e\right)}{\|\mu_{Z_Y}+\lambda_K\cdot e\|} + \lambda_K\cdot h\left(e\right) \\ &\qquad = h\left(\mu_{Z_Y}+\lambda_K\cdot e\right) - K\cdot\frac{h\left(\mu_{Z_Y}+\lambda_K\cdot e\right)}{\|\mu_{Z_Y}+\lambda_K\cdot e\|} \\ &\qquad = 0,\end{aligned}$$

where we have used that $\mathbf{HB}_{\mathbf{HB}_x}\left(h\right) = \frac{h(x)}{\|x\|}$ (cf. (2.8)).

Letting $K \to \infty$ finally gives that $\lambda_K \to \infty$, and

$$\begin{aligned}\frac{\mathbf{HB}_{\mu_{Z_Y}+\lambda_K\cdot e}}{\mathbf{HB}_{\mu_{Z_Y}+\lambda_K\cdot e}\left(e\right)} &= \frac{\mathbf{HB}_{\frac{\mu_{Z_Y}}{\lambda_K}+e}}{\mathbf{HB}_{\frac{\mu_{Z_Y}}{\lambda_K}+e}\left(e\right)} \\ &\xrightarrow[K\to\infty]{} \frac{\mathbf{HB}_e}{\mathbf{HB}_e\left(e\right)} \\ &= \frac{\mathbf{HB}_e}{\|e\|},\end{aligned}$$

because we assume $x \mapsto \mathbf{HB}_x$ continuous in a neighbourhood of e. □

7. Optimal Transport, Revisited

We have seen in previous sections that discrete measures are dense with respect to the Wasserstein metric. So it is intuitively clear that mass-points cannot be shifted to an arbitrary place without significantly increasing the distance to the initial measure.

The next theorem reveals that in the modified measure, there will be again a mass-point close to another mass-point in the original distribution, and its maximal transportation distance is given. The result is a by-product of previous investigations in this work.

Theorem 7.1. *Consider the discrete measures* $\mathbb{P} := \sum_s \mathbb{P}_s \cdot \delta_{\omega_s}$ *and* $\mathbb{Q} = \sum_t \mathbb{Q}_t \cdot \delta_{\tilde{\omega}_t}$. *Then, for every* s *with* $\mathbb{P}_s > 0$ *there exists* t *with* $\mathbb{Q}_t > 0$ *such that*

$$d(\omega_s, \tilde{\omega}_t) \le \frac{\mathsf{d}_r(\mathbb{P}, \mathbb{Q})}{\sqrt[r]{\mathbb{P}_s}}.$$

Proof. Problem (6.2) may be restated as

$$\begin{array}{ll} \underset{(\text{in } \pi)}{\text{minimize}} & \sum_t \mathbb{Q}_t \cdot Y(\tilde{\omega}_t) \qquad (= \mathbb{E}_{\mathbb{Q}}[Y]) \\ \text{subject to} & \sum_{s,t} \pi_{s,t} d(\omega_s, \tilde{\omega}_t)^r \le K^r, \\ & \sum_{s,t} \pi_{s,t} = 1, \\ & \sum_s \pi_{s,t} = \mathbb{Q}_t, \\ & \sum_t \pi_{s,t} = \mathbb{P}_s, \\ & \pi_{s,t} \ge 0. \end{array}$$

The marginal measure $\mathbb{Q}_t$ is given by the transport plan, the problem thus rewrites

$$\begin{array}{ll} \underset{(\text{in } \pi)}{\text{minimize}} & \sum_{s,t} \pi_{s,t} \cdot Y(\tilde{\omega}_t) \\ \text{subject to} & \sum_{s,t} \pi_{s,t} = 1, \\ & \sum_t \pi_{s,t} = \mathbb{P}_s, \\ & \sum_{s,t} \pi_{s,t} d(\omega_s, \tilde{\omega}_t)^r \le K^r, \\ & \pi_{s,t} \ge 0. \end{array}$$

This is a linear program in π, its dual thus is

$$\begin{array}{ll} \underset{(\text{in } \gamma, \lambda_s, \mu)}{\text{maximize}} & \gamma + \sum_s \mathbb{P}_s \lambda_s - \mu K^r \\ \text{subject to} & \gamma + \lambda_s - \mu d(\omega_s, \tilde{\omega}_t)^r \le Y(\tilde{\omega}_t), \\ & \mu \ge 0. \end{array}$$

Notice, that both, γ and λ_s are completely free, and as $\sum_s \mathbb{P}_s = 1$ we may replace $\gamma \leftarrow \gamma - c$ and at the same time all $\lambda_s \leftarrow \lambda_s + c$ without affecting the constraints, nor changing the objective for any arbitrary constant c. So we may in particular eliminate γ and simplify the dual as

$$\begin{array}{rl} \text{maximize} & \\ (\text{in } \lambda_s, \mu) & \sum_s \mathbb{P}_s \lambda_s - \mu K^r \\ \text{subject to} & \lambda_s - \mu d\left(\omega_s, \tilde{\omega}_t\right)^r \leq Y\left(\tilde{\omega}_t\right), \\ & \mu \geq 0. \end{array} \tag{7.1}$$

Now note that the constraints force

$$\mu \geq \max_{\omega_s \neq \tilde{\omega}_t} \frac{\lambda_s - Y\left(\tilde{\omega}_t\right)}{d\left(\omega_s, \tilde{\omega}_t\right)^r}$$

for any index s.

We may fix an index s_0 now and at the same time choose λ_{s_0} big enough, such that equality is obtained for this particular index, that is

$$\mu = \max_{\omega_{s_0} \neq \tilde{\omega}_t} \frac{\lambda_{s_0} - Y\left(\tilde{\omega}_t\right)}{d\left(\omega_{s_0}, \tilde{\omega}_t\right)^r},$$

and moreover such that μ is feasible, that is $\mu \geq 0$.

Recall the objective

$$\begin{aligned} \sum_s \mathbb{P}_s \lambda_s - \mu K^r &\geq \sum_s \mathbb{P}_s \lambda_s - K^r \max_{\tilde{\omega}_{s_0} \neq \omega_t} \frac{\lambda_{s_0} - Y\left(\tilde{\omega}_t\right)}{d\left(\omega_{s_0}, \tilde{\omega}_t\right)^r} \\ &\geq \lambda_{s_0}\left(\mathbb{P}_{s_0} - K^r \max_{\tilde{\omega}_{s_0} \neq \omega_t} \frac{\lambda_{s_0} - Y\left(\omega_t\right)}{d\left(\tilde{\omega}_{s_0}, \omega_t\right)^r}\right). \end{aligned}$$

Letting $\lambda_{s_0} \to \infty$, the objective is growing with slope

$$\mathbb{P}_{s_0} - \max_{\omega_{s_0} \neq \tilde{\omega}_t} \frac{K^r}{d\left(\omega_{s_0}, \tilde{\omega}_t\right)^r}.$$

If this latter quantity were positive, then the objective tends to $+\infty$ as well – this, however, is impossible, as the objective is bound (for example) by $\mathbb{E}_{\mathbb{P}}[Y]$. The quantity thus is non-positive, that is

$$\mathbb{P}_{s_0} - \max_{\omega_s \neq \tilde{\omega}_t} \frac{K^r}{d\left(\omega_{s_0}, \tilde{\omega}_t\right)^r} \leq 0.$$

To put this in different words: For any s_0

$$\min_{\omega_{s_0} \neq \tilde{\omega}_t} d\left(\omega_{s_0}, \tilde{\omega}_t\right)^r \leq \frac{K^r}{\mathbb{P}_{s_0}},$$

whence there is an index t such that

$$d\left(\omega_{s_0}, \tilde{\omega}_t\right) \leq \frac{K}{\sqrt[r]{\mathbb{P}_{s_0}}},$$

which is the desired assertion. $\square$

Part II.

Asymptotics Of Quantizers

8. Asymptotic Quantizers

The initial chapters of this theses have been dedicated to elaborating that discrete measures are dense in the Wasserstein metric. Then very concrete continuity properties have been investigated, in particular continuity properties of acceptability functionals with respect to the Wasserstein distance.

For potential computational purposes we want to combine these efforts now and elaborate further on the distances of a given measure and a discrete approximation. We want to know, how many mass-points are needed, and the appropriate weights assigned to these points in order to get a sufficiently good approximation of the probability measure in consideration. As the quality of the approximation increases, the required mass-points will increase as well – in this situation we are of course interested in the asymptotics and clarify for example, how many mass-points are asymptotically necessary to obtain a desired approximation quality.

To this end we will judge some concrete results from the literature, and add some further results.

8.1. Quantization Preliminaries

Quantization has been studied in detail by Graf and Luschgy in [21]; another very compelling reference is the book by Rachev and Rüschendorf, [45].

Definition 8.1 (Terminology)**.** Tessellation and Voronoi[1]-Tessellation.

- ▹ A finite collection of measurable sets $(\Omega_\omega)_{\omega\in\mathcal{Q}}$ is a $\mathbb{P}$-*tessellation* of Ω, provided that
  - (i) $\mathbb{P}\left[\bigcup_{\omega\in\mathcal{Q}}\Omega_\omega\right]=1$ and
  - (ii) $\mathbb{P}\left[\Omega_\omega\cap\Omega_{\omega'}\right]=0$ for $\omega\neq\omega'$.
- ▹ A $\mathbb{P}$-*tessellation* $(\Omega_\omega)_\omega$ is a *Voronoi-tessellation* provided that $d(x,\omega)\leq d(x,\omega')$ for all $x\in\Omega_\omega$ and $\omega'\in\mathcal{Q}$.
- ▹ A finite measure $\mathbb{P}^{\mathcal{Q}}:=\sum_{\omega\in\mathcal{Q}}p_\omega\delta_\omega$ is called a quantizer.

Remark 8.2. As already anticipated in the notation just introduced we usually consider the index set $\mathcal{Q}$ itself a finite collection of samples from Ω, $\mathcal{Q}\subseteq\Omega$; and moreover we shall assume $\omega\in\Omega_\omega$, that is to say the set $\mathcal{Q}$ simply collects some representative elements of any of the fragments Ω_ω.

[1]Georgi Feodosjewitsch Woronoi (1868 - 1908) was a Russian mathematician, born in Ukrain.

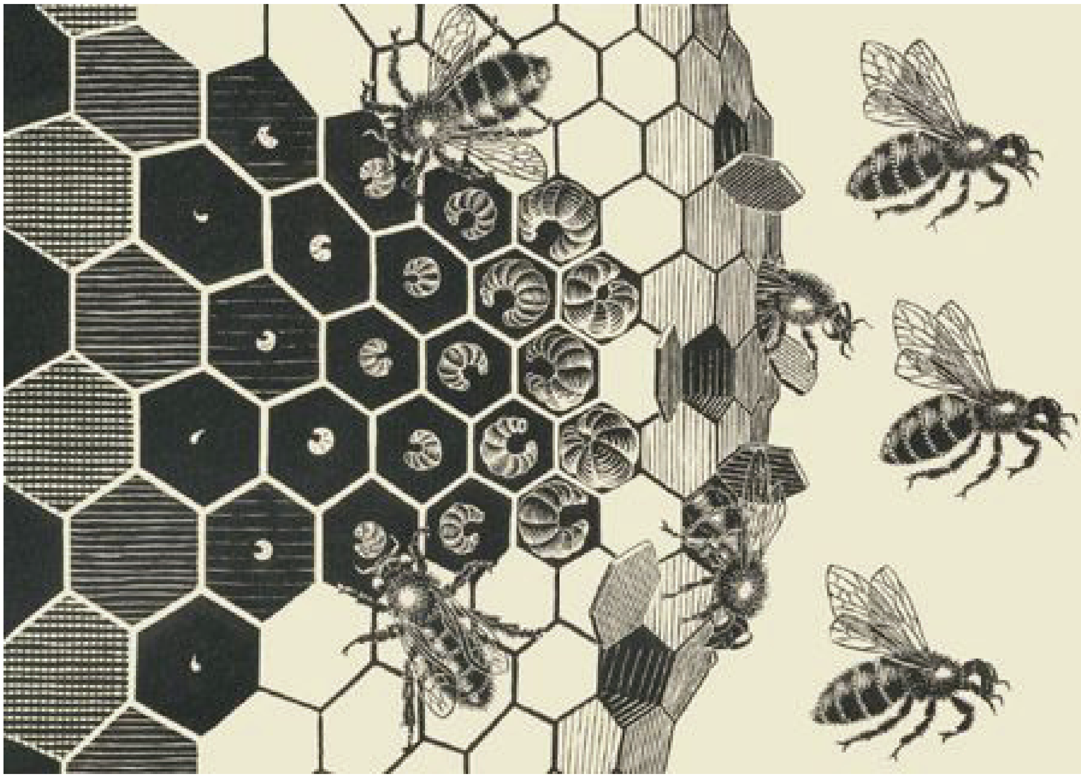

Figure 8.1.: M. C. Escher (1898 - 1972): Metamorphose II. Quantizers with the ambition for higher dimensions.

Remark 8.3. Particularly in situations where

$$\mathbb{P}^{\mathcal{Q}} = \sum_{\omega\in\mathcal{Q}} p_\omega \delta_\omega = \sum_{\omega\in\mathcal{Q}} \mathbb{P}\left[\Omega_\omega\right] \delta_\omega$$

for an appropriate tessellation $(\Omega_\omega)_{\omega\in\mathcal{Q}}$, then $\mathcal{Q}$ and $\mathbb{P}^{\mathcal{Q}}$ are simultaneously called a quantizer.

Lemma 8.4. *Let $\mathcal{Q} \subseteq \Omega$ be a finite set and consider the measure $\mathbb{P}_{\mathcal{Q}} := \sum_{\omega\in\mathcal{Q}} p_\omega \delta_\omega$ (with $\sum_{q\in\mathcal{Q}} p_q = 1$). Then the following holds true:*

(i) Let $(\Omega_q)_{q\in\mathcal{Q}}$ be a tessellation of Ω and $p_q := \mathbb{P}\left[\Omega_q\right]$, then

$$\mathsf{d}_r \left(\mathbb{P}, \mathbb{P}^{\mathcal{Q}}\right)^r \le \sum_{q\in\mathcal{Q}} \int_{\Omega_q} d\left(\omega,\, q\right)^r \mathbb{P}\left[\mathrm{d}\omega\right];$$

(ii) If $(\Omega_q)_q$ is a Voronoi-tessellation and $p_q = \mathbb{P}\left[\Omega_q\right]$, then

$$\mathsf{d}_r \left(\mathbb{P}, \mathbb{P}_{\mathcal{Q}}\right)^r = \int \min_{q\in\mathcal{Q}} d\left(\omega, q\right)^r \mathbb{P}\left[\mathrm{d}\omega\right]. \tag{8.1}$$

Remark 8.5. Given a finite number of support points $\mathcal{Q}$, then the weights $p_q = \mathbb{P}[\Omega_q]$ – $(\Omega_q)_q$ the corresponding Voronoi tessellation – are the *optimal choice* to approximate $\mathbb{P}$ by a finite measure with respect to the Wasserstein distance d_r for *any* $r \geq 1$; the latter Lemma states moreover, that the minimal distance may be evaluated explicitly by computing (8.1).

Notice that the problem of finding good approximations thus reduces to finding good locations, that is to minimize

$$(z_s)_{s=1}^n \mapsto \int \min_s d(\omega, z_s)^r \, \mathbb{P}[\mathrm{d}\omega].$$

This problem, however, is not convex and not trivial in general.

Proof. Let π be any probability measure on $\Omega \times \Omega$ with marginals $\mathbb{P}$ and $\mathbb{P}^{\mathcal{Q}}$. Then

$$\begin{aligned}
&\int_{\Omega\times\Omega} d(\omega, \tilde{\omega})^r \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}] \\
&\qquad = \int_{\Omega\times Q} d(\omega, \tilde{\omega})^r \, \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}] \\
&\qquad \geq \int_{\Omega\times Q} \min_{q\in\mathcal{Q}} d(\omega, q)^r \, \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}] \\
&\qquad = \int_{\Omega} \min_{q\in\mathcal{Q}} d(\omega, q)^r \, \mathbb{P}[\mathrm{d}\omega].
\end{aligned}$$

To prove the 1^{st} assertion define the transport plan

$$\pi[A\times B] := \sum_{q\in\mathcal{Q}\cap B} \mathbb{P}[A\cap\Omega_q] = \sum_{q\in\mathcal{Q}} \mathbb{P}[A\cap\Omega_q]\,\delta_q[B]$$

and observe the marginals

$$\pi[\Omega\times B] = \sum_{q\in\mathcal{Q}} \mathbb{P}[\Omega_q]\,\delta_q[B] = \sum_{q\in\mathcal{Q}} p_q\delta_q[B] = \mathbb{P}^{\mathcal{Q}}[B]$$

and

$$\pi[A\times\Omega] = \sum_{q\in\mathcal{Q}} \mathbb{P}[A\cap\Omega_q] = \mathbb{P}[A],$$

as Ω_q form a tessellation. Moreover, $\pi[\mathrm{d}x\times\{q\}] = \begin{cases}\mathbb{P}[\mathrm{d}x\cap\Omega_q] & \text{if } q\in\mathcal{Q}\\ 0 & \text{if } q\notin\mathcal{Q}\end{cases}$. Thus,

$$\begin{aligned}
\mathsf{d}_r\left(\mathbb{P}, \mathbb{P}^{Q}\right)^r &\leq \int\int_{\Omega\times\Omega} d(x,q)^r \, \pi[\mathrm{d}x, \mathrm{d}q] \\
&= \sum_{q\in\mathcal{Q}} \int_{\Omega} d(x,q)^r \, \mathbb{P}[\mathrm{d}x\cap\Omega_q] \\
&= \sum_{q\in\mathcal{Q}} \int_{\Omega_q} d(x,q)^r \, \mathbb{P}[\mathrm{d}x].
\end{aligned}$$

□

The proof of the latter lemma considerably simplifies in case $r = 1$, as the function $x \mapsto \min_{q \in \mathcal{Q}} d(x, q)$ is Lipschitz-continuous with Lipschitz constant 1. In this situation the Theorem of Kantorovich-Rubinstein provides a significant simplification (cf. [44]).

8.2. Asymptotic Quantizers

The notation in the literature – probably for historical changes – is not unique, different authors sometimes use even conflicting notations. However, there seems to be a certain trend, or let's say tendency, to unify the notation and we adopt the symbols used here to comply with recent works in the area. For the sake of completeness and to simplify further reading we give the symbols used in different occurrences as well.

Definition 8.6 (Quantization)**.** The following characteristics will be essential:

- $e_{n,r}(\mathbb{P}) := \inf_{|\mathcal{Q}| \le n} \mathsf{d}_r\left(\mathbb{P}, \mathbb{P}^{\mathcal{Q}}\right)$, the infimum being computed over all quantizers $\mathbb{P}^{\mathcal{Q}}$ with not more than n atoms, $|\mathcal{Q}| \le n$ (other authors use the term n^{th}-*quantization error* for $\mathbb{P}$ of order r and the definition $V_{n,r}(\mathbb{P}) := e_{n,r}(\mathbb{P})^r$.)
- $J_{r,m} := \inf_{n \ge 1} n^{\frac{r}{m}} \cdot e_{n,r}\left(U[0,1]^m\right)^r$, where $U([0,1]^m)$ is the m-dimensional, uniform distribution on the unit cube;
- $Q_r(\mathbb{P}) := \lim_{n \to \infty} n^{\frac{r}{d}} \cdot e_{n,r}(\mathbb{P})^r$ is called r^{th}-*quantization coefficient* of the probability measure $\mathbb{P}$; notice, that $J_{r,m} \le Q_r\left(U[0,1]^m\right)$; however, below both quantities will be found to coincide.

The following stability result holds:

Lemma 8.7. $|e_{n,r}(\mathbb{P}_1) - e_{n,r}(\mathbb{P}_2)| \le \mathsf{d}_r(\mathbb{P}_1, \mathbb{P}_2)$.

Proof. Choose an optimal $\mathcal{Q}$ minimizing $\mathsf{d}_r\left(\mathbb{P}_2, \mathbb{P}^{\mathcal{Q}}\right)$ and let π denote the respective optimal, bi-variate measure. Then

$$\begin{aligned} e_{n,r}(\mathbb{P}_1) - e_{n,r}(\mathbb{P}_2) &= e_{n,r}(\mathbb{P}_1) - \mathsf{d}_r\left(\mathbb{P}_2, \mathbb{P}^{\mathcal{Q}}\right) \\ &\le \mathsf{d}_r\left(\mathbb{P}_1, \mathbb{P}^{\mathcal{Q}}\right) - \mathsf{d}_r\left(\mathbb{P}_2, \mathbb{P}^{\mathcal{Q}}\right) \\ &\le \mathsf{d}_r(\mathbb{P}_1, \mathbb{P}_2) \end{aligned}$$

by the triangle inequality. Revering the role of $\mathbb{P}_1$ and $\mathbb{P}_2$ proves the Lemma. □

Some first results on the speed of convergence have been achieved by [15]. In the center of quantization, however, is Zador's result, which is contained in [60]:

Theorem 8.8 (Zador-Gersho formula)**.** *Let $\mathbb{P}$ satisfy the moment-like condition $\int \|x\|^{r+\eta}\, \mathbb{P}\,[\mathrm{d}x]$ for some $\eta > 0$. Then*

$$Q_r\,(\mathbb{P}) = J_{r,m} \cdot \left\| \frac{\mathrm{d}\mathbb{P}_a}{\mathrm{d}\lambda^m} \right\|_{\frac{m}{m+r}},$$

where $\mathbb{P}_a$ denotes the absolutely continuous part of $\mathbb{P}$ with respect to the Lebesgue measure λ^m on $\mathbb{R}^m$.

Proof. A proof in six steps of this fundamental result on the asymptotic quantization error is provided in [21], Theorem 6.2. □

Remark 8.9. Some remarks seem to be appropriate:

(i) The theorem links the quantization of an arbitrary, m-dimensional distribution $\mathbb{P}$ with the problem of quantization of the uniform distribution on the unit cube $[0,1]^m$;

(ii) The norm of the initial space finds its way into $J_{r,m}$, but it is *not* reflected in the quantity $\left\| \frac{\mathrm{d}\mathbb{P}_a}{\mathrm{d}\lambda^m} \right\|_{\frac{m}{m+r}}$.

(iii) In particular notice that the theorem provides that

$$\begin{aligned} Q_r\left([0,1]^m\right) &:= \lim_{n\to\infty} n^{\frac{r}{m}} \cdot e_{n,r}\left([0,1]^m\right)^r \\ &= \inf_{n\geq 1} n^{\frac{r}{m}} \cdot e_{n,r}\left(U\,[0,1]^m\right)^r \\ &=: J_{r,m}. \end{aligned}$$

(iv) Moreover, $\frac{m}{m+r} < 1$ – contrary to the notation – $\|.\|_{\frac{m}{m+r}}$ does *not* represent a norm, because it does not satisfy the triangle inequality; however, this provides a comprehensive and convenient notation.

Theorem 8.10. *The r^{th}-quantization coefficient satisfies*

$$1 \leq \frac{Q_r\left([0,1]^m\right)}{M_r\left(B\,(0,1)\right)} \leq \Gamma\left(2 + \frac{r}{m}\right), \tag{8.2}$$

where $B\,(0,1)$ is the unit ball of the underlying norm, and

$$M_r\left(B\,(0,1)\right) = \frac{m}{(m+r)\,\lambda^m\left(B\,(0,1)\right)^{\frac{r}{m}}}.$$

Proof. As for the proof we refer again to [21], Proposition 8.3 and Proposition 9.3. □

Remark 8.11 (Volume of the unit cube).

▷ The Lebesgue-volume of the d-dimensional unit ball with respect to the equally-weighted l^p-norm $x \mapsto \left(\sum |x_i|^p\right)^{\frac{1}{p}}$ is $\lambda\left(B\,(0,1)\right) = \frac{\left(2\Gamma\left(1+\frac{1}{p}\right)\right)^m}{\Gamma\left(1+\frac{m}{p}\right)}$ (cf. Pisier, [43]).

▷ Recall that $\Gamma\left(2+\frac{r}{m}\right) = 1 + \mathcal{O}\left(\frac{1}{m}\right)$, so in high dimensions $m \gg 1$ (8.2) provides a very sharp bound for $J_{r,m}$, which is usually sufficiently good – good enough for numerical evaluations.

8.3. Scenario generation

What the latter theorem provides is an upper and a lower bound for the optimal quantizing measure. To find such an optimal measure is, as discussed, difficult, and the theorem, nor its proof reveal a strategy how to obtain such a measure, that is the quantizing points.

Thus it is *still* of fundamental interest how to obtain such optimal quantizing points, or to find – at least – an approximation which is acceptably good, in a sense which needs to be specified.

The procedure to find quantizers is often called *scenario generation* (a sample point $\omega \in \mathcal{Q} \subseteq \Omega$ in the sample space then is called a *scenario*).

We will give in the sequel a *constructive* upper bound for an approximation by unit cubes for the uniform distribution in m dimensions, and we will show that this upper bound for typical applications is acceptably good. Then we will generalize the procedure to other measures, keeping the quality of the approximation, and so finally we have established then a procedure to *generate* scenarios.

Theorem 8.12 (Quantization by unit cubes)**.** *Let $\|x\|_p$ denote the weighted l^p-norm $\|x\|_p := \left(\sum_{i=1}^m w_i \, |x_i|^p\right)^{\frac{1}{p}}$ on $\mathbb{R}^m$ and $\bar{w} := \prod_{j=1}^m w_j^{\frac{1}{m}}$ the geometric mean of the weights. Then*

$$J_{r,m} \leq \frac{1}{2^r}\left(\frac{m\cdot\bar{w}}{1+p}\right)^{\frac{r}{p}}\left(1+\mathcal{O}\left(\frac{1}{m}\right)\right);$$

more explicitly and precisely

$$J_{r,m} \leq \frac{1}{2^r}\left(\frac{m\bar{w}}{1+p}\right)^{\frac{r}{p}} + \frac{m^{\frac{r}{p}-1}}{2^r\,(2p+1)}\begin{cases}0, & r \leq p\\ (r-p)\,p\left(\frac{\bar{w}}{1+p}\right)^{\frac{r}{p}}, & p \leq r \leq 2p\\ 1-(1+r)\left(\frac{\bar{w}}{1+p}\right)^{\frac{r}{p}}, & 2p \leq r.\end{cases}$$

Remark 8.13. Recall, that $J_{r,m} = Q_r\left([0,1]^m\right)$. We thus may compare the upper bound in the latter theorem with the bounds obtained by Graf and Luschgy, (8.2). The main difference, however, is given by the fact that the bound in the latter theorem relies on an immediate decomposition of the unit cube $[0,1]^m$.

In the charts below we have depicted

$$1,\ \Gamma\left(2+\frac{r}{m}\right)$$

in purple and blue, and the upper bound from the latter theorem

$$\frac{\bar{J}_r\left([0,1]^m\right)}{M_r\left(B\left(0,1\right)\right)}$$

(i.e. the right hand side) in red and dashed.

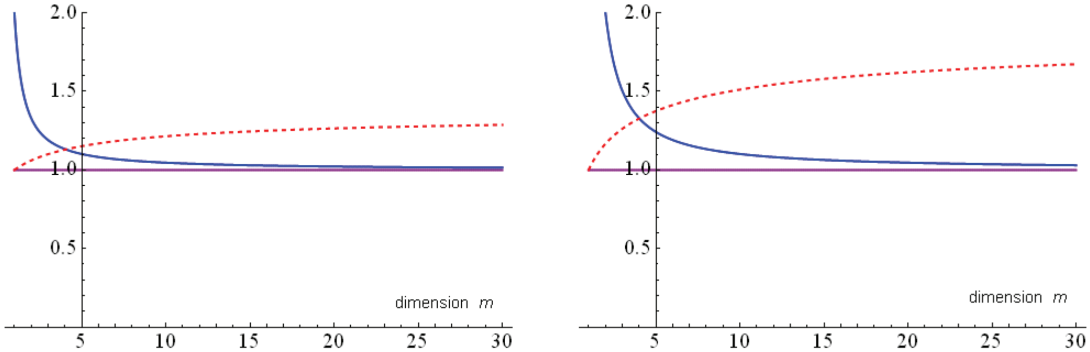

Figure 8.2.: $p = 1$: l^1-distance $\sum |x_i|$: Upper and lower bound of the real quantization coefficient and the unit cube approximation (red and dashed; $r = 1$ left, $r = 2$ right) for increasing dimension m (ordinate).
Unit cube approximation requires up to 60% more quantization points.

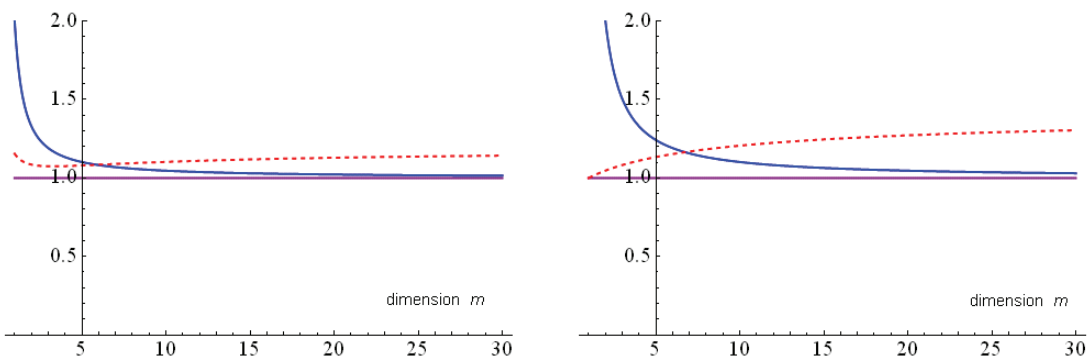

Figure 8.3.: $p = 2$: l^2-distance $\left(\sum |x_i|^2\right)^{\frac{1}{2}}$: Upper and lower bound of the real quantization coefficient and the unit cube approximation (right and dashed; $r = 1$ left, $r = 2$ right).
Unit cube approximation requires up to 20% more quantization points.

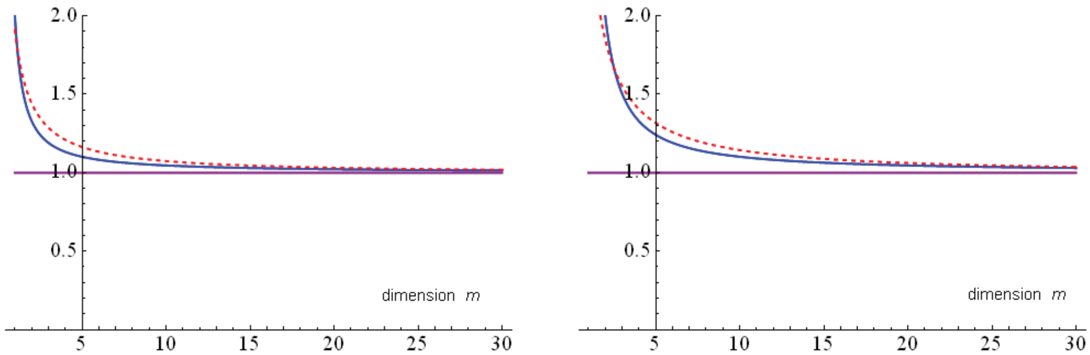

Figure 8.4.: $p = \infty$: l^∞-distance $\max |x_i|$: Upper and lower bound of the real quantization coefficient and the unit cube approximation (right and dashed; $r = 1$ left, $r = 2$ right).
Unit cube approximation is basically the best choice.

Remark 8.14. In many situations the weights w_i have a natural interpretation as time steps, that is $w_i = \Delta t_i = t_i - t_{i-1}$. Involving the inequality of arithmetic and geometric means observe that

$$m \cdot \bar{w} \leq m \cdot \frac{1}{m} \sum_{i=1}^{m} w_i = \sum_{i=1}^{m} t_i - t_{i-1} = t_m - t_0 =: T,$$

the constant thus *stays bounded* on finite time horizons when improving (i.e. decreasing) the discretisation steps: we find, notably irrespective of the discretisation chosen,

$$J_{r,m} \leq \frac{1}{2^r} \left(\frac{T}{1+p} \right)^{\frac{r}{p}} \left(1 + \mathcal{O}\left(\frac{1}{m} \right) \right).$$

Proof. First define the numbers $n_i := \left\lceil n^{\frac{1}{m}} \frac{w_i^{\frac{1}{p}}}{\prod_j w_j^{\frac{1}{mp}}} \right\rceil$ and observe that $n \leq \prod_{i=1}^{m} n_i$ [2].

The unit cube $[0,1]^m$ can be covered by $\bar{n} = \prod_i n_i$ translates of the cube $\times_{i=1}^{m} \left[0, \frac{1}{n_i}\right]$, all of them anchored at the center points $\left(\frac{i_1 - \frac{1}{2}}{n_1}, \frac{i_2 - \frac{1}{2}}{n_2}, \dots \frac{i_m - \frac{1}{2}}{n_m} \right)$. Thus,

[2] $\lceil . \rceil$ is the integer-valued ceiling function satisfying $x \leq \lceil x \rceil < x + 1$

$$\begin{aligned}
J_{r,m} &= \inf_n n^{\frac{r}{m}} \cdot e_{n,r} \left(U\left[0,1\right]^m\right)^r \\
&\leq \bar{n}^{\frac{r}{m}} \cdot e_{\bar{n},r} \left(U\left[0,1\right]^m\right)^r \\
&\leq \bar{n}^{\frac{r}{m}} \cdot \bar{n} \cdot \int_0^{\frac{1}{n_1}} \dots \int_0^{\frac{1}{n_m}} \left\| x - \left(\frac{1}{2n_1}, \frac{1}{2n_2}, \dots \frac{1}{2n_m} \right) \right\|_p^r \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \bar{n}^{\frac{r}{m}} \cdot \bar{n} \cdot 2^m \cdot \int_0^{\frac{1}{2n_1}} \dots \int_0^{\frac{1}{2n_m}} \left\| x \right\|_p^r \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \bar{n}^{\frac{r}{m}} \cdot \bar{n} \cdot 2^m \cdot \int_0^{\frac{1}{2n_1}} \dots \int_0^{\frac{1}{2n_m}} \left(\sum_{i=1}^m w_i \left| x_i \right|^p \right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m .
\end{aligned}$$

Now substitute $x_i \leftarrow \frac{x_i}{2n_i}$ on each of all m axis to transform integration to the unit cube again,

$$\begin{aligned}
J_{r,m} &\leq \bar{n}^{\frac{r}{m}} \cdot \bar{n} \cdot 2^m \cdot \frac{1}{2^m \bar{n}} \cdot \int_0^1 \dots \int_0^1 \left(\sum_{i=1}^m w_i \left| \frac{x_i}{2n_i} \right|^p \right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \frac{\bar{n}^{\frac{r}{m}}}{2^r} \cdot \int_0^1 \dots \int_0^1 \left(\sum_{i=1}^m w_i \left| \frac{x_i}{n_i} \right|^p \right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \frac{1}{2^r} \cdot \int_0^1 \dots \int_0^1 \left(\sum_{i=1}^d \frac{w_i}{n_i^p} \bar{n}^{\frac{p}{m}} \left| x_i \right|^p \right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m
\end{aligned}$$

Now observe, that

$$\begin{aligned}
\frac{w_i}{n_i^p} \bar{n}^{\frac{p}{m}} &\leq \frac{w_i}{n^{\frac{p}{m}} w_i} \bar{n}^{\frac{p}{m}} \cdot \prod_j w_j^{\frac{1}{m}} \\
&= \left(\frac{\bar{n}}{n} \right)^{\frac{p}{m}} \cdot \prod_j w_j^{\frac{1}{m}} \\
&\to \prod_j w_j^{\frac{1}{m}} = \bar{w},
\end{aligned}$$

because $\frac{\bar{n}}{n} \to 1$, as n tends towards infinity. Using the geometric mean $\bar{w} = \prod_{j=1}^m w_j^{\frac{1}{m}}$ we may thus continue with the observation

$$J_{r,m} \leq \frac{\bar{w}^{\frac{r}{p}}}{2^r} \cdot \int_0^1 \dots \int_0^1 \left(\sum_{i=1}^m \left| x_i \right|^p \right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m .$$

To give a sufficiently good proxy for this integral we deduce a useful upper bound

$$\varphi(x) := x^{\frac{r}{p}} \leq \tilde{\varphi}\left(x\right), \tag{8.3}$$

where

$$\begin{aligned}\tilde{\varphi}(x) &:= \left(\frac{1}{1+p}\right)^{\frac{r}{p}} + \frac{r}{p}\left(\frac{1}{1+p}\right)^{\frac{r}{p}-1}\left(x-\frac{1}{1+p}\right) + \\ &+\max\left\{0, \frac{r-p}{p}\left(\frac{1}{1+p}\right)^{\frac{r}{p}-2}, \left(\frac{1+p}{p}\right)^2 - \frac{1+r}{p^2}\left(\frac{1}{1+p}\right)^{\frac{r}{p}-2}\right\}\left(x-\frac{1}{1+p}\right)^2.\end{aligned}$$

By Taylor series expansion at the point $x_0 := \frac{1}{1+p}$,

$$\varphi(x) = \left(\frac{1}{1+p}\right)^{\frac{r}{p}} + \frac{r}{p}\left(\frac{1}{1+p}\right)^{\frac{r}{p}-1}\left(x-\frac{1}{1+p}\right) + R_1(x),$$

the remaining term taking the explicit and exact form

$$R_1(x) = \int_{\frac{1}{1+p}}^{x} \frac{r}{p}\left(\frac{r}{p}-1\right) t^{\frac{r}{p}-2}(x-t)\,\mathrm{d}t.$$

- ▷ $1 \le r \le p$: In this case it is obvious that $R_1(x) \le 0$, the function φ is concave and the above inequality (8.3) holds true;
- ▷ $r = p$: here, $\varphi(x) = \tilde{\varphi}(x)$;
- ▷ $1 \le p \le r \le 2p$: Using integration by parts the remainder term rewrites as

$$\begin{aligned}R_1(x) &= \int_{\frac{1}{1+p}}^{x} \frac{r}{p}\left(\frac{r}{p}-1\right) t^{\frac{r}{p}-2}(x-t)\,\mathrm{d}t \\ &= \frac{r}{p}\left(\frac{r}{p}-1\right)\left(\frac{1}{1+p}\right)^{\frac{r}{p}-2}\frac{\left(x-\frac{1}{1+p}\right)^2}{2} \\ &\quad + \int_{\frac{1}{1+p}}^{x} \frac{r}{p}\left(\frac{r}{p}-1\right)\left(\frac{r}{p}-2\right) t^{\frac{r}{p}-3}\frac{(x-t)^2}{2}\mathrm{d}t \\ &\le \left(\frac{r}{p}-1\right)\left(\frac{1}{1+p}\right)^{\frac{r}{p}-2}\left(x-\frac{1}{1+p}\right)^2,\end{aligned}$$

  the desired inequality;
- ▷ $r = 2p$: again, $\varphi(x) = \tilde{\varphi}(x)$;
- ▷ $2p < r$: The remainder for the 2$^{\text{nd}}$ term is $R_2(x) = \frac{1}{2}\int_{\frac{1}{1+p}}^{x} \frac{r}{p}\left(\frac{r}{p}-1\right)\left(\frac{r}{p}-2\right) t^{\frac{r}{p}-3}(x-t)^2\,\mathrm{d}t$, and $\frac{R_2(x)}{\left(x-\frac{1}{1+p}\right)^3}$ is strictly positive. Subtracting $\alpha\left(x-\frac{1}{1+p}\right)^2$ thus turns R_2 negative in a neighbourhood of $\frac{1}{1+p}$, and as α increases, this neighbourhood increases as well. Choosing α big enough, such that $R_2(1) + \alpha\left(1-\frac{1}{1+p}\right)^2 = 0$, gives the result.

This at hand we turn back to the original statements we are interested in.

Let $(U_i)_{i=1}$ denote independent, uniformly distributed random variables, and $Z_m := \frac{1}{m}\sum_{i=1}^m U_i^p$. We thus may rewrite

$$\begin{aligned}
J_{r,m} &\leq \frac{\bar{w}^{\frac{r}{p}}}{2^r} \cdot \int_0^1 \dots \int_0^1 \left(\sum_{i=1}^m |x_i|^p\right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \frac{(\bar{w}\, m)^{\frac{r}{p}}}{2^r} \cdot \int_0^1 \dots \int_0^1 \left(\frac{1}{m}\sum_{i=1}^m x_i^p\right)^{\frac{r}{p}} \mathrm{d}x_1 \dots \mathrm{d}x_m \\
&= \frac{(\bar{w}\, m)^{\frac{r}{p}}}{2^r} \cdot \mathbb{E}\left[Z_m^{\frac{r}{p}}\right] \\
&\leq \frac{(\bar{w}\, m)^{\frac{r}{p}}}{2^r} \cdot \mathbb{E}\left[\tilde{\varphi}\left(Z_m\right)\right].
\end{aligned}$$

Some straight forward computations show that

$$\begin{aligned}
\mathbb{E}\left[Z_m\right] &= \mathbb{E}\left[U_i^p\right] = \frac{1}{p+1}, \\
\mathbb{E}\left[\left(Z_m - \mu_p\right)^2\right] &= \frac{1}{m}\mathbb{E}\left[\left(U_i^p - \mu_p\right)^2\right] = \frac{1}{m} \cdot \frac{1}{2p+1}\left(\frac{p}{p+1}\right)^2.
\end{aligned}$$

As $\tilde{\varphi}$ is a function involving exactly these two moments the desired result follows. □

8.4. Caught In The Curse Of Dimensionality[3]

The theorem in the latter section gives the upper bound of the unit cube approximation for $U\left([0,1]^m\right)$. We shall elaborate now how this may be expanded naturally to any non-atomic distribution on $\mathbb{R}^m$ by involving the Zador-Gersho formula.

For $m = 1$ it seems likely (and this is proved in [21]) that the asymptotically optimal quantizers are the quantiles of the distribution $\mathbb{P}_r$ with density

$$f_r := \frac{f^{\frac{m}{m+r}}}{\|f\|_{\frac{m}{m+r}}^{\frac{m}{m+r}}},$$

where $f = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}$ is the m-dimensional Lebesgue density of $\mathbb{P}$. To obtain the (asymptotically) optimal quantizers thus the equations

$$\int_{-\infty}^{\omega_i} f^{\frac{1}{1+r}}\left(\omega\right) \mathrm{d}\omega = \frac{2i+1}{2n} \int_{-\infty}^{\infty} f^{\frac{1}{1+r}}\left(\omega\right) \mathrm{d}\omega$$

[3]Cf. [29] on a recent survey on the *Curse Of Dimensionality*.

Algorithm 8.1 Quantization Algorithm

As for higher dimensions the quantization algorithm generalizes as follows:

(i) Consider the projection onto the first dimension $\mathbb{P}_1[A] := \mathbb{P}_r[A \times \mathbb{R}^{m-1}]$. This problem has dimension one and we may proceed as described to obtain points $\omega_{i_1}^1$ and a Voronoi tessellation $\Omega_{i_1}^{;1}$ of $\mathbb{R}$.

(ii) Next consider the measures $\mathbb{P}_2^{i_1}[A] := \mathbb{P}_r\left[\Omega_{i_1}^{;1} \times A \times \mathbb{R}^{m-2}\right]$ for any of $i_1 \in \{1, \dots n_1\}$. Again we may obtain quantization points $\omega_{i_2}^{i_1;2}$ and a Voronoi tessellation $\Omega_{i_2}^{i_1;2}$ of $\mathbb{R}$.

(iii) Repeat step 2 dimension after dimension.

(iv) The measure ($n = n_1 \cdot \dots n_m$)

$$\mathbb{P}_n := \sum_{i_1, i_2, \dots i_n} \mathbb{P}\left[\Omega_{i_1}^{;1} \times \Omega_{i_2}^{i_1;2} \times \dots \Omega_{i_m}^{i_1,\dots i_{m-1};m}\right] \cdot \delta_{\left(\omega_{i_1}^{;1}, \omega_{i_2}^{i_1;2}, \dots \omega_{i_m}^{i_1,\dots i_{m-1};m}\right)}$$

does the required job in $\mathbb{R}^d$ and is an approximation for $\mathbb{P}$.

---

have to be solved to obtain the quantizers ω_i. Together with the Voronoi-weights

$$p_i := \int_{\frac{1}{2}(\omega_{i-1}+\omega_i)}^{\frac{1}{2}(\omega_{i+1}+\omega_{i+1})} f(\omega)\, \mathrm{d}\omega$$

the measures

$$\mathbb{P}_n := \sum_{i=1}^{n} p_i \delta_{\omega_i}$$

then are asymptotically optimal.

To transfer this procedure to higher dimensions Algorithm 8.1 is just adequate, as it respects the findings of Theorem 8.12 and its proof.

8.5. The Zador-Gersho Density

The latter procedure already brings the other quantity in Zador's Theorem (8.8) $\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}$ to light, which involves the density function of the (absolutely continuous part of the) initial probability measure $\mathbb{P}$. We try to investigate this quantity now in various dimensions further.

To this end we split the space $\mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ into two subspaces ($m = m_1 + m_2$) and the multivariate probability measure accordingly. This is reflected in the distributions

- ▷ $\mathbb{P}_1[A_1] := \mathbb{P}[A_1 \times \mathbb{R}^{m_2}]$ and $\mathbb{P}_2[A_2|x_1]$.

The following density functions account for the distributions involved:

▷ $f := \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}$, thus $\mathbb{P}[A] = \int_A f(x)\lambda^m[\mathrm{d}x]$

▷ $f_1(x_1) := \int f(x_1, x_2)\,\mathrm{d}x_2$, that is $\mathbb{P}_1[A_1] = \mathbb{P}[A_1 \times \mathbb{R}^{m_2}] = \int_{A_1} f_1(x_1)\lambda^{m_1}[\mathrm{d}x_1]$

▷ $f_{2|1}(x_2|x_1) := \frac{f(x_1,x_2)}{f_1(x_1)}$, or $\mathbb{P}_2[A_2|\,x_1] = \int_{A_2} f_{2|1}(x_2|\,x_1)\,\lambda^{m_2}[\mathrm{d}x_2]$.

The following proposition clarifies the quantity $\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}$ in Zador's theorem (8.8) in higher dimensions:

Proposition 8.15. *Let $0 < r \le r'$, then*

$$\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}^{\frac{1}{r}} \le \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r'}}^{\frac{1}{r'}},$$

and

$$\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^d}\right\|_{\infty}^{-\frac{1}{m}} \le \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}^{\frac{1}{r}}.$$

Remark. It is worth mentioning that the presumption $1 \le \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}^{\frac{1}{r}}$ somehow seems likely, but the statement is not correct.

Remark. Notice, that $r \mapsto \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}}^{\frac{1}{r}}$ is an increasing function (in r). If in additionally the support of $\mathbb{P}$ is unbounded, then we will find $r > 0$ big enough such that $\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r}} > c^r > 1$. Hence, $\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r'}} > c^{r'}$, which shows that tails get a high attention.

Proof. The proof relies on an application of Hölder's inequality (cf. Proposition 10.1 in the Appendix) in its interpolation form with $\theta = \frac{r}{r'}$: Notice, that $\frac{1}{\frac{m}{m+r}} = \frac{1-\frac{r}{r'}}{1} + \frac{\frac{r}{r'}}{\frac{m}{m+r'}}$, thus

$$\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^d}\right\|_{\frac{m}{m+r}} \le \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{1}^{1-\frac{r}{r'}} \cdot \left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_{\frac{m}{m+r'}}^{\frac{r}{r'}}.$$

Since $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}$ is a density function, $\left\|\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda^m}\right\|_1 = 1$ and the second inequality follows.

As for the other statement notice that

$$\begin{aligned} 1 &= \int f^{-\frac{r}{r+m}} \cdot f^{\frac{r}{r+m}}\,\mathrm{d}\mathbb{P} \\ &\le \int f^{-\frac{r}{r+m}}\,\mathrm{d}\mathbb{P} \cdot \sup f^{\frac{r}{r+m}} \\ &= \int f^{1-\frac{r}{r+m}}\,\mathrm{d}\lambda \cdot \sup f^{\frac{r}{r+m}} \\ &= \|f\|_{\frac{m}{m+r}}^{\frac{m}{m+r}} \cdot \sup \|f\|^{\frac{r}{r+m}}, \end{aligned}$$

and so the statement is immediate. □

Example 8.16. To get a better understanding of this quantity here is the math for some selected distributions:

- ▷ 1-dimensional exponential distribution with inverse scale λ: $\left\| x \mapsto \lambda e^{-\lambda x} \right\|_{\frac{1}{1+r}} = \lambda \cdot \left(\frac{1+r}{\lambda}\right)^{r+1}$,
- ▷ d-dimensional Gaussian distribution with covariance matrix Σ and density function $\frac{1}{\sqrt{(2\pi)^m \det \Sigma}} e^{-\frac{1}{2}x^\top \Sigma^{-1} x}$: $\left\| \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda} \right\|_{\frac{m}{m+r}} = (2\pi)^{\frac{r}{2}} (\det \Sigma)^{\frac{r}{2m}} \left(\frac{m+r}{m}\right)^{\frac{m+r}{2}}$ – this quantity grows rapidly, as r or m increase.
- ▷ The distribution with symmetric density $f(x) = \frac{\alpha \sin \frac{\pi}{\alpha}}{2\pi} \frac{1}{1+(x^2)^{\frac{\alpha}{2}}}$ has heavy tails, particularly for small $\alpha > 1$. The coefficient, which has the closed form $\|f\|_{\frac{1}{1+r}} = \frac{\alpha \sin \frac{\pi}{\alpha}}{2\pi} \left(\frac{\frac{2\alpha}{\alpha-1-r}}{\left(\frac{\frac{1}{1+r}}{\frac{1}{\alpha}}\right)} \right)^{r+1}$, has a pole at $\alpha - 1$ as r approaches from below and tends to infinity.

Proposition 8.17. *Let the density functions be dissected as above, so they satisfy*

$$\int_{A_1 \times A_2} f(x_1, x_2) \lambda^m [\mathrm{d}x_1, \mathrm{d}x_1] = \int_{A_1} f_1(x_1) \int_{A_2} f_{2|1}(x_2 | x_1) \lambda^{m_1} [\mathrm{d}x_2] \lambda^{m_1} [\mathrm{d}x_1].$$

Then,

$$\|f\|^m_{\frac{m}{m+r}} \leq \|f_1\|^{m_1}_{\frac{m_1}{m_1+r}} \cdot \sup_{x_1} \left\| f_{2|1}(.|x_1) \right\|^{m_2}_{\frac{m_2}{m_2+r}}$$

and

$$\|f\|^m_{\frac{m}{m+r}} \leq \|f_1\|^{m_1}_{\frac{m_1}{m_1+r}} \cdot \left(\int f_1(x_1) \cdot f_{2|1}(x_2 | x_1)^{\frac{m_2}{m_2+r}} \mathrm{d}x_2 \mathrm{d}x_1 \right)^{m_2}.$$

The proof relays on some applications of generalizations of Hölder's inequality (cf. proposition (10.1)).

Proof. We apply Hölder's inequality to the triple $\frac{1}{\frac{1}{m+r}} = \frac{1}{\frac{1}{m_1+r}} + \frac{1}{\frac{1}{m_2}}$:

$$\begin{aligned}
\|f\|^{m}_{\frac{m}{m+r}} &= \left(\int f^{\frac{m}{m+r}}\right)^{m+r} = \|f^m\|_{\frac{1}{m+r}} \\
&= \left\| f_1^{m_1} f_{2|1}^{m\frac{m_1+r}{m+r}} f_1^{m_2} f_{2|1}^{m\frac{m_2}{m+r}} \right\|_{\frac{1}{d+r}} \\
&\leq \left\| f_1^{m_1} f_{2|1}^{m\frac{m_1+r}{m+r}} \right\|_{\frac{1}{m_1+r}} \cdot \left\| f_1^{m_2} f_{2|1}^{m\frac{m_2}{m+r}} \right\|_{\frac{1}{m_2}} \\
&= \left(\int f_1^{\frac{m_1}{m_1+r}} \int f_{2|1}^{\frac{m}{m+r}} \mathrm{d}\lambda_2 \mathrm{d}\lambda_1\right)^{m_1+r} \cdot \left(\int f_1 \int f_{2|1}^{\frac{m}{m+r}} \mathrm{d}\lambda_2 \mathrm{d}\lambda_1\right)^{m_2} \\
&\leq \left(\int f_1^{\frac{m_1}{m_1+r}} \mathrm{d}\lambda_1 \cdot \sup_{x_1} \int f_{2|1}^{\frac{m}{m+r}} \mathrm{d}\lambda_2\right)^{m_1+r} \cdot \left(\int f_1 \mathrm{d}\lambda_1 \cdot \sup_{x_1} \int f_{2|1}^{\frac{m}{m+r}} \mathrm{d}\lambda_2\right)^{m_2} \\
&= \|f_1\|^{m_1}_{\frac{m_1}{m_1+r}} \sup_{x_1} \left\| f_{2|1}(.|x_1) \right\|^{m\frac{m_1+r}{m+r}}_{\frac{m}{m+r}} \cdot 1 \cdot \sup_{x_1} \left\| f_{2|1}(.|x_1) \right\|^{m\frac{m_2}{m+r}}_{\frac{m}{m+r}} \\
&= \|f_1\|^{m_1}_{\frac{m_1}{m_1+r}} \sup_{x_1} \left\| f_{2|1}(.|x_1) \right\|^{m}_{\frac{m}{m+r}} .
\end{aligned}$$

Now notice, that $\frac{1}{\frac{m}{m+r}} = \frac{1-\frac{m_1}{m}}{\frac{m_2}{m_2+r}} + \frac{\frac{m_1}{m}}{1}$. Thus, using Hölder's interpolation formula (cf. appendix, $\theta = \frac{m_1}{m}$),

$$\left\| f_{2|1}(.|x_1) \right\|^{m}_{\frac{m}{m+r}} \leq \left\| f_{2|1}(.|x_1) \right\|^{\left(1-\frac{m_1}{m}\right)m}_{\frac{m_2}{m_2+r}} \left\| f_{2|1}(.|x_1) \right\|^{\frac{m_1}{m}m}_{1},$$

and as $x_2 \to f(x_2|\, x_1)$ is a density function,

$$\left\| f_{2|1}(.|x_1) \right\|^{m}_{\frac{m}{m+r}} \leq \left\| f_{2|1}(.|x_1) \right\|^{m_2}_{\frac{m_2}{m_2+r}} .$$

This completes the proof of the first assertion.

As for the 2nd notice, that $\frac{1}{\frac{m}{m+r}} = \frac{1}{\frac{m}{m_1}} + \frac{1}{\frac{m}{m_2+r}}$, whence

$$\begin{aligned}
\|f\|^{m}_{\frac{m}{m+r}} &\leq \left\| f_1^q f_{2|1}^{\frac{m_1}{m}} \right\|^{m}_{\frac{m}{m_1}} \cdot \left\| f_1^{1-q} f_{2|1}^{\frac{m_2}{m}} \right\|^{m}_{\frac{m}{m_2+r}} \\
&= \left(\int\int f_1^{q\frac{m}{m_1}} f_{2|1} \mathrm{d}x_2 \mathrm{d}x_1\right)^{m_1} \cdot \left(\int\int f_1^{(1-q)\frac{m}{m_2+r}} f_{2|1}^{\frac{m_2}{m_2+r}} \mathrm{d}x_2 \mathrm{d}x_1\right)^{m_2+r}
\end{aligned}$$

for every choice of q.

From the definition of $f_{2|1}$ it follows that $\int f_{2|1}(x_2|\, x_1)\, \mathrm{d}x_2 = 1$ for every x_1 and f_1 does not depend on x_2, we thus may continue

$$\begin{aligned}
&= \left(\int f_1^{q\frac{m}{m_1}} \mathrm{d}x_1\right)^{m_1} \cdot \left(\int f_1^{(1-q)\frac{m}{m_2+r}} \left\| f_{2|1}(.|\, x_1) \right\|^{\frac{m_2}{m_2+r}}_{\frac{d_2}{d_2+r}} \mathrm{d}x_1\right)^{m_2+r} \\
&= \|f_1\|^{qm}_{q\frac{m}{m_1}} \cdot \left(\int f_1^{\frac{r}{m_2+r}\frac{m_1}{m_1+r}} \cdot f_1^{(1-q)\frac{m}{m_2+r} - \frac{r}{m_2+r}\frac{m_1}{m_1+r}} \left\| f_{2|1}(.|\, x_1) \right\|^{\frac{m_2}{m_2+r}}_{\frac{m_2}{m_2+r}} \mathrm{d}x_1\right)^{m_2+r}
\end{aligned}$$

Next we fix $q := \frac{m_1}{m}\frac{m_1}{m_1+r}$ and apply Hölder's inequality again for the pair $1 = \frac{1}{\frac{m_2+r}{r}} + \frac{1}{\frac{m_2+r}{m2}}$, thus

$$\begin{aligned}
&= \|f_1\|_{q\frac{m}{m_1}}^{qm} \cdot \left(\int f_1^{\frac{r}{m_2+r}\frac{m_1}{m_1+r}}(x_1) \cdot f_1^{\frac{m2}{m_2+r}}(x_1) \left\| f_{2|1}(.|\, x_1) \right\|_{\frac{m_2}{m_2+r}}^{\frac{m_2}{m_2+r}} \mathrm{d}x_1 \right)^{m_2+r} \\
&\leq \|f_1\|_{q\frac{m}{m_1}}^{qm} \cdot \left(\left\| f_1^{\frac{r}{m_2+r}\frac{m_1}{m_1+r}} \right\|_{\frac{m_2+r}{r}} \cdot \left\| f_1^{\frac{m_2}{m_2+r}} \left\| f_{2|1} \right\|_{\frac{m_2}{m_2+r}}^{\frac{m_2}{m_2+r}} \right\|_{\frac{m_2+r}{m_2}} \right)^{m_2+r} \\
&= \|f_1\|_{q\frac{m}{m_1}}^{qm} \cdot \|f_1\|_{\frac{d_1}{d_1+r}}^{\frac{m_1}{m_1+r}\frac{r}{m_2+r}(m_2+r)} \cdot \left(\int f_1(x_1) \left\| f_{2|1} \right\|_{\frac{m_2}{m_2+r}} \mathrm{d}x_1 \right)^{m_2+r} \\
&= \|f_1\|_{\frac{m_1}{m_1+r}}^{qm} \cdot \|f_1\|_{\frac{m_1}{m_1+r}}^{\frac{m_1 r}{m_1+r}} \left(\int f_1(x_1) \left\| f_{2|1}(.|\, x_1)\, \mathrm{d}x_1 \right\|_{\frac{m_2}{m_2+r}} \right)^{m_2+r} \\
&= \|f_1\|_{\frac{m_1}{m_1+r}}^{m_1} \cdot \left(\int f_1(x_1) \left\| f_{2|1}(.|\, x_1) \right\|_{\frac{m_2}{m_2+r}} \mathrm{d}x_1 \right)^{m_2+r}.
\end{aligned}$$

Now observe that $\varphi(x) := x^{\frac{m_2+r}{m_2}}$ is convex, and Jensen's inequality may be rewritten as $\varphi\left(\int f(x)\, g(x)\, \mathrm{d}x\right) \leq \int f(x)\, \varphi(g(x))\, \mathrm{d}x$. To continue,

$$\begin{aligned}
\|f\|_{\frac{m}{m+r}}^{m} &\leq \|f_1\|_{\frac{m_1}{m_1+r}}^{m_1} \cdot \left(\int f_1(x_1) \left\| f_{2|1}(.|\, x_1) \right\|_{\frac{m_2}{m_2+r}}^{\frac{m_2+r}{m_2}} \mathrm{d}x_1 \right)^{m_2} \\
&= \|f_1\|_{\frac{m_1}{d_1+r}}^{m_1} \cdot \left(\int f_1(x_1) \cdot f_{2|1}(x_2|\, x_1)^{\frac{m_2}{m_2+r}} \mathrm{d}x_2 \mathrm{d}x_1 \right)^{m_2},
\end{aligned}$$

which finally is the desired assertion. □

8.6. Impact On Multistage Scenarios

As a closing remark we want to exhibit the reverse Young inequality, which is of use as well to estimate the Zador-Gersho density, applicable for convoluted densities

$$(f_1 * f_2)(x) := \int_{\mathbb{R}^m} f_1(y)\, f_2(x-y)\, \mathrm{d}y.$$

Recall that the convolution describes the density of a sum $X_1 + X_2$ of random variables, this inequality thus can be applied in situations which allow an accordant decomposition.

Theorem 8.18 (Reverse Young inequality)**.** *Let $0 < p, q, r \leq 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ (and f_1 and f_2 be non-negative). Then*

$$\|f_1 * f_2\|_r \geq C^m \cdot \|f_1\| \cdot \|f_2\|,$$

where $C = \frac{C_p C_q}{C_r}$ for $C_s^2 = \frac{|s|^{\frac{1}{s}}}{|s'|^{\frac{1}{s'}}}$.

As for a proof we refer to [10].

It should be mentioned as well that other inequalities, which might be appropriate in a given, very concrete situation, are contained [19], especially some Sobolev-type inequalities.

We shall now close with a slightly modified problem, which derives from multistage scenarios, and involve the time component more explicitly.

Notation

Consider the distance $d^t(x,y)^p := \sum_{t'\le t} w_{t'}\cdot d_{t'}(u_{t'},v_{t'})^p$ on the space $(\Omega^t,d^t) := \times_{t'\le t}(\Omega_t,d_t)$, let $\mathbb{P}^t$ be a probability measure on Ω^t such that $\mathbb{P}^{t'}[A] = \mathbb{P}^t[A\times\Omega_{t'}\times\dots\Omega_t]$. Then let $\mathbb{P}_t$ denote the conditional measure on Ω_t such that

$$\mathbb{P}^{t_{i+1}}\left[A^{t_i}\times B_{t_{i+1}}\right] = \int_{A^{t_i}} \mathbb{P}_{t_{i+1}}\left[B_{t_{i+1}}|\,u^{t_i}\right]\mathbb{P}^{t_i}\left[\mathrm{d}u^{t_i}\right].$$

Such a decomposition exists according to the disintegration theorem (cf. [14], or [16]).

Then the following holds true (cf. Mirkov, Pflug [33, 32] for an initial result in a similar direction):

Theorem 8.19. *Let $\mathbb{P}$ and $\mathbb{Q}$ be measures on Ω^t such that*

$$d_p\left(\mathbb{P}_t\left[.|u^{t-1}\right],\mathbb{Q}_t\left[.|v^{t-1}\right]\right)\le \tau_t+\kappa_t\cdot d^{t-1}\left(u^{t-1},v^{t-1}\right).$$

Then

$$d_p\left(\mathbb{P}^t,\mathbb{Q}^t\right)\le d_p\left(\mathbb{P}^{t_1},\mathbb{Q}^{t_1}\right)\cdot\prod_{t'\le t}\left(1+w_{t'}\kappa_{t'}\right)+ \\ +\sum_{t'\le t}\tau_{t'}w_{t'}\prod_{t''>t'}\left(1+w_{t''}\kappa_{t''}\right).$$

Proof. Let π^t denote the optimal measure for $d_p(\mathbb{P}^t,\mathbb{Q}^t)$ and moreover $\pi_{t+1}[.|u,v]$ the optimal measure for $d_p(\mathbb{P}_{t+1}[.|u],\mathbb{Q}_{t+1}[.|v])$ and define

$$\pi^{t+1}\left[A_t\times B^t, C_t\times D^t\right] := \int_{A_t\times C_t}\pi_t\left[B^t\times D^t|u^t,v^t\right]\pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right].$$

Then

$$\begin{aligned}
d_p\left(\mathbb{P}^{t+1},\mathbb{Q}^{t+1}\right)^p \leq & \int d^{t+1}\left(u^{t+1},v^{t+1}\right)^p \pi^{t+1}\left[\mathrm{d}u^{t+1},\mathrm{d}v^{t+1}\right] \\
= & \int \left(d^t\left(u^t,v^t\right)^p + w_{t+1}d_{t+1}\left(u_{t+1},v_{t+1}\right)^p\right) \\
& \quad \pi_t\left[\mathrm{d}u_{t+1},\mathrm{d}v_{t+1}|u^t,v^t\right]\pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] \\
= & \int d^t\left(u^t,v^t\right)^p \pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] + \\
& \quad + w_{t+1}\int d_{t+1}\left(u_{t+1},v_{t+1}\right)^p \pi_t\left[\mathrm{d}u_{t+1},\mathrm{d}v_{t+1}|u^t,v^t\right]\pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] \\
= & \int d^t\left(u^t,v^t\right)^p \pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] + \\
& \quad + w_{t+1}\int d_p\left(\mathbb{P}_{t+1}\left[.|u^t\right],\mathbb{Q}_{t+1}\left[.|v^t\right]\right)^p \pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] \\
\leq & \int d^t\left(u^t,v^t\right)^p \pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] + \\
& \quad + w_{t+1}\int \tau_{t+1} + \kappa_{t+1}d^t\left(u^t,v^t\right)^p \pi^t\left[\mathrm{d}u^t,\mathrm{d}v^t\right] \\
= & d_p\left(\mathbb{P}^t,\mathbb{Q}^t\right)^p + w_{t+1}\left(\tau_{t+1} + \kappa_{t+1}d_p\left(\mathbb{P}^t,\mathbb{Q}^t\right)^p\right) \\
= & w_{t+1}\tau_{t+1} + \left(1 + w_{t+1}\kappa_{t+1}\right)d_p\left(\mathbb{P}^t,\mathbb{Q}^t\right)^p
\end{aligned}$$

The assertion of the statement then follows. □

Notice, that the quantities in the latter theorem are often available, and they are not too strict. We thus consider this result as well-adapted for the approximation of multistage problems.

9. Summary and Acknowledgment

In this work we try to investigate features and properties, as they appear natural in the context of stochastic programming. A special focus is given to study the influence of the underlying probability measure to the solution of the problem.

It turns out that the Wasserstein metric is well-adapted for the problems in consideration, and we can give very precise bounds when employing the Wasserstein distance.

The theory developed then is applied to acceptability functions, and precise Lipschitz-bounds can be given for this situation. As acceptability functions are in the center of recent investigations, the continuity properties developed justify the importance and relevance of these functionals and their deployment for financial management.

In this context we prove for example, that a uniform investment – as already propose by Markowitz in contrast to his own model – is optimal, whenever we allow the underlying probability measure to take all potential deviations within a given radius.

The second part addresses the problem of giving good, and well-adapted approximations of a given measure.

As the best approximation quality is known since a few years and their asymptotics as well, we provide a useful and constructive way to find approximations of a given measure, especially for higher dimensions. This is the basis for scenario generation, which is an essential tool in stochastic programming. In this context I we are able to give quantitative results which justify a bundle of methods, as they are often used in stochastic programming intuitively.

Last, but not least, I want to repeat what I mentioned already in the introduction and address two essential words to Prof. Plug:

Thank You.

Remark. All charts have been produced by use of Mathematica.

10. Appendices

10.1. Hölder's Inequality

Hölder's inequality is used in to a high extend, particularly in a generalized framework. As we require Hölder's inequality for non-standard indexes as well we give a proof of the generalized inequality.

Proposition 10.1 (Hölder's inequality)**.** *Let f and g be measurable functions on a measure space. Then*

$$\|f \cdot g\|_p \leq \|f\|_{p_1} \cdot \|g\|_{p_2},$$

provided that $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$ and $0 < p, p_1, p_2 < \infty$. Equality holds, if $|g| \propto |f|^{\frac{p_1}{p_2}}$ almost everywhere.

Moreover (interpolation),

$$\|f\|_{p_\theta} \leq \|f\|_{p_0}^{1-\theta} \cdot \|f\|_{p_1}^{\theta}, \tag{10.1}$$

provided that $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$ and $0 < p_0, p_\theta, p_1 < \infty$.

Remark 10.2. The interpolation result – equation (10.1) – is sometimes called *Lyapunov inequality.*

Proof. Hölder's inequality is well-known for the case $p = 1$ (see, for instance, [59]) and we reduce the general assertion to this particular one. Set $\tilde{p_1} := \frac{p_1}{p}$ and $\tilde{p_2} := \frac{p_2}{p}$ and notice that $\tilde{p_1} \geq 1$, $\tilde{p_2} \geq 1$ and $\frac{1}{\tilde{p_1}} + \frac{1}{\tilde{p_2}} = 1$. Then,

$$\begin{aligned} \|fg\|_p^p &= \|f^p \cdot g^p\|_1 \\ &\leq \|f^p\|_{\tilde{p_1}} \cdot \|g^p\|_{\tilde{p_2}} \\ &= \|f\|_{p_1}^p \cdot \|g^p\|_{p_2}^p. \end{aligned}$$

Raising to the power $\frac{1}{p}$ gives the statement.

To prove the 2nd assertion we apply the 1st one to $f \leftarrow f^{1-\theta}$ and $g \leftarrow f^{\theta}$. That is

$$\begin{aligned} \|f\|_{p_\theta} &= \left\|f^{1-\theta} \cdot f^{\theta}\right\|_{p_\theta} \\ &\leq \left\|f^{1-\theta}\right\|_{\frac{p_0}{1-\theta}} \cdot \left\|f^{\theta}\right\|_{\frac{p_1}{\theta}} \\ &= \|f\|_{p_0}^{1-\theta} \cdot \|f\|_{p_1}^{\theta}, \end{aligned}$$

because $\frac{1}{p_\theta} = \frac{1}{\frac{p_0}{1-\theta}} + \frac{1}{\frac{p_1}{\theta}}$. □

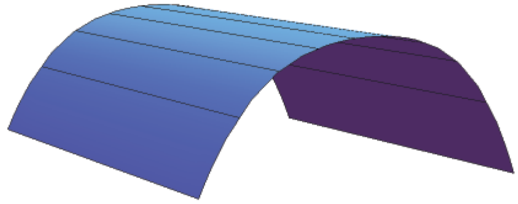

Figure 10.1.: $\mathcal{A}$ is affine linear on any parallel subspace and $\mathcal{A}'_x(d) = e$.

10.2. Translation Equivariance

When introducing the property *translation equivariance* it was stated in Remark 3.6 that this property holds true, if just one single element has the accordant property. Here we shall justify this statement:

Proposition 10.3. *Let $\mathcal{Y}$ be convex, $\mathcal{A} : \mathcal{Y} \to \bar{\mathbb{R}}$ concave and $\mathcal{A}(Y_0 + \mu \cdot d) \geq \mathcal{A}(Y_0) + \mu \cdot e$ for some Y_0 and any $\mu \in \mathbb{R}$. Then*

$$\mathcal{A}(Y + \mu \cdot d) = \mathcal{A}(Y) + \mu \cdot e$$

for all Y in the interior of $\mathcal{A}$'s domain, $Y \in \operatorname{int}\operatorname{dom}(\mathcal{A})$.

Proof. As Y is in the interior there is $Y_1 \in \mathcal{Y}$ such that $Y = (1 - \lambda')\, Y_0 + \lambda' Y_1$ for some $\lambda' \in (0, 1)$. For any $\mu \in \mathbb{R}$ and $0 < \lambda \leq 1$ thus

$$\begin{aligned}
\mathcal{A}(Y + \mu \cdot d) &= \mathcal{A}\left((1-\lambda)\, Y + \lambda\left(Y + \frac{\mu}{\lambda} d\right)\right) \\
&\geq (1-\lambda)\,\mathcal{A}(Y) + \lambda \mathcal{A}\left(Y + \frac{\mu}{\lambda} d\right) \\
&= (1-\lambda)\,\mathcal{A}(Y) + \lambda \mathcal{A}\left((1-\lambda')\, Y_0 + \lambda' Y_1 + \frac{\mu}{\lambda} d\right) \\
&= (1-\lambda)\,\mathcal{A}(Y) + \lambda \mathcal{A}\left((1-\lambda')\left(Y_0 + \frac{\mu}{\lambda(1-\lambda')} d\right) + \lambda' Y_1\right) \\
&\geq (1-\lambda)\,\mathcal{A}(Y) + \lambda(1-\lambda')\,\mathcal{A}\left(Y_0 + \frac{\mu}{\lambda(1-\lambda')} d\right) + \lambda\lambda' \mathcal{A}(Y_1) \\
&\geq (1-\lambda)\,\mathcal{A}(Y) + \lambda(1-\lambda')\left(\mathcal{A}(Y_0) + \frac{\mu}{\lambda(1-\lambda')} e\right) + \lambda\lambda' \mathcal{A}(Y_1) \\
&= (1-\lambda)\,\mathcal{A}(Y) + \lambda(1-\lambda')\,\mathcal{A}(x_0) + \mu \cdot e + \lambda\lambda' \mathcal{A}(Y_1).
\end{aligned}$$

Now let $\lambda \to 0$ to obtain

$$\mathcal{A}(Y + \mu d) \geq \mathcal{A}(Y) + \mu \cdot e \tag{10.2}$$

for any $\mu \in \mathbb{R}$ and $Y \in \operatorname{int}\operatorname{dom}(\mathcal{A})$.

By convexity and (10.2) moreover

$$\begin{aligned}\mathcal{A}(Y) &= \mathcal{A}\left((1-\lambda)(Y+\mu d)+\lambda\left(Y-\frac{1-\lambda}{\lambda}\mu d\right)\right)\\ &\geq (1-\lambda)\mathcal{A}(Y+\mu d)+\lambda\mathcal{A}\left(Y-\frac{(1-\lambda)\mu}{\lambda}d\right)\\ &\geq (1-\lambda)\mathcal{A}(Y+\mu d)+\lambda\left(\mathcal{A}(Y)-\frac{(1-\lambda)\mu}{\lambda}e\right),\end{aligned}$$

that is

$$\mathcal{A}(Y) \geq (1-\lambda)\mathcal{A}(Y+\mu d)+\lambda\mathcal{A}(Y)-(1-\lambda)\mu\cdot e.$$

Now again let $\lambda\to 0$ and thus

$$\mathcal{A}(Y) \geq \mathcal{A}(Y+\mu d)-\mu\cdot e,$$

that is

$$\mathcal{A}(Y+\mu d)\leq \mathcal{A}(Y)+\mu\cdot e.$$

Together with (10.2) this gives

$$\mathcal{A}(Y+\mu d)=\mathcal{A}(x)+\mu\cdot e,$$

which is the assertion. □

10.3. Duality In Optimization

This is a very comprehensive exposition of materials collected in various documents, among them is the (modern) book [9].

Any real-valued function L on $D\times\Lambda$ satisfies the *max-min-inequality*

$$\underbrace{\sup_{\lambda\in\Lambda}\underbrace{\inf_{x\in D}L(x;\lambda)}_{d(\lambda)}}_{d^*}\leq\underbrace{\inf_{x\in D}\sup_{\lambda\in\Lambda}L(x;\lambda)}_{p^*}.$$

- ▷ the inequality $d^*\leq p^*$ is called *weak duality*, and
- ▷ $p^*-d^*\geq 0$ is the *duality gap*;
- ▷ in case of $d^*=p^*$ L is said to have the *strong max-min* property, *strong duality* or *saddle-point* property;
- ▷ the function
$$d(\lambda):=\inf_{x\in D}L(x;\lambda) \tag{10.3}$$
is called *dual function*. Obviously $d(\lambda)\leq d^*\leq p^*$.

A point (x^*, λ^*) is a *saddle-point* if

$$L(x^*; \lambda) \leq L(x; \lambda^*)$$

for all x and λ (in this case $L(x^*; \lambda) \leq L(x^*; \lambda^*) \leq L(x; \lambda^*)$).

The existence of a saddle-point implies the strong max-min property and $d^* = d(\lambda^*) = L(x^*; \lambda^*) = p^*$, because

$$\begin{aligned} p^* &= \inf_{x\in D} \sup_{\lambda\in\Lambda} L(x; \lambda) \\ &\leq \sup_{\lambda\in\Lambda} L(x^*; \lambda) \\ &\leq \inf_{x\in D} L(x; \lambda^*) = d(\lambda^*) \\ &\leq \sup_{\lambda\in\Lambda} \inf_{x\in D} L(x; \lambda) = d^* \end{aligned} \tag{10.4}$$

(By convention, $\inf\{\} = +\infty$, $\sup\{\} = -\infty$, resp.)

Theorem 10.4 (Sion's Minimax Theorem, cf. [54]). *Let*

(i) D and Λ be convex and (at least) one of these sets be compact,

(ii) $x \mapsto L(x, \lambda)$ (quasi-)convex[1] and lsc. for any $\lambda \in \Lambda$ and

(iii) $\lambda \mapsto L(x, \lambda)$ (quasi-)concave and usc. for any $x \in D$,

then L has the strong max-min property.

10.3.1. Lagrangian

To investigate the *primal problem*

$$\text{(P)} \quad \begin{array}{ll} \text{minimize (in x)} & f(x) \\ \text{subject to} & g_j(x) \leq 0 \\ & h_i(x) = 0 \\ & x \in D \end{array}$$

define the *Lagrange-function* on $D \times \{\lambda_i \in \mathbb{R}\} \times \{\mu_j \geq 0\}$ as

$$L(x; \lambda, \mu) := f(x) + \sum_i \lambda_i h_i(x) + \sum_j \mu_j g_j(x).$$

The *Lagrange dual function*, as defined in (10.3), is the *concave* function

$$d(\lambda, \mu) := \inf_{x\in D} L(x; \lambda, \mu).$$

The (unconstrained) *Lagrange dual problem* is the *concave* problem

$$\text{(D)} \quad \begin{array}{ll} \text{maximize (in } \lambda, \mu) & d(\lambda, \mu) \\ \text{subject to} & \mu_j \geq 0. \end{array}$$

[1] f is quasi-convex iff $f((1-\lambda)x_0 + \lambda x_1) \leq \max\{f(x_0), f(x_1)\}$

Theorem 10.5. *(x^*, λ^*, μ^*) is a saddle point for the Lagrangian L iff*

(i) x^* is primal optimal,

(ii) (λ^*, μ^*) is dual optimal and

(iii) strong duality is obtained.

In addition, $d^* = d(\lambda^*, \mu^*) = L(x^*; \lambda^*, \mu^*) = f(x^*) = p^*$ and ${\mu_j^*}^\top g_j(x^*) = 0$ (complementary slackness).

Corollary. *Let x^* be primal optimal and (λ^*, μ^*) dual optimal, but with strictly positive duality gap. Then there does not exist any saddle-point, but the following inequalities hold:*

$$d^* = d(\lambda^*, \mu^*) \leq L(x^*; \lambda^*, \mu^*) = L(x^*; 0, \mu^*) \leq f(x^*) = p^*,$$

and consequently $0 \leq -{\mu^*}^\top g(x^*) \leq p^* - d^*$. The saddle point inequality rewrites

$$L(x^*; \lambda, \mu) - f(x^*) \leq 0 \leq L(x; \lambda^*, \mu^*) - d(\lambda^*, \mu^*)$$

for all x, λ and $\mu \geq 0$.

Proof. Let (x^*, λ^*, μ^*) be a saddle point, then

$$\begin{aligned} d(\lambda, \mu) &\leq L(x^*; \lambda, \mu) \\ &\leq \inf_{x \in D} L(x; \lambda^*, \mu^*) \\ &= d(\lambda^*, \mu^*), \end{aligned}$$

which shows that (λ^*, μ^*) is optimal for the dual.

Strong duality follows via (10.4) since we assume a saddle-point.

In addition

$$\begin{aligned} f(x^*) + \lambda^\top h(x^*) + \mu^\top g(x^*) &= L(x^*; \lambda, \mu) \\ &\leq L(x^*; \lambda^*, \mu^*) \\ &= f(x^*) + \lambda^{*\top} h(x^*) + \mu^{*\top} g(x^*) \end{aligned}$$

for all λ and $\mu \geq 0$, whence $h(x^*) = 0$ and $g(x^*) \leq 0$, which shows that x^* is feasible for the primal problem. Consequently $\mu^\top g(x^*) \leq \mu^{\top *} g(x^*) \leq 0$ for all $\mu \geq 0$, so we deduce $\mu_j^* g_j(x^*) = 0$ (complementary slackness).

Again from the saddle-point-property

$$\begin{aligned} f(x^*) &= L(x^*; \lambda^*, \mu^*) \\ &\leq L(x; \lambda^*, \mu^*) \\ &= f(x) + \lambda^{*\top} \underbrace{h(x)}_{=0} + \underbrace{\mu^{*\top} g(x)}_{\leq 0} \end{aligned}$$

and so it follows that x^* is indeed optimal for the primal problem.

Conversely, observe that

$$\sup_{\lambda,\mu\geq 0}\underbrace{f(x)+\lambda^\top h(x)+\mu^\top g(x)}_{L(x;\lambda,\mu)}=\begin{cases} f(x) & \text{if } h_i(x)=0 \text{ and } g_j(x)\leq 0,\\ \infty & \text{else.}\end{cases}$$

Thus, as x^* is primal optimal and (λ^*,μ^*) dual optimal, $d^*=d(\lambda^*,\mu^*)\leq L(x^*;\lambda^*,\mu^*)\leq f(x^*)=p^*$, and consequently $0\leq -{\mu^*}^\top g(x^*)\leq p^*-d^*$.

Moreover,

$$\begin{aligned} L(x^*;\lambda,\mu) &\leq \sup_{\lambda,\mu\geq 0} L(x^*;\lambda,\mu)\\ &= \inf_{x\in D}\sup_{\lambda,\mu\geq 0} L(x;\lambda,\mu)\\ &= \sup_{\lambda,\mu\geq 0}\inf_{x\in D} L(x;\lambda,\mu)+p^*-d^*\\ &= \sup_{\lambda,\mu\geq 0} d(\lambda,\mu)+p^*-d^*\\ &= d(\lambda^*,\mu^*)+p^*-d^*\\ &\leq L(x,\lambda^*,\mu^*)+p^*-d^* \end{aligned}$$

for all x, λ and $\mu\geq 0$, establishing the saddle-point inequality. □

10.3.2. Fenchel-Transform

It is useful here to naturally extend the (concave) Lagrange-dual function by

$$d(\lambda,\mu)=\begin{cases} d(\lambda,\mu) & \text{if } \mu_j\geq 0\\ -\infty & \text{else.}\end{cases}$$

We may state the dual problem (D) equivalently as

$$\begin{array}{ll} \text{minimize (in } \lambda,\mu) & -d(\lambda,\mu)\\ \text{subject to} & -\mu_j\leq 0, \end{array}$$

(the same form as (P) without h, but $g(\mu)=-\mu$) and start from this problem as initial problem: The Lagrangian is $\tilde{L}(\lambda,\mu;y)=-d(\lambda,\mu)-y^\top\mu$, the corresponding concave dual function is

$$\begin{aligned} \tilde{d}(y) &= \inf_{\lambda,\mu\geq 0}\tilde{L}(\lambda,\mu;y)\\ &= \inf_{\lambda,\mu} -y^\top\mu-d(\lambda,\mu)\\ &= -\sup_{\lambda,\mu}(0,y)^\top\begin{pmatrix}\lambda\\ \mu\end{pmatrix}+d(\lambda,\mu)\\ &= -(-d)^*(0,y), \end{aligned}$$

where

$$f^*(y) := \sup_x y^\top x - f(x)$$

(cf. [49]) is f's *convex conjugate* function (f^* is always convex and lsc.; other names are Fenchel transform, *Legendre-Fenchel transform*; note the *Fenchel-Young* inequality $x^\top y \le f(x) + f^*(y)$; we shall call $-(-f)^*$ *concave conjugate*).

The *dual-dual* problem, in view of (D), thus is

$$\text{(DD)}\quad \begin{array}{ll} \text{maximize (in y)} & \tilde{d}(y) \\ \text{subject to} & y_j \ge 0. \end{array}$$

We may start here *again* with the Lagrangian $\tilde{\tilde{L}}(y;\tilde{\mu}) = -\tilde{d}(y) - \tilde{\mu}^\top y$, the corresponding dual function thus is $\tilde{\tilde{d}}(\tilde{\mu}) = \inf_{y\ge 0} -\tilde{d}(y) - \tilde{\mu}^\top y = -\sup_y \tilde{\mu}^\top y + \tilde{d}(y) = -\left(-\tilde{d}\right)^*(\tilde{\mu})$, and the *dual-dual-dual* thus is

$$\text{(DDD)}\quad \begin{array}{ll} \text{maximize (in } \tilde{\mu}) & -\left(-\tilde{d}\right)^*(\tilde{\mu}) \\ \text{subject to} & \tilde{\mu}_j \ge 0. \end{array}$$

This is the same as (DD), but $\tilde{d}$ replaced by its concave conjugate $-\left(-\tilde{d}\right)^*$. The difference to the dual (D) is that we finally got rid of λ.

Repeating the procedure will lead us back to the (DD), as for convex (lsc.) functions $\left(-\tilde{d}\right)^{**} = -\tilde{d}$. Note the optimal values $d^* = d(\lambda^*, \mu^*) = \tilde{d}(y^*) = -\left(-\tilde{d}\right)^*(\tilde{\mu}^*)$ etc..

10.3.3. Linear Program

These following linear programs are dual – in the sense described – to each other:

Linear Program (primal)	Dual Program
minimize (in x) $c^\top x$ subject to $Ax \ge b$ $x \ge 0$	maximize (in μ) $\mu^\top b$ subject to $\mu^\top A \le c^\top$ $\mu \ge 0$,
minimize (in x) $c^\top x$ subject to $Ax = b$ $x \ge 0$	maximize (in λ) $\lambda^\top b$ subject to $\lambda^\top A \le c^\top$,
minimize (in x) $c^\top x$ subject to $Ax \ge b$	maximize (in μ) $\mu^\top b$ subject to $\mu^\top A = c^\top$ $\mu \ge 0$,
minimize (in x) $c^\top x$ subject to $A_1 x = b_1$ $A_2 x \ge b_2$	maximize (in λ,μ) $\lambda^\top b_1 + \mu^\top b_2$ subject to $\lambda^\top A_1 + \mu^\top A_2 = c^\top$ $\mu \ge 0$,
minimize (in x) $c^\top x$ subject to $A_1 x = b_1$ $A_2 x \ge b_2$ $x \ge 0$	maximize (in λ,μ) $\lambda^\top b_1 + \mu^\top b_2$ subject to $\lambda^\top A_1 + \mu^\top A_2 \le c^\top$ $\mu \ge 0$.

10.3.4. Karush–Kuhn–Tucker (KKT)

Let L be differentiable in the saddle point (x^*, λ^*, μ^*), then $\nabla L(x^*, \lambda^*, \mu^*) = 0$ (notice the *simultaneous* differentiation with respect to all 3 variables).

From Theorem 10.5 we deduce: For *any* optimization problem with *differentiable objective and constraint functions* for which strong duality obtains, any pair of primal and dual optimal points must satisfy the conditions (KKT; cf. [28])

(i) Stationarity: $0 \in \partial f(x^*) + \sum_i \lambda_i^* \cdot \partial h_i(x^*) + \sum_j \mu_j^* \cdot \partial g_j(x^*)$ $(0 = \nabla_x L)$,

(ii) Primal feasibility: $h_i(x^*) = 0$, $g_j(x^*) \leq 0$ $(\nabla_\lambda L = 0,\ \nabla_\mu L = 0)$,

(iii) Dual feasibility: $\mu_j^* \geq 0$ and

(iv) Complementary slackness: $\mu_j^* \cdot g_j(x^*) \geq 0$.

An element u^* out of the (locally convex) linear space's dual is called *sub-gradient*, iff the *sub-gradient inequality* $u^* \in \partial f(x) :\Longleftrightarrow f(z) \geq f(x) + u^*(z - x)$ holds for all z.

$\partial f(x)$, the (convex and closed) set of all sub-gradients in x, is called *sub-differential*.

Theorem. *Let x^* be primal optimal for the primal* (P) *(plus some regularity conditions), then there exist λ^* and μ^* such that (KKT).*

Remark. If the primal problem is convex, then (KKT) are also *sufficient* conditions for optimality of x^*, (λ^*, μ^*).

For differentiable f, g and h the problem

$$
\text{(WD)}\quad \begin{array}{ll} \text{maximize (in } \lambda, \mu) & L(x; \lambda, \mu) \\ \text{subject to} & \mu_j \geq 0, \\ & \nabla_x L(x; \lambda, \mu) = 0 \end{array}
$$

is called *Wolfe dual problem*.

10.3.5. Derivative

Theorem 10.6. *Consider the function*

$$
f(x) := \min\{f(x, y) :\ g(x, y) \leq 0 \text{ and } h(x, y) = 0\}
$$

with minimizing argument $y(x)$ satisfying (KKT) for any x. Let f, g, h and $y \in C^1$, then

$$
f'(x) = f_x(x, y(x)) + \lambda(x)^\top h_x(x, y(x)) + \mu(x)^\top g_x(x, y(x))
$$

with respective Lagrange multipliers (dependent on x).

Proof. As for the proof notice first that $h(x, y(x)) = 0$, thus $h_x + h_y y'(x) = 0$. Then we find either $g_i(x, y(x)) = 0$ or $g_i(x, y(x)) < 0 \wedge \mu_i = 0$ by complementary slackness, that is again $\mu_i(g_{i,x} + g_{i,y} y'(x)) = 0$ and $\mu^\top(g_x + g_y y'(x)) = 0$.

Now recall that $f(x) = f(x, y(x))$ and $f_y + \lambda^\top h_y + \mu^\top g_y = 0$ (KKT). Thus

$$\begin{aligned} f'(x) &= f_x + f_y y'(x) \\ &= f_x - \lambda^\top h_y y'(x) - \mu^\top g_y y'(x) \\ &= f_x + \lambda^\top h_x + \mu^\top g_x. \end{aligned}$$

□

Bibliography

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient flows in metric spaces and in the space of probability measures. *Séminaire de probabilités*, 2005. 18, 84, 86

[2] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203– 228, 1999. 33

[3] Philippe Artzner, Freddy Delbaen, and David Heath. Thinking coherently. *Risk*, 10:68–71, November 1997. 33

[4] Jean-Pierre Aubin and Ivar Ekeland. *Applied Nonlinear Analysis*. John Wiley and Sons, New York, 1984. 48

[5] Stefan Banach. Sur les fonctionelles linéaires. *Studia Mathematica*, 1:214– 229, 1929. 26

[6] Mathias Beiglboeock, Christian Léonard, and Walter Schachermayer. A general duality theorem for the Monge-Kantorovich transport problem. 2009. 21

[7] Francois Bolley. Separability and completeness for the wasserstein distance. *Séminaire de probabilités*, XLI, 2008. 18, 20

[8] Jonathan M. Borwein. Automatic continuity and openness of convex relations. *Proceedings of the American Mathematical Society*, 99(1):49–55, January 1987. 50

[9] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. cambridge university press, 2004. 125

[10] Herm Jan Brascamp and Elliott H. Lieb. Best constants in young's inequality, its converse, and its generalization to more than three functions. *Adv. Math.*, 20:151–173, 1976. 119

[11] Giuseppe Carlo Calafiore and Marco Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742 – 753, 2006. 96

[12] Marco C. Campi and Simone Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM J. Optimization*, 19:1211 – 1230, 2008. 96

[13] Rose-Anne Dana. A representation result for concave Schur concave functions. *Math. Finance*, 15:613–634, 2005. 41

[14] C. Dellacherie and P.-A. Meyer. *Probabilities and potential.* North-Holland Mathematics Studies, North-Holland Publishing Co., Amsterdam., 1978. 119

[15] R. M. Dudley. The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40 Nr. 1:40 – 50, 1969. 106

[16] Richard A. Durrett. *Probability. Theory and Examples.* Duxbury Press, Belmont, CA, second edition, 2004. 119

[17] Thomas S. Ferguson. *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press, New York, 1967. 31

[18] Hans Föllmer and Alexander Schied. *Stochastic Finance - An Introduction in Discrete Time.* de Gruyter Studies in Mathematics 27. de Gruyter, 2004. 33

[19] Richard J. Gardner. The Brunn-Minkowski inequality. *Bulletin of the American Mathematical Society*, 39:355–405, 2002. 119

[20] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70 (3):419 – 435, 2002. 12

[21] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics.* Springer-Verlag Berlin Heidelberg, 2000. 103, 107, 113

[22] Hans Hahn. Über lineare Gleichungssysteme in linearen Räumen. *Journal für die reine und angewandte Mathematik*, 157:214–229, 1927. 26

[23] Wassilij Hoeffding. Maßstabinvariante Korrelationstheorie. *Schriften Math. Inst. Univ. Berlin*, 5:181–233, 1940. 36

[24] Elyès Jouini, Walter Schachermayer, and Nizar Touzi. Law invariant risk measures have the Fatou property. *Advances in mathematical economics.*, 9:49–71, 2006. 49, 52, 66

[25] Elyès Jouini, Walter Schachermayer, and Nizar Touzi. Optimal risk sharing with law invariant monetary utility functions. *Mathematical Finance*, 18(2):269–292, 2008. 35

[26] Leonid Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942. 12

[27] Bangwon Ko, Ralph P. Russo, and Nariankadu D. Shyamalkumar. A note on the nonparametric estimation of the CTE. *Astin Bulletin*, 39(2):717–734, 2009. 96

[28] Harold W. Kuhn and Albert W. Tucker. Nonlinear programming. Proceedings of 2nd Berkeley Symposium, pages 481–492. Berkeley: University of California Press, 1951. 130

[29] Frances Y. Kuo and Ian H. Sloan. Lifting the curse of dimensionality. *Notices of the AMS*, 52:1320–1328, 2005. 113

[30] Shiego Kusuoka. On law invariant risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001. 52

[31] Harry Max Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952. 69

[32] Radoslava Mirkov. *Tree Approximations of Dynamic Stochastic Programs*. PhD thesis, Universität Wien, 2008. 119

[33] Radoslava Mirkov and Georg Ch. Pflug. Tree approximations of dynamic stochastic programs. *SIAM Journal on Optimization*, 18, no. 3:1082 – 1105, 2007. 119

[34] Gaspard Monge. Mémoire sue la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781. 12

[35] Roger B. Nelsen. *An Introduction to Copulas*. Number 139 in Lecture Notes in Statistics. Springer, 1998. 35

[36] Georg Ch. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*, pages 272–281. Kluwer Academic Publishers, Dordrecht, 2000. 43

[37] Georg Ch. Pflug. On distortion functionals. *Statistics and Decisions*, 24:45–60, 2006. 41

[38] Georg Ch. Pflug. Subdifferential representations of risk measures. *Math. Program., Ser. B.*, 108:339–354, 2006. 67, 68

[39] Georg Ch. Pflug and Werner Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, August 2007. 31, 33, 66, 68, 69

[40] Georg Ch. Pflug and David Wozabal. *Quantitative Fund Management*. Financial Mathematics Series. Chapman & Hall/ CRC, 2009. 63

[41] Alois Pichler. Construction of life tables. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, 23:107–119, 1997. 12

[42] Alois Pichler. Deferred life annuities. (Anwartschaftsrenten). *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, 24:541–546, 2000. (German). 35

[43] Gilles Pisier. *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge University Press, Cambridge, 1989. 107

[44] Svetlozar T. Rachev. *Probability metrics and the stability of stochastic models*. John Wiley and Sons Ltd., West Sussex PO19, 1UD, England, 1991. 11, 106

[45] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems Vol. I: Theory, Vol. II: Applications*. Probability and its applications. Springer-Verlag, New York, 1998. 12, 103

[46] A. Rockafellar and Y. Wardi. Nondifferentiability of th esteady-state function in discrete event dynamic systems. *IEEE Transactions on Automatic Control*, 39:1707–1711, 1994. 40

[47] R. Tyrrell Rockafellar. Conjugate duality and optimization. *CBMS-NSF Regional Conference Series in Applied Mathematics*, 16, 1974. 45, 48

[48] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21– 41, 2000. 43, 46

[49] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis.* Springer Verlag, 1997. 129

[50] Andrzej Ruszczyński and Alexander Shapiro. Optimization of risk measures. In G. Calafiore and F Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 17–158. Springer, 2005. 50

[51] Andrzej Ruszczyński and Alexander Shapiro. Conditional risk mappings. *Mathematics of operations research*, 31:544–556, 2006. 50

[52] Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31:433–452, 2006. 50, 52

[53] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming.* MPS-SIAM Series on Optimization, 2009. 12, 31, 50

[54] Maurice Sion. On general minimax theorems. *Pacific J. Math*, 8:171–176, 1958. 126

[55] Gregor Svindland. Continuity properties of law-invariant (quasi-)convex risk functions on l1. *Mathematics and Financial Economics*, 2010. 52

[56] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, 2003. 12, 18, 22, 86

[57] Cédric Villani. *Optimal transport, old and new.* 2008. 12

[58] David Williams. *Probability with Martingales.* Cambridge University Press, 1991. 82

[59] Przemyslaw Wojtaszczyk. *Banach Spaces for Analysts.* Cambridge University Press, Cambridge, 1991. 123

[60] Paul L. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory*, 28:139–149, 1982. 106

Abstract

Acceptability Functionals

Acceptability Functionals – or likewise Risk Functionals – have gained and attracted interest since they have been introduced a decade ago approximately. Their key properties, which are set in an axiomatic way, are intuitively straightforward and very natural; therefore it does not come as a big surprise that various compelling properties can be derived to hold in such an environment.

Applications

Acceptability functionals have found their way in many industry applications, particularly in the financial sector. They may be used – for example – to base an investment decision. And recently US and Canadian insurance supervisory authorities are employing risk functionals as well to measure the risk within a given company, so they have become an element of the actuarial profession as well.

Properties

Various key properties relay on the fact that these functionals are convex (concave). So the entire and well developed theory of convex functions and convex optimization can be applied here to derive general results, and this has been exploited in the past to a high extent.

The convexity property, however, often does not touch the fact that acceptability functionals are typically defined in some probabilistic environment, so they obey stochastic properties as well. So the question arises, if those acceptability functionals are continuous with respect to these underlying probability measures?

Results

To answer this question is a key driver of the present work. Continuity is proven for adequate and fitting distances of probability measures, and even Lipschitz-continuity is established: so acceptability functionals obey quite strong continuity properties.

This is good news: As the probability distribution often is available just from observations as an empirical measure, we may thus trust that evaluating the risk

functionals, based on some observations, is a good proxy for the same risk functional, but evaluated in its original distribution.

An additional result is rather curious: it states that it is optimal to equally distribute ones funds in all available assets, if the objective is to maximize the return, given that the risk has to be accepted by an enlarging class of distributions.

Kurzbeschreibung

Bewertung von Risiko

Funktionale, die Risiko bewerten, haben in der Dekade seit ihrer Einführung große Aufmerksamkeit erreicht. Ihre axiomatischen Haupteigenschaften sind sehr natürlich, aus denen sich dann weitere, sehr schöne und überzeugende mathematische Eigenschaften ableiten lassen.

Anwendungen

Risikofunktionale haben ihr weitestes Anwendungsgebiet in der Finanzbranche, weil sie sehr leicht zur Beurteilung von finanziellen Risiken eingesetzt werden können. Allerdings nicht nur zum Fondsmanagement, denn kürzlich haben die amerikanische sowie die kanadische Versicherungsaufsicht begonnen, gleichfalls Risikofunktionale zur Bewertung von Versicherungsportefeuilles einzusetzen, und spätestens damit wurden Risikofunktionale ein wichtiges Element auch der aktuariellen Zunft.

Eigenschaften

Eine der wichtigsten Eigenschaft eines Risikofunktionals ist natürlich die Konvexität, mit der zugehörigen Dualitätstheorie lassen sich viele weitere Eigenschaften gut beschreiben.

Eine zusätzliche Eigenschaft ist aber ihr aleatorischer Charakter, denn ein Risikofunktional quantifiziert eben das Risiko, das es einem Wahrscheinlichkeitsmaß zumisst. Damit drängt sich die Frage auf, wie denn die Resultate eines Risikofunktionals variieren, wenn sich das zu Grunde liegende Maß ändert?

Ergebnisse

Diese Frage ist zentral in der vorliegenden Arbeit. Es wird gezeigt, dass die Ergebnisse tatsächlich stetig vom Maß abhängen, wenn die Distanz richtig und passend gewählt wird. Ja es gilt sogar Lipschitz-Stetigkeit, und die entsprechende Konstante wird für gängige Risikofunktionale auch konkret angegeben.

In einer umgekehrten Untersuchung wird jenes Wahrscheinlichkeitsmaß eruiert, das, ein gewisses Risiko tolerierend, vom ursprünglichen Maß möglichst weit entfernt liegt. Die Ergebnisse haben eine frappierende Ähnlichkeit zu anderen Ergebnissen in der Transporttheorie.

Diese Fragen sind deshalb interessant, weil das zu Grunde liegende Wahrscheinlichkeitsmaß in aller Regel nicht bekannt ist, weil es beispielsweise nur als Näherung, oder als empirisches Maß aus konkreten Beobachtungen zur Verfügung steht.

Ein weiteres Ergebnis befasst sich mit einer Frage der robusten Optimierung, wie nämlich eine Investmententscheidung aussieht, wenn man weitere Maß zulässt. Das Ergebnis bestätigt die Intuition, dass in diesem Fall eine gleichmäßige Aufteilung der Mittel auf alle bestehenden Investitionstitel optimal ist.

Alois Pichler Curriculum Vitae

PERSONAL INFORMATION	
Name	**PICHLER,** DIPL.- ING. MAG. **Alois**
Nationality	Austria
Date of birth	16. 4. 1969 in Altenberg/ Linz
Children	Louise (31. 10. 1994), Paul Ludwig (10. 6. 1997)
EXPERIENCE	
2009, March	Department of Statistics and Decision Support Systems.
2004	Raiffeisen International, VP
	Responsible to develop Bancassurance.
2003	Kontinuita, Slovak Insurance Company (member of VIG):
	Member of the board.
2001	Compensa Życie, Polish Life Insurance Company (member of VIG):
	Member of the board. Responsible for the actuarial department, re-insurance and IT.
1994	Actuarial department in Vienna Insurance Group (VIG).
EDUCATION AND TRAINING	
1999	Actuary in the Austrian Society of Actuaries (AVÖ).
1993	Post-graduate study visit in Paris, Université Pierre et Marie Curie.
1993	Diploma examination with commendation, degree Magister rerum naturalium (Mag.) and graduate in civil engineering (Dipl.-Ing.). Thesis in applied (numerical) mathematics on the topic: Extraction with chemical reaction in counter current flow columns.
1987 – 1992	Studies of mathematics and lectureship for mathematics and physics at Johannes Keplers University in Linz.
1987, June	General qualification for university entrance with commendation.
1979 – 1987	Grammar school in Linz.
1975 – 1979	Elementary school in Altenberg bei Linz.
LANGUAGES	
Mother Tongue	German
other Languages	English (excellent), French (good), Polish (basic)